J. Chosun Natural Sci.
Vol. 14, No. 3 (2021) pp. 123 – 129
https://doi.org/10.13160/ricns.2021.14.3.123

Effect of Positively Skewed Distribution on the Two sample t-test: Based on Chi-square Distribution

Sunyeong Heo†
Department of statistics, Changwon National University, Changwon 51140, Korea

Abstract

This research examines the effect of positively skewed population distribution on the two sample t-test through simulation. For simulation work, two independent samples were selected from the same chi-square distributions with 3, 5, 10, 15, 20, 30 degrees of freedom and sample sizes 3, 5, 10, 15, 20, 30, respectively. Chi-square distribution is largely skewed to the right at small degrees of freedom and getting symmetric as the degrees of freedom increase. Simulation results show that the sampled populations are distributed positively skewed like chi-square distribution with small degrees of freedom, the F-test for the equality of variances shows poor performances even at the relatively large degrees of freedom and sample sizes like 30 for both, and so it is recommended to avoid using F-test. When two population variances are equal, the skewness of population distribution does not affect on the t-test in terms of the confidence level. However even though for the highly positively skewed distribution and small sample sizes like three or five the t-test achieved the nominal confidence level, the error limits are very large at small sample size. Therefore, if the sampled population is expected to be highly skewed to the right, it will be recommended to use relatively large sample size, at least 20.

Keywords: Chi-square distribution, equal variance, small samples, power of test, two sample *t*-test.

(Recceived August 4, 2021 Revised August 10, 2021 Accepted August 15, 2021)

1. Introduction

Two sample *t*-test is generally used to test the population mean differences for small samples of size less than 30 when the population distributions are normal and the population variances are unknown.

However for small data sets of size 10 or less it is not easy to say whether the sampled populations are normal or not. This research examines the effect of non-normal population on the two sample *t*-test, especially when the distribution of a sampled population is chi-square distribution, through computer simulation. For simulation R 3.6.3 was used.

2. Two sample *t*-test

It is easy to find the definition of the two sample t-test in many statistics textbooks. It is as follows. It first assumes that

(1) $Y_1, Y_2, \cdots, Y_{n_1}$ consists a random sample of size n_1 from a normal population with a mean μ_1 and a variance σ_1^2, and (μ_1, σ_1^2) are unknown

(2) $Y_1, Y_2, \cdots, Y_{n_2}$ consists a random sample of size n_2 from a normal population with a mean μ_2 and a variance σ_2^2, and (μ_2, σ_2^2) are unknown,

(3) Two sampled populations are independent and two samples are selected independently from each population.

(4) Both sample sizes n_1, n_2 are small (<30).

Under these assumptions when one can assume that $\sigma_1 = \sigma_2$, a $(1 - \alpha)100\%$ confidence interval for $\mu_1 = \mu_2$ is given by

†Corresponding author : syheo@changwon.ac.kr

$$(\bar{Y}_1-\bar{Y}_2)\pm t_{\alpha/2,n_1+n_2-2}\cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad (1)$$

where

$$S_p=[\{(n_1-1)S_1^2+(n_2-1)S_2^2\}/(n_1+n_2-2)]^{1/2},$$

and $\bar{Y}_1$, $\bar{Y}_2$ are sample means and S_1^2, S_2^2 the sample variances from each random sample, and $t_{\alpha/2,m+n-2}$ is the upper $(\alpha/2)100$th percentile of $t(m+n-2)$, t-distribution with degrees of freedom n_1+n_2-2.

On the other hand when one can not assume that $\sigma_1=\sigma_2$, a $(1-\alpha)100\%$ confidence interval for $\mu_1=\mu_2$ is given by

$$(\bar{Y}_1-\bar{Y}_2)\pm t_{\alpha/2,\phi}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \quad (2)$$

where the degrees of freedom

$$\phi=\left(\frac{S_1^2}{m}+\frac{S_2^2}{n}\right)^2\Big/\left\{\frac{(S_1^2/m)^2}{m-1}+\frac{(S_2^2/n)^2}{n-1}\right\}$$

(Casella and Berger, 1990; freund, 1992; Satterthwaite, 1946).

3. Test of mean differences based on chi-square distribution

3.1. Skewness and Kurtosis of Chi-square distribution

Skewness and Kurtosis are often used to see how much a distribution is similar as normal distribution.

Skewness is a measure of the lack of symmetric of a distribution. The skewness of a random variable X is given by

$$\alpha_3=\frac{E(X-\mu)^3}{\sigma^3}$$

where μ, σ are the population mean and standard deviation of X. For normal population $\alpha_3=0$.

Kurtosis is a measure of the peakness or flatness of a distribution. The kurtosis of a random variable X is given by

$$\alpha_4=\frac{E(X-\mu)^4}{\sigma^4},$$

and $\alpha_4=3$ for normal distribution (Casella and Berger, 1990; Freund and Walpole, 1980; Ruppert, 1987).

For a random variable X having a chi-square distribution with p degrees of freedom, χ_p^2, the nth moment about origin of X is

$$E(X^n)=\prod_{i=1}^{n}\{p+2(i-1)\}$$

and the skewness and kurtosis of X are as follows

$$\alpha_3=\sqrt{\frac{8}{p}}\ \rightarrow 0 \text{ as } p\rightarrow\infty$$

$$\alpha_4=3+\frac{12}{p}\ \rightarrow 3 \text{ as } p\rightarrow\infty.$$

So, a chi-square random variable has a approximately normal distribution when degrees of freedom p is large.

Table 1 shows mean, variance, skewness and kurtosis of chi-square distribution with degrees of freedom $p=3$, 5. 10, 15, 20, 30.

3.2. Simulation design

In this paper, the effect of non-normal population on the two sample t-test was examined by simulation work. Two independent random samples were selected from chi-square distributions with degrees of freedom p, $p=3$, 5. 10, 15, 20, 30, and sample size n, $n=3$, 5. 10, 15, 20, 30. For each of (p, n) combinations,

I selected two independent random samples 100,000 times, and calculated the proportion of confidence inter-

Table 1. Mean, variance, skewness and kurtosis of chi-square distribution with degrees of freedom $p=3$, 5. 10, 15, 20, 30

degrees of freedom (p)	mean	variance	standard deviation	skewness	kurtosis
3	3	6	2.449	1.633	7
5	5	10	3.162	1.265	5.4
10	10	20	4.472	0.894	4.2
15	15	30	5.477	0.730	3.8
20	20	40	6.325	0.632	3.6
30	30	60	7.746	0.516	3.4

Table 2. Proportions of the 95% confidence intervals including the true $\mu_1 - \mu_2 = 0$ out of 100,000 replications when the equation (1) was used

degrees of freedom (p)	sample size (n)					
	3	5	10	15	20	30
3	0.956	0.958	0.955	0.954	0.952	0.952
5	0.953	0.956	0.954	0.953	0.951	0.950
10	0.952	0.952	0.951	0.951	0.951	0.951
15	0.951	0.952	0.950	0.951	0.951	0.951
20	0.951	0.951	0.951	0.950	0.951	0.951
30	0.950	0.951	0.950	0.950	0.951	0.951

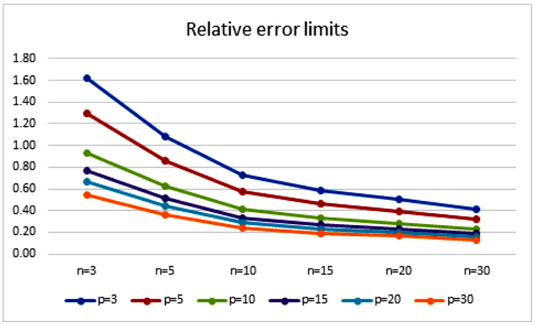

Fig. 1. Relative error limits by degrees of freedom (p) and sample size (n).

vals including $\mu_1 - \mu_2 = 0$. For example, when $p = 5$ and $n = 3$ a random sample of size is selected from χ^2_5, chi-square distribution with degrees of freedom 5, and another random sample with the same sample size three is also selected independently from the same χ^2_5. For each pair of random samples, first, the equality of variances was tested. And then, if the equality of variances was accepted as the result, a 95% confidence interval for $\mu_1 - \mu_2$ was calculated using equation (1), and if not, using equation (2). Finally for each confidence interval it was checked whether it contained the true $\mu_1 - \mu_2 = 0$ or not, and counted the number of confidence interval containing $\mu_1 - \mu_2 = 0$ out of 100,000 confidence interval.

3.3. Simulation results

Based on this simulation design, two random samples had been selected from the same χ^2_p, and both samples had the same sample size. Therefore $\mu_1 - \mu_2 = 0$ and $\sigma_1 = \sigma_2$ are true. Table 2 shows the proportions including true $\mu_1 - \mu_2 = 0$ out of 100,000 95% confidence intervals which were calculated using equation (1) because $\sigma_1 = \sigma_2$ is true. In Table 2 we can see that all proportions are greater than or equal to the nominal confidence level, 95%, for the selected sample sizes and degrees of freedoms. From Table 2 we can see that even though the two sample *t*-test is defined conditional on normal population, it is robust for some non-normal population like having chi-square distribution with small degrees of freedom which is highly skewed to the right.

However it needs to be noticed that the proportions are a little bigger when the sample sizes and degrees of freedoms are small, for example $p = 3$, 5 and $n = 3$, 5, than when the sample sizes and degrees of freedom are large.

Table 3 shows detail simulation results. Table 3 shows the proportions of accepting the true H_0: $\sigma_1 = \sigma_2$ out of 100,000 F-tests for the equality of variances at the 5% level of significance, and the proportions of 95% confidence intervals including $\mu_1 - \mu_2 = 0$ by whether accepting H_0: $\sigma_1 = \sigma_2$ or not, for each combination of (p, n), $p = 3$, 5. 10, 15, 20, 30 and 3, 5. 10, 15, 20, 30.

The F-test concerning the equality of variances assumes that two independent samples are selected from normal populations. We can refer to many statistics textbooks for the details of F-test (e.g. Casella and Berger, 1990; freund, 1992).

In the third column of Table 3 we can see that all the proportions are less than 0.95, the nominal confidence level. We also see that for all selected sizes of sample the proportions are increasing as the degrees of freedoms are increasing. However for a given degrees of freedom, the proportions are decreasing as the sample sizes are increasing. The smallest value is 0.781 at $p = 3$, $n = 30$. From the result, we can know that when the population distribution is not normal and positively

Table 3. Proportions of accepting the true H_0: $\sigma_1 = \sigma_2$ out of 100,000 F-tests for the equality of variances at the 5% level of significance, and the proportions of confidence intervals including the true $\mu_1 - \mu_2 = 0$ out of 100,000 95% confidence intervals by whether the true H_0: $\sigma_1 = \sigma_2$ were accepted or not

degrees of freedom (p)	sample size (n)	Proportion of accepting H_0: $\sigma_1 = \sigma_2$	Proportion of including $\mu_1 - \mu_2 = 0$		
			when H_0: $\sigma_1 = \sigma_2$ accepted (using eq. (1))	when H_0: $\sigma_1 = \sigma_2$ rejected (using eq. (2))	combined
3	3	0.912	0.956	0.990	0.959
	5	0.875	0.964	0.958	0.963
	10	0.827	0.970	0.901	0.958
	15	0.806	0.974	0.884	0.956
	20	0.794	0.975	0.870	0.953
	30	0.781	0.978	0.863	0.953
5	3	0.931	0.953	0.989	0.956
	5	0.910	0.959	0.959	0.959
	10	0.875	0.963	0.905	0.956
	15	0.861	0.966	0.879	0.954
	20	0.851	0.966	0.869	0.952
	30	0.839	0.968	0.856	0.950
10	3	0.941	0.952	0.987	0.954
	5	0.932	0.954	0.964	0.955
	10	0.912	0.956	0.912	0.952
	15	0.905	0.958	0.898	0.952
	20	0.899	0.959	0.885	0.951
	30	0.893	0.960	0.882	0.951
15	3	0.944	0.951	0.987	0.953
	5	0.938	0.953	0.966	0.953
	10	0.927	0.954	0.920	0.951
	15	0.921	0.955	0.911	0.951
	20	0.916	0.956	0.896	0.951
	30	0.912	0.956	0.896	0.951
20	3	0.946	0.951	0.986	0.953
	5	0.942	0.952	0.966	0.953
	10	0.933	0.953	0.927	0.951
	15	0.929	0.953	0.919	0.951
	20	0.925	0.955	0.908	0.951
	30	0.922	0.955	0.903	0.951
30	3	0.947	0.950	0.987	0.952
	5	0.944	0.951	0.965	0.952
	10	0.939	0.952	0.930	0.951
	15	0.937	0.952	0.928	0.950
	20	0.934	0.953	0.919	0.951
	30	0.932	0.954	0.919	0.951

Table 4. Proportions of 95% confidence intervals including the true $\mu_1 - \mu_2 = 0$, the average of error limits, and the relative error limits, by whether the true H_0: $\sigma_1 = \sigma_2$ were accepted or not

degrees of freedom (p)	sample size (n)	when H_0: $\sigma_1 = \sigma_2$ accepted			when H_0: $\sigma_1 = \sigma_2$ rejected		
		Proportion of including $\mu_1 - \mu_2 = 0$	average error limit	relative error limit	Proportion of including $\mu_1 - \mu_2 = 0$	average error limit	relative error limit
3	3	0.956	4.879	1.626	0.990	9.040	3.013
	5	0.964	3.262	1.087	0.958	4.563	1.521
	10	0.970	2.172	0.724	0.901	2.557	0.852
	15	0.974	1.753	0.584	0.884	1.967	0.656
	20	0.975	1.514	0.505	0.870	1.654	0.551
	30	0.978	1.234	0.411	0.863	1.312	0.437
5	3	0.953	6.478	1.296	0.989	11.334	2.267
	5	0.959	4.316	0.863	0.959	5.821	1.164
	10	0.963	2.851	0.570	0.905	3.314	0.663
	15	0.966	2.295	0.459	0.879	2.543	0.509
	20	0.966	1.976	0.395	0.869	2.136	0.427
	30	0.968	1.607	0.321	0.856	1.693	0.339
10	3	0.952	9.352	0.935	0.987	15.214	1.521
	5	0.954	6.217	0.622	0.964	7.924	0.792
	10	0.956	4.089	0.409	0.912	4.576	0.458
	15	0.958	3.281	0.328	0.898	3.545	0.355
	20	0.959	2.819	0.282	0.885	2.990	0.299
	30	0.960	2.287	0.229	0.882	2.376	0.238
15	3	0.951	11.530	0.769	0.987	18.344	1.223
	5	0.953	7.663	0.511	0.966	9.497	0.633
	10	0.954	5.032	0.335	0.920	5.528	0.369
	15	0.955	4.033	0.269	0.911	4.300	0.287
	20	0.956	3.463	0.231	0.896	3.630	0.242
	30	0.956	2.806	0.187	0.896	2.897	0.193
20	3	0.951	13.348	0.667	0.986	21.054	1.053
	5	0.952	8.873	0.444	0.966	10.814	0.541
	10	0.953	5.823	0.291	0.927	6.314	0.316
	15	0.953	4.665	0.233	0.919	4.933	0.247
	20	0.955	4.006	0.200	0.908	4.167	0.208
	30	0.955	3.244	0.162	0.903	3.333	0.167
30	3	0.950	16.396	0.547	0.987	25.663	0.855
	5	0.951	10.897	0.363	0.965	13.037	0.435
	10	0.952	7.146	0.238	0.930	7.650	0.255
	15	0.952	5.723	0.191	0.928	5.986	0.200
	20	0.953	4.912	0.164	0.919	5.064	0.169
	30	0.954	3.977	0.133	0.919	4.061	0.135

skewed like chi-square distribution with small degrees of freedom the F-test for the equality of variances gives no good performance, and so wee need to find any other method, not the F-test.

The last three columns in Table 3 show the proportions of 95% confidence intervals including the true $\mu_1 - \mu_2 = 0$ by the following three cases; when H_0: $\sigma_1 = \sigma_2$ was accepted (fourth column), when H_0: $\sigma_1 = \sigma_2$ was rejected (fifth column), and when both cases were combined (sixth column). When H_0: $\sigma_1 = \sigma_2$ was accepted the 95% confidence interval was obtained by using equation (1), and when H_0: $\sigma_1 = \sigma_2$ rejected by equation (2).

The last column was obtained by combining both cases. For example, when $p = 3$ and $n = 10$, the number of F-tests which accepted the true H_0: $\sigma_1 = \sigma_2$ was 82,714 out of 100,000 tests. And the number of the 95% confidence intervals which included the true $\mu_1 - \mu_2 = 0$ was 80,253 out of the 82,714 confidence intervals which were calculated using equation (1). On the other hand, there were 17,286 F-tests which rejected the true H_0: $\sigma_1 = \sigma_2$, the number of the 95% confidence intervals which included $\mu_1 - \mu_2 = 0$ was 15,570 out of the 17,286 confidence intervals from equation (2). And then, at $p = 3$ and $n = 10$ the combined proportion accepted $\mu_1 - \mu_2 = 0$ becomes (80253+15570)/100000 = 0.958.

Table 3 shows that when H_0: $\sigma_1 = \sigma_2$ was accepted all the proportions are greater than or equal to 0.95, the nominal confidence level, and as the sample sizes are increasing the proportions are also increasing a little. However, the degrees of freedoms are increasing the proportions are decreasing.

Meanwhile when H_0: $\sigma_1 = \sigma_2$ was wrongly rejected, for a given degrees of freedom the proportions are decreasing as the sample sizes are increasing, which trends are the similar as the results of F-tests, and the proportions are achieved the 95% nominal level only when the sample sizes are small like $n = 3$, 5.

When combined both cases, all the proportions are achieving 0.95, the nominal confidence level. Even though this, as the degrees of freedom and sample sizes are increasing the proportions are decreasing.

Table 4 gives the proportions of the 95% confidence intervals including the true $\mu_1 - \mu_2 = 0$, the average of error limits, and the relative error limits, by whether the true H_0: $\sigma_1 = \sigma_2$ were accepted or not.

In Table 4 the values of the third column are equal to those of the fourth column in Table 3 and the values of the sixth column equal to those of the fifth column.

The fourth and seventh columns in Table 4 give the averages of error limits. For example, when $p = 3$, $n = 10$ and H_0: $\sigma_1 = \sigma_2$ accepted, there were 82,714 95% confidence intervals obtained (see Table 3), and so 82,714 error limits were calculated because one error limit is defined per confidence interval. The average of 82,714 error limits is 2.172.

The relative error limit in Table 4 was obtained by the average error limit divided by population mean. For example, when $p = 3$, $n = 10$ and H_0: $\sigma_1 = \sigma_2$ accepted, the relative error limit is 2.172/3=0.724. Since the mean and variance of chi-square distribution increase in proportion to the degrees of freedom, the error limit also increases as the degrees of freedom increases. Therefore in the case, the relative error limit will be better to compare their efficiency. Table 4 show that the relative error limits are decreasing as the degrees of freedom and sample sizes are increasing.

From Table 3 and Table 4, we can see that when the small degrees of freedom and sample sizes like $p = 3$, 5 and $n = 3$, 5 even though the proportions of accepting the true H_0: $\sigma_1 = \sigma_2$ and of including the true $\mu_1 - \mu_2 = 0$ are much bigger than when p, n are large, since the error limits and relative error limits at those small (p, n) are much bigger than at large p, n, it can be said that the test results are uninformative and so useless.

4. Conclusion

Two sample t-test for population mean differences assumes that two independent sampled populations are normally distributed, and population variances are unknown.

The general order of the two sample t-test is as follows. First, we conduct F-test for the equality of variances, and then by depending on the test result we select one of two formulas of the two-sample t-test (equation (1) or (2)).

However, when the sampled population has chi-square distribution with small degrees of freedom, highly skewed distribution to the right, the F-test shows poor performances. Therefore before one tests the equality of variances the test of normality is required, e.g., Shapiro-Wilk test (Shapiro and Wilk, 1965). Once the test rejects the normality of the sampled population distribution, then one needs to find any other proper test for the equality of variances, not the F-test.

When the two population variances are equal and the equality of variances is accepted as the result of a test of equal variances, the positively skewness of population distribution does not affect much on the two sample t-test, which uses equation (1), in terms of the confidence level. Even though the degrees of freedom and sample sizes are small like $p = 3, 5$ and $n = 3, 5$, the two sample t-test achieves the nominal confidence level. However the error limits and the relative error limits are large at small degrees of freedom and sample sizes like $p = 3, 5$ and $n = 3, 5$. Therefore if the sampled population is expected to be largely skewed to the right, it will be recommended to use large sample size, at least 20. Refer to Table 4.

When the two population variances are equal but the equality of variances is wrongly rejected, the two sample t-test using equation (2) shows poor performances. The simulation results can be summarized as follows: First, when the degrees of freedom and sample sizes are small, the test achieves the nominal confidence level but the error limits are very large. Second, when the degrees of freedom are small and the sample sizes are relatively large like 30, the empirical confidence levels of the test are much smaller than the nominal level. Third, even though the degrees of freedom and sample sizes are large like $p = 3, 5$, $n = 3, 5$, the test does not achieve the nominal confidence level.

References

[1] Casella, G. and Berger, R. L., Statistical Inference (Brooks/Cole Publishing Company, 1990).

[2] Freund, J. E., Mathematical Statistics (5th ed) (Prentice-Hall International, Inc., 1992)

[3] Satterthwaite, F. E., “An approximate distribution of estimates of variance components”, Biometrics Bullein, Vol 2, 110-114 (1946).

[4] Freund, J. E. and Walpole, R. E., Mathematical Statistics (3rd ed) (Prentice-Hall International, Inc., 1980)

[5] Ruppert, D., “What is kurtois?: An infulence function approach”, The American Statistician, Vol 41, 1-5 (1987).

[6] Shapiro, S. S. and Wilk, M. B., “An analysis of variance test for normality (complete samples)”, Biometrika, Vol. 52, 591-611 (1965).

[7] The Korean Statistical Society, Statistical Glossary, Freeacademy Inc, Seoul (1987).