العنوان:	VIOLATION OF STATISTICAL ASSUMPTIONS IN THE TWO SAMPLE T-TEST
المصدر:	المجلة العلمية للاقتصاد والتجارة
الناشر:	جامعة عين شمس - كلية التجارة
المؤلف الرئيسي:	Ibrahim, Abrahim Hassan
المجلد/العدد:	ع1
محكمة:	نعم
التاريخ الميلادي:	1992
الصفحات:	69 - 82
رقم MD:	664064
نوع المحتوى:	بحوث ومقالات
قواعد المعلومات:	EcoLink
مواضيع:	الفروض الاحصائية، التجارب الاحصائية، التحليل الاحصائي، الاختبارات
رابط:	http://search.mandumah.com/Record/664064

VIOLATION OF STATISTICAL ASSUMPTIONS IN THE TWO SAMPLE T-TEST

Dr. IBRAHIM HASSAN IBRAHIM

Faculty of Commerce - Tanta University, Egypt

Key words and Phrases : *Normality assumption; Homogeneity assumption; Simulation; Power of the test; t-test; Satterthwaite's test* .

ABSTRACT

The main concern of this article is to study the effect of violating the basic assumptions when using the t-test for testing the equality of two population means . These assumptions are the normality and homogeneity . Violating homogeneity is rather serious than normality assumption . That is because the normality assumption can be waved if sample size is large . However, if the variances of the two populations are not equal , it is suggested that the Satterthwaite's test should be used . The power of the t-test and the Satterthwaite's test was studied extensively using simulation techniques . A comparison among the t-test power and the Satterthwaite's test power proved that the sample sizes should also be taken into consideration when deciding which test should be used even if the homogeneity assumption was violated .

I. INTRODUCTION

It is known in statistical literature that the regular t-test for testing the difference between two population means requires some important assumptions such as independence, normality and homogeneity . This article is concentrating on the homogeneity assumption only . That is because if we are having two dependent samples, the t-test for matched pairs can be used in this regard . While, the normality assumption can be waved if sample size was large enough . Where the t distribution can be approximated by the normal distribution .

The major problem exits if the assumption if homogeneity was violated . This means that the underlying samples were selected from populations with different variances . Consequently, the Satterthwaite's test is suggested instead of the regular t-test . However, there are some argument about the use of the Satterthwaite's test in the situation of none homogeneous variances .

The purpose of this article is to study the effect of violating homogeneity assumption when testing the equality of two population means using the regular t-test . A comparison between the power function of the t-test versus the Satterthwaite's test was done when the assumption of homogeneity exists, and when we have none homogeneous variances i.e., the assumption of homogeneity was violated . This goal was accomplished through simulation techniques for both tests in different situations .

II. DEFINITIONS AND THEORETICAL SETUP

For comparison purpose the t-test statistic was defined to be T_1 and the Satterthwaite's statistic was defined to be T_2 . Also, since we are dealing with a two sample situation, we need to define the original populations of these two samples as follows :

Let $X_1, X_2, X_3, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta_1,\theta_2)$, where θ_1 is the population mean and θ_2 is the population variance of the random variable X .

Similarly, let $Y_1, Y_2, Y_3, \ldots, Y_m$ denote a random sample from a distribution that is $N(\theta_3,\theta_4)$, where θ_3 is the population mean and θ_4 is the population variance of the random variable Y . Where X and Y are two stochastically independent random variables .

Define $\hat{\theta}_1$ and $\hat{\theta}_3$ as the estimated sample means of θ_1 and θ_3 respectively, while $\hat{\theta}_2$ and $\hat{\theta}_4$ are the estimated sample variances of θ_2 and θ_4 .

Let the parameter space be $\Omega = \{ (\theta_1,\theta_2,\theta_3,\theta_4) ; -\infty < \theta_1,\theta_3 < \infty , 0 < \theta_2,\theta_4 < \infty \}$ Then the simple hypothesis $H_0 : \theta_1 = \theta_3$ is to be tested against all alternatives in H_1 . Then $\omega = \{ (\theta_1,\theta_2,\theta_3,\theta_4) ; -\infty < \theta_1 = \theta_3 < \infty , 0 < \theta_2,\theta_4 < \infty \}$, where the set ω is a subset of Ω . Hence the null hypothesis can be described as $H_0 : (\theta_1,\theta_2,\theta_3,\theta_4) \, \varepsilon \, \omega$.

The test statistic for testing the simple null hypothesis $H_0 : \theta_1 = \theta_3$ against the composite hypothesis is the likelihood ratio defined by Hogg and Craig (1978) as follows :

$$\lambda = L(\hat{\omega}) / L(\hat{\Omega}) \qquad \ldots (1)$$

Where the likelihood functions in this case are :

$$L(\omega) = (1/2\pi\theta_2)^{n/2} (1/2\pi\theta_4)^{m/2} \exp\{-\{[\Sigma(x_i - \theta_1)^2]/2\theta_2 + [\Sigma(y_j - \theta_1)^2]/2\theta_4\}\}$$
$$\text{for} \quad (\theta_1,\theta_2,\theta_3,\theta_4) \,\varepsilon\, \omega$$

and

$$L(\Omega) = (1/2\pi\theta_2)^{n/2} (1/2\pi\theta_4)^{m/2} \exp\{-\{[\Sigma(x_i - \theta_1)^2]/2\theta_2 + [\Sigma(y_j - \theta_3)^2]/2\theta_4\}\}$$
$$\text{for} \quad (\theta_1,\theta_2,\theta_3,\theta_4) \,\varepsilon\, \Omega$$

The likelihood ratio λ in equation (1) is defined to be a t distribution with ($n + m - 2$) degrees of freedom. While the null hypothesis $H_0 : \theta_1 = \theta_3$ is to be rejected if $|T_1| \geq c_1$, and if and only if $H_0 : \theta_2 = \theta_4$, i.e., we have homogeneous variances. Where the significance level of the test can be defined as :

$$\alpha_1 = \Pr(\,|T_1| \geq c_1 \mid H_0 : \theta_1 = \theta_3 \text{ and } H_0 : \theta_2 = \theta_4) \qquad ...(2)$$

where c_1 is the critical t value obtained from the regular t - table with degrees of freedom $\nu_1 = n + m - 2$ and α_1 level of significance.

That is because the statistic T_1 which is the likelihood ratio test is defined to be :

$$T_1 = \frac{Q_1}{\{Q_2 / (n+m-2)\}^{1/2}} \qquad ...(3)$$

where

$$Q_1 = \frac{(\hat{\theta}_1 - \hat{\theta}_3)}{\sigma\,(1/n + 1/m)^{1/2}} \qquad ...(4)$$

and

$$Q_2 = [\,\Sigma(x_i - \hat{\theta}_1)^2 + \Sigma(y_j - \hat{\theta}_3)^2\,] \,/\, \sigma^2 \qquad ...(5)$$

The distribution of Q_1 is $N(\,\delta = \dfrac{(\theta_1 - \theta_3)}{\sigma\,(1/n + 1/m)^{1/2}}\,, 1)$ and Q_2 is $\chi^2\,(n+m-2)$,

while Q_1 and Q_2 are stochastically independent and δ is the noncentrality parameter. Also σ^2 is the common variance of the two populations which is usually unknown and in practical applications the estimated pooled variance s_p^2 is used where,

$$s_p^2 = [\,(n-1)\,\hat{\theta}_2 + (m-1)\,\hat{\theta}_4\,] \,/\, \nu_1 \qquad ...(6)$$

However, if the assumption of homogeneity was violated . In other word if the null hypothesis $H_0 : \theta_2 = \theta_4 = \sigma^2$ was not true then the test statistic T_1 in equation (3) can not be used because Q_1 and Q_2 are not distributed as previously stated . As a solution for this problem Satterthwaite suggested the statistic T_2 which depends on the use of the individual variances :

$$T_2 = \frac{(\hat{\theta}_1 - \hat{\theta}_3)}{(\theta_2 / n + \theta_4 / m)^{1/2}} \qquad \text{...(7)}$$

The test statistic T_1 in equation (3) is distributed as t distribution with ν_1 degrees of freedom, where $\nu_1 = n + m - 2$. However, the test statistic T_2 in equation (7) is distributed as t distribution with ν_2 degrees of freedom, where

$$\nu_2 = \frac{(u_1 + u_2)^2}{(u_1)^2/(n-1) + (u_2)^2/(m-1)} \qquad \text{...(8)}$$

where $u_1 = \theta_2 / n$ and $u_2 = \theta_4 / m$. If we let $\eta = \theta_4 / \theta_2$ to be the ratio of the two variances, then equation (8) can be rewritten in the following form :

$$\nu_2 = \frac{(1/n + \eta/m)^2}{[1/n^2(n-1)] + [\eta^2/m^2(m-1)]} \qquad \text{...(9)}$$

In this case the null hypothesis $H_0 : \theta_1 = \theta_3$ is to be rejected if $|T_2| \geq c_2$, and if $H_0 : \theta_2 = \theta_4$ was previously rejected , i.e., we have none homogeneous variances . Where the significance level of this test can be defined as :

$$\alpha_2 = \Pr(\ |T_2| \geq c_2 \mid H_0 : \theta_1 = \theta_3 \text{ and } H_1 : \theta_2 \neq \theta_4) \qquad \text{...(10)}$$

where c_2 is the critical t value obtained from the regular t - table with degrees of freedom ν_2 as defined in equation (8) or equation (9) and α_2 is the level of significance . It worth noting that if the real variances θ_2 and θ_4 are not known which is very often the case then, their equivalent estimated values can be used instead .

To test homogeneity assumption $H_0 : \theta_2 = \theta_4$ against $H_1 : \theta_2 \neq \theta_4$ for two independently normally distributed random samples, the following test statistic can be used :

$$F^* = \hat{\theta}_4 / \hat{\theta}_2 \qquad ...(11)$$

The null hypothesis is to be rejected if :

$$F^* > F_{\alpha/2;\,(m-1),(n-1)} \qquad \text{OR} \qquad F^* < F_{(1-\alpha/2);\,(m-1),(n-1)}$$

where $F_{\alpha/2;\,(m-1),(n-1)}$ and $F_{(1-\alpha/2);\,(m-1),(n-1)}$ denote the upper $\alpha/2$ and lower $1 - (\alpha/2)$ percentage points of the F distribution with (n - 1) and (m - 1) degrees of freedom . And $F_{(1-\alpha/2);\,(m-1),(n-1)} = 1 / [\, F_{\alpha/2;\,(n-1),(m-1)} \,]$.

III. COMPUTING THE POWER OF THE TESTS THROUGH SIMULATION TECHNIQUE

The power of any test in general is defined to be the probability of rejecting the null hypothesis given it was not true . But for the underlying problem we have to distinguish between the cases of homogeneity and none homogeneity of variances . Hence the power functions may be defined as follows for the regular t-test statistic T_1 and for Satterthwaite's test statistic T_2 .

(1) The power of the regular t-test statistic (T_1) :

$$\begin{aligned} P_t = \; & \Pr(\text{ Rejecting } H_0 : \theta_1 = \theta_3 \mid H_1 \text{ is true and variances are equal}) \\ & + \Pr(\text{ Rejecting } H_0 : \theta_1 = \theta_3 \mid H_1 \text{ is true and variances are not equal}) \end{aligned}$$

For simplicity, the alternative hypothesis H_1 which is a composite hypothesis will be considered as a one tail hypothesis $H_1 : \theta_1 > \theta_3$. Then the power function can be written as follows :

$$\begin{aligned} P_t = \; & \Pr(T_1 \geq c_1 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_0 : \theta_2 = \theta_4) \\ & + \Pr(T_1 \geq c_1 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_1 : \theta_2 \neq \theta_4) \end{aligned} \qquad ...(12)$$

The simulation procedures depends on replicating the test statistic T_1 as many as N number of replications . The number of times $H_0 : \theta_1 = \theta_3$ is rejected - in both cases equal or none equal variances - divide by N is the going to be the power of the test . Hence, equation (12) can be written as follows :

$$P_t = \#[(T_1 \geq c_1 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_0 : \theta_2 = \theta_4)]/N$$
$$+ \#[(T_1 \geq c_1 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_1 : \theta_2 \neq \theta_4)]/N \quad ...(13)$$

(2) The power of the Satterthwaite's test statistic (T_2) :

In a similar way the power function of the Satterthwaite's test statistic can be defined as follows :

$$P_{st} = \Pr(T_2 \geq c_2 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_0 : \theta_2 = \theta_4)$$
$$+ \Pr(T_2 \geq c_2 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_1 : \theta_2 \neq \theta_4) \quad ...(14)$$

and for simulation purpose the power is defined to be :

$$P_{st} = \#[(T_2 \geq c_2 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_0 : \theta_2 = \theta_4)]/N$$
$$+ \#[(T_2 \geq c_2 \mid H_1 : \theta_1 > \theta_3 \text{ and } H_1 : \theta_2 \neq \theta_4)]/N \quad ...(15)$$

A MINITAB program was designed to compute the power functions defined in equations (13) and (15) . The number of replications used in all computations was $N = 100$ and the significance level used in all cases was $\alpha = 5\%$. The algorithm required for computing the power functions is summarized in the following steps :

Step-1 : Generate a random sample $X_1, X_2, X_3, \ldots, X_n$ from a distribution that is normal, i.e., X is $N(\theta_1,\theta_2)$, where θ_1 is the population mean and θ_2 is the population variance of the random variable X .

Step-2 : Independently, generate another random sample $Y_1, Y_2, Y_3, \ldots, Y_m$ from a distribution that is normal, i.e., Y is $N(\theta_3,\theta_4)$, where θ_3 is the population mean and θ_4 is the population variance of the random variable Y .

Step-3 : Compute the estimated sample means of θ_1 and θ_3 . Also for the same samples compute the estimated sample variances θ_2 and θ_4 .

Step-4 : Substitute in equation (3) to find the calculated test statistic T_1 and in equation (7) to find the test statistic T_2 .

Step-5 : Repeat step-1 to step-4 as many as N times from the same normal distributions . In every replication the statistics T_1 and T_2 are calculated .

Step-6 : Count the number of times the null hypothesis was rejected when using the statistics T_1 and T_2 . Then use equations (13) and (15) to find the power of both tests .

IV. STATISTICAL FINDINGS

The simulation procedures were processed on different sample sizes generated from normal distributions with different means and variances as explained in the following possible six different cases :

Case - 1 : $n = \text{or} \neq m$, $\theta_1 = \theta_3$ and $\theta_2 = \text{or} \neq \theta_4$
Case - 2 : $n = m$, $\theta_1 > \theta_3$ and $\theta_2 = \theta_4$ or $\theta_2 < \theta_4$ or $\theta_2 > \theta_4$
Case - 3 : $n \neq m$, $\theta_1 > \theta_3$ and $\theta_2 = \theta_4 \Leftrightarrow \eta = 1$
Case - 4 : $n > m$, $\theta_1 > \theta_3$ and $\theta_2 < \theta_4 \Leftrightarrow \eta > 1$
Case - 5 : $n < m$, $\theta_1 > \theta_3$ and $\theta_2 > \theta_4 \Leftrightarrow \eta < 1$
Case - 6 : $n < m$, $\theta_1 > \theta_3$ and $\theta_2 < \theta_4 \Leftrightarrow \eta > 1$

TABLE-I demonstrates results of the first case where we have samples with equal means . The power in all cases is the level of the test as previously stated in equations (2) and (10), where $P_t = P_{st} = \alpha_1 = \alpha_2 = \alpha = 0.05$. Even when we have different sample sizes or none equal variances, results remain the same . These results prove the validity of the simulation program designed for this study .

TABLE-I

Case - 1 : $n = \text{or} \neq m$, $\theta_1 = \theta_3$ and $\theta_2 = \text{or} \neq \theta_4$

n	m	θ_1	θ_3	θ_2	θ_4	ν_1	ν_2	c_1	c_2	P_t	P_{st}
20	20	6	6	1	1	38	38	1.68	1.68	0.05	0.05
20	20	8	8	2	2	38	38	1.68	1.68	0.05	0.05
10	20	6	6	3	3	28	18	1.70	1.73	0.05	0.05
20	10	5	5	2	5	28	13	1.70	1.77	0.05	0.05

The second case is shown in TABLE-II where we have equal sample sizes with different means while the variance ratio have been changed from one to less than one and to greater than one, i.e., $\eta = 1, \eta < 1$ and $\eta > 1$. For example, the first case of TABLE-II where $\eta = 1$, the power $P_t = P_{st} = 100\%$. While the variance ratio of the fourth case of the same table $\eta = 5/2$ and the power values remain the same where, $P_t = P_{st} = 93\%$. Also, the last case of the same table $\eta = 1/6$ and the power values remain the same where, $P_t = P_{st} = 97\%$.

It is quit obvious that the power of both tests, the t-test and the Satterthwaite's test are the same in all cases whenever sample sizes are equal, except when the variance ratio gets very large, i.e., larger than 10 or 20, where $P_t = 84\%$ and $P_{st} = 83\%$ for $\eta = 10$. While $P_t = 68\%$ and $P_{st} = 67\%$ for $\eta = 20$.

TABLE-II

Case - 2 : $n = m$, $\theta_1 > \theta_3$ and $\theta_2 = \theta_4$ or $\theta_2 < \theta_4$ or $\theta_2 > \theta_4$

n	m	θ_1	θ_3	θ_2	θ_4	ν_1	ν_2	c_1	c_2	P_t	P_{st}
20	20	8	6	1	1	38	38	1.68	1.68	1.00	1.00
20	20	8	6	2	2	38	38	1.68	1.68	1.00	1.00
20	20	8	6	1	2	38	34	1.68	1.67	1.00	1.00
20	20	8	6	2	5	38	32	1.68	1.69	0.93	0.93
20	20	8	6	1	3	38	30	1.68	1.70	1.00	1.00
20	20	8	6	1	4	38	28	1.68	1.70	0.98	.098
20	20	8	6	1	6	38	25	1.68	1.71	0.94	0.94
20	20	8	6	1	8	38	24	1.68	1.71	0.91	0.91
20	20	8	6	1	10	38	23	1.68	1.71	0.84	0.83
20	20	8	6	1	20	38	21	1.68	1.72	0.68	0.67
20	20	8	6	5	2	38	32	1.68	1.69	0.93	0.93
20	20	8	6	6	1	38	25	1.68	1.71	0.97	0.97

The third case is presented in TABLE-III where sample sizes are different with different means but with equal variances, which means that the variance ratio equals to one, i.e., $\eta = 1$ in all cases . Results shown in TABLE-III prove that both t-test and Satterthwaite's test are having the same power in detecting possible differences between population means, that is because $P_t = P_{st}$ whenever $\eta = 1$.

TABLE-III

Case - 3 : $n \neq m$, $\theta_1 > \theta_3$ and $\theta_2 = \theta_4 \Leftrightarrow \eta = 1$

n	m	θ_1	θ_3	θ_2	θ_4	ν_1	ν_2	c_1	c_2	P_t	P_{st}
20	10	8	6	1	1	28	18	1.70	1.73	1.00	1.00
20	10	8	6	2	2	28	18	1.70	1.73	0.98	0.98
10	20	8	6	3	3	28	18	1.70	1.73	0.89	0.89

The fourth case is shown in TABLE-IV where the first sample (n > m) is larger than the second while on the contrary the variance of the smaller sample ($\theta_2 < \theta_4$) size is greater than the variance of the larger sample .

TABLE-IV

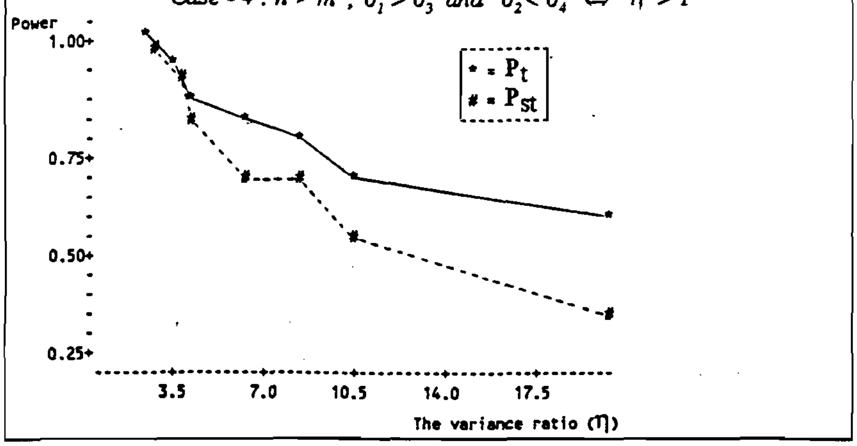
Case - 4 : n > m , $\theta_1 > \theta_3$ and $\theta_2 < \theta_4 \Leftrightarrow \eta > 1$

n	m	θ_1	θ_3	θ_2	θ_4	ν_1	ν_2	c_1	c_2	P_t	P_{st}
20	10	8	6	1	2	28	14	1.70	1.76	1.00	0.99
20	10	8	6	2	5	28	13	1.70	1.77	0.91	0.88
20	10	8	6	1	3	28	12	1.70	1.78	0.96	0.94
20	10	8	6	1	4	28	11	1.70	1.80	0.92	0.83
20	10	8	6	1	6	28	11	1.70	1.80	0.86	0.71
20	10	8	6	1	8	28	10	1.70	1.81	0.80	0.68
20	10	8	6	1	10	28	10	1.70	1.81	0.71	0.56
20	10	8	6	1	20	28	9	1.70	1.83	0.60	0.35

Results in TABLE-IV prove that the regular t-test is more powerful where P_t is greater than P_{st} in all cases although we are having several none homogeneous cases . Because, if we use the homogeneity test statistic F^* for testing $H_0 : \theta_2 = \theta_4$ against $H_1 : \theta_2 \neq \theta_4$, the null hypothesis is to be rejected if :

FIGURE-I

Case - 4 : n > m , $\theta_1 > \theta_3$ and $\theta_2 < \theta_4 \Leftrightarrow \eta > 1$



$F^* > F_{\alpha/2;\,(m-1),(n-1)}$ OR $F^* < F_{(1-\alpha/2);\,(m-1),(n-1)}$

$= F_{0.025;9,19}$ OR $= F_{0.975;9,19}$

$= 2.88$ OR $= 0.27$

Noting that from equation (11), $F^* = \eta = \theta_4 / \theta_2$. We can find in TABLE-IV that for $\eta > 2.88$ we are having none homogeneous cases . These cases are when η = 3, 4, 6, 8, 10 and 20 . This case is shown in FIGURE-I where the power of both tests are almost the same when the variance ratio is near to one but the gap between them gets bigger as the variance ratio gets greater than one .

Similar results are shown in TABLE-V where the smaller sample has a greater variance than the larger sample .

TABLE-V

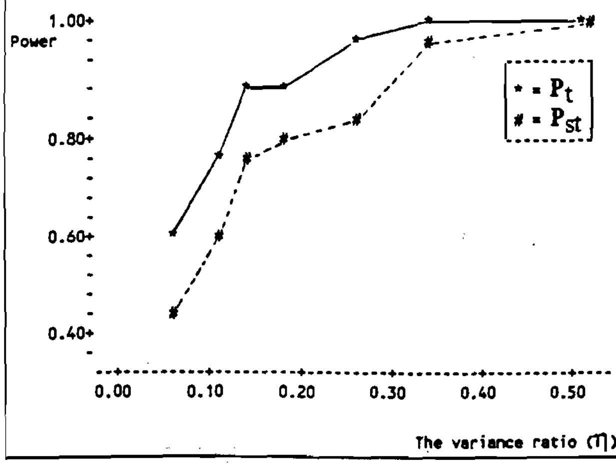
Case - 5 : $n < m$, $\theta_1 > \theta_3$ and $\theta_2 > \theta_4$ $\Leftrightarrow$ $\eta < 1$

n	m	θ_1	θ_3	θ_2	θ_4	ν_1	ν_2	c_1	c_2	P_t	P_{st}
10	20	8	6	2	1	28	14	1.70	1.76	0.99	0.99
10	20	8	6	3	1	28	12	1.70	1.78	1.00	0.95
10	20	8	6	4	1	28	11	1.70	1.80	0.96	0.85
10	20	8	6	6	1	28	11	1.70	1.80	0.87	0.78
10	20	8	6	8	1	28	10	1.70	1.81	0.87	0.74
10	20	8	6	10	1	28	10	1.70	1.81	0.77	0.60
10	20	8	6	20	1	28	9	1.70	1.83	0.60	0.42

There are some none homogeneous cases in TABLE-V . These are the cases when $(1/\eta) = F^*$ = 3, 4, 6, 8, 10 and 20 because we are having the same critical values given before for TABLE-IV . The reciprocal of the ratio η was used because it was defined to be θ_4 / θ_2 while the F^* ratio should be the ratio of the larger variance to the smaller one . But regardless of the homogeneity situation, the power of the regular t-test is greater than the power of the Satterthwaite's test as shown in FIGURE-II . The power in this case is an increasing function of the variance ratio .

FIGURE-II

Case - 5 : $n < m$, $\theta_1 > \theta_3$ and $\theta_2 > \theta_4$ $\Leftrightarrow$ $\eta < 1$



A different case is shown in TABLE-VI where we have different sample sizes but the smaller sample has the smaller variance . Also, the larger sample has the greater variance where , $n < m$ and $\theta_2 < \theta_4$ which is equivalent to $\eta > 1$.

For the given sample sizes and variances the null hypothesis $H_0 : \theta_2 = \theta_4$ against $H_1 : \theta_2 \neq \theta_4$, is to be rejected if $\eta = F^* > 3.69$ or $\eta = F^* < 0.347$.

TABLE-VI

Case - 6 : $n < m$, $\theta_1 > \theta_3$ and $\theta_2 < \theta_4$ $\Leftrightarrow$ $\eta > 1$

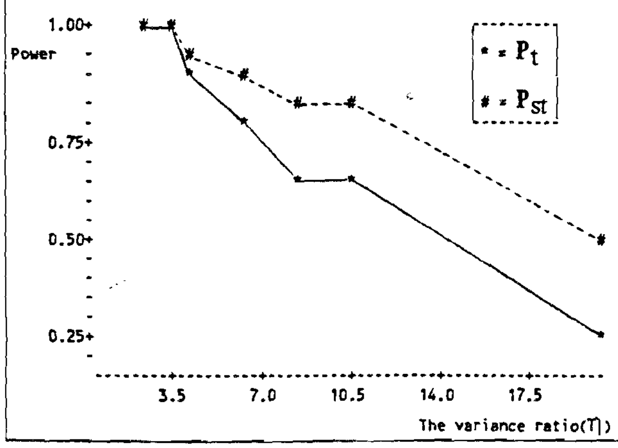
n	m	θ_1	θ_3	θ_2	θ_4	ν_1	ν_2	c_1	c_2	P_t	P_{st}
10	20	8	6	1	2	28	24	1.70	1.71	1.00	1.00
10	20	8	6	1	3	28	27	1.70	1.70	0.99	1.00
10	20	8	6	1	4	28	28	1.70	1.70	0.89	0.95
10	20	8	6	1	6	28	27	1.70	1.70	0.78	0.91
10	20	8	6	1	8	28	26	1.70	1.71	0.66	0.86
10	20	8	6	1	10	28	25	1.70	1.71	0.65	0.86
10	20	8	6	1	20	28	23	1.70	1.71	0.27	0.52

The first two cases of TABLE-VI represent homogeneous cases, because $\eta = 2 < 3.69$ for the first case and $\eta = 3 < 3.69$ for the second case . That is why the power values are about the same . However, in the remaining cases the power values are for the Satterthwaite's test where the power P_{st} is greater than the power P_t of the regular t-test . That is because $\eta > 3.69$ for the remaining cases, where $\eta = 4, 6, 8, 10$ and 20 respectively .

FIGURE-III represents the last case where the larger sample has the greater variance . The power of the t-test is smaller than the power of the Satterthwaite's test. In this case it is better to use the Satterthwaite's test because the regular t-test behaves poorly especially if the variance ratio is increasing and getting much far from unity.

FIGURE-III

Case - 6 : $n < m$, $\theta_1 > \theta_3$ *and* $\theta_2 < \theta_4 \Leftrightarrow \eta > 1$



In general, the power of any one of the two mentioned tests decreases if the variance ratio is as much large as 10 or 20 or more . As its shown in TABLE-IV the power dropped from P_{st} = 99% to P_{st} = 35% , while P_t = 100% dropped to P_t = 60% as the variance ratio increased from $\eta = 2$ to $\eta = 20$.

Also, in TABLE-V the power dropped from P_{st} = 99% to P_{st} = 42%, while P_t = 99% dropped to P_t = 60% . And in TABLE-VI the power dropped from P_{st} = 100% to P_{st} = 52%, while P_t = 100% dropped to P_t = 27% .

V. CONCLUSION

The regular t-test for testing equality of two population means requires some basic assumptions . The most important one is the homogeneity assumption, because if variances are different the power of detecting a difference between means decreases as the difference between variances increases .

The Satterthwaite's test is known as an alternative test for the t-test when the assumption of homogeneity was violated . The power of both t-test and Satterthwaite's test was studied extensively in this work through simulation techniques . The comparison between the power of the two tests revealed that sample size plays an important role in determining which test should be used especially when the homogeneity assumption was violated, as indicated in the following results :

(1) If sample sizes are equal, the power of both t-test and Satterthwaite's test are the same as long as the variance ratio is near one and less than or equal to eight . However, this power decreases sharply if the variance ratio is greater than 20 or more .

(2) If there is no difference between population variances, the power of detecting differences between population means when using any of the two tests remains the same .

(3) The power of the regular t-test is greater than the Satterthwaite's test if sample sizes are different and the larger sample is the sample with the smaller variance .

(4) The power of the Satterthwaite's test is greater than the regular t-test if sample sizes are different and the smaller sample is the sample with the smaller variance .

BIBLIOGRAPHY

Berenson, L ., Levine D ., and Goldstein M . (1983), *Intermediate Statistical Methods and Applications, A Computer Package Approach*, New Jersey : Prentice-Hall, Inc .

Cochran, W . G . (1947), "Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied," *Biometrics*, 3, 22-38 .

Conover, W . J ., Johnson M . E . , and Johnson M . M . (1981), "A Comparative Study of Tests for Homogeneity of Variances, with Applications to Outer Continental Shelf Bidding Data," *Technometrics*, 23, 351-361 .

Hogg, R . V ., and Craig, A . T . (1978), *Introduction to Mathematical Statistics* , Fourth edition, New York : Macmilan Publishing Co., Inc.

Ibrahim, H . I . (1991), "Evaluating the Power of the Mann-Whitney Test Using the Bootstrap Method," *Communications in Statistics - Theory and Methods*, 20 , 2919-2931 .

Montgomery, D . C . (1984), *Design and Analysis of Experiments* , Second edition, New York : John Wiley & Sons .

Moser, B. K ., and Stevens, G . R . (1992), "Homogeneity of Variance in the Two Sample Means Test," *The American Statistician*, 46, 19-21 .

Moser, B. K ., Watts, C . L ., and Stevens, G . R . (1989), "The Two Sample t Test Versus Satterthwaite's Approximate F Test," *Communications in Statistics - Theory and Methods*, 18 .

Welch, B . L . (1937), "The Significance of the Difference Between Two Means When the Population Variances are Unequal," *Biometrika*, 29, 350-362 .

Wilks, S . S . (1962), *Mathematical Statistics* , Second edition, New York : John Wiley & Sons .