Statistical Methodology 8 (2011) 528–534

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

A note on a two-sample T test with one variance unknown

Liqian Peng [a], Tiejun Tong [b,*]

[a] *Department of Physics, University of Colorado, Boulder, CO 80309, USA*

[b] *Department of Mathematics, Hong Kong Baptist University, Hong Kong*

ARTICLE INFO

Article history:
Received 28 January 2011
Received in revised form
26 April 2011
Accepted 2 July 2011

Keywords:
Behrens–Fisher
Bias correction
Welch–Satterthwaite approximation
Student's t distribution
Type I error

ABSTRACT

This note revisits Maity and Sherman's two-sample testing problem with one variance known but the other one unknown [A. Maity, M. Sherman, The two-sample t test with one variance unknown, The American Statistician 60 (2006) 163–166]. Inspired by the fact that the number of degrees of freedom used in their testing method is overestimated, we propose in this note a new testing method by introducing an unbiased estimator of the number of degrees of freedom. Simulation studies indicate that the proposed testing method provides a more accurate control than Maity and Sherman's method.

1. Introduction

The two-sample comparison is a frequently encountered problem in applied statistics and is introduced in most introductory statistics textbooks. One main purpose in a two-sample comparison is to make inferences about the means of the two populations. As a common practice, it is often assumed that both samples are independent and normally distributed. If not, one may perform a certain normalization procedure to the samples before the comparison.

Let $Y_{11}, \ldots, Y_{1n_1}$ be a random sample of size n_1 from Normal(μ_1, σ_1^2), and $Y_{21}, \ldots, Y_{2n_2}$ be a random sample of size n_2 from Normal(μ_2, σ_2^2). In this note we are interested in testing $H_0 : \mu_1 - \mu_2 = \mu_0$, for μ_0 a fixed difference of interest. For ease of notation, let $\bar{Y}_1 = \sum_{i=1}^{n_1} Y_{1i}/n_1$ and $\bar{Y}_2 = \sum_{i=1}^{n_2} Y_{2i}/n_2$ be the sample means, and $S_1^2 = \sum_{i=1}^{n_1}(Y_{1i} - \bar{Y}_1)^2/(n_1 - 1)$ and $S_2^2 = \sum_{i=1}^{n_2}(Y_{2i} - \bar{Y}_2)^2/(n_2 - 1)$ be the sample variances.

* Corresponding author.
E-mail addresses: tongt@hkbu.edu.hk, tiejun.tong@colorado.edu (T. Tong).

1572-3127/$ – see front matter
doi:10.1016/j.stamet.2011.07.001

When both σ_1^2 and σ_2^2 are known, the following z statistic can be used:

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2 - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Under H_0, Z follows a standard normal distribution and that makes the test very straightforward. However, if the two variances are unknown but equal, one can use the pooled t statistic

$$T_1 = \frac{\bar{Y}_1 - \bar{Y}_2 - \mu_0}{\sqrt{S_{\text{pool}}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $S_{\text{pool}}^2 = \{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2\}/(n_1 + n_2 - 2)$ is the pooled estimate of the common variance. It is known that under H_0, T_1 has an exact t distribution with $n_1 + n_2 - 2$ degrees of freedom. A more general situation arises when the two variances are unknown and unequal. This is known as the "Behrens–Fisher" problem [5,1]. In 1938, Welch suggested using

$$T_2 = \frac{\bar{Y}_1 - \bar{Y}_2 - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

and proposed an approximate t test with the estimated number of degrees of freedom

$$\text{d.f.} = \left\{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}\right\}^{-1} \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2.$$

It is of interest that the above Welch's approximate t test has been proposed by other researchers in different contexts [6,3,4].

In addition to the above situations, another interesting scenario arises when one variance is known but the other one unknown. This situation occurs, for instance, when a new drug is compared to a routinely used standard drug. Given the amount of historical data, the variance of the standard drug can be treated as known, while for the new drug, the variance is assumed to be unknown because of insufficient data. Additionally, it cannot be assumed that the two drugs have a common variance due to possible formulation differences. This situation was first studied by Maity and Sherman [2] who proposed a new test statistic for the comparison. Specifically, a method analogous to

Welch's t test was used to establish an approximate t distribution for the null hypothesis (see more detail in Section 2).

This note revisits Maity and Sherman's two-sample testing problem with one variance known but the other one unknown. In Section 2, we review Maity and Sherman's testing method. Inspired by the fact that the number of degrees of freedom used in their testing method is overestimated, we propose in Section 3 a new testing method by introducing an unbiased estimator of the number of degrees of freedom. We then conclude the note in Section 4 with a simulation study that verifies the superiority of the proposed method.

2. Maity and Sherman's testing method

Maity and Sherman [2] considered the situation where one variance is known but the other one unknown. Without loss of generality, we assume that the first variance, σ_1^2, is known. Maity and Sherman proposed the following test statistic:

$$T_3 = \frac{\bar{Y}_1 - \bar{Y}_2 - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

Note that T_3 does not follow an exact t distribution since the term

$$\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{-1}\left(\frac{\sigma_1^2}{n_1}+\frac{S_2^2}{n_2}\right) \tag{1}$$

is not chi-square distributed. Like Welch [7] and Satterthwaite [3], Maity and Sherman proposed an approximation to the exact distribution of (1) by using a chi-square distribution with γ degrees of freedom,

$$\gamma\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^{-1}\left(\frac{\sigma_1^2}{n_1}+\frac{S_2^2}{n_2}\right)\sim\chi_\gamma^2,\ \text{approximately},$$

where the notation $\sim$ means "follows the distribution of". The value of γ is obtained by matching the variances of both sides of the above equation. Specifically, it gives

$$\gamma=\left\{\frac{(\sigma_2^2/n_2)^2}{n_2-1}\right\}^{-1}\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^2. \tag{2}$$

Further, Maity and Sherman replaced the unknown σ_2^2 in (2) by its sample estimate S_2^2, which leads to

$$\hat{\gamma}=\left\{\frac{(S_2^2/n_2)^2}{n_2-1}\right\}^{-1}\left(\frac{\sigma_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2. \tag{3}$$

Throughout this note, we take the integer part of $\hat{\gamma}$ whenever necessary. Then under H_0, we have

$$T_3=\frac{\bar{Y}_1-\bar{Y}_2-\mu_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim t_{\hat{\gamma}},\ \text{approximately}.$$

Let $t_{\alpha,\nu}$ denote the upper αth quantile of the student t distribution with ν degrees of freedom. On the basis of the above approximate t distribution, the level-α tests conducted are given as follows.

Alternative hypothesis	Rejection criterion
$H_1:\mu_1-\mu_2\neq\mu_0$	$T_3>t_{\alpha/2,\hat{\gamma}}$ or $T_3<-t_{\alpha/2,\hat{\gamma}}$
$H_1:\mu_1-\mu_2>\mu_0$	$T_3>t_{\alpha,\hat{\gamma}}$
$H_1:\mu_1-\mu_2<\mu_0$	$T_3<-t_{\alpha,\hat{\gamma}}$

3. Unbiased estimation of the number of degrees of freedom

In this section, we first point out that the approximated number of degrees of freedom, $\hat{\gamma}$, in (3) is overestimated. By the fact that $(n_2-1)S_2^2/\sigma_2^2$ is chi-square distributed with n_2-1 degrees of freedom, we have

$$\frac{\sigma_2^2}{(n_2-1)S_2^2}\sim\text{Inv-}\chi^2_{n_2-1}, \tag{4}$$

where Inv-$\chi^2_{n_2-1}$ is the inverse-chi-square distribution with n_2-1 degrees of freedom. By (4), it is easy to see that for any $n_2>5$,

$$E\left(\frac{1}{S_2^2}\right)=\frac{n_2-1}{(n_2-3)\sigma_2^2}, \tag{5}$$

$$E\left(\frac{1}{(S_2^2)^2}\right)=\frac{(n_2-1)^2}{(n_2-3)(n_2-5)\sigma_2^4}. \tag{6}$$

Further, we have

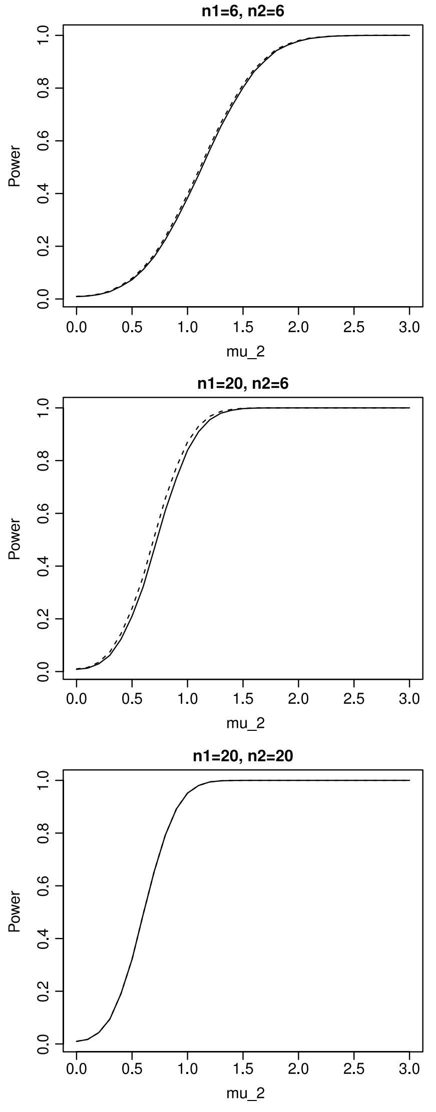

Fig. 1. Power functions for the two methods with $\sigma_2 = 1/3$, where the solid line corresponds to the new method and the dashed line corresponds to the method of Maity and Sherman.

$$
\begin{aligned}
E(\hat{\gamma}) &= E\left[\left\{\frac{(S_2^2/n_2)^2}{n_2-1}\right\}^{-1}\left(\frac{\sigma_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2\right] \\
&= (n_2-1)\left\{\frac{n_2^2(n_2-1)^2\sigma_1^4}{n_1^2(n_2-3)(n_2-5)\sigma_2^4}+\frac{2n_2(n_2-1)\sigma_1^2}{n_1(n_2-3)\sigma_2^2}+1\right\} \\
&> (n_2-1)\left\{\frac{n_2^2\sigma_1^4}{n_1^2\sigma_2^4}+\frac{2n_2\sigma_1^2}{n_1\sigma_2^2}+1\right\} \\
&= \gamma.
\end{aligned}
\tag{7}
$$

This indicates that the estimated number of degrees of freedom (3) is positively biased.

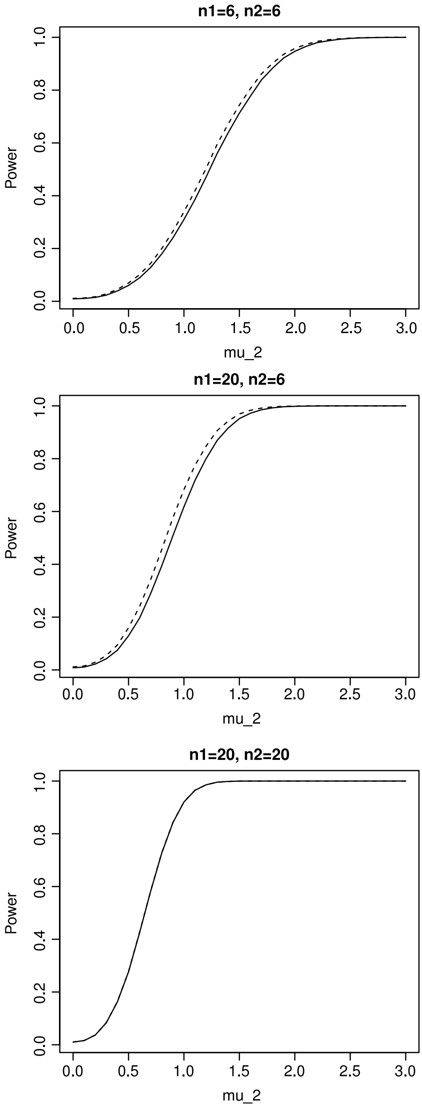

Fig. 2. Power functions for the two methods with $\sigma_2 = 1/2$, where the solid line corresponds to the new method and the dashed line corresponds to the method of Maity and Sherman.

Note that for any fixed significance level $\alpha < 0.5$, the upper αth quantile of the student t distribution, $t_{\alpha,\nu}$, is a decreasing function of the number of degrees of freedom ν. We conclude that an overestimated $\hat{\gamma}$ will lead to an underestimated threshold value $t_{\alpha,\hat{\gamma}}$, especially when n_2 is small. For instance, when $n_1 = n_2 = 6$ and $\sigma_1^2 = \sigma_2^2 = 1$, by (2) the true number of degrees of freedom is given as $\gamma = 20$. If we take $\alpha = 0.01$, then the theoretical threshold is $t_{0.01,20} = 2.528$, while for $\hat{\gamma}$, by (7) we have $E(\hat{\gamma}) = 190/3$. Thus, on average, the estimated threshold is $t_{0.01,63} = 2.387$ which is smaller than 2.528. As a consequence, the type I error of the conducted test may not be controlled.

Motivated by the above finding, we propose in this note an unbiased estimator of γ. Let

$$\tilde{\gamma} = (n_2 - 1)\left\{\frac{n_2^2(n_2-3)(n_2-5)\sigma_1^4}{n_1^2(n_2-1)^2}\frac{1}{(S_2^2)^2} + \frac{2n_2(n_2-3)\sigma_1^2}{n_1(n_2-1)}\frac{1}{S_2^2} + 1\right\}. \tag{8}$$

Table 1
Average type I errors of Maity and Sherman's method (M&S) and the new method for $n_1 = n_2 = 6$.

α		$\sigma_2 = 1/3$	$\sigma_2 = 1/2$	$\sigma_2 = 1$	$\sigma_2 = 2$	$\sigma_2 = 3$
0.001	M&S	0.0011	0.0010	0.0017	0.0031	0.0032
	New	0.0009	0.0008	0.0009	0.0017	0.0018
0.01	M&S	0.0100	0.0102	0.0121	0.0149	0.0149
	New	0.0091	0.0085	0.0086	0.0105	0.0113
0.05	M&S	0.0496	0.0498	0.0523	0.0560	0.0554
	New	0.0478	0.0460	0.0447	0.0483	0.0496

By (5) and (6), it is easy to verity that $\tilde{\gamma}$ is an unbiased estimator of γ. In addition, it can be shown that

$$\mathrm{Var}(\tilde{\gamma}) \le \mathrm{Var}(\hat{\gamma}).$$

This indicates that our proposed $\tilde{\gamma}$ has a smaller mean squared error than Maity and Sherman's estimator $\hat{\gamma}$. Or equivalently, the estimator $\hat{\gamma}$ is inadmissible under the commonly used quadratic loss function $L(\hat{\gamma}) = (\hat{\gamma} - \gamma)^2$.

Finally, with the proposed unbiased estimator $\tilde{\gamma}$, we conduct tests as follows.

Alternative hypothesis	Rejection criterion
$H_1 : \mu_1 - \mu_2 \neq \mu_0$	$T_3 > t_{\alpha/2,\tilde{\gamma}}$ or $T_3 < -t_{\alpha/2,\tilde{\gamma}}$
$H_1 : \mu_1 - \mu_2 > \mu_0$	$T_3 > t_{\alpha,\tilde{\gamma}}$
$H_1 : \mu_1 - \mu_2 < \mu_0$	$T_3 < -t_{\alpha,\tilde{\gamma}}$

We reiterate here that, to make this note short, we have assumed that $n_2 > 5$ for the proposed unbiased estimator $\tilde{\gamma}$. When the sample size is at most 5, to improve the performance of T_3 it might be necessary to adopt another remedy, e.g., estimating $1/\sigma_2^2$ by the median or the mode of the inverse-chi-square distribution.

4. Simulation studies

In this section, we conduct simulations to evaluate the performance of T_3 with the proposed $\tilde{\gamma}$ in (8). The first study is to compare the type I errors of the two methods and check how well they behave under the nominal level α. The second study is to compare their corresponding powers. Without loss of generality, we set $\mu_1 = \mu_2 = 0$ and $\sigma_1 = 1$. We consider three different combinations of (n_1, n_2): $(6, 6)$, $(20, 6)$ and $(20, 20)$.

To assess the type I errors under various settings, we consider five different values of the unknown variance, $\sigma_2 = 1/3, 1/2, 1, 2$ and 3, to represent different levels of discrepancy, apart from σ_1. Then for each σ_2 value, we simulate the data $Y_{11}, \ldots, Y_{1n_1}$ from Normal(μ_1, σ_1^2), and $Y_{21}, \ldots, Y_{2n_2}$ from Normal(μ_2, σ_2^2). Finally, to test the following hypothesis, we consider three different significance levels of α at 0.001, 0.01 or 0.05, respectively:

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

We repeat the above procedure 1000,000 times for each setting and report the average type I errors in Table 1 for $(n_1, n_2) = (6, 6)$, in Table 2 for $(n_1, n_2) = (20, 6)$, and in Table 3 for $(n_1, n_2) = (20, 20)$. As anticipated in Section 3, the simulated type I errors of Maity and Sherman [2] exceed the nominal level at α in most settings, especially when n_2 is small and/or when σ_2 is large. For the new method, the simulated type I errors are always close to or below the given nominal level. In addition, we observe that when the sample sizes are large, e.g., when $(n_1, n_2) = (20, 20)$, the two methods give a similar performance. Overall, it is evident that the proposed method with $\tilde{\gamma}$ provides a more accurate control than the testing method of Maity and Sherman.

For the power comparisons, we fix $\mu_1 = 0$ without loss of generality. We choose μ_2 to be non-zero, ranging from 0 to 3, to represent different levels of effect size. All other settings are the same as before. Recall that the method of Maity and Sherman is anti-conservative for large σ_2 values. To make the comparison meaningful, we report the power functions only for $\sigma_2 = 1/3$ in Fig. 1 and for

Table 2
Average type I errors of Maity and Sherman's method (M&S) and the new method for $n_1 = 20$ and $n_2 = 6$.

α		$\sigma_2 = 1/3$	$\sigma_2 = 1/2$	$\sigma_2 = 1$	$\sigma_2 = 2$	$\sigma_2 = 3$
0.001	M&S	0.0011	0.0015	0.0030	0.0031	0.0024
	New	0.0007	0.0009	0.0016	0.0019	0.0016
0.01	M&S	0.0101	0.0113	0.0148	0.0143	0.0128
	New	0.0081	0.0082	0.0104	0.0111	0.0109
0.05	M&S	0.0501	0.0515	0.0552	0.0547	0.0525
	New	0.0450	0.0443	0.0475	0.0500	0.0498

Table 3
Average type I errors of Maity and Sherman's method (M&S) and the new method for $n_1 = n_2 = 20$.

α		$\sigma_2 = 1/3$	$\sigma_2 = 1/2$	$\sigma_2 = 1$	$\sigma_2 = 2$	$\sigma_2 = 3$
0.001	M&S	0.0010	0.0011	0.0011	0.0011	0.0011
	New	0.0010	0.0010	0.0010	0.0011	0.0011
0.01	M&S	0.0099	0.0100	0.0100	0.0104	0.0102
	New	0.0099	0.0099	0.0098	0.0101	0.0100
0.05	M&S	0.0502	0.0500	0.0500	0.0506	0.0502
	New	0.0502	0.0499	0.0495	0.0501	0.0499

$\sigma_2 = 1/2$ in Fig. 2. In both scenarios, the method of Maity and Sherman provides a slightly larger power than the new method. This is the price that we pay for having a more accurate control of the type I error.

References

[1] S. Kim, A.S. Cohen, On the Behrens–Fisher problem: a review, Journal of Educational and Behavioral Statistics 23 (1998) 356–377.
[2] A. Maity, M. Sherman, The two-sample t test with one variance unknown, The American Statistician 60 (2006) 163–166.
[3] F.E. Satterthwaite, Synthesis of variance, Psychometrika 6 (1941) 309–316.
[4] F.E. Satterthwaite, An approximate distribution of estimates of variance components, Biometric Bulletin 6 (1946) 110–114.
[5] H. Scheffé, Practical solutions of the Behrens–Fisher problem, Journal of the American Statistical Association 65 (1970) 1501–1508.
[6] H.F. Smith, The problem of comparing the results of two experiments with unequal errors, Journal of the Council for Scientific and Industrial Research 9 (1936) 211–212.
[7] B.L. Welch, The significance of the difference between two means when the population variances are unequal, Biometrika 29 (1938) 350–362.