Malaysian Journal of Science **26** (SKSM14 Special Issue): 105 - 109 (2007)

Testing the Equality of the Means of Two Populations with Unequal Variances

Pan Wei Yeing[1] and **Pooi Ah Hin**[2]

Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia.
[1]wypan@.um.edu.my, [2]ahpooi@um.edu.my (corresponding author)

ABSTRACT The Behrens–Fisher problem is the problem of comparing the means when the assumption of normality for the distributions is made and the variances are different. A widely used solution to this problem is Welch's test. Presently an alternative test is proposed. The alternative test is found to be comparable to Welch's test in terms of Type-I and Type-II errors.

ABSTRAK Masalah Behrens-Fisher merupakan masalah yang berkaitan dengan perbandingan min apabila taburan adalah normal dan varian berbeza. Penyelesaian yang digunakan secara meluas bagi masalah ini ialah ujian Welch. Kini, satu ujian alternatif dicadangkan. Didapati bahawa ujian alternatif ini hampir sama dengan ujian Welch kerana kedua-dua ujian mempunyai ralat Jenis-I dan Jenis –II yang berhampiran.

(Hypothesis testing, normality, unequal variances, Type-I and Type-II errors)

INTRODUCTION

Suppose we have independent random samples $x_1, x_2, ..., x_m$ from the $N(\mu_x, \sigma_x^2)$ distribution and $y_1, y_2, ..., y_n$ from the $N(\mu_y, \sigma_y^2)$ distribution. The problem of testing the null hypothesis $H_0 : \mu_x - \mu_y = 0$ in the presence of the nuisance parameters $\sigma_x{}^2$ and $\sigma_y{}^2$ is called the Behrens-Fisher problem.

An early solution to this problem was proposed by Behrens [1], and this solution was endorsed by Fisher [2]. A disadvantage of this solution is that the Type-I error is often less than the desired size.

A widely used solution is called Welch's approximate solution [3]. Welch's approximate solution makes use of the statistic

$$V = \frac{(\bar{x} - \bar{y})}{\left[(s_x^2 / m) + (s_y^2 / n)\right]^{1/2}} \qquad (1)$$

where $\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$, $\bar{y} = \frac{\sum_{j=1}^{n} y_j}{n}$,

$$s_x^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1}, \text{ and } s_y^2 = \frac{\sum_{j=1}^{n}(y_j - \bar{y})^2}{n-1}$$

The distribution of V can be approximated by a t_f distribution in which the data dependent degrees of freedom are

$$f = \frac{\left((s_x^2 / m) + (s_y^2 / n)\right)^2}{\{[s_x^4 /(m^3 - m^2)] + [s_y^4 /(n^3 - n^2)]\}} \qquad (2)$$

[3]. We call this test, using V and the t distribution with fractional degrees of freedom, the V-test. Wang's work showed that the actual significant level of the V-test is very close to the nominal significance level [4]. Best and Rayner, on the other hand, showed that the power of the V-test is similar to those of the likelihood ratio test and score test [5]. Efficient experimental design may be constructed for inference in the Behrens-Fisher problem if the experimenter is

able to specify a specific region for the ratio σ_y^2/σ_x^2 of the population variances [6]. Presently an alternative solution of the Behrens – Fisher problem is considered in this paper. The performance of Welch's test and that of the alternative test are found to be comparable.

AN ALTERNATIVE SOLUTION OF THE BEHRENS–FISHER PROBLEM

First let $T = T(\bar{x}, \bar{y}) = (\bar{x} - \bar{y})^2$. It can be shown that under $H_0 : \mu_x - \mu_y = 0$, the first four moments of T are given by

$$E[T] = \left(\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}\right), \qquad (3)$$

$$E_2 = E[T - E(T)]^2 = 2\left(\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}\right)^2, \qquad (4)$$

$$E_3 = E[T - E(T)]^3 = 8\left(\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}\right)^3, \qquad (5)$$

and

$$E_4 = E[T - E(T)]^4 = 60\left(\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}\right)^4. \qquad (6)$$

The values of σ_x^2 and σ_y^2 are unknown. But based on the data $\underset{\sim}{x}$ and $\underset{\sim}{y}$, we can estimate the unknown variances by using $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ given by

$$\hat{\sigma}_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2 \qquad (7)$$

and

$$\hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2. \qquad (8)$$

Let $\hat{E}_k$ be the value of E_k when σ_x^2 and σ_y^2 in the right sides of equations (3) - (6) are replaced by $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$, respectively.

Let λ_1, λ_2 and λ_3 be constants and $E \sim \mathrm{N}(0,1)$. In [7] the random variable ε given by the nonlinear function

$$\varepsilon = -\lambda_2\frac{1+\lambda_3}{2} + \left[\lambda_1 + \frac{2}{7}\lambda_2 - \frac{2}{7}\lambda_2\lambda_3\right]E + \left[\frac{\lambda_2}{2} + \frac{\lambda_2\lambda_3}{2}\right]E^2 + \left[\frac{41}{180}\lambda_2 - \frac{41}{180}\lambda_2\lambda_3\right]E^3 + \left[-\frac{1}{72}\lambda_2 + \frac{1}{72}\lambda_2\lambda_3\right]E^5 + \left[\frac{1}{2520}\lambda_2 - \frac{1}{2520}\lambda_2\lambda_3\right]E^7 \qquad (9)$$

was introduced. When $\lambda_2 = 0$, the random variable will have the N(0, λ_1^2) distribution. When $\lambda_2 \neq 0$ and $\lambda_3 = -1$, the random variable will have a symmetrical non-normal distribution. When $\lambda_2 \neq 0$ and $\lambda_3 \neq -1$, the random variable will have a skewed distribution. The random variable ε is said to have a quadratic-normal distribution with parameters 0 and $\underset{\sim}{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^T$ $\left(\varepsilon \sim \mathrm{QN}\left(0, \underset{\sim}{\lambda}\right)\right)$.

Let $m_k = \mathrm{E}(\varepsilon^k)$ be the k^{th} moment of ε. We find the value $\underset{\sim}{\lambda} = \underset{\sim}{\hat{\lambda}}$ such that

$$\hat{\mathrm{E}}_k = m_k, \quad k = 2,3,4. \qquad (10)$$

Then the estimate of the (1-α) quantile of T when $\mu_x = \mu_y$, $\sigma_x^2 = \hat{\sigma}_x^2$ and $\sigma_y^2 = \hat{\sigma}_y^2$ is

$$q(\underset{\sim}{x}, \underset{\sim}{y}) = \left(\frac{\hat{\sigma}_x^2}{m} + \frac{\hat{\sigma}_y^2}{n}\right) + \hat{\lambda}_1(Z_\alpha) + \hat{\lambda}_2\left(Z_\alpha^2 - \frac{1+\hat{\lambda}_3}{2}\right)$$

where Z_α is the (1-α) quantile of the standard normal distribution.

Now we define $\underset{\sim}{\tilde{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m)^T$ and $\underset{\sim}{\tilde{y}} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n)^T$ to be independent random samples from the $N(\mu_x, \hat{\sigma}_x^2)$ and $N(\mu_y, \hat{\sigma}_y^2)$ distributions, respectively. Furthermore, let $\bar{\tilde{x}} = \frac{1}{m}\sum_{i=1}^{m}\tilde{x}_i$, $\bar{\tilde{y}} = \frac{1}{n}\sum_{j=1}^{n}\tilde{y}_j$ and $\tilde{T} = (\bar{\tilde{x}} - \bar{\tilde{y}})^2$.

Malaysian Journal of Science **26** (SKSM14 Special Issue): 105 - 109 (2007)

The multiplication factor c = $c(\underset{\sim}{x},\underset{\sim}{y})$ is defined to be a constant such that

$$P\left\{T(\bar{\tilde{x}},\bar{\tilde{y}})\le c\left(\underset{\sim}{x},\underset{\sim}{y}\right)q\left(\underset{\sim}{\tilde{x}},\underset{\sim}{\tilde{y}}\right) \mid \mu_x=\mu_y, \sigma_x^2=\hat{\sigma}_x^2, \sigma_y^2=\hat{\sigma}_y^2\right\}=0.95 \qquad (11)$$

[**Note** : In equation (11), $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are treated as constants.] To find the multiplication factor $c(\underset{\sim}{x},\underset{\sim}{y})$, we perform the following simulation:

We generate M sets of values of $(\underset{\sim}{\tilde{x}},\underset{\sim}{\tilde{y}})$ using the equations

$$\tilde{x}_i \sim N(\hat{\mu},\hat{\sigma}_x^2), \qquad i=1,2,\ldots,m;$$

$$\tilde{y}_j \sim N(\hat{\mu},\hat{\sigma}_y^2), \qquad j=1,2,\ldots,n.$$

For each generated value of $(\underset{\sim}{\tilde{x}},\underset{\sim}{\tilde{y}})$ we find

$$\tilde{\sigma}_x^2=\frac{1}{m-1}\sum_{i=1}^{m}\left(\tilde{x}_i-\bar{\tilde{x}}\right)^2,$$

$$\tilde{\sigma}_y^2=\frac{1}{n-1}\sum_{j=1}^{n}\left(\tilde{y}_j-\bar{\tilde{y}}\right)^2,$$

$$\tilde{E}_2=2\left(\frac{\tilde{\sigma}_X^2}{m}+\frac{\tilde{\sigma}_Y^2}{n}\right)^2, \quad \tilde{E}_3=8\left(\frac{\tilde{\sigma}_X^2}{m}+\frac{\tilde{\sigma}_Y^2}{n}\right)^3,$$

$$\tilde{E}_4=60\left(\frac{\tilde{\sigma}_X^2}{m}+\frac{\tilde{\sigma}_Y^2}{n}\right)^4.$$

Let $\underset{\sim}{\tilde{\lambda}}$ be the value of $\underset{\sim}{\lambda}$ such that

$$\tilde{E}_k=m_k, \qquad k=2,3,4.$$

Furthermore, let $c^*=1$. We then find the proportion of $(\underset{\sim}{\tilde{x}},\underset{\sim}{\tilde{y}})$ (out of M values of $(\underset{\sim}{\tilde{x}},\underset{\sim}{\tilde{y}})$) for which

$$T(\bar{\tilde{x}},\bar{\tilde{y}})\le c^* q\left(\underset{\sim}{\tilde{x}},\underset{\sim}{\tilde{y}}\right).$$

If this proportion is approximately equal to 0.95, then c is chosen to be equal to c^*. If not, the value of c^* is varied until the proportion is approximately equal to 0.95.

When m=12 and n=10, some values of $c(\underset{\sim}{x},\underset{\sim}{y})$ are shown in Table 1.

Table 1. The values of the multiplication factor when m = 12 and n = 10

$\hat{\sigma}_x$	$\hat{\sigma}_y$	$c(\underset{\sim}{x},\underset{\sim}{y})$
0.05	0.30	1.20
0.10	0.30	1.20
0.15	0.30	1.30
0.20	0.30	1.25
0.25	0.30	1.00
0.30	0.30	1.10
0.35	0.30	1.25

$\hat{\sigma}_x$	$\hat{\sigma}_y$	$c(\underset{\sim}{x},\underset{\sim}{y})$
0.40	0.30	1.20
0.45	0.30	1.27
0.50	0.30	1.13
0.55	0.30	1.15
0.60	0.30	1.20
0.65	0.30	1.20
0.70	0.30	1.20

$\hat{\sigma}_x$	$\hat{\sigma}_y$	$c(\underset{\sim}{x},\underset{\sim}{y})$
0.75	0.30	1.25
0.80	0.30	1.20
0.85	0.30	1.30
0.90	0.30	1.25
0.95	0.30	1.15
1.00	0.30	1.30

$\hat{\sigma}_x$	$\hat{\sigma}_y$	$c(\underset{\sim}{x},\underset{\sim}{y})$
0.05	0.60	1.40
0.10	0.60	1.25
0.15	0.60	1.30
0.20	0.60	1.40
0.25	0.60	1.25
0.30	0.60	1.35
0.35	0.60	1.17

$\hat{\sigma}_x$	$\hat{\sigma}_y$	$c(\underset{\sim}{x},\underset{\sim}{y})$
0.40	0.60	1.35
0.45	0.60	1.23
0.50	0.60	1.20
0.55	0.60	1.35
0.60	0.60	1.20
0.65	0.60	1.10
0.70	0.60	1.35

$\hat{\sigma}_x$	$\hat{\sigma}_y$	$c(\underset{\sim}{x},\underset{\sim}{y})$
0.75	0.60	1.20
0.80	0.60	1.20
0.85	0.60	1.30
0.90	0.60	1.20
0.95	0.60	1.05
1.00	0.60	1.20

We next generate N sets of value of $(\underset{\sim}{x},\underset{\sim}{y})$ using the equations

$$x_i \sim N(\mu_x, \sigma_x^2), \quad i = 1,2,\ldots,m;$$
$$y_j \sim N(\mu_y, \sigma_y^2), \quad j = 1,2,\ldots,n$$

For each generated value of $(\underset{\sim}{x},\underset{\sim}{y})$, we find $\hat{\mu} = \dfrac{\sum_{i=1}^{m} x_i + \sum_{j=1}^{n} y_j}{m+n}, \hat{\sigma}_x^2, \hat{\sigma}_y^2, \hat{E}_2, \hat{E}_3, \hat{E}_4, \underset{\sim}{\hat{\lambda}}$ and $q(\underset{\sim}{x},\underset{\sim}{y})$.

We approximate the probability $P^{(a)}(\sigma_x, \sigma_y)$ of the acceptance region of the alternative test by the proportion of $(\underset{\sim}{x},\underset{\sim}{y})$ (out of N values of $(\underset{\sim}{x},\underset{\sim}{y})$) for which

$$T(\bar{x},\bar{y}) \le c\left(\underset{\sim}{x},\underset{\sim}{y}\right)q\left(\underset{\sim}{x},\underset{\sim}{y}\right).$$

The approximate value of the probability $P^{(v)}(\sigma_x, \sigma_y)$ of the acceptance region of the *V*-test can be found in a similar way.

When m = 12, n = 10, we can use simulation based on N = M = 10,000 to estimate the probability of the acceptance region of the *V*-test and that of the alternative test. Some of the results are shown in Table 2.2. These results show that the probabilities of the acceptance regions for the *V*-test and the alternative test are both quite close to the target value 0.95.

COMPARISON OF THE POWERS OF THE *V*-TEST AND THE ALTERNATIVE TEST

As both the *V*-test and the alternative test have approximately the same size, it is meaningful to compare the powers of the two tests on the fair basis of equal sizes of the Type-I error.

The power function $\pi(\mu_x, \mu_y, \sigma_x, \sigma_y)$ of a test is defined to be the probability of the rejection region of the test when the values of the parameters are μ_x, μ_y, σ_x and σ_y, respectively. The power functions of the two tests can be written respectively as

$$\pi^{(v)}(\mu_x, \mu_y, \sigma_x, \sigma_y) = P\left(|V| > t_{\alpha/2} \middle| \mu_x, \mu_y, \sigma_x, \sigma_y\right) \qquad (12)$$

$$\pi^{(a)}(\mu_x, \mu_y, \sigma_x, \sigma_y) = P\left(T(\bar{X},\bar{Y}) > c(\underset{\sim}{x},\underset{\sim}{y})q(\underset{\sim}{x},\underset{\sim}{y}) \middle| \mu_x, \mu_y, \sigma_x, \sigma_y\right) \qquad (13)$$

where $t_{\alpha/2}$ is the 1-(α/2) quantile of the t-distribution with f (see equation (2)) degrees of freedom.

When the values of μ_x, μ_y, σ_x and σ_y are given, it is possible to find the approximate values of $\pi^{(v)}(\mu_x, \mu_y, \sigma_x, \sigma_y)$ and $\pi^{(a)}(\mu_x, \mu_y, \sigma_x, \sigma_y)$ by using simulation. When m = 12, n = 10, the simulated results based on M = N = 10,000 are shown in Table 3.

Table 2. The probability of the acceptance region when m = 12 and n = 10

σ_x	σ_y	$P^{(v)}(\sigma_x,\sigma_y)$	$P^{(a)}(\sigma_x,\sigma_y)$
0.05	0.60	0.9496	0.9481
0.15	0.60	0.9531	0.9527
0.25	0.60	0.9471	0.9443
0.35	0.60	0.9468	0.9460
0.45	0.60	0.9503	0.9493

σ_x	σ_y	$P^{(v)}(\sigma_x,\sigma_y)$	$P^{(a)}(\sigma_x,\sigma_y)$
0.55	0.60	0.9519	0.9512
0.65	0.60	0.9564	0.9564
0.75	0.60	0.9554	0.9555
0.85	0.60	0.9501	0.9518
0.95	0.60	0.9543	0.9543

Table 3. Powers of the tests when m = 12, n = 10, μ_x = 0.0 and nominal value of α is 0.05

σ_x	σ_y	μ_y	$\pi^{(v)}$	$\pi^{(a)}$
0.1	0.3	-0.10	0.1522	0.1691
0.2	0.3	-0.10	0.1329	0.1357
0.3	0.3	-0.10	0.1152	0.1205
0.4	0.3	-0.10	0.0911	0.0901
0.5	0.3	-0.10	0.0831	0.0826

σ_x	σ_y	μ_y	$\pi^{(v)}$	$\pi^{(a)}$
0.6	0.3	-0.10	0.0739	0.0733
0.7	0.3	-0.10	0.0691	0.0683
0.8	0.3	-0.10	0.0643	0.0624
0.9	0.3	-0.10	0.0635	0.0614
1.0	0.3	-0.10	0.0610	0.0596

Table 3.1 shows that the powers of the two tests are not very much different.

CONCLUDING REMARKS

The present paper deals with the case when there are two independent normal distributions $N(\mu_x,\sigma_x^2)$ and $N(\mu_y,\sigma_y^2)$. A more general version of the problem is the case when there are a $(a \geq 3)$ independent normal distributions $N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$, ..., $N(\mu_a,\sigma_a^2)$. The ideas used in the present paper can also be applied to devise a test for testing the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_a$ [8].

The even more general situation is the case when we have a $(a \geq 3)$ quadratic-normal distributions $\mathrm{QN}\left(\mu_1, \underset{\sim}{\lambda}^{(1)}\right)$, $\mathrm{QN}\left(\mu_2, \underset{\sim}{\lambda}^{(2)}\right)$, ..., $\mathrm{QN}\left(\mu_a, \underset{\sim}{\lambda}^{(a)}\right)$. The ideas contained in this paper can again be used to find a test for the hypothesis that the a means $\mu_1, \mu_2, \ldots, \mu_a$ are equal [9].

REFERENCES

1. Behrens, W. V. (1929). Ein Beitrag Zur Fehlerberechung Beiwenigen Beobachtungen. *Landwirtsch* **68**: 807-837.
2. Fisher, R. A. (1939). The comparison of samples with possibly unequal variances. *Annals of Eugenics* **9**: 174-180.
3. Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**: 350-362.
4. Wang, Y.Y. (1971). Probabilities of the type-I error of the Welch tests for the Behrens-Fisher problem. *Journal of the American Statistical Association* **66**: 605-608.
5. D.J. Best and J.C.W. Rayner (1987). Welch's approximate solution for the Behrens-Fisher problem. *Technometrics* **29**: 205-210.
6. Dette, Holger and O'Brien, Timothy E. (2004). Efficient experimental design for the Behrens-Fisher problem with application to bioassay. *American Statistician* **58** (2): 138-143.
7. Pooi, A.H. (2003). Effects of non-normality on confidence intervals in linear model. *Technical Report No. 6/2003, Institute of Mathematical Sciences, University of Malaya.*
8. Chew, S.M. and Pooi, A.H. (2007). Testing the equality of the means of three or more normal distributions. *Technical Report No. 2/2007, Institute of Mathematical Sciences, University of Malaya.*
9. Chew, S.M. and Pooi, A.H. (2007). Testing the equality of the means of three or more quadratic-normal distributions. *Technical Report No. 3/2007, Institute of Mathematical Sciences, University of Malaya.*