COMMUN. STATIST.—SIMULA., 26(1), 333-353 (1997)

# ROBUST TESTS FOR EQUALITY OF TWO POPULATION MEANS UNDER THE NORMAL MODEL

Srabashi Basu

Stat—Math Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035 India

Sahadeb Sarkar

Department of Statistics, Oklahoma State University, Stillwater, OK 74078 USA

Ayanendranath Basu

Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035 India

*Key Words and Phrases: Asymptotic full efficiency; Hellinger distance; minimum disparity estimation; weighted likelihood estimation; robust tests.*

## ABSTRACT

A robust procedure is developed for testing the equality of means in the two sample normal model. This is based on the weighted likelihood estimators of Basu et al. (1993). When the normal model is true the tests proposed have the same asymptotic power as the two sample Student's t–statistic in the equal variance case. However, when the normality assumptions are only approximately true the proposed tests can be substantially more powerful than the classical tests. In a Monte Carlo study for the equal variance case under various outlier models the proposed test using Hellinger distance based weighted likelihood estimator compared favorably with the classical test as well as the robust test proposed by Tiku (1980).

## 1. INTRODUCTION

Let $X_{i1}, \ldots, X_{i,n_i}$, i= 1, 2, be independent random samples from populations with mean $\mu_i$ and variance $\sigma_i^2$. Testing equality of $\mu_1$ and $\mu_2$ is a common statistical problem. Under the assumption of normality and equal (and unknown) variances one can use the classical Student's two sample t test. But, in practice, the population distributions may not be exactly normal. Usually, the Student's t test is sensitive to the departures from normality. As stated by Tiku et al. (1986, p. 120): the Student's "t test is asymptotically robust, and for finite $n_1$ and $n_2$ it possesses type I error robustness if $n_1 = n_2$ or if the distribution is symmetric; if $n_1 \neq n_2$ and the distribution is skew, the effects of departure from normality may be considerable". Tiku (1980) proposed a robust test statistic based on modified maximum likelihood (MML) estimators of the parameters. Tiku's test statistic (Tiku et al. 1986, equation (4.3.1)) has considerably higher power than the Student's t test under nonnormal distributions while having only slightly lower power under normal distributions. Tiku et al. (1986, pp. 122–130) discuss how this test compares to other robust tests available in the literature.

If the underlying population distributions are normal with unequal (and unknown) variances, the Welch's t–statistic (Welch 1937) is usually used. This test statistic is nonrobust under most nonnormal distributions (Yuen 1974, Table 1). Tiku and Singh (1982, equation (7)) proposed a test statistic analogous to Welch's statistic based on the MML estimators of the parameters $\mu_1$, $\mu_2$, $\sigma_1$ and $\sigma_2$. The test of Tiku and Singh has generally higher power than the Welch's t test under nonnormal distributions with a little loss in asymptotic power (compared to Welch's t) under normal distributions. For details on this and a literature review on robust tests see Section 4.5 of Tiku et al. (1986).

In this paper we present robust alternatives to the two sample t tests based on minimum disparity estimators (MDEs) of Basu and Lindsay (1994) and weighted likelihood estimators (WLEs) of Basu et al. (1993). While the MDEs are asymptotically fully efficient at the model, their efficiency depends on the choice of transparent kernels (see Section 2.1 below) which may not be available for a nonnormal model. The WLEs, on the other hand, have efficiency and robustness properties similar to those of the MDEs but their efficiency does not depend on the availability of a transparent kernel. In addition, the WLEs are considerably simpler to compute than the MDEs, Thus, in practice, one can use test statistics based on the WLEs rather than MDEs. Under normal distributions the tests of Tiku (1980) are asymptotically slightly less powerful than the Student's t statistic (Tiku et al. 1986, p. 122) but our proposed test statistics are asymptotically as powerful as the Student's t statistic (Lemma 1 below). For the equal variance case we have conducted a Monte Carlo study for different sample sizes under various *outlier models*, i.e., where the data are generated by a mixture distribution with a large proportion of the normal distribution and a small proportion of a heavier tailed distribution. The simulation results show that the empirical powers of our proposed test using the Hellinger distance based WLE compare favorably with those of Tiku's test (1980). Investigation of the robust alternatives in the unequal variances case is an ongoing project and will be reported in a sequel paper.

The remainder of this paper is organized as follows. In Section 2 we give a brief review of the MDEs and WLEs. In Section 3 we define robust t statistics based on MDEs and WLEs and discuss their properties. Monte Carlo results are presented in Section 4. Concluding remarks are given in Section 5.

## 2. MINIMUM DISPARITY AND WEIGHTED LIKELIHOOD ESTIMATION

### 2.1. Minimum disparity estimation

Minimum disparity estimation (Lindsay 1994, Basu and Lindsay 1994) is an efficient and robust estimation method in parametric models. Let $m_\beta(x)$ represent the density of a continuous parametric family of models, indexed by an unknown parameter vector $\beta$. Given a sample of n independent and identically distributed observations $X_1,\ldots,X_n$, construct a nonparametric kernel density estimator $f^*(x)$, as $f^*(x) = \int k(x;y,h)dF_n(y)$, where $F_n$ is the empirical distribution function and k is the kernel function with mean y and standard deviation h. Let $M_\beta$ be the cumulative distribution function of the model. Next applying the same smoothness to the model, we get $m^*_\beta(x) = \int k(x;y,h)dM_\beta(y)$. Now we can construct a density based distance between $f^*(x)$ and $m^*_\beta(x)$, like the squared Hellinger distance

$$HD = \int[(f^*(x))^{1/2} - (m^*_\beta(x))^{1/2}]^2 dx, \tag{2.1}$$

which may be minimized to obtain the minimum Hellinger distance estimator.

Define the *Pearson residual*, a standardized version of the residual as $\delta^*(x) = [f^*(x) - m^*_\beta(x)]/m^*_\beta(x)$. For an arbitrary real valued thrice differentiable convex function G with $G(0) = 0$, define a disparity measure $\rho$ between $f^*(x)$ and $m^*_\beta(x)$ as

$$\rho(f^*, m^*_\beta) = \int G(\delta^*(x))m^*_\beta(x)dx. \tag{2.2}$$

The value of $\beta$ which minimizes the above disparity is called the MDE of $\beta$. If G is strictly convex the MDE is Fisher consistent. The obvious analog of the maximum likelihood estimator in this case is the "MLE*", which is the value of $\beta$ that minimizes the likelihood disparity

$$LD = \int f^*(x) \ln[f^*(x)/m^*_\beta(x)]dx. \tag{2.3}$$

Both (2.1) and (2.3) can be written in the form of (2.2).

The approach of smoothing the model (i.e., using $m^*_\beta$ in place of $m_\beta$) has several advantages over the conventional methods which do not smooth the model before the disparity is constructed. It does not require consistency or rate of convergence results for the nonparametric density estimators. Also, the MDEs are consistent and asymptotically normally distributed for any *fixed* bandwidth h. In some situations a kernel can be chosen so that the MLE* equals the ordinary maximum likelihood estimator regardless of the bandwidth, such kernels being known as the *transparent* kernels. Therefore, under transparent kernels all other MDEs are asymptotically equivalent to the MLE* and are first order efficient. See Basu and Lindsay (1994) for more details on transparent kernels.

Under differentiability of the model, minimization of the disparity measure $\rho$ with respect to $\beta$ corresponds to solving a set of estimating equations of the form:

$$-\frac{\partial\rho}{\partial\beta} = \int A(\delta^*(x))\,\frac{\partial m^*_\beta(x)}{\partial\beta}\,dx = 0. \tag{2.4}$$

The function $A(\delta) = (1+\delta)[\frac{d}{d\delta}G(\delta)] - G(\delta)$, $\delta \in \mathbb{R}$, is unique to the particular disparity measure and can be centered and rescaled so that it takes the value zero at $\delta = 0$ and its derivative at $\delta = 0$ is one. $A(\delta)$ is called the residual adjustment function of the disparity. For the squared Hellinger distance

$$G(\delta) = [(\delta+1)^{1/2} - 1]^2,\;\; A(\delta) = 2[(\delta+1)^{1/2} - 1].$$

For robust disparities like the Hellinger distance the residual adjustment function can severely downweight observations with large Pearson residuals. In this respect it is almost exactly like the $\psi$ function of the M–estimation approach. For the likelihood disparity $A(\delta) = \delta$ and the estimating equation has the form

$$\int \frac{f^*(x)}{m^*_\beta(x)}\frac{\partial}{\partial\beta}\, m^*_\beta(x)\, dx = \int \left[\frac{f^*(x)}{m^*_\beta(x)} - 1\right]\frac{\partial}{\partial\beta} m^*_\beta(x)\, dx = 0 \tag{2.5}$$

Computation of the MDE is done using numerical evaluation of integrals and iterative techniques. We next describe a modification of the above estimation method which is computationally much simpler and which does not sacrifice efficiency or robustness properties. For more details on this see Basu et al. (1993).

### 2.2. Weighted likelihood estimation

Note that the estimating equation (2.4) can be expressed as

$$\int\left[\frac{A(\delta^*(x))+1}{\delta^*(x)+1}\right](\delta^*(x)+1)\left[\frac{\partial}{\partial\beta}\, m^*_\beta(x)\right] dx = 0$$

i.e.,

$$\int w(\delta^*(x)) \left[ \frac{\frac{\partial}{\partial \beta} m^*_\beta(x)}{m^*_\beta(x)} \right] dF^*(x) = 0, \tag{2.6}$$

where $w(\delta^*(x)) = [A(\delta^*(x))+1]/[\delta^*(x)+1]$ and $F^*$ is the distribution function corresponding to $f^*$. Equation (2.6) is a weighted version of the estimating equations (2.5) of the MLE* with weights $w(\delta^*(x))$. By replacing the smoothed model density $m^*_\beta$ and the smoothed empirical distribution $F^*$ in equation (2.6) with their corresponding unsmoothed versions $m_\beta$ and $F_n$ respectively (except in the weight part), one can define the following estimating equation for $\beta$

$$\int w(\delta^*(x)) \left[ \frac{\frac{\partial}{\partial \beta} m_\beta(x)}{m_\beta(x)} \right] dF_n(x) = 0$$

i.e.,

$$n^{-1} \sum_{j=1}^{n} w(\delta^*(X_j))\, u(X_j, \beta) = 0 \tag{2.7}$$

where $u(x,\beta) = \frac{\partial}{\partial \beta} \ln m_\beta(x)$ is the usual maximum likelihood score function and $F_n$ is the empirical distribution function for the data $X_1,\ldots,X_n$. Note that the kernel smoothing appears in equation (2.7) only through the weight part $w(\delta^*(X_j))$ and not through the score part $u(X_j,\beta)$ of the equation. As in the case of iteratively reweighted least squares estimation method (Beaton and Tukey 1974; Holland and Welsch 1977), the WLE of $\beta$ is obtained by solving equation (2.7) iteratively. For a given set of initial estimates of $\beta$ one can construct weights $w(\delta^*(X_j))$, then solve equation (2.7) for an improved estimate of $\beta$; this is then used to construct new weights and this iterative procedure is continued until a suitable convergence criterion is met.

From the results of Basu et al. (1993) it follows that the WLEs are asymptotically fully efficient at the model. Under the model as the sample size increases the weights $w(\delta^*(X_i))$ tend to one and the equation (2.7) behaves like the maximum likelihood score equation for all the disparities. In addition, WLEs generated by disparities like the Hellinger distance have good robustness properties since the weight $w(\delta^*(X_i)) = [A(\delta^*(X_i)) + 1]/(\delta^*(X_i) + 1)$ will be significantly downweighted from 1 for an observation $X_i$ with a large Pearson residual. Unlike the MDEs, the WLEs do not require a transparent kernel to achieve full asymptotic efficiency. Note that the estimating equation (2.7) of the WLE is a sum over the data points $X_1,\dots,X_n$ rather than an integral over the entire support of $m_\beta$ as in the estimating equation (2.6) for the MDE. Consequently, the evaluation of the WLE requires no numerical integration and becomes computationally much simpler.

## 3. ROBUST TESTS

For a random sample $X_{i1}, \dots, X_{i,n_i}$ from $N(\mu_i,\sigma^2)$, i=1,2, let $\hat{F}_i$ be the corresponding empirical distribution function. We are interested in testing the null hypothesis

$$H_0: \mu_1 = \mu_2. \tag{3.1}$$

Basu and Lindsay (1994) showed that for the univariate normal problem a transparent kernel for the model is given by $k(x,y,h) = N(y, h^2)$ density. Let $m_{\beta_i}(x)$ represent the $N(\mu_i,\sigma^2)$ density, and $\beta_i = (\mu_i,\sigma^2)'$, i=1,2. Let $f_i^*$ be the kernel density estimates for the i–th sample and let $m^*_{\beta_i}$ be the corresponding kernel smoothed version of the model density. For a given disparity measure $\rho$, let $\rho(f_i^*,m^*_{\beta_i})$ be the disparity constructed over the i–th sample. Then, the minimum disparity estimation procedure of the one sample case can be extended in a straightforward manner to the two sample case with the overall disparity

$$\rho_O = (n_1+n_2)^{-1}[n_1\rho(f_1^*,m^*_{\beta_1}) + n_2\rho(f_2^*,m^*_{\beta_2})] \tag{3.2}$$

playing the role of the disparity to be minimized. For the specific case of the Hellinger distance, (3.2) is the form of overall disparity that was considered by Simpson (1989) for the two sample problem. Sarkar and Basu (1995) also used this form of the overall disparity to construct tests of hypothesis involving multiple discrete populations. In the two sample normal case the MLE*s (i.e., the minimizers of the above overall disparity when $\rho$ is the likelihood disparity) of $\mu_1$, $\mu_2$, $\sigma^2$ are again equal to the corresponding ordinary maximum likelihood estimators.

In the two sample case the estimating equations of the WLE can again be obtained by replacing the smoothed model and the smoothed empirical distributions by their corresponding unsmoothed versions in the score part of the estimating equation of the MDE. This yields the three estimating equations

$$n_i \int w_i(\delta_i^*(x)) \left[ \frac{\frac{\partial}{\partial \mu_i} m_{\beta_i}(x)}{m_{\beta_i}(x)} \right] d\hat{F}_i(x) = \sum_{j=1}^{n_i} w_i(\delta_i^*(X_{ij})) u_\mu(X_{ij}, \beta_i) = 0, \quad i=1,2$$

$$\sum_{i=1}^{2} n_i \int w_i(\delta_i^*(x)) \left[ \frac{\frac{\partial}{\partial \sigma^2} m_{\beta_i}}{m_{\beta_i}} \right] d\hat{F}_i(x) = \sum_{i=1}^{2} \sum_{j=1}^{n_i} w_i(\delta_i^*(X_{ij})) u_{\sigma^2}(X_{ij}, \beta_i) = 0$$

where

$$w_i(\delta_i^*(x)) = [A(\delta_i^*(x)) + 1]/[\delta_i^*(x) + 1], \quad \delta_i^*(x) = [f_i^*(x) - m_{\beta_i}^*(x)]/m_{\beta_i}^*(x),$$

and $u_\mu$ and $u_{\sigma^2}$ are the score functions with respect to $\mu$ and $\sigma^2$ respectively in the normal $N(\mu, \sigma^2)$ model. The above equations can be simplified to give

$$\sum_{j=1}^{n_i} w_i(\delta_i^*(X_{ij}))\,(X_{ij} - \mu_i) = 0, \quad i=1,2,$$

$$\sum_{i=1}^{2} \sum_{j=1}^{n_i} w_i(\delta_i^*(X_{ij}))\,[(X_{ij} - \mu_i)^2 - \sigma^2] = 0. \tag{3.3}$$

The three equations in (3.3) are solved iteratively by creating new weights $w_i(\delta_i^*(X_{ij}))$ at every stage. Following Basu et al. (1993) one can verify that these estimators are first order efficient independently of the choice of the kernel and the smoothing parameter.

For testing (3.1) the classical Student's t statistic is defined as

$$T_1 = \frac{\bar{X}_1 - \bar{X}_2}{s(1/n_1 + 1/n_2)^{1/2}}$$

where $s^2 = [(n_1-1)s_1^2 + (n_2-1)s_2^2]/(n_1+n_2-2)$, $\bar{X}_i$ and $s_i^2$ are respectively the i–th sample mean and variance for $i = 1, 2$. Next we define Tiku's (1980) test statistic. Let $(\mu_i\dagger, \sigma_i\dagger)$, i=1,2, be the MML estimators of $(\mu_i, \sigma)$ based on the symmetric type II censored i–th sample with $100\times r_i\%$ censoring on each side where $r_i = [0.5 + 0.1n_i]$ denotes the integer part of $(0.5 + 0.1n_i)$. These estimators are known as the MML10 estimators (Tiku 1980). Let the combined estimator of $\sigma^2$ be

$$\sigma\dagger^2 = [(A_1-1)\sigma_1^{\dagger 2} + (A_2-1)\sigma_2^{\dagger 2}]/(A_1+A_2-2)$$

where $A_i = n_i - 2r_i$. Then Tiku's (1980) test statistic is defined as

$$T_2 = \frac{\mu_1^\dagger - \mu_2^\dagger}{\sigma\dagger(1/m_1 + 1/m_2)^{1/2}}$$

where $m_i = n_i - 2r_i + 2r_i\beta_i$, i=1,2, and $\beta_i$ is a constant given by Tiku (1967, Table 1). Finally, let $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}^2)$ and $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2)$ denote the MDEs and the WLEs of the parameters $(\mu_1, \mu_2, \sigma)$ respectively. Let $T_3 = (\tilde{\mu}_1 - \tilde{\mu}_2)/[\tilde{\sigma}(1/n_1 + 1/n_2)^{1/2}]$ and $T_4 = (\hat{\mu}_1 - \hat{\mu}_2)/[\hat{\sigma}(1/n_1 + 1/n_2)^{1/2}]$.

Next we present a result which states that under the assumption of normality $T_1$, $T_3$ and $T_4$ are slightly more powerful than $T_2$ when a large, equal sample size n is used. The result follows from the arguments of Tiku et al. (1986, p. 122) and the asymptotic equivalence of the maximum likelihood estimator, MDE and WLE (Basu et al. 1993).

**Lemma 1.** *Suppose the population distributions are normal and* $n_1 = n_2 = n$. *Let* $d = \mu_1 - \mu_2$, *and* $\Delta = n^{1/2}d/[2^{1/2}\sigma]$. *Then for testing* $H_0$: $d = 0$ *against* $H_a$: $d \neq 0$ *at level* $\alpha$ *the power of the test statistics* $T_1$, $T_3$ *and* $T_4$ *are approximately equal to* $P(|Z + \Delta| \geq z_0)$ *for large* n, *where* Z *is the standard normal random variable and* $z_0$ *is its* $100(1-\alpha/2)$*–th percentile point. The power of* $T_2$ *is given by* $P(|Z + 0.983\Delta| \geq z_0)$ *so that the other 3 tests are asymptotically slightly more powerful than* $T_2$.

To carry out the above test procedures we need the distributions of the test statistics under the null hypothesis. Let $t(\nu)$ denote a central t distribution with $\nu$ degrees of freedom. Under the normality assumption, it is well known that the null distribution of $T_1$ is $t(n_1+n_2-2)$, and by Theorem 4.3.1 of Tiku et al. (1986) the null distribution of $T_2$ in large samples is approximately $t(A_1+A_2-2)$, where $A_1$, $A_2$ are as defined above. Lemma 1 above motivates us to approximate the null distribution of $T_3$ and $T_4$ by the $t(n_1+n_2-2)$ distribution.

## 4. MONTE CARLO STUDY

We conducted a Monte Carlo experiment for the equal (and unknown) variances case and compared the performance of $T_1$, Student's t test, $T_2$, Tiku's robust test, and $T_4$(HD), the $T_4$ statistic using the Hellinger distance, and $T_4$(LD), the $T_4$ statistic using the likelihood disparity. Note that $T_4$(LD) is exactly like $T_1$ except that in the pooled estimator of $\sigma^2$ the divisor is $(n_1+n_2)$ instead of $(n_1+n_2-2)$.

The tests were computed for nominal levels 10%, 5% and 1% and for equal sample sizes ($n_1 = n_2 = n$) 20, 40, 50 and 75. In each case five thousand pairs of independent pseudo random samples were generated. The pairs of population distributions considered are:

$$\text{Distribution (4.1):}\quad N(\mu_i, \sigma^2 = 1),\ i=1,2;$$

$$\text{Distribution (4.2):}\quad (1-\epsilon_i)N(\mu_i,\sigma^2=1) + \epsilon_i N(\mu_i,64),\ i=1,2;$$

$$\text{Distribution (4.3):}\quad (1-\epsilon_i)N(\mu_i,\sigma^2=1) + \epsilon_i t(1),\ i=1,2;$$

$$\text{Distribution (4.4):}\quad (1-\epsilon_i)N(\mu_i,\sigma^2=1) + \epsilon_i[N(\mu_1,\sigma^2=1)/U(0,1)],\ i=1,2,$$

where $\epsilon_1$ and $\epsilon_2$ are the contaminating proportions for the first and second population distributions respectively. Distribution (4.1) generates uncontaminated data, distributions (4.2) – (4.4) create heavier tails relative to the true population distributions. The empirical levels of the tests were computed by setting the true parameters $\mu_1$ and $\mu_2$ equal to zero, and the empirical powers of the tests were computed for $\mu_1 = 0$ and $\mu_2 = 0.5$. We did the computations for the one sample contamination case using $(\epsilon_1, \epsilon_2) = (0.1, 0)$ as well as for the two sample contamination case using $(\epsilon_1, \epsilon_2) = (0.1, 0.1)$.

In the computation of the test statistics $T_4(HD)$ and $T_4(LD)$ the initial estimates of $\mu_1$, $\mu_2$, $\sigma^2$ were defined as $\hat{\mu}_{1(0)}$ = median $(X_{11},\ldots,X_{1n})$, $\hat{\mu}_{2(0)}$ = median $(X_{21},\ldots,X_{2n})$, $\hat{\sigma}^2_{(0)} = 1.48 \times M$ where M = median $(|X_{11}-\hat{\mu}_{1(0)}|, \ldots, |X_{1n}-\hat{\mu}_{1(0)}|, |X_{21}-\hat{\mu}_{2(0)}|, \ldots, |X_{2n}-\hat{\mu}_{2(0)}|)$. For sample i we chose the bandwidth h of the $N(0,h^2)$ kernel function to be $0.5 \times M_i$ where $M_i$ = Median $(|X_{i1}-\hat{\mu}_{i(0)}|, \ldots, |X_{in}-\hat{\mu}_{i(0)}|)$.

The simulation results are presented in Tables I – VIII. Tables I–IV present the empirical powers and levels for the one sample contamination case, and Tables V–VIII present the empirical powers and levels for the two sample contamination case. In all the tables the levels are given in parentheses beside the corresponding powers. Note that the empirical powers of the $T_4(HD)$ are higher than those of Tiku's test for all the sample sizes considered. We have to be cautious about our conclusions, however, since the empirical levels of the $T_4(HD)$ in small samples are a little higher than the nominal levels. We have deliberately chosen the bandwidth h to be suitable small multiple of $M_i$ to get good robustness properties. The WLEs are asymptotically efficient estimates so that the empirical levels of $T_4(HD)$ do converge to the nominal levels for large n. However, for n as large as 75, where the empirical levels of $T_4(HD)$ are very close to the nominal levels, the $T_4(HD)$ has significantly higher powers than Tiku's test. Note also that both the Student's t test and $T_4(LD)$ lose power rapidly under contamination.

## 5. CONCLUDING REMARKS

Testing equality of two normal means is a well known problem in statistics. In this paper we have presented a test statistic with good power robustness. The statistic based on the Hellinger distance has comparable

TABLE I

Empirical powers (levels) of the test statistics under one sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 20$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.457 (0.099) | 0.325 (0.051) | 0.138 (0.009) |
| $T_2$ | 0.464 (0.104) | 0.332 (0.053) | 0.137 (0.014) |
| $T_4$(HD) | 0.495 (0.116) | 0.365 (0.061) | 0.175 (0.015) |
| $T_4$(LD) | 0.475 (0.111) | 0.346 (0.056) | 0.152 (0.011) |
| | Distribution (4.2) | | |
| $T_1$ | 0.293 (0.085) | 0.181 (0.033) | 0.057 (0.004) |
| $T_2$ | 0.414 (0.118) | 0.293 (0.058) | 0.115 (0.012) |
| $T_4$(HD) | 0.485 (0.125) | 0.367 (0.065) | 0.173 (0.017) |
| $T_4$(LD) | 0.309 (0.095) | 0.199 (0.038) | 0.066 (0.004) |
| | Distribution (4.3) | | |
| $T_1$ | 0.402 (0.084) | 0.283 (0.041) | 0.115 (0.006) |
| $T_2$ | 0.438 (0.100) | 0.315 (0.050) | 0.131 (0.010) |
| $T_4$(HD) | 0.493 (0.127) | 0.376 (0.065) | 0.179 (0.016) |
| $T_4$(LD) | 0.406 (0.092) | 0.293 (0.047) | 0.126 (0.008) |
| | Distribution (4.4) | | |
| $T_1$ | 0.369 (0.081) | 0.256 (0.039) | 0.102 (0.006) |
| $T_2$ | 0.404 (0.118) | 0.289 (0.062) | 0.106 (0.013) |
| $T_4$(HD) | 0.486 (0.121) | 0.373 (0.063) | 0.174 (0.015) |
| $T_4$(LD) | 0.384 (0.089) | 0.272 (0.043) | 0.114 (0.007) |

TABLE II

Empirical powers (levels) of the test statistics under one sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 40$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.731 (0.108) | 0.609 (0.057) | 0.355 (0.011) |
| $T_2$ | 0.703 (0.100) | 0.582 (0.051) | 0.336 (0.009) |
| $T_4$(HD) | 0.744 (0.113) | 0.633 (0.057) | 0.388 (0.014) |
| $T_4$(LD) | 0.736 (0.113) | 0.618 (0.060) | 0.366 (0.012) |
| | Distribution (4.2) | | |
| $T_1$ | 0.379 (0.101) | 0.278 (0.047) | 0.116 (0.005) |
| $T_2$ | 0.644 (0.114) | 0.526 (0.055) | 0.298 (0.010) |
| $T_4$(HD) | 0.717 (0.121) | 0.606 (0.062) | 0.358 (0.015) |
| $T_4$(LD) | 0.385 (0.105) | 0.286 (0.051) | 0.122 (0.005) |
| | Distribution (4.3) | | |
| $T_1$ | 0.562 (0.094) | 0.447 (0.041) | 0.241 (0.008) |
| $T_2$ | 0.673 (0.102) | 0.556 (0.051) | 0.306 (0.013) |
| $T_4$(HD) | 0.720 (0.114) | 0.606 (0.058) | 0.371 (0.014) |
| $T_4$(LD) | 0.564 (0.097) | 0.445 (0.044) | 0.249 (0.008) |
| | Distribution (4.4) | | |
| $T_1$ | 0.534 (0.092) | 0.418 (0.042) | 0.211 (0.007) |
| $T_2$ | 0.637 (0.117) | 0.503 (0.062) | 0.264 (0.014) |
| $T_4$(HD) | 0.723 (0.116) | 0.607 (0.060) | 0.360 (0.015) |
| $T_4$(LD) | 0.541 (0.095) | 0.428 (0.045) | 0.221 (0.007) |

TABLE III

Empirical powers (levels) of the test statistics under one sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 50$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.794 (0.099) | 0.696 (0.048) | 0.455 (0.010) |
| $T_2$ | 0.784 (0.101) | 0.681 (0.051) | 0.442 (0.010) |
| $T_4$(HD) | 0.802 (0.116) | 0.713 (0.057) | 0.478 (0.011) |
| $T_4$(LD) | 0.797 (0.103) | 0.704 (0.051) | 0.463 (0.011) |
| | Distribution (4.2) | | |
| $T_1$ | 0.406 (0.096) | 0.306 (0.038) | 0.137 (0.006) |
| $T_2$ | 0.719 (0.116) | 0.603 (0.057) | 0.377 (0.013) |
| $T_4$(HD) | 0.788 (0.118) | 0.693 (0.067) | 0.458 (0.014) |
| $T_4$(LD) | 0.412 (0.100) | 0.311 (0.040) | 0.144 (0.007) |
| | Distribution (4.3) | | |
| $T_1$ | 0.603 (0.086) | 0.497 (0.039) | 0.296 (0.007) |
| $T_2$ | 0.771 (0.111) | 0.671 (0.058) | 0.418 (0.011) |
| $T_4$(HD) | 0.789 (0.116) | 0.685 (0.068) | 0.458 (0.012) |
| $T_4$(LD) | 0.609 (0.088) | 0.504 (0.041) | 0.299 (0.007) |
| | Distribution (4.4) | | |
| $T_1$ | 0.580 (0.080) | 0.477 (0.038) | 0.275 (0.007) |
| $T_2$ | 0.715 (0.120) | 0.601 (0.058) | 0.343 (0.014) |
| $T_4$(HD) | 0.795 (0.109) | 0.695 (0.061) | 0.459 (0.014) |
| $T_4$(LD) | 0.585 (0.084) | 0.484 (0.041) | 0.283 (0.008) |

TABLE IV

Empirical powers (levels) of the test statistics under one sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 75$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.917 (0.099) | 0.857 (0.052) | 0.671 (0.010) |
| $T_2$ | 0.909 (0.100) | 0.849 (0.050) | 0.652 (0.011) |
| $T_4$(HD) | 0.921 (0.109) | 0.863 (0.057) | 0.687 (0.011) |
| $T_4$(LD) | 0.919 (0.102) | 0.860 (0.053) | 0.677 (0.011) |
| | Distribution (4.2) | | |
| $T_1$ | 0.497 (0.106) | 0.384 (0.048) | 0.195 (0.007) |
| $T_2$ | 0.859 (0.123) | 0.781 (0.061) | 0.559 (0.012) |
| $T_4$(HD) | 0.904 (0.116) | 0.836 (0.060) | 0.646 (0.014) |
| $T_4$(LD) | 0.500 (0.108) | 0.388 (0.050) | 0.200 (0.008) |
| | Distribution (4.3) | | |
| $T_1$ | 0.692 (0.076) | 0.604 (0.034) | 0.409 (0.006) |
| $T_2$ | 0.902 (0.104) | 0.834 (0.053) | 0.639 (0.011) |
| $T_4$(HD) | 0.914 (0.106) | 0.859 (0.056) | 0.676 (0.011) |
| $T_4$(LD) | 0.702 (0.078) | 0.614 (0.035) | 0.418 (0.006) |
| | Distribution (4.4) | | |
| $T_1$ | 0.632 (0.080) | 0.543 (0.038) | 0.365 (0.006) |
| $T_2$ | 0.860 (0.124) | 0.778 (0.069) | 0.566 (0.015) |
| $T_4$(HD) | 0.910 (0.114) | 0.848 (0.058) | 0.662 (0.012) |
| $T_4$(LD) | 0.635 (0.083) | 0.549 (0.038) | 0.369 (0.006) |

TABLE V

Empirical powers (levels) of the test statistics under two sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0.1$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 20$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.457 (0.099) | 0.325 (0.051) | 0.138 (0.009) |
| $T_2$ | 0.464 (0.104) | 0.332 (0.053) | 0.137 (0.014) |
| $T_4$(HD) | 0.495 (0.116) | 0.365 (0.061) | 0.175 (0.015) |
| $T_4$(LD) | 0.475 (0.111) | 0.346 (0.056) | 0.152 (0.011) |
| | Distribution (4.2) | | |
| $T_1$ | 0.293 (0.086) | 0.176 (0.032) | 0.038 (0.003) |
| $T_2$ | 0.358 (0.114) | 0.254 (0.053) | 0.093 (0.010) |
| $T_4$(HD) | 0.469 (0.128) | 0.356 (0.071) | 0.168 (0.019) |
| $T_4$(LD) | 0.311 (0.094) | 0.190 (0.039) | 0.044 (0.003) |
| | Distribution (4.3) | | |
| $T_1$ | 0.293 (0.079) | 0.196 (0.036) | 0.071 (0.006) |
| $T_2$ | 0.374 (0.113) | 0.266 (0.061) | 0.103 (0.011) |
| $T_4$(HD) | 0.440 (0.132) | 0.329 (0.068) | 0.154 (0.015) |
| $T_4$(LD) | 0.304 (0.087) | 0.209 (0.041) | 0.081 (0.008) |
| | Distribution (4.4) | | |
| $T_1$ | 0.338 (0.071) | 0.226 (0.033) | 0.078 (0.004) |
| $T_2$ | 0.425 (0.104) | 0.308 (0.053) | 0.113 (0.012) |
| $T_4$(HD) | 0.478 (0.121) | 0.355 (0.062) | 0.170 (0.017) |
| $T_4$(LD) | 0.356 (0.078) | 0.244 (0.037) | 0.087 (0.006) |

TABLE VI

Empirical powers (levels) of the test statistics under two sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0.1$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 40$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.731 (0.108) | 0.609 (0.057) | 0.355 (0.011) |
| $T_2$ | 0.703 (0.100) | 0.582 (0.051) | 0.336 (0.009) |
| $T_4$(HD) | 0.744 (0.113) | 0.633 (0.057) | 0.388 (0.014) |
| $T_4$(LD) | 0.736 (0.113) | 0.618 (0.060) | 0.366 (0.012) |
| | Distribution (4.2) | | |
| $T_1$ | 0.431 (0.108) | 0.303 (0.045) | 0.106 (0.003) |
| $T_2$ | 0.564 (0.124) | 0.447 (0.060) | 0.238 (0.011) |
| $T_4$(HD) | 0.697 (0.127) | 0.581 (0.066) | 0.360 (0.016) |
| $T_4$(LD) | 0.441 (0.108) | 0.312 (0.045) | 0.112 (0.003) |
| | Distribution (4.3) | | |
| $T_1$ | 0.384 (0.075) | 0.285 (0.030) | 0.125 (0.005) |
| $T_2$ | 0.611 (0.106) | 0.484 (0.055) | 0.260 (0.011) |
| $T_4$(HD) | 0.648 (0.113) | 0.531 (0.059) | 0.294 (0.014) |
| $T_4$(LD) | 0.392 (0.080) | 0.291 (0.033) | 0.131 (0.005) |
| | Distribution (4.4) | | |
| $T_1$ | 0.480 (0.080) | 0.356 (0.029) | 0.170 (0.004) |
| $T_2$ | 0.656 (0.098) | 0.530 (0.046) | 0.291 (0.008) |
| $T_4$(HD) | 0.725 (0.120) | 0.615 (0.060) | 0.372 (0.014) |
| $T_4$(LD) | 0.487 (0.085) | 0.367 (0.032) | 0.178 (0.004) |

TABLE VII

Empirical powers (levels) of the test statistics under two sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0.1$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 50$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.794 (0.099) | 0.696 (0.048) | 0.455 (0.010) |
| $T_2$ | 0.784 (0.101) | 0.681 (0.051) | 0.442 (0.010) |
| $T_4$(HD) | 0.802 (0.116) | 0.713 (0.057) | 0.478 (0.011) |
| $T_4$(LD) | 0.797 (0.103) | 0.704 (0.051) | 0.463 (0.011) |
| | Distribution (4.2) | | |
| $T_1$ | 0.486 (0.099) | 0.349 (0.044) | 0.140 (0.006) |
| $T_2$ | 0.635 (0.118) | 0.522 (0.061) | 0.298 (0.011) |
| $T_4$(HD) | 0.757 (0.124) | 0.654 (0.068) | 0.432 (0.017) |
| $T_4$(LD) | 0.493 (0.102) | 0.358 (0.046) | 0.145 (0.006) |
| | Distribution (4.3) | | |
| $T_1$ | 0.422 (0.080) | 0.318 (0.034) | 0.153 (0.005) |
| $T_2$ | 0.694 (0.101) | 0.577 (0.049) | 0.334 (0.012) |
| $T_4$(HD) | 0.730 (0.114) | 0.625 (0.058) | 0.386 (0.013) |
| $T_4$(LD) | 0.429 (0.081) | 0.325 (0.036) | 0.160 (0.005) |
| | Distribution (4.4) | | |
| $T_1$ | 0.491 (0.075) | 0.385 (0.033) | 0.202 (0.004) |
| $T_2$ | 0.739 (0.098) | 0.629 (0.051) | 0.390 (0.009) |
| $T_4$(HD) | 0.782 (0.115) | 0.689 (0.062) | 0.464 (0.013) |
| $T_4$(LD) | 0.496 (0.078) | 0.391 (0.035) | 0.209 (0.005) |

TABLE VIII

Empirical powers (levels) of the test statistics under two sample contamination (i.e., $\epsilon_1 = 0.1$, $\epsilon_2 = 0.1$) for different values of nominal level $\alpha$ and for $n_1 = n_2 = 75$

| Test statistic | $\alpha$=0.10 | $\alpha$=0.05 | $\alpha$=0.01 |
|---|---|---|---|
| | Distribution (4.1) | | |
| $T_1$ | 0.917 (0.099) | 0.857 (0.052) | 0.671 (0.010) |
| $T_2$ | 0.909 (0.100) | 0.849 (0.050) | 0.652 (0.011) |
| $T_4$(HD) | 0.921 (0.109) | 0.863 (0.057) | 0.687 (0.011) |
| $T_4$(LD) | 0.919 (0.102) | 0.860 (0.053) | 0.677 (0.010) |
| | Distribution (4.2) | | |
| $T_1$ | 0.608 (0.098) | 0.482 (0.047) | 0.230 (0.006) |
| $T_2$ | 0.792 (0.125) | 0.700 (0.064) | 0.480 (0.015) |
| $T_4$(HD) | 0.883 (0.121) | 0.818 (0.062) | 0.625 (0.015) |
| $T_4$(LD) | 0.611 (0.101) | 0.488 (0.049) | 0.237 (0.006) |
| | Distribution (4.3) | | |
| $T_1$ | 0.472 (0.076) | 0.371 (0.033) | 0.211 (0.004) |
| $T_2$ | 0.837 (0.107) | 0.739 (0.052) | 0.510 (0.011) |
| $T_4$(HD) | 0.856 (0.117) | 0.783 (0.058) | 0.566 (0.012) |
| $T_4$(LD) | 0.476 (0.077) | 0.375 (0.034) | 0.215 (0.004) |
| | Distribution (4.4) | | |
| $T_1$ | 0.556 (0.072) | 0.452 (0.029) | 0.258 (0.003) |
| $T_2$ | 0.886 (0.103) | 0.811 (0.051) | 0.608 (0.011) |
| $T_4$(HD) | 0.907 (0.118) | 0.840 (0.060) | 0.660 (0.013) |
| $T_4$(LD) | 0.560 (0.073) | 0.455 (0.030) | 0.262 (0.003) |

power to Student's t in the uncontaminated case, achieves high power and preserves the level in the contaminated cases also. In addition, it is simple to compute and can be of substantial practical value. The authors are currently studying the performance of this approach to the problem of testing the equality of normal means under unequal variances.

## ACKNOWLEDGEMENTS

The research of the second author was partly supported by a Grant from the College of Arts and Sciences at Oklahoma State University. The research of the third author was partly supported by a Mathematics Department Research Award, University of Texas at Austin. The authors thank Bruce G. Lindsay for some helpful comments.

## BIBLIOGRAPHY

Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for *continuous models: Efficiency, distributions and robustness. Ann. Inst. Statist. Math.* **46**, 683–705.

Basu, A., Markatou, M. and Lindsay, B. G. (1993). Robustness via maximum likelihood estimating equations. *Technical Report #* **93–01**, Department of Statistics, Columbia University, New York, NY.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band spectroscopic data. *Technometrics* **16**, 1470–1485.

Holland, P. W. and Welsch, R. E. (1977). *Robust regression using iteratively reweighted least squares. Commun. Statist.* A **6**, 813–827.

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist.* **22**, 1081–1114.

Sarkar, S. and Basu, A. (1995). On disparity based robust tests for two discrete populations. *Sankhya* B **57**, Part 3, 353–364.

Simpson, D. G. (1989). Hellinger deviance test: Efficiency, breakdown points and examples. *J. Amer. Statist. Assoc.* **84**, 107–113.

Tiku, M. L. (1967). Estimating the mean and standard deviation from censored normal samples. *Biometrika* **54**, 155–165.

Tiku, M. L. (1980). Robustness of MML estimators based on censored samples and robust test statistics. *J. Statist. Plann. Inf.* 4, 123–143.

Tiku, M. L. and Singh, M. (1982). Robust statistics for testing mean vectors of multivariate distributions. *Commun. Statist. – Theory and Methods* A **11**, 985–1001.

Tiku, M.L., Tan, W.Y. and Balakrishnan, N. (1986). *Robust inference.* New York and Basel: Marcel Dekker, Inc. .

Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29**, 2057–2071.

Yuen, K. K. (1974). The two sample trimmed t for unequal population variances. *Biometrika* **61**, 165–170.

Received June, 1995; Revised August, 1996.