On Difference of Means with Incomplete Data
Author(s): Pi-Erh Lin and Lawrence E. Stivers
Source: *Biometrika*, Vol. 61, No. 2 (Aug., 1974), pp. 325-334
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2334361

Biometrika (1974), **61**, 2, *p.* 325
Printed in Great Britain

On difference of means with incomplete data

By PI-ERH LIN and LAWRENCE E. STIVERS
Department of Statistics, The Florida State University, Tallahassee

Summary

An estimate of the difference of means is obtained when sampling from a bivariate normal distribution with variances σ_1^2 and σ_2^2 and correlation ρ, where some observations on either of the variables are missing. It is shown that this estimate has desirable properties. In this paper a test of the hypothesis of the equality of means is also considered. The above estimate is adopted and three new statistics based on the difference of sample means are proposed for the test. Their empirical powers are computed for different values of ρ and σ_1^2/σ_2^2.

Some key words: Asymptotic efficiency; Conservative test; Empirical powers; Likelihood ratio test; Maximum likelihood estimate; Missing data; Preferred test; Probabilities of type I error; Unbiasedness; Weak consistency; Welch approximate test.

1. Introduction

Let $(x_1, x_2)'$ be a bivariate vector normally distributed with means μ_1 and μ_2 and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Let $(x_{1\alpha}, x_{2\alpha})'$ $(\alpha = 1, \ldots, n)$ be n pairs of observations on $(x_1, x_2)'$; $x_{1,n+j}$ $(j = 1, \ldots, n_1)$ be n_1 additional observations on x_1; $x_{2,n+k}$ $(k = 1, \ldots, n_2)$ be n_2 additional observations on x_2. Based on these observations we wish to make inferences about the difference of means $\delta = \mu_1 - \mu_2$. It is assumed that $(x_{1\alpha}, x_{2\alpha})'$, $x_{1,n+j}$ and $x_{2,n+k}$ are mutually independent for $\alpha = 1, \ldots, n$, $j = 1, \ldots, n_1$ and $k = 1, \ldots, n_2$. Without loss of generality, the data may be arranged as follows:

$$\begin{matrix} x_{11}, \ldots, x_{1n}, & x_{1,n+1}, \ldots, x_{1,n+n_1} & \\ x_{21}, \ldots, x_{2n}, & & x_{2,n+1}, \ldots, x_{2,n+n_2}, \end{matrix} \tag{1·1}$$

where $(x_{1\alpha}, x_{2\alpha})'$ will be referred to as a paired observation while $x_{1,n+j}$ and $x_{2,n+k}$ will be referred to as incomplete or unpaired observations. Data as that in (1·1) may be called missing or incomplete and arise frequently. Occasionally, however, the missing observation may occur on only one of the variables, e.g. $n_2 = 0$. This is an important special pattern of incomplete data. Any experimental situation where the treatment is susceptible to subject mortality will often encounter incomplete data of this type. For $n_2 = 0$, Anderson (1957) obtained the maximum likelihood estimate of $\mu = (\mu_1, \mu_2)'$. Mehta & Gurland (1969*a*) obtained the maximum likelihood estimate of δ assuming $\sigma_1^2 = \sigma_2^2$, and Lin (1971) obtained a class of estimates for δ assuming $\sigma_1^2 \neq \sigma_2^2$. For testing the equality of means, Mehta & Gurland (1969*b*, 1973), Morrison (1973) and Lin (1973) each proposed a statistic using all available data. For the incomplete data pattern (1·1), Rao (1952, pp. 161–3) obtained the maximum likelihood estimates of μ and Σ implicitly. A simple iterative procedure for finding these estimates is described by Orchard & Woodbury (1970, Example 2).

In this paper we propose noniterative procedures to estimate δ and test the hypothesis $\delta = 0$ using all available data.

2. THE ESTIMATION OF $\mu_1 - \mu_2$

When Σ is known, the maximum likelihood estimate of μ is $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)'$, where

$$\hat{\mu}_1 = h[n(n+n_2)\,\bar{x}_1^{(n)} + n_1\{n+n_2(1-\rho^2)\}\,\bar{x}_1^{(n_1)} - nn_2\beta_{12}(\bar{x}_2^{(n)} - \bar{x}_2^{(n_2)})],$$

$$\hat{\mu}_2 = h[n(n+n_1)\,\bar{x}_2^{(n)} + n_2\{n+n_1(1-\rho^2)\}\,\bar{x}_2^{(n_2)} - nn_1\beta_{21}(\bar{x}_1^{(n)} - \bar{x}_1^{(n_1)})],$$

with $\beta_{21} = \rho\sigma_2/\sigma_1$, $\beta_{12} = \rho\sigma_1/\sigma_2$ being the regression coefficients,

$$\bar{x}_i^{(n)} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \quad \bar{x}_i^{(n_i)} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{i,n+j} \quad (i = 1, 2) \tag{2·1}$$

and

$$h = \frac{1}{(n+n_1)(n+n_2) - n_1 n_2 \rho^2}. \tag{2·2}$$

It can be shown that $\hat{\mu}$ is a sufficient statistic for μ and that $\hat{\mu}$ has the distribution $N(\mu, R)$, where, with $\sigma_{12} = \rho\sigma_1\sigma_2$,

$$R = h\begin{bmatrix} \{n+n_2(1-\rho^2)\}\,\sigma_1^2 & n\sigma_{12} \\ n\sigma_{12} & \{n+n_1(1-\rho^2)\}\,\sigma_2^2 \end{bmatrix}. \tag{2·3}$$

Hence the maximum likelihood estimate of δ is

$$\hat{\delta} = a\bar{x}_1^{(n)} + (1-a)\,\bar{x}_1^{(n_1)} - b\bar{x}_2^{(n)} - (1-b)\,\bar{x}_2^{(n_2)}, \tag{2.4}$$

where

$$a = nh(n+n_2+n_1\beta_{21}), \quad b = nh(n+n_1+n_2\beta_{12}). \tag{2·5}$$

It is clear that $\hat{\delta}$ has the distribution $N(\delta, \gamma^2)$, where

$$\begin{aligned}\gamma^2 &= \left\{\frac{a^2}{n} + \frac{(1-a)^2}{n_1}\right\}\sigma_1^2 - \frac{2ab}{n}\sigma_{12} + \left\{\frac{b^2}{n} + \frac{(1-b)^2}{n_2}\right\}\sigma_2^2 \\ &= h[\{n+n_2(1-\rho)^2\}\,\sigma_1^2 - 2n\sigma_{12} + \{n+n_1(1-\rho^2)\}\,\sigma_2^2].\end{aligned} \tag{2·6}$$

Inferences regarding δ may be drawn using the above $\hat{\delta}$ and its distribution.

When Σ is unknown, the maximum likelihood estimate of δ is obtained by replacing ρ, β_{12} and β_{21} by their respective maximum likelihood estimates in the expression of $\hat{\delta}$ in (2·4). Since the maximum likelihood estimate of Σ, from which the maximum likelihood estimates of ρ, β_{12} and β_{21} may be derived, does not have an explicit form (Rao, 1952, pp. 161–3), the maximum likelihood estimate of δ must be obtained implicitly. A simple iterative procedure described by Orchard & Woodbury (1970, Example 2) may be applied to find the numerical value of this estimate. When $n_2 = 0$, Anderson (1957) has shown the maximum likelihood estimate of β_{21} to be the usual estimate of the regression coefficient, which does not depend on the incomplete observations. For this reason, we will adopt the maximum likelihood estimates of ρ, β_{12} and β_{21} based only on the individuals for which no observation is missing, as estimates of ρ, β_{12} and β_{21}, respectively, and propose a new noniterative estimate of δ as follows: Define

$$a_{ij} = \sum_{k=1}^{n} (x_{ik} - \bar{x}_i^{(n)})(x_{jk} - \bar{x}_j^{(n)}) \quad (i, j = 1, 2), \tag{2·7}$$

$$r = a_{12}/(a_{11}a_{22})^{\frac{1}{2}}, \quad v = a_{12}/a_{11}, \quad w = a_{12}/a_{22}. \tag{2·8}$$

Then the new estimate of δ is given by

$$\delta^* = \hat{a}\bar{x}_1^{(n)} + (1-\hat{a})\,\bar{x}_1^{(n_1)} - \hat{b}\bar{x}_2^{(n)} - (1-\hat{b})\,\bar{x}_2^{(n_2)}, \tag{2·9}$$

where

$$\hat{a} = n\hat{h}(n + n_2 + n_1 v), \quad \hat{b} = n\hat{h}(n + n_1 + n_2 w), \tag{2·10}$$

with

$$\hat{h} = \frac{1}{(n+n_1)(n+n_2) - n_1 n_2 r^2}. \tag{2·11}$$

Now δ^* is the maximum likelihood estimate of δ in the following cases: (i) incomplete observations on only one of the variables, and (ii) no incomplete observations.

In the sequel we will denote the statement '$n \to \infty$, $n_i/n \to f_i$, $0 < f_i < \infty$, for $i = 1, 2$,' by $n \to^* \infty$.

Since $\hat{\delta}$ has the distribution $N(\delta, \gamma^2)$, that is, $\hat{\delta}$ is unbiased and is a complete sufficient statistic, it is the minimum variance unbiased estimate of δ with variance γ^2 given by (2·6). The new estimate δ^* given by (2·9) has the following properties: (i) unbiasedness; (ii) weak consistency in the sense that $\text{var}(\delta^*) \to 0$ as $n \to^* \infty$; (iii) asymptotic efficiency in the sense that $\text{var}(\delta^*)/\text{var}(\hat{\delta}) \to 1$ as $n \to^* \infty$; and (iv) an asymptotic normal distribution with mean δ and variance γ^{*2}, as $n \to^* \infty$, where

$$n\gamma^{*2} = \{a^{*2} + (1-a^*)^2/f_1\}\,\sigma_1^2 - 2a^*b^*\sigma_{12} + \{b^{*2} + (1-b^*)^2/f_2\}\,\sigma_2^2$$

with

$$a^* = \frac{1 + f_2 + f_1\beta_{21}}{(1+f_1)(1+f_2) - f_1 f_2 \rho^2}, \quad b^* = \frac{1 + f_1 + f_2\beta_{12}}{(1+f_1)(1+f_2) - f_1 f_2 \rho^2}.$$

Property (i) may be proved using the fact that a_{11}, a_{12} and a_{22} are independent of $\bar{x}_i^{(n)}$ and $\bar{x}_i^{(n_i)}$ for $i = 1$ and 2. Properties (ii) and (iii) follow by applying the Lebesgue Dominated Convergence Theorem to $E(\gamma_1^2)$ so that

$$\lim_{n \to^* \infty} nE(\gamma_1^2) = E(\lim_{n \to^* \infty} n\gamma_1^2),$$

where γ_1^2 is obtained by replacing a and b by $\hat{a}$ and $\hat{b}$, respectively, in the first expression of γ^2 in (2·6). Finally, if we let

$$\hat{\delta}^* = a^*\bar{x}_1^{(n)} + (1-a^*)\,\bar{x}_1^{(n_1)} - b^*\bar{x}_2^{(n)} - (1-b^*)\,\bar{x}_2^{(n_2)},$$

then property (iv) is proved by noting that $\delta^* \to \hat{\delta}^*$ in probability, and hence in distribution, as $n \to^* \infty$, and that $\hat{\delta}^* \sim N(\delta, \gamma^{*2})$.

3. Testing for equality of means

3·1. *General comments*

In this section the hypothesis

$$H_0: \delta = 0 \quad \textit{versus} \quad H_1: \delta > 0 \tag{3·1}$$

is tested. This is a special case of the hypothesis $C\mu = 0$ versus $C\mu \neq 0$, where C is a $q \times 2$ known matrix and 0 is a $q \times 1$ null vector ($q \leqslant 2$). The likelihood ratio statistic for this hypothesis does not have a closed form and thus must be obtained iteratively. Furthermore, the null distribution of the likelihood ratio statistic is unknown except when n, n_1 and n_2 are sufficiently large, i.e. as $n \to^* \infty$. In this case the likelihood ratio statistic is asymptotically chi-squared distributed with one degree of freedom. On the other hand, it is shown in §2 that δ^* is asymptotically distributed as $N(\delta, \gamma^{*2})$ as $n \to^* \infty$. Therefore a test based on δ^* is appropriate when n, n_1 and n_2 are sufficiently large. Specifically, let

$$Z = \frac{\delta^* - \delta}{\hat{\gamma}}, \tag{3·2}$$

where

$$\hat{\gamma}^2 = \left\{\frac{\hat{a}^2}{n} + \frac{(1-\hat{a})^2}{n_1}\right\}\frac{a_{11}}{n-1} - \frac{2\hat{a}\hat{b}}{n}\frac{a_{12}}{n-1} + \left\{\frac{\hat{b}^2}{n} + \frac{(1-\hat{b})^2}{n_2}\right\}\frac{a_{22}}{n-1}$$

with $\hat{a}$ and $\hat{b}$ given by (2·10). Then Z is asymptotically distributed as $N(0, 1)$ as $n \to^* \infty$. Since Z is obtained noniteratively and its asymptotic distribution known, it may be used in lieu of the likelihood ratio statistic to test the hypothesis (3·1) when n, n_1 and n_2 are sufficiently large. When the sample sizes are small, however, the exact distribution of Z is too complicated to be of value. A referee has suggested that the distribution of Z be approximated by a Student's t with n degrees of freedom; the authors are indebted to the referee for this suggestion. This aspect is investigated in §§3·3 and 3·4. It turns out that this approximation is quite satisfactory and is therefore adopted for testing (3·1).

It appears that, when sample sizes are small, no exact test can be constructed utilizing all available data except for the case $\rho = 0$ and $c^2 = \sigma_1^2/\sigma_2^2 = 1$. It is well known that the two-sample t test is appropriate when $\rho = 0$ and $c^2 = 1$, and that Welch's approximate t test (Welch, 1947) is used when $\rho = 0$ and $c^2 \neq 1$, which of course is the Behrens–Fisher problem. In this section we will propose approximate test procedures for the hypothesis (3·1) when sample sizes are moderately small whether $\rho = 0$ or not. Like the above existing t statistics, the proposed statistics are based on the difference of sample means

$$\bar{x}_1^{(n+n_1)} - \bar{x}_2^{(n+n_2)},$$

where for $i = 1, 2$

$$\bar{x}_i^{(n+n_i)} = \frac{1}{n+n_i}\sum_{j=1}^{n+n_i} x_{ij}. \tag{3·3}$$

These approximate tests use all the data and are simple in computation, though in some circumstances they may be less efficient than the likelihood ratio test.

3·2. *Some new tests*

If one is allowed to discard partial data, an exact test procedure which retains all paired observations will be provided by

$$T_1 = \frac{\bar{x}_1^{(n)} - \bar{x}_2^{(n)} - \delta}{\sqrt{\left\{\frac{a_{11} - 2a_{12} + a_{22}}{n(n-1)}\right\}}},$$

where $\bar{x}_i^{(n)}$ and a_{ij} $(i, j = 1, 2)$ are given by (2·1) and (2·7), respectively. Here T_1, of course, is the 'paired t' statistic with $n-1$ degrees of freedom, and is obtained by deleting all unpaired observations from (1·1). In this subsection we will construct three new statistics without discarding any data.

Define $\lambda_i = n/(n+n_i)$ $(i = 1, 2)$. Then the difference of sample means $\bar{x}_1^{(n+n_1)} - \bar{x}_2^{(n+n_2)}$ may be written as

$$(\lambda_1 \bar{x}_1^{(n)} - \lambda_2 \bar{x}_2^{(n)}) + (1-\lambda_1)\,\bar{x}_1^{(n_1)} - (1-\lambda_2)\,\bar{x}_2^{(n_2)},$$

which is a linear function of three independent normal random variables. Using these quantities the following test procedures are proposed.

When $c = 1$ and ρ unknown. Recall that $c = \sigma_1/\sigma_2$. When c is known, one may assume without loss of generality that $c = 1$. Define

$$T_2 = \frac{\bar{x}_1^{(n+n_1)} - \bar{x}_2^{(n+n_2)} - \delta}{\sqrt{\left\{\dfrac{1}{n+n_1} + \dfrac{1}{n+n_2} - \dfrac{2nr}{(n+n_1)(n+n_2)}\right\}} \sqrt{\left(\dfrac{a_{11}^* + b_{22}}{N-2}\right)}},$$

where $N = n + n_1 + n_2$ and

$$a_{11}^* = \sum_{j=1}^{n+n_1} (x_{1j} - \bar{x}_1^{(n+n_1)})^2,$$

$$b_{ii} = \sum_{j=n+1}^{n+n_i} (x_{ij} - \bar{x}_i^{(n_i)})^2 \quad (i = 1, 2).$$

If r were replaced by ρ in the expression of T_2, the exact distribution of T_2 would be Student's t with $N-2$ degrees of freedom. By a reduction in the degrees of freedom we claim that T_2 is approximately t distributed with $N-4$ degrees of freedom for $5 \leqslant n \leqslant 20$. This claim is well supported by the simulation study to be discussed in § 3·3.

A conservative test. Define for $i = 1, 2$

$$\begin{aligned} \Delta^2 = \Delta_0^2 &= (\lambda_1^2 a_{11} - 2\lambda_1 \lambda_2 a_{12} + \lambda_2^2 a_{22})/(n-1), \\ \Delta_i^2 &= (1-\lambda_i)^2 b_{ii}/(n_i - 1). \end{aligned} \tag{3·4}$$

Let

$$T_3 = \frac{\bar{x}_1^{(n+n_1)} - \bar{x}_2^{(n+n_2)} - \delta}{\sqrt{\left(\sum_{i=0}^{2} \dfrac{t_i^2 \Delta_i^2}{n_i}\right)}},$$

where $n_0 = n$ and $t_i = t_{n_i-1}(\alpha)$ is the upper 100α % critical level of a t distribution with $n_i - 1$ degrees of freedom for $i = 0, 1, 2$. Then a conservative test of (3·1) is to reject H_0 if $T_3 \geqslant 1$. The test is conservative in the sense that the probability of type I error is bounded above by α. The conservativeness of this test is an immediate consequence of Banerjee's (1960) result.

Welch type of approximate test. Define

$$T_4 = \frac{\bar{x}_1^{(n+n_1)} - \bar{x}_2^{(n+n_2)} - \delta}{\sqrt{\left(\sum_{i=0}^{2} \dfrac{\Delta_i^2}{n_i}\right)}},$$

where $n_0 = n$ and Δ_i $(i = 0, 1, 2)$ are given by (3·4). Then T_4 is approximately t distributed with the degrees of freedom estimated by

$$f = \frac{\left(\sum_{i=0}^{2} \dfrac{\Delta_i^2}{n_i}\right)^2}{\sum_{i=0}^{2} \dfrac{\Delta_i^4}{n_i^2(n_i-1)}}.$$

This is an application of Welch's (1947) approximation. The accuracy of the Welch approximation for the case of two independent normal random samples has been extensively studied. Wang (1971) calculates the probabilities of type I error of Welch's approximate test. The results show satisfactory agreement at the 0·05 and 0·01 nominal levels. It is suggested by the same paper that, in practice, one may use the usual t table to carry out the Welch approximate test without much loss of accuracy.

3·3. *Probabilities of type I error*

The test based on T_1 is well-known. However, further study was needed for Z defined by (3·2) and the appropriate tests proposed in § 3·2. For this reason, one thousand simulated random samples were generated from a bivariate normal distribution with $\mu_1 = \mu_2 = 0$ for each of the following values of ρ and $c^2 = \sigma_1^2/\sigma_2^2$

$$\begin{array}{llllllll} \rho = -0\cdot9, & -0\cdot5, & -0\cdot1, & 0, & 0\cdot1, & 0\cdot5, & 0\cdot9, \\ c^2 = 0\cdot25, & 0\cdot5, & 1, & 2, & 4, \end{array} \tag{3·5}$$

and the following combinations of n, n_1, n_2

$$\begin{array}{rccccc} n: & 5 & 8 & 10 & 15 & 20 \\ n_1: & 5 & 9 & 5 & 5 & 5. \\ n_2: & 10 & 3 & 10 & 10 & 10 \end{array} \tag{3·6}$$

The hypothesis (3·1) was tested at the 0·05 and 0·01 levels using Z, T_1, T_2, T_3 and T_4. We recall that Z is approximated by a t distribution with n degrees of freedom. The relative frequencies of the tests exceeding the nominal α levels, i.e. the proportions of tests for which H_0 was rejected, were recorded. These proportions are called the empirical α levels. An approximate test is said to be satisfactory if its nominal α level is within two standard deviations of the empirical α level, i.e. within an approximately 95% confidence interval of the nominal α.

The results for Z when $n \geqslant 10$ are very satisfactory for almost all values of ρ and c^2 with exceptions occurring at $n = 15$, $\rho \leqslant -0\cdot9$ and $2 \leqslant c^2 \leqslant 4$. When $n < 10$, the empirical α levels are erratic. They are mostly different from the nominal levels by more than two standard deviations. For this reason, it is suggested that Z be approximated by a t distribution with n degrees of freedom when $n \geqslant 10$. No satisfactory approximation has been found when $n < 10$.

The results for T_1 are as expected. The nominal α levels are within two standard deviations of the empirical α levels for all values of ρ and c^2.

For T_2, the empirical levels are generally very close to the nominal levels when $c^2 = 1$. For c^2 different from one, however, such is not the case. Significant deviations from the nominal levels occur whenever (i) $n = 5$, $|\rho| \geqslant 0\cdot9$ or $n = 5$, $c^2 = 4$; (ii) $n = 10$, $c^2 < 0\cdot25$ or $c^2 > 4$; (iii) $n = 15$, $c^2 < 0\cdot5$ or $c^2 > 2$; and (iv) $n = 20$, $c^2 \neq 1$. This is not surprising since T_2 is the test specifically formulated for the case of equal variances. It is also apparent that T_2 becomes unstable whenever $n \geqslant 20$ and $c^2 \neq 1$. For this reason it is suggested that T_2 be used when there are a moderate number of paired observations, i.e. $5 \leqslant n \leqslant 20$, and c^2 is near unity.

Analysis by T_3 always results in an empirical α level much smaller than the nominal level. At a nominal level of 0·05, the empirical level is usually about 0·03, at 0·01, the empirical level is usually 0·003, with some instances of 0·000.

The results for T_4 parallel those for T_2 when $n < 20$ and c^2 is near unity. Unlike T_2, however, the nominal levels for T_4 are within two standard deviations of the empirical levels when $c^2 \neq 1$.

In general, when testing (3·1), Z, with $n \geqslant 10$, and the paired t procedures give the most satisfactory results for all values of ρ and c^2 as far as the empirical level is concerned with exceptions for Z as noted earlier when $n = 15$, $\rho \leqslant -0{\cdot}9$ and $2 \leqslant c^2 \leqslant 4$; T_2 was good for $|\rho| \leqslant 0{\cdot}5$, $0{\cdot}5 \leqslant c^2 \leqslant 2$ when $5 \leqslant n < 20$ and $c^2 = 1$ when $n = 20$; and T_3 is always conservative. The Welch type of procedure is satisfactory for all values of c^2 when $5 \leqslant n < 20$.

3·4. *Powers of the tests*

If the approximate tests are satisfactory then it is desirable to compare their powers and decide which is the most powerful test, or the preferred test, among them. To do so, one thousand samples were generated from a bivariate normal distribution for each of seven values of ρ and five values of c^2 as given by (3·5). Since T_3 was so conservative, it was decided to exclude it from the power study. Powers were calculated for $\delta = 1, 3$, in fact for $\mu_1 = 1, 3$ and $\mu_2 = 0$, using the combinations of (n, n_1, n_2) listed in (3·6) except for (8, 9, 3). Based on the simulation study the regions in which each of the statistics is preferred are summarized below:

(i) Z is preferred when

(*a*) $n = 10, \quad \rho \geqslant 0{\cdot}9;$
(*b*) $n = 15, \quad |\rho| \geqslant 0{\cdot}5;$
(*c*) $n = 20, \quad |\rho| \geqslant 0{\cdot}5, \quad c^2 = 1;$
(*d*) $n = 20, \quad c^2 \neq 1.$

(ii) The paired t is preferred when $n = 5$, $\rho \geqslant 0{\cdot}9$, $0{\cdot}5 \leqslant c^2 \leqslant 4$.

(iii) T_2 is preferred when

(*a*) $n = 5, \quad |\rho| \leqslant 0{\cdot}5, \quad 0{\cdot}5 \leqslant c^2 \leqslant 2;$
(*b*) $n = 10, \quad \rho < 0{\cdot}9, \quad 0{\cdot}25 \leqslant c^2 \leqslant 2;$
(*c*) $n = 15, \quad |\rho| \leqslant 0{\cdot}5, \quad 0{\cdot}5 \leqslant c^2 \leqslant 1;$
(*d*) $n = 20, \quad -0{\cdot}5 \leqslant \rho \leqslant 1, \quad c^2 = 1.$

(iv) T_4 is preferred when

(*a*) $n = 5, \quad c^2 = 0{\cdot}25;$
(*b*) $n = 5, \quad |\rho| \leqslant 0{\cdot}5, \quad c^2 = 0{\cdot}5;$
(*c*) $n = 5, \quad \rho \leqslant 0{\cdot}5, \quad c^2 = 4;$
(*d*) $n = 10, \quad \rho = 0{\cdot}5, \quad c^2 = 1;$
(*e*) $n = 10, \quad \rho = 0{\cdot}1, \quad c^2 = 2;$
(*f*) $n = 10, \quad -0{\cdot}1 \leqslant \rho \leqslant 0{\cdot}5, \quad c^2 = 4;$
(*g*) $n = 15, \quad |\rho| \leqslant 0{\cdot}5, \quad c^2 = 0{\cdot}25;$
(*h*) $n = 15, \quad |\rho| \leqslant 0{\cdot}1, \quad 2 \leqslant c^2 \leqslant 4;$
(*i*) $n = 20, \quad |\rho| \leqslant 0{\cdot}1, \quad c^2 = 2.$

In the above summary there are regions in which two statistics are both preferred. For example, when $n = 15$, $\rho = 0{\cdot}5$ and $0{\cdot}5 \leqslant c^2 \leqslant 1$ both Z and T_2 are preferred. This may occur because either (1) one is preferred at the 5 % level and the other 1 %, or (2) one is more powerful while its empirical level is also greater than the other at both the 5 % and 1 % nominal levels.

Based on the simulation study, it is observed that the powers of the tests under consideration are increasing functions of ρ for a fixed value of c^2. For fixed ρ, the powers decrease as c^2 deviates from one. Moreover, it is clear that Z is the best criterion among those compared for reasonably correlated data when there are sufficiently large number n of complete observations. The simulations indicate $n = 15$. When $5 \leqslant n < 15$, T_2 and T_4 are useful simple tests for ρ^2 low and c^2 near unity, and when $n = 5$, the paired t is preferred for $\rho \geqslant 0{\cdot}9$.

Finally, we present in Table 1 empirical levels and powers ($\delta = 1, 3$) of the tests for $c^2 = 0{\cdot}5$, 1, and 2 when $n = 15$, $n_1 = 5$, $n_2 = 10$. This table should provide the reader some idea about the simulations.

Table 1. *Empirical levels and powers* ($\times$ 1000) *of* Z, T_1, T_2 *and* T_4, *when* $\delta = 1, 3$ *for* $n = 15$, $n_1 = 5$, $n_2 = 10$ ($c^2 = 0{\cdot}5$, 1, 2)

Statistics		Z		T_1		T_2		T_4	
α levels		5%	1%	5%	1%	5%	1%	5%	1%
$\rho = -0{\cdot}9$	$c^2 = 0{\cdot}5$	057	009	060	013	056	012	058	011
		692	373	447	197	688	426	664	373
		1000	1000	998	982	1000	1000	1000	1000
	$c^2 = 1$	058	009	058	013	051	009	061	011
		804	549	571	304	802	537	793	529
		1000	1000	1000	998	1000	1000	1000	1000
	$c^2 = 2$	064*	008	057	014	046	008	058	009
		678	370	451	199	611	318	648	360
		1000	1000	998	982	1000	1000	1000	1000
$\rho = -0{\cdot}5$	$c^2 = 0{\cdot}5$	050	011	050	014	068*	016	053	011
		733	429	525	237	768	488	741	440
		1000	1000	1000	994	1000	1000	1000	1000
	$c^2 = 1$	057	011	055	015	063	009	058	010
		867	608	668	374	868	638	866	619
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 2$	061	010	054	014	055	008	055	010
		719	425	525	248	693	388	715	416
		1000	1000	999	993	1000	1000	1000	1000
$\rho = -0{\cdot}1$	$c^2 = 0{\cdot}5$	048	011	047	012	051	014	048	009
		818	517	642	325	860	605	828	556
		1000	1000	1000	998	1000	1000	1000	1000
	$c^2 = 1$	052	011	048	015	050	013	048	009
		924	715	793	504	926	769	926	747
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 2$	055	011	050	014	050	009	053	010
		783	497	630	338	788	484	795	513
		1000	1000	1000	999	1000	1000	1000	1000
$\rho = 0{\cdot}0$	$c^2 = 0{\cdot}5$	042	012	052	010	054	013	046	008
		842	551	683	358	883	634	856	589
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 1$	048	012	049	015	049	014	049	010
		941	743	824	546	943	795	945	772
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 2$	055	012	050	014	049	010	052	011
		801	523	667	366	811	517	819	548
		1000	1000	1000	1000	1000	1000	1000	1000

Table 1 (*cont.*)

Statistics		Z		T_1		T_2		T_4	
α levels		5%	1%	5%	1%	5%	1%	5%	1%
$\rho = 0{\cdot}1$	$c^2 = 0{\cdot}5$	043	010	053	010	054	013	045	006
		861	598	732	401	900	676	880	616
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 1$	048	012	046	014	051	012	050	011
		954	792	863	602	955	829	958	816
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 2$	055	011	051	013	047	010	050	012
		828	560	712	415	838	558	844	578
		1000	1000	1000	1000	1000	1000	1000	1000
$\rho = 0{\cdot}5$	$c^2 = 0{\cdot}5$	053	009	046	012	053	013	039	007
		969	814	893	683	967	838	961	803
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 1$	049	012	051	009	045	008	046	007
		989	947	975	875	985	945	986	942
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 2$	048	012	048	015	042	010	043	009
		945	767	896	662	924	744	927	735
		1000	1000	1000	1000	1000	1000	1000	1000
$\rho = 0{\cdot}9$	$c^2 = 0{\cdot}5$	057	009	044	010	056	012	049	007
		1000	1000	1000	998	998	975	997	951
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 1$	055	007	047	012	046	006	047	007
		1000	1000	1000	1000	1000	996	1000	993
		1000	1000	1000	1000	1000	1000	1000	1000
	$c^2 = 2$	049	012	049	014	046	011	042	010
		1000	996	1000	996	984	932	982	898
		1000	1000	1000	1000	1000	1000	1000	1000

* Three entries in each cell are empirical α-level, power at $\delta = 1$ and power at $\delta = 3$. The observed relative frequency was different from the nominal α level by more than $2\sqrt{\{\alpha(1-\alpha)/1000\}}$.

The authors are grateful to the referees for their comments and suggestions. The research for the first author was supported in part by the National Institute of General Medical Sciences.

References

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Am. Statist. Assoc.* **52**, 200–4.

Banerjee, S. K. (1960). Approximate confidence interval for linear functions of means of k populations when the population variances are not equal. *Sankhyā* **22**, 357–8.

Lin, P. E. (1971). Estimation procedures for difference of means with missing data. *J. Am. Statist. Assoc.* **66**, 634–6.

Lin, P. E. (1973). Procedures for testing the difference of means with incomplete data. *J. Am. Statist. Assoc.* **68**, 699–703.

Mehta, J. S. & Gurland, J. (1969*a*). Some properties and an application of a statistic arising in testing correlation. *Ann. Math. Statist.* **40**, 1736–45.

MEHTA, J. S. & GURLAND, J. (1969*b*). Testing the equality of means in the presence of correlation. *Biometrika* **56**, 119–26.

MEHTA, J. S. & GURLAND, J. (1973). A test for equality of means in the presence of correlation and missing data. *Biometrika* **60**, 211–3.

MORRISON, D. M. (1973). A test for equality of means of correlated variates with missing data on one response. *Biometrika* **60**, 101–5.

ORCHARD, T. & WOODBURY, M. A. (1970). A missing information principle: theory and applications. *Proc. 6th Berkeley Symp.* **1**, 697–715.

RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: Wiley.

WANG, Y. Y. (1971). Probabilities of type I errors of Welch tests for the Behrens–Fisher problem. *J. Am. Statist. Assoc.* **66**, 605–7.

WELCH, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika* **34**, 28–35.

[*Received March* 1973. *Revised November* 1973]