Statistica Sinica **8**(1998), 1115-1130

THE EFFECT OF SAMPLING DESIGN ON ANDERSON'S EXPANSION OF THE DISTRIBUTION OF FISHER'S SAMPLE DISCRIMINANT FUNCTION

Kam-Wah Tsui and Ching-Ho Leu

University of Wisconsin-Madison and National Cheng-Kung University

Abstract: In the problem of discriminant analysis using Fisher's sample discriminant function W, Anderson (1973) obtained an asymptotic expansion of the distribution of W in the case of simple random samples. The approximation to the rate of misclassification resulting from this asymptotic expansion is very good. This paper examines the situation where the sample selection can depend on the values of auxiliary variables of the population under study, for example stratified random sampling. We show that Anderson's asymptotic expansion has an error of order $O(1)$ under disproportionate stratified random sampling and we provide a reasonable correction which takes into account such sampling design effect.

Key words and phrases: Asymptotic expansion, discriminant function, sampling design effect, stratified random sampling.

1. Introduction

Statistical procedures in multivariate analysis are usually based on a sample consisting of independently and identically distributed (i.i.d.) observations from some hypothetical superpopulation. When the samples are collected via a complex survey design employing stratification, multistage selection or unequal probability selection using auxiliary information, procedures need to be modified to take into account the sampling design if one is to draw valid inferences. Nathan and Holt (1980) show how procedures in regression analysis can be adjusted for the situation where sample selection depends on an auxiliary variable; Rao, Sutradhar and Yue (1993) study regression analysis with two-stage cluster samples; Skinner, Holmes and Smith (1986) consider the effects of a certain sampling design in principal component analysis; Skinner, Holmes and Smith (1989) and Lehtonen and Pahkinen (1995) cover many of the important advances in multivariate analysis under both aggregated and disaggregated approaches.

Discriminant analysis is a procedure for classifying units into mutually exclusive and exhaustive groups on the basis of x, a vector of characteristics, called the discriminator. The usual procedure in the two-group problem is to classify a new unit with a discriminator x_0 according to Fisher's sample discriminant function, W, defined by

$$W(x_0) = (x_0 - (\bar{x}_1 + \bar{x}_2)/2)^T S_{11}^{-1}(\bar{x}_1 - \bar{x}_2). \tag{1.1}$$

Here $\bar{x}_i$ is the sample mean of the discriminators of the training sample from group i, $i = 1, 2$, and S_{11} is the usual pooled-sample covariance matrix of the discriminator (see Section 2 for a more detailed description). Recently Leu and Tsui (1997) examined the performance of several discriminant functions when parameters are estimated by conventional estimators, maximum likelihood estimators (MLE), probability-weighted estimators and conditionally unbiased estimators, under a superpopulation model and a sample design similar to that of Nathan and Holt (1980) and Skinner, Holmes and Smith (1986). They discussed the rates of misclassification of four discriminant functions and the effect of the sample design on their rates via a simulation study. The discriminant function using MLE performed the best and Fisher's discriminant function the worst under a stratified sampling design with increasing allocation. In this article we provide a theoretical result concerning the effect of the sample design on Fisher's sample discriminant function.

Okamoto (1963) (with correction, Okamoto (1968)) and Anderson (1973) obtained asymptotic expansions of the distribution of $W(x_0)$ under simple random samples (SRS). An approximation of the rate of misclassification can be obtained by replacing the unknown parameters in an asymptotic expansion with corresponding sample estimates. These approximations are generally more accurate than others in the literature. Here we use the same set-up as in Leu and Tsui (1997) to consider the situation where the sample selection can depend on the value of an auxiliary variable. We obtain an asymptotic expansion of the distribution of W analogous to that of Anderson (1973) in this general situation.

Our result shows that Anderson's asymptotic expansion has an error of order $O(1)$ under disproportionate stratified random sampling. We propose a reasonable correction to take into account such sampling effect.

We describe the superpopulation model for the discriminant analysis problem and then the sampling design in Section 2. Section 3 contains our main results.

2. The Superpopulation Model

Consider the problem of classifying a unit into one of two groups (finite populations) G_1 or G_2, where G_1 and G_2 consist of N_1 and N_2 identifiable units, labeled $i = 1, \ldots, N_1$, and $j = 1, \ldots, N_2$, respectively. Let $N = N_1 + N_2$. Suppose that associated with each unit i in G_α is a discriminator $x_{\alpha i} = (x_{\alpha i,1}, \ldots, x_{\alpha i,p})^T$, a $p \times 1$ vector to be measured in the survey, and $z_{\alpha i} = (z_{\alpha i,1}, \ldots, z_{\alpha i,q})^T$, a $q \times 1$ vector of known values to be used in sample selection. We assume that

$x_{\alpha i}$ and $z_{\alpha i}$ are realizations of the random vectors $X_{\alpha i}$ and $Z_{\alpha i}$ respectively, $i = 1, \ldots, N_\alpha$, and $\alpha = 1, 2$. Furthermore we assume that the random vectors $(X_{\alpha i}^T, Z_{\alpha i}^T)^T$, $i = 1, \ldots, N_\alpha$, and $\alpha = 1, 2$, are independent and that each vector has a multivariate normal distribution

$$\begin{pmatrix} X_{\alpha i} \\ Z_{\alpha i} \end{pmatrix} \sim N_{p,q}\Big[\begin{pmatrix} \mu_\alpha \\ \mu_z \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\Big], \tag{2.1}$$

where Ω_{22} is assumed to be positive definite. Note that we assume a common covariance matrix (Ω) for the two groups and that the means of random vectors Z_{1i}, Z_{2i} are equal to μ_z. Hence only the random vectors $X_{\alpha i}$, $i = 1, \ldots, N_\alpha$, and $\alpha = 1, 2$, appear to be useful for discriminating the two groups.

Let $\mathbf{x}_\alpha = (x_{\alpha 1}, \ldots, x_{\alpha N_\alpha})$, $\mathbf{z}_\alpha = (z_{\alpha 1}, \ldots, z_{\alpha N_\alpha})$, $\mathbf{X}_\alpha = (X_{\alpha 1}, \ldots, X_{\alpha N_\alpha})$ and $\mathbf{Z}_\alpha = (Z_{\alpha 1}, \ldots, Z_{\alpha N_\alpha})$, $\alpha = 1, 2$. The auxiliary information $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2)$ and the population sizes N_1 and N_2 are known at the beginning of the survey. The assumption of common means for the auxiliary variables Z_{1i} and Z_{2i} can be tested by using the usual T^2-statistic test (See for example Johnson and Wichern (1992), Chapter 6). If the test rejects it, the problem of examining the sampling design effect on misclassification rates becomes more complicated and is beyond the scope of this paper. For the rest of the paper we assume common means of the auxiliary variables Z_{1i} and Z_{2i}.

Two samples $s_1 = (i_1, \ldots, i_{n_1})$ of n_1 distinct units from group G_1 and $s_2 = (j_1, \ldots, j_{n_2})$ of n_2 distinct units from group G_2 are selected independently by prechosen randomized sampling designs $p(s_1|\mathbf{z}_1)$ and $p(s_2|\mathbf{z}_2)$, respectively. We assume the sampling designs $p(s_\alpha|\mathbf{z}_\alpha), \alpha = 1, 2$, depend on the auxiliary information $\mathbf{z} = (\mathbf{z_1}, \mathbf{z_2})$, but not on the $x_{\alpha i}$. The auxiliary information $\mathbf{z}$ can serve as a size variable for probability-proportional-to-size sampling, or as a grouping variable for stratified or cluster sampling. We assume that the sample sizes n_1 and n_2 of s_1 and s_2, respectively, are fixed. Denote the sample data by $(\mathbf{x}_{s1}, \mathbf{x}_{s2}, \mathbf{z}_1, \mathbf{z}_2, s_1, s_2)$ and let $\mathbf{x}_{s\alpha}$ and $\mathbf{X}_{s\alpha}$ be the subvectors of $\mathbf{x}_\alpha$ and $\mathbf{X}_\alpha$ with indices in s_α, respectively.

Let $\sum_s$ denote the summation over the units in the sample and $\sum_{N_\alpha}$ denote the summation over group G_α, with

$$\bar{x}_1 = n_1^{-1} \sum_{s1} x_{1i}, \qquad \bar{X}_1 = N_1^{-1} \sum_{N_1} x_{1i}, \tag{2.2a}$$

$$\bar{x}_2 = n_2^{-1} \sum_{s2} x_{2j}, \qquad \bar{X}_2 = N_2^{-1} \sum_{N_2} x_{2j}, \tag{2.2b}$$

$$\bar{z}_1 = n_1^{-1} \sum_{s1} z_{1i}, \quad \bar{z}_2 = n_2^{-1} \sum_{s2} z_{2j}, \quad \bar{Z} = N^{-1}\Big(\sum_{N_1} z_{1i} + \sum_{N_2} z_{2j}\Big), \tag{2.2c}$$

$$m_{1s1} = (n_1 - 1)^{-1} \sum_{s1} (x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)^T, m_{2s1} = (n_1 - 1)^{-1} \sum_{s1} (z_{1i} - \bar{z}_1)(z_{1i} - \bar{z}_1)^T,$$

$$m_{1s2} = (n_2-1)^{-1}\sum_{s2}(x_{2j}-\bar{x}_2)(x_{2j}-\bar{x}_2)^T, m_{2s2} = (n_2-1)^{-1}\sum_{s2}(z_{2j}-\bar{z}_2)(z_{2j}-\bar{z}_2)^T.$$

Let

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \tag{2.2d}$$

where S is the usual pooled estimator of Ω, and note that S is unbiased under the independent simple random sampling design.

Using the usual Fisher's sample discriminant function in (1.1), we can classify the unit with discriminator x_0 to G_1 if $W(x_0) > c$ and to G_2 otherwise, where c is a given constant which depends on the prior probabilities and the costs of misclassification. For simplicity, we assume equal prior probabilities and equal cost of misclassification, i.e., c is zero.

The statistic W is important since it is the usual discriminant rule under the i.i.d. model in standard discriminant analysis. Under the i.i.d. model, the limiting distribution of W as $n_1 \to \infty$ and $n_2 \to \infty$ is normal with variance Δ^2 and mean $\Delta^2/2$ if the new unit belongs to G_1 and mean $-\Delta^2/2$ if it belongs to G_2, where

$$\Delta^2 = (\mu_1 - \mu_2)^T \Omega_{11}^{-1}(\mu_1 - \mu_2) \tag{2.3}$$

is the Mahalanobis distance, estimated by

$$D^2 = (\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(\bar{x}_1 - \bar{x}_2). \tag{2.4}$$

The asymptotic expansion of the distribution of the studentized W, $(W\pm\frac{1}{2}D^2)/D$, gives approximate evaluations of the pair of misclassification rates. Note that the limiting distribution of both $(W - \frac{1}{2}D^2)/D$ and $(W + \frac{1}{2}D^2)/D$ are $N(0,1)$ when the new unit belongs to G_1 and G_2, respectively, under the i.i.d. model. The usual approximation of the corresponding misclassification rates are, respectively, $\Phi(D/2)$ and $\Phi(-D/2)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variate, under the i.i.d. model. We investigate the sampling design effect on the asymptotic expansion of the distribution of W under our sample design.

3. Main Results

Let $n = n_1 + n_2 - 2$. Okamoto (1963) obtained the asymptotic expansion of the distribution of W up to terms of order n^{-2} by using the properties of the characteristic function of the standard normal distribution. Anderson (1973) proposed another asymptotic expansion of the distribution for the studentized W by using Taylor's expansion of the cumulative distribution function of the standard normal distribution. Both asymptotic expansions provide more accurate approximations of the rates of misclassification than the usual ones, with an error of order n^{-2}. Note that both procedures are based on the i.i.d. model. In this paper, we only consider the sampling design effect on Anderson's approximation because his expansion formula is simpler than that of Okamoto (1963). Following Anderson (1973) closely, we derive an asymptotic expansion of the distribution for the studentized W with an error of order $O(n^{-3/2})$ under the superpopulation model given in (2.1) and the proposed sampling scheme in Section 2. This contrasts with Anderson's (1973) asymptotic expansion that has an error of order $O(1)$ under disproportionate stratified random sampling.

Assume that $n_2/n_1 \to k$, a finite positive constant. We write

$$W - D^2/2 = (x_0 - \bar{x}_1)^T S_{11}^{-1} (\bar{x}_1 - \bar{x}_2).$$

Then,

$$P\Big[\frac{W-D^2/2}{D} \le u\Big] = E_{s,\mathbf{z}} P\Big[\frac{W-D^2/2}{D} \le u|s,\mathbf{z}\Big] = E_{s,\mathbf{z}} P[(\bar{x}_1-\bar{x}_2)^T S_{11}^{-1}(x_0-\mu)$$
$$\le u((\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(\bar{x}_1 - \bar{x}_2))^{1/2} + (\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(\bar{x}_1 - \mu)|s,\mathbf{z}]. \tag{3.1}$$

Note that x_0 has a multivariate normal distribution, $N(\mu, \Omega_{11})$, independent of $\bar{x}_1, \bar{x}_2, S_{11}, s$ and $\mathbf{z}$. The conditional distribution of $(\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(x_0 - \mu)$ given $\bar{x}_1, \bar{x}_2, S_{11}, s$ and $\mathbf{z}$ is $N[0, \tilde{\Omega}_{11}]$, where $\tilde{\Omega}_{11} = (\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1} \Omega_{11} S_{11}^{-1}(\bar{x}_1 - \bar{x}_2)$, and

$$\gamma = \frac{(\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(x_0 - \mu)}{\tilde{\Omega}_{11}^{1/2}}$$

has a standard normal distribution. Then

$$\begin{aligned} P\Big[\frac{W - D^2/2}{D} \le u\Big] &= E_{\bar{x}_1,\bar{x}_2,S_{11},s,\mathbf{z}} P\Big[\frac{W - D^2/2}{D} \le u|\bar{x}_1, \bar{x}_2, S_{11}, s, \mathbf{z}\Big] \\ &= E_{s,\mathbf{z}} E_{\bar{x}_1,\bar{x}_2,S_{11}|s,\mathbf{z}} P\Big[\gamma \le \frac{uD + (\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(\bar{x}_1 - \mu)}{\tilde{\Omega}_{11}^{1/2}}\Big] \\ &= E_{s,\mathbf{z}} E_{\bar{x}_1,\bar{x}_2,S_{11}|s,\mathbf{z}} \Phi\Big[\frac{uD + (\bar{x}_1 - \bar{x}_2)^T S_{11}^{-1}(\bar{x}_1 - \mu)}{\tilde{\Omega}_{11}^{1/2}}\Big]. \end{aligned} \tag{3.2}$$

The distributions of W and D^2 are invariant under any non-singular linear transformation of the following type:

$$\begin{pmatrix} x \\ z \end{pmatrix} \to \begin{pmatrix} A_{11} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} + \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}.$$

We can choose A_{11}, Q_1 and Q_2 to transform Ω_{11} to I, $\mu_1 - \mu_2$ to $\delta = (\Delta, 0, \ldots, 0)^T$, and μ_z to 0, where $\Delta^2 = (\mu_1 - \mu_2)^T \Omega_{11}^{-1} (\mu_1 - \mu_2)$ is the Mahalanobis distance.

Let

$$\beta = A_{11}\Omega_{12}\Omega_{22}^{-1}, \quad \psi = \beta S_{22} \beta^T \quad \eta = \beta \Omega_{22} \beta^T. \tag{3.3}$$

Denote the i,jth element of matrix [ ] by $[\ ]_{ij}$, and denote the ith coordinate of a vector [ ] by $[\ ]_i$. We have the following theorem (the proof appears in the Appendix).

Theorem 1. *The conditional distribution of the studentized classification statistic W given s and $\mathbf{z}$, when $E[x_0] = \mu_1$, is*

$$\begin{aligned}
&P\Big[\frac{W - D^2/2}{D} \leq u|s, \mathbf{z}\Big] \\
&= \Phi(u) + \phi(u)[\frac{u}{2}(\psi_{11} - \eta_{11}) + [\beta \bar{z}_1]_1] + \frac{1}{n}\phi(u)\Big\{\frac{1}{\Delta}\Big[\frac{n}{n_1}(p-1) \\
&\quad -\frac{n}{n_1}\sum_{i=2}^{p}\eta_{ii} + n\sum_{i=2}^{p}[\beta(\bar{z}_1 - \bar{z}_2)]_i[\beta\bar{z}_1]_i + nu\sum_{i=2}^{p}(\psi_{1i} - \eta_{1i})[\beta(\bar{z}_1 - \bar{z}_2)]_i\Big] \\
&\quad +u\Big[(\frac{3}{4} - p) + (\frac{1}{2} - p)(\psi_{11} - \eta_{11}) - (\psi_{11} - \eta_{11})\sum_{i=2}^{p}(\psi_{ii} - \eta_{ii}) \\
&\quad -(n+1)\sum_{i=1}^{p}(\psi_{1i} - \eta_{1i})^2 - \sum_{i=2}^{p}(\psi_{ii} - \eta_{ii}) + \sum_{i=2}^{p}\psi_{1i}^2 + \psi_{11}\sum_{i=1}^{p}\psi_{ii}\Big] \\
&\quad -\frac{u^3}{8}[2 + 4(\psi_{11} - \eta_{11}) - 2\psi_{11}^2 + (n+2)(\psi_{11} - \eta_{11})^2] \\
&\quad -\frac{u^2}{2}n[\beta\bar{z}_1]_1(\psi_{11} - \eta_{11}) - \frac{u}{2}(\frac{n}{n_1} - \frac{n}{n_1}\eta_{11} + n[\beta\bar{z}_1]_1^2)\Big\} + O(n^{-3/2}), \qquad (3.4)
\end{aligned}$$

where $\phi(\ )$ is the p.d.f. of the standard normal distribution. We have

$$P\Big[\frac{W - D^2/2}{D} \leq u\Big] = E_{s,\mathbf{z}}P\Big[\frac{W - D^2/2}{D} \leq u|s, \mathbf{z}\Big]. \qquad (3.5)$$

3.1. I.I.D. sampling results

Under the i.i.d. model, nS_{22} has central Wishart distribution. The means and covariances of elements of S_{22} are given by $E(S_{22ij}|s, \mathbf{z}) = \Omega_{22ij}$ and

$$\text{Cov}\,(S_{22ij}, S_{22kl}|s, \mathbf{z}) = (\Omega_{22ik}\Omega_{22jl} + \Omega_{22il}\Omega_{22jk})/n \qquad (3.6)$$

(see e.g. Anderson (1957), p.161.). Since $\bar{z}_1, \bar{z}_2$ and S_{22} are mutually independent under the i.i.d. model design, we have

$$E_{s,\mathbf{z}}([\beta\bar{z}_1]_i^2) = \text{Var}_{\,s,\mathbf{z}}([\beta\bar{z}_1]_i) + (E_{s,\mathbf{z}}[\beta\bar{z}_1]_i)^2 = \mu_z^2 + \frac{1}{n_1}\eta_{ii}. \qquad (3.7)$$

By (3.6),

$$E_{s,\mathbf{z}}[\psi_{1i}^2] = \text{Var}_{\,s,\mathbf{z}}(\psi_{1i}) + [E_{s,\mathbf{z}}(\psi_{1i})]^2 = (\eta_{1i}^2 + \eta_{11}\eta_{ii})/n + \eta_{1i}^2 \quad i = 1, \ldots, p. \qquad (3.8)$$

Hence, the right hand side of (3.5) (excluding the remainder term) is

$$\Phi(u)-\phi(u)\Big\{\frac{1}{n_1}\Big[\frac{u}{2}-\frac{p-1}{\Delta}\Big]+\frac{1}{n}\Big[\frac{u^3}{4}+\Big[p-\frac{3}{4}\Big]u\Big]\Big\}. \tag{3.9}$$

This is Anderson's (1973) asymptotic expansion (excluding the remainder term). However if the sampling design depends on some auxiliary information, as discussed in Section 2, we claim that Anderson's asymptotic expansion may have error $O(1)$.

3.2. Stratified sampling results

Suppose that the auxiliary variable $z_{\alpha i}$ is one-dimensional. Suppose further that L equal-sized strata are formed according to increasing values of $z_{\alpha i}$ $i = 1,\ldots,N_\alpha, \alpha = 1,2$, for groups G_1 and G_2. Then a stratified simple random sample with $\tau_h n_\alpha$ units from the hth stratum of group G_α is drawn, where $\alpha = 1,2, \tau_h > 0$ and $\sum \tau_h = 1$. Let $m = n_1 + n_2$. Following Skinner, Holmes and Smith (1986), for each $h = 1,\ldots,L$, we may treat the $z'_{hi}s$ from the hth stratum as a simple random sample from a normal distribution truncated at the $100(h-1)/L$ and $100h/L$ percentage points. Suppose that $\tau_h{}'s$ are fixed as $n_1 \to \infty$ and $n_2 \to \infty$. Then analogous to Skinner, Holmes and Smith (1986) for the one population case, it can be proved that $E_{s,\mathbf{z}}(S_{22}) = plim(S_{22}) + O(m^{-1})$ and

$$plim(S_{22}) = \Omega_{22}\Big[\sum \tau_h((\mu_h-\tilde{\mu})^2+\sigma_h^2)\Big], \tag{3.10}$$

where

$$\tilde{\mu} = \sum_h \tau_h\mu_h. \tag{3.11}$$

Here μ_h is the mean and σ_h is the standard deviation of the standard normal distribution truncated at its $100(h-1)/L$ and $100h/L$ percentage points. Since $\sum_h \mu_h = 0$ and $\sum_h(\mu_h^2+\sigma_h^2)/L = 1$, (3.10) shows that under proportionate stratified random sampling $\tilde{\mu} = 0$, and hence the estimator S_{22} is consistent. If the allocation is disproportionate, however, S_{22} has a bias of $O(1)$. Hence, ψ_{ij} is not a consistent estimator of η_{ij} and has a bias $O(1)$ under the same sampling scheme, i.e., we have $E_{s,\mathbf{z}}(\psi_{11}) = plim(\psi_{11}) + O(n^{-1})$ and $plim(\psi_{11}-\eta_{11}) = O(1)$. This implies $\phi(u)[\frac{u}{2}E(\psi_{11}-\eta_{11})] = O(1)$. By comparing the expectation of (3.4) (with respect to $s,\mathbf{z}$) with (3.9), it appears that Anderson's asymptotic expansion has a remainder term with order $O(1)$ and hence is not good for estimating the distribution of the studentized W under disproportionate allocation.

The situation can be modified as follows. Following the arguments above, we have $\mathrm{Var}_{s,\mathbf{z}}(\bar{z}_1) = \sum_h \tau_h\sigma_h^2/n_1$ and $\mathrm{Var}_{s,\mathbf{z}}(\psi_{11}) = O(n^{-1})$. Hence the asymptotic expansion of the distribution of the studentized W under stratified simple random sampling is

$$
\begin{aligned}
&P\Big[\frac{W-D^2/2}{D}\le u|x_0\in G_1\Big]\\
&=\Phi(u)+\phi(u)\Big\{\frac{1}{n_1}\Big[\frac{p-1}{\Delta}-\frac{u}{2}\Big]+\frac{u}{2}\zeta_{11}+[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}\\
&+\frac{1}{\Delta n_1}\sum_{i=2}^{p}[\beta(\sum_h\tau_h\sigma_h^2-1)\Omega_{22}\beta^T]_{ii}-\frac{u^3}{8}\zeta_{11}^2-\frac{u^2}{2}[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}\zeta_{11}\\
&-\frac{u}{2}\Big[\frac{1}{n_1}+\frac{1}{n_1}[\beta(\sum_h\tau_h\sigma_h^2-1)\Omega_{22}\beta^T]_{11}+[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}^2\Big]\Big\}+O(n^{-1})\\
&=\Phi(u)+\phi(u)\Big\{\frac{u}{2}\zeta_{11}+[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}-\frac{u}{2}\Big[\frac{u}{2}\zeta_{11}+[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}\Big]^2\Big\}+O(n^{-1}).
\end{aligned}\tag{3.12}
$$

In (3.12), ζ_{11} is the $1,1$th element of $\zeta=\beta\Omega_{22}^*\beta^T=\beta\Omega_{22}[(\sum\tau_h[(\mu_h-\tilde{\mu})^2+\sigma_h^2])-1]\beta^T$, since $1/n_1=O(n^{-1})$. The expression in (3.12) depends on the unknown parameters. Replacing these unknown parameters by sample estimates, we obtain the approximate rates of misclassification.

If $x_o\in G_2$, interchanging n_1 and n_2 gives

$$
\begin{aligned}
&P\Big[-\frac{W+D^2/2}{D}\le u|s,\mathbf{z},x_0\in G_2\Big]\\
&=\Phi(u)+\phi(u)[\frac{u}{2}(\psi_{11}-\eta_{11})+[\beta\bar{z}_2]_1]+\frac{1}{n}\phi(u)\Big\{\frac{1}{\Delta}\Big[\frac{n}{n_2}(p-1)\\
&-\frac{n}{n_2}\sum_{i=2}^{p}\eta_{ii}+n\sum_{i=2}^{p}[\beta(\bar{z}_2-\bar{z}_1)]_i[\beta\bar{z}_2]_i+nu\sum_{i=2}^{p}(\psi_{1i}-\eta_{1i})[\beta(\bar{z}_2-\bar{z}_1)]_i\Big]\\
&+u\Big[(\frac{3}{4}-p)+(\frac{1}{2}-p)(\psi_{11}-\eta_{11})-(\psi_{11}-\eta_{11})\sum_{i=2}^{p}(\psi_{ii}-\eta_{ii})\\
&-(n+1)\sum_{i=1}^{p}(\psi_{1i}-\eta_{1i})^2-\sum_{i=2}^{p}(\psi_{ii}-\eta_{ii})+\sum_{i=2}^{p}\psi_{1i}^2+\psi_{11}\sum_{i=1}^{p}\psi_{ii}\Big]\\
&-\frac{u^3}{8}[2+4(\psi_{11}-\eta_{11})-2\psi_{11}^2+(n+2)(\psi_{11}-\eta_{11})^2]\\
&-\frac{u^2}{2}n[\beta\bar{z}_2]_1(\psi_{11}-\eta_{11})-\frac{u}{2}(\frac{n}{n_2}-\frac{n}{n_2}\eta_{11}+n[\beta\bar{z}_2]_1^2)\Big\}+O(n^{-3/2}).
\end{aligned}\tag{3.13}
$$

Hence,

$$
\begin{aligned}
&P\Big[-\frac{W+D^2/2}{D}\le u|x_0\in G_2\Big]\\
&=\Phi(u)+\phi(u)\Big\{\frac{u}{2}\zeta_{11}+[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}-\frac{u}{2}\Big[\frac{u}{2}\zeta_{11}+[\beta\tilde{\mu}\Omega_{22}^{1/2}]_{11}\Big]^2\Big\}+O(n^{-1}),
\end{aligned}\tag{3.14}
$$

since $1/n_2 = O(n^{-1})$.

3.3. A numerical example

An example will show how the rates of misclassification of Fisher's sample discriminant function are affected by various allocations of stratum sample sizes in stratified random sampling. Our example also shows the accuracy of the corrected expansion formula (3.14) under several stratified sampling designs, and shows that the error of Anderson's expansion (3.9) can be very large if a highly disproportionate sampling design is treated as a simple random sample design.

The numerical example is partly based on some of the results of the simulation study described in Leu and Tsui (1997). Two finite populations, each having 5000 units, were created using the following superpopulation model. The associated vector for each unit, $(\mathbf{x}, \mathbf{z})$, was generated from (2.1) with $\mu_1 = (1, 2)^T$, $\mu_2 = (5, 4)^T$, $\mu_z = 3$ and a common covariance matrix equal to

$$\begin{pmatrix} 8 & 2 & 5 \\ 2 & 10 & 5 \\ 5 & 5 & 10 \end{pmatrix}.$$

The Mahalanobis distance, Δ^2, is 2.105. Each finite population was stratified into five equal strata of same size according to the increasing values of $z_{\alpha i}, i = 1, 2$(see Section 2). The notation $(m_1, \ldots, m_5)$denotes a stratified random sampling design with units selected from the hth stratum, $h = 1, ..., 5$. Seven sampling designs were considered: design D1(100) using independent simple random sampling (SRS) of size 100; design D2(20,20,20,20,20) using proportional allocation; designs with increasing allocation: D3(5,15,20,25,35), D4(5,5,10,30,50), and D5(1,2,5,16,75); U-shaped allocation designs D6(30,15,10,15,30) and D7(44,5,2, 5,44). The same sampling design was applied to each of the two populations with sample sizes of 100 for each population.

When n_1 and n_2, are equal and the costs of misclassification for the two populations are the same, the optimal choice of k in the sample discriminant rule, $W(x_0) \leq (>)k$, is $k = 0$. Since $\Delta^2 = 2.105$, the usual rates of misclassification, $p(1|2)$ and $p(2|1)$, of Fisher's sample discriminant function are equal to

$$p(2|1) = p(1|2) = \Phi(-\Delta/2) = \Phi(-0.7254) = 0.234.$$

The transformation matrix A_{11} and β given in Theorem 1 are

$$A_{11} = \begin{pmatrix} 0.15684 & 0.25377 \\ 0.32708 & -0.20215 \end{pmatrix}$$

and $\beta = (0.2053, 0.0625)^T$. Since the true Mahalanobis distance Δ^2 is not known in practice, we replace it by many possible sample estimates of the Mahalanobis distance, D^2, around the true value $\Delta^2 = 2.105$, in the expansions (3.9), (3.12) and (3.14). The estimated rates of misclassification under different sampling designs are displayed in Table 1. We use Anderson's asymptotic expansion (3.9) for design D1 and our correction (3.12) or (3.14) for other designs. Anderson's result for design D1 can also be regarded as the "naive" estimator for the other designs.

Table 1. Achieved rates of misclassification (in %) based on asymptotic expansion (3.9) for design D1 and asymptotic expansion (3.14) for designs D2, ..., D7. Here D^2 is the estimate of the Mahalanobis distance.

-D/2	D1(100) Anderson	D2(20,20, 20,20,20)	D3(5,15, 20,25,35)	D4(5,5, 10,30,50)	D5(1,3, 5,16,75)	D6(30,15 10,15,30)	D7(44,5 2,5,44)
-1.725	4.523	4.226	8.901	12.331	18.658	3.075	2.044
-1.575	6.105	5.763	11.488	15.645	23.101	4.384	3.031
-1.425	8.093	7.708	14.563	19.483	28.051	6.112	4.427
-1.325	9.670	9.259	16.891	22.325	31.595	7.534	5.638
-1.225	11.463	11.029	19.441	25.381	35.298	9.196	7.108
-1.125	13.483	13.030	22.209	28.633	39.123	11.117	8.868
-1.025	15.736	15.268	25.184	32.059	43.031	13.309	10.943
-0.925	18.226	17.748	28.353	35.633	46.978	15.782	13.351
-0.825	20.949	20.469	31.696	39.325	50.922	18.540	16.107
-0.775	22.397	21.917	33.426	41.205	52.879	20.025	17.616
-0.725	23.900	23.423	35.190	43.101	54.819	21.579	19.212
-0.675	25.458	24.984	36.986	45.010	56.737	23.201	20.894
-0.625	27.067	26.599	38.809	46.926	58.629	24.888	22.660
-0.525	30.432	29.979	42.522	50.764	62.316	28.450	26.433
-0.425	33.972	33.542	46.298	54.578	65.847	32.238	30.502
-0.325	37.661	37.259	50.103	58.334	69.195	36.221	34.827
-0.225	41.467	41.099	53.902	61.998	72.342	40.358	39.359
-0.125	45.355	45.026	57.661	65.541	75.272	44.607	44.039
-0.025	49.289	49.003	61.348	68.936	77.979	48.918	48.804
0.025	51.261	50.997	63.154	70.572	79.248	51.082	51.197
0.125	55.191	54.974	66.675	73.705	81.618	55.393	55.961
0.225	59.070	58.901	70.052	76.644	83.770	59.642	60.641
0.375	64.713	64.617	74.797	80.668	86.609	65.792	67.365
0.525	70.048	70.021	79.109	84.213	89.021	71.550	73.567

Using the substitution principle (see Arnold (1981), p.402) in their simulation study, Leu and Tsui (1997) find that the rates of misclassification of sampling designs D1, D2, D3, D4, D5, D6 and D7 are 23.8 %, 23.8 %, 33.2 %, 40.3 %, 49.3 %, 23.8 % and 23.7 %, respectively. These rates are regarded as the true rates of misclassification for the above designs. Comparing the rates of misclassification in Table 1 and the simulation results in Leu and Tsui (1997), we observe that the absolute error of the estimated rate of misclassification using our corrected expansion is 0.1% for design D1, 0.4% for design D2 and 2% to 5.5% for designs D3 to D7 when $-D/2 = -.725$ is close to the true $-\Delta/2$ value. However, under designs with increasing allocation (D3, D4, and D5), it appears that the higher the increasing rate the higher the design effect on the rates of misclassification. The absolute error of Anderson's asymptotic expansion is seen to be as high as 49.3% - 23.9% = 25.4% under design D5 when $-D/2 = -.725$. Under proportional allocation design (D2) and U-shaped allocation designs (D6 and D7), it appears that the design effect is not as serious, and Anderson's approximation for the rates of misclassification may be appropriate in these situations.

We conclude that the complex survey design can affect the distribution of the discriminant function and also the rates of misclassification. We have shown that, under a given model, Anderson's (1973) asymptotic expansion of the distribution of W has an error of order $O(1)$, not of order $O(n^{-2})$, in a disproportional stratified random sampling design. The expansion formulas in (3.12) and (3.14) take into account auxiliary information and provide a reasonable correction of Anderson's expansion. The design effect on discriminant analysis is an interesting subject for further research.

Appendix: The Proof of Theorem 1

We first consider the asymptotic expansion of the conditional expectation, $E_{\bar{x}_1,\bar{x}_2,S_{11}|s,\mathbf{z}}$ of $\Phi(\cdot)$, in (3.2). Recall that we can choose A_{11}, Q_1 and Q_2 so as to transform Ω_{11} to I, $\mu_1 - \mu_2$ to $\delta = (\Delta, 0, \ldots, 0)^T$ and μ_z to 0, where $\Delta^2 = (\mu_1 - \mu_2)^T \Omega_{11}^{-1}(\mu_1 - \mu_2)$ is the Mahalanobis distance.

Let A, B, V be defined by

$$\bar{x}_1 - \bar{x}_2 = \delta + \frac{1}{(n)^{1/2}} A, \qquad \bar{x}_1 = \frac{1}{(n)^{1/2}} B, \tag{A.1}$$

$$S_{11} = \Omega_{11.2} + \frac{1}{(n)^{1/2}} V, \tag{A.2}$$

where $\Omega_{11.2} = A_{11}(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})A_{11}^T = I - A_{11}\Omega_{12}\Omega_{22}^{-1}\Omega_{21}A_{11}^T = I - \beta\Omega_{22}\beta^T = I - \eta$, and β, ψ, η are as given in (3.3). The (joint) conditional distribution of $(A^T, B^T)^T$ given s and $\mathbf{z}$ is

$$N\Big[\begin{pmatrix} (n)^{1/2}\beta(\bar{z}_1 - \bar{z}_2) \\ (n)^{1/2}\beta\bar{z}_1 \end{pmatrix}, \begin{pmatrix} n(\frac{1}{n_1} + \frac{1}{n_2})\Omega_{11.2}, & \frac{n}{n_1}\Omega_{11.2} \\ \frac{n}{n_1}\Omega_{11.2} & \frac{n}{n_1}\Omega_{11.2} \end{pmatrix}\Big]. \tag{A.3}$$

Then (3.2) becomes

$$
\begin{aligned}
&P\Big[\frac{W-D^2/2}{D}\le u\Big]\\
&=E_{s,\mathbf{z}}E_{\bar{x}_1,\bar{x}_2,S_{11}|s,\mathbf{z}}\Phi\Big\{\Big[u[(\delta+\frac{A}{(n)^{1/2}})^T(I+\frac{V}{(n)^{1/2}})^{-1}(\delta+\frac{A}{(n)^{1/2}})]^{1/2}\\
&\qquad +(\delta+\frac{A}{(n)^{1/2}})^T(I+\frac{V}{(n)^{1/2}})^{-1}\frac{B}{(n)^{1/2}}\Big]\\
&\div[(\delta+\frac{A}{(n)^{1/2}})^T(I+\frac{V}{(n)^{1/2}})^{-2}(\delta+\frac{A}{(n)^{1/2}})]^{1/2}\Big\}. \qquad (A.4)
\end{aligned}
$$

We can write

$$
\begin{aligned}
(I+\frac{V}{(n)^{1/2}})^{-1}&=I-\frac{1}{(n)^{1/2}}V+\frac{1}{n}V^2-\frac{1}{n(n)^{1/2}}V^3+\frac{1}{n^2}V^4\\
&\quad-\frac{1}{n^{5/2}}V^5(I+\frac{V}{(n)^{1/2}})^{-1}, \qquad (A.5)\\
(I+\frac{V}{(n)^{1/2}})^{-2}&=I-\frac{2}{(n)^{1/2}}V+\frac{3}{n}V^2-\frac{4}{n(n)^{1/2}}V^3+\frac{5}{n^2}V^4\\
&\quad-\frac{1}{n^{5/2}}(6V^5+\frac{5}{(n)^{1/2}}V^6)(I+\frac{V}{(n)^{1/2}})^{-2}. \qquad (A.6)
\end{aligned}
$$

Plugging (A.5) and (A.6) into (A.4), some simple algebra as in Anderson (1973) reduces the argument of $\Phi(\cdot)$ in (A.4) to

$$
\begin{aligned}
&u+\frac{1}{(n)^{1/2}}\Big[\frac{u}{2\Delta^2}\delta^TV\delta+\frac{1}{\Delta}\delta^TB\Big]+\frac{1}{n}\Big[\frac{u}{\Delta^2}(\delta^TVA-\delta^TV^2\delta)\\
&+\frac{u}{\Delta^4}(-\delta^TA\delta^TV\delta+\frac{7}{8}(\delta^TV\delta)^2)+\frac{1}{\Delta}(A^TB-\delta^TVB)\\
&+\frac{1}{\Delta^3}(-\delta^TAB^T\delta+\delta^TB\delta^TV\delta)\Big]+R_{1n}(A,B,V). \qquad (A.7)
\end{aligned}
$$

Here $R_{1n}(A,B,V)$ is a remainder term consisting of $\frac{1}{n(n)^{1/2}}$ times a homogeneous polynomial (not depending on n) of degree 3 in the elements of A, B and V, plus $\frac{1}{n^2}$ times a homogeneous polynomial of degree 4 plus a remainder term which is $O(n^{-5/2})$ for fixed A, B and V.

Let

$$C(B,V)=\frac{u}{2\Delta^2}\delta^TV\delta+\frac{1}{\Delta}\delta^TB \qquad (A.8)$$

$$
\begin{aligned}
D(A,B,V)&=\frac{u}{\Delta^2}(\delta^TVA-\delta^TV^2\delta)+\frac{u}{\Delta^4}(-\delta^TA\delta^TV\delta+\frac{7}{8}(\delta^TV\delta)^2)\\
&\quad+\frac{1}{\Delta}(A^TB-\delta^TVB)+\frac{1}{\Delta^3}(-\delta^TAB^T\delta+\delta^TB\delta^TV\delta). \qquad (A.9)
\end{aligned}
$$

A Taylor's expansion of $\Phi(\ )$ in (A.4) gives

$$\begin{aligned}&\Phi\Big[u+\frac{1}{(n)^{1/2}}C(B,V)+\frac{1}{n}D(A,B,V)+R_{1n}(A,B,V)\Big]\\&=\Phi(u)+\phi(u)\Big[\frac{1}{(n)^{1/2}}C(B,V)+\frac{1}{n}[D(A,B,V)-\frac{u}{2}C^2(B,V)]\Big]\\&\quad+\frac{1}{n(n)^{1/2}}R_2(A,B,V)+\frac{1}{n^2}R_3(A,B,V)+R_{4n}(A,B,V),\end{aligned}\qquad\text{(A.10)}$$

where $R_2(A,B,V)$ is a homogeneous polynomial (not depending on n but depending on u) of degree 3 and $R_3(A,B,V)$ is a polynomial of degree greater than or equal to 4 in the elements of A,B and V. $R_{4n}(A,B,V)$ is a remainder term which is $O(n^{-5/2})$ for fixed A,B and V.

Since the third-order and fourth-order absolute moments of A,B and V exist and are bounded, the contribution of $\frac{1}{n(n)^{1/2}}R_2(A,B,V)$ is $O(n^{-3/2})$. Similarly, the contribution of $n^{-2}R_3(A,B,V)$ is $O(n^{-2})$. Thus, the conditional expectation, $E_{\bar{x}_1,\bar{x}_2,S_{11}|s,\mathbf{z}}$ of $\Phi(\cdot)$, can be expressed as

$$\Phi(u)+\phi(u)\Big[\frac{1}{(n)^{1/2}}E[C(B,V)]+\frac{1}{n}E[D(A,B,V)-\frac{u}{2}C^2(B,V)]\Big]+O(n^{-3/2}).\qquad\text{(A.11)}$$

Note that nS_{11} given s and $\mathbf{z}$ has a non-central Wishart distribution with covariance matrix $\Omega_{11.2}$ and non-centrality parameter $n\beta S_{22}\beta^T$. (The proof of this property follows from Skinner (1982).) The conditional mean of S_{11} is therefore

$$E[S_{11}|s,\mathbf{z}]=\Omega_{11.2}+\beta(S_{22}-\Omega_{22})\beta^T.\qquad\text{(A.12)}$$

The covariance of the elements of S_{11} is given by the following lemma in Skinner (1982).

Lemma. (Skinner (1982))

$$\begin{aligned}\mathrm{Cov}\,(S_{11ij},S_{11kl}|s,\mathbf{z})=(&\Omega_{11.2ik}\Omega_{11.2jl}+\Omega_{11.2il}\Omega_{11.2jk}+\Omega_{11.2jl}\psi_{ik}+\Omega_{11.2jk}\psi_{ik}\\&+\Omega_{11.2il}\psi_{jk}+\Omega_{11.2ik}\psi_{jl})/n.\end{aligned}\qquad\text{(A.13)}$$

Alternatively,

$$\mathrm{Cov}\,(S_{11ij},S_{11kl}|s,\mathbf{z})=(\Omega^*_{11ik}\Omega^*_{11jl}+\Omega^*_{11il}\Omega^*_{11jk}-\psi_{ik}\psi_{jl}-\psi_{il}\psi_{jk})/n,\qquad\text{(A.14)}$$

where $\Omega^*_{11}=\Omega_{11}+\beta(S_{22}-\Omega_{22})\beta^T$.

Since (A,B) and V are independent and Ω_{11} has been transformed to I, using the Lemma and (A.12), we have (A.15)-(A.25).

$$E[\delta^TV\delta|s,\mathbf{z}]=\Delta^2(n)^{1/2}[\beta(S_{22}-\Omega_{22})\beta^T]_{11}=\Delta^2(n)^{1/2}(\psi_{11}-\eta_{11}),\qquad\text{(A.15)}$$

$$E[\delta^T B|s,\mathbf{z}] = \Delta(n)^{1/2}[\beta\bar{z}_1]_1, \tag{A.16}$$

$$\begin{aligned} E[\delta^T VA|s,\mathbf{z}] &= \Delta n \sum_{i=1}^{p}[\beta(S_{22}-\Omega_{22})\beta^T]_{1i}[\beta(\bar{z}_1-\bar{z}_2)]_i, \\ &= \Delta n \sum_{i=1}^{p}(\psi_{1i}-\eta_{1i})[\beta(\bar{z}_1-\bar{z}_2)]_i, \end{aligned} \tag{A.17}$$

$$\begin{aligned} E[\delta^T V^2\delta|s,\mathbf{z}] &= \Delta^2\sum_{i=1}^{p}E[v_{1i}^2|s,\mathbf{z}] = \Delta^2\sum_{i=1}^{p}[\mathrm{Var}\,(v_{1i}|s,\mathbf{z}) + (E(v_{1i}|s,\mathbf{z}))^2] \\ &= \Delta^2\sum_{i=1}^{p}(\Omega^*_{111i}\Omega^*_{11i1} + \Omega^*_{1111}\Omega^*_{11ii} - \psi_{1i}\psi_{i1} - \psi_{11}\psi_{ii} \\ &\qquad + n[\beta(S_{22}-\Omega_{22})\beta^T]_{1i}[\beta(S_{22}-\Omega_{22})\beta^T]_{1i}) \\ &= \Delta^2\Big[1+p+(p+2)(\psi_{11}-\eta_{11}) + \sum_{i=1}^{p}(\psi_{ii}-\eta_{ii}) + (\psi_{11}-\eta_{11})\sum_{i=1}^{p}(\psi_{ii}-\eta_{ii}) \\ &\qquad - \sum_{i=1}^{p}\psi_{1i}^2 - \psi_{11}\sum_{i=1}^{p}\psi_{ii} + (n+1)\sum_{i=1}^{p}(\psi_{1i}-\eta_{1i})^2\Big], \end{aligned} \tag{A.18}$$

$$\begin{aligned} E[\delta^T A\delta^T V\delta|s,\mathbf{z}] &= E[\delta^T A|s,\mathbf{z}]E[\delta^T V\delta|s,\mathbf{z}] = n\Delta^3[\beta(\bar{z}_1-\bar{z}_2)]_1[\beta(S_{22}-\Omega_{22})\beta^T]_{11} \\ &= n\Delta^3(\psi_{11}-\eta_{11})[\beta(\bar{z}_1-\bar{z}_2)]_1, \end{aligned} \tag{A.19}$$

$$\begin{aligned} E[(\delta^T V\delta)^2|s,\mathbf{z}] &= \Delta^4[\mathrm{Var}\,(v_{11}) + (Ev_{11})^2] \\ &= \Delta^4(2\Omega^{*2}_{1111} - 2\psi_{11}^2 + n[\beta(S_{22}-\Omega_{22})\beta^T]_{11}^2) \\ &= \Delta^4[2+4(\psi_{11}-\eta_{11}) - 2\psi_{11}^2 + (n+2)(\psi_{11}-\eta_{11})^2], \end{aligned} \tag{A.20}$$

$$\begin{aligned} E[A^T B|s,\mathbf{z}] &= \mathrm{tr}[\mathrm{Cov}\,(A,B|s,\mathbf{z}) + E(B|s,\mathbf{z})E(A^T|s,\mathbf{z})] \\ &= \mathrm{tr}[\frac{n}{n_1}\Omega_{11.2} + n\beta\bar{z}_1(\bar{z}_1-\bar{z}_2)^T\beta^T] \\ &= \frac{n}{n_1}p - \frac{n}{n_1}\sum_{i=1}^{p}\eta_{ii} + n\sum_{i=1}^{p}[\beta(\bar{z}_1-\bar{z}_2)]_i[\beta\bar{z}_1]_i, \end{aligned} \tag{A.21}$$

$$E[\delta^T VB|s,\mathbf{z}] = \Delta n\sum_{i=1}^{p}[\beta(S_{22}-\Omega_{22})\beta^T]_{1i}[\beta\bar{z}_1]_i = \Delta n\sum_{i=1}^{p}(\psi_{1i}-\eta_{1i})[\beta\bar{z}_1]_i, \tag{A.22}$$

$$E[\delta^T AB^T\delta|s,\mathbf{z}] = \Delta^2(\frac{n}{n_1} - \frac{n}{n_1}\eta_{11} + n[\beta(\bar{z}_1-\bar{z}_2)]_1[\beta\bar{z}_1]_1), \tag{A.23}$$

$$\begin{aligned} E[\delta^T B \delta^T V \delta)|s, \mathbf{z}] &= \Delta^3 n [\beta \bar{z}_1]_1 [\beta(S_{22} - \Omega_{22})\beta^T]_{11} \\ &= \Delta^3 n [\beta \bar{z}_1]_1 (\psi_{11} - \eta_{11}), \end{aligned} \tag{A.24}$$

and

$$E[\delta^T B B^T \delta | s, \mathbf{z}] = \Delta^2 (\frac{n}{n_1} - \frac{n}{n_1}\eta_{11} + n[\beta \bar{z}_1]_1^2). \tag{A.25}$$

Taking the expectation of (A.8) and using (A.15) and (A.16), we have

$$E[C(B, V)|s, \mathbf{z}] = (n)^{1/2} \frac{u}{2} (\psi_{11} - \eta_{11}) + (n)^{1/2} [\beta \bar{z}_1]_1. \tag{A.26}$$

Also, taking the expectation of (A.9) and using results of (A.17)-(A.24), we have

$$\begin{aligned} &E[D(A, B, V)|s, \mathbf{z}] \\ &= \frac{un}{\Delta} \sum_{i=1}^{p} (\psi_{1i} - \eta_{1i})[\beta(\bar{z}_1 - \bar{z}_2)]_i - u[1 + p + (p+2)(\psi_{11} - \eta_{11}) \\ &\quad + \sum_{i=1}^{p} (\psi_{ii} - \eta_{ii}) + (\psi_{11} - \eta_{11}) \sum_{i=1}^{p} (\psi_{ii} - \eta_{ii}) - \sum_{i=1}^{p} \psi_{1i}^2 - \psi_{11} \sum_{i=1}^{p} \psi_{ii} \\ &\quad +(n+1) \sum_{i=1}^{p} (\psi_{1i} - \eta_{1i})^2] - n\frac{u}{\Delta}(\psi_{11} - \eta_{11})[\beta(\bar{z}_1 - \bar{z}_2)]_1 \\ &\quad + \frac{7u}{8}[2 + 4(\psi_{11} - \eta_{11}) - 2\psi_{11}^2 + (n+2)(\psi_{11} - \eta_{11})^2] \\ &\quad + \frac{1}{\Delta}(\frac{n}{n_1}p - \frac{n}{n_1}\sum_{i=1}^{p} \eta_{ii} + n\sum_{i=1}^{p} [\beta(\bar{z}_1 - \bar{z}_2)]_i [\beta \bar{z}_1]_i) - n \sum_{i=1}^{p} (\psi_{1i} - \eta_{1i})[\beta \bar{z}_1]_i \\ &\quad - \frac{1}{\Delta}(\frac{n}{n_1} - \frac{n}{n_1}\eta_{11} + n[\beta(\bar{z}_1 - \bar{z}_2)]_1 [\beta \bar{z}_1]_1) + n[\beta \bar{z}_1]_1 (\psi_{11} - \eta_{11}) \\ &= \frac{1}{\Delta}\Big[\frac{n}{n_1}(p-1) - \frac{n}{n_1}\sum_{i=2}^{p} \eta_{ii} + n\sum_{i=2}^{p} [\beta(\bar{z}_1 - \bar{z}_2)]_i [\beta \bar{z}_1]_i + nu \sum_{i=2}^{p} (\psi_{1i} - \eta_{1i})[\beta(\bar{z}_1 - \bar{z}_2)]_i\Big] \\ &\quad + u\Big[(\frac{3}{4} - p) + (\frac{1}{2} - p)(\psi_{11} - \eta_{11}) - (\psi_{11} - \eta_{11}) \sum_{i=2}^{p} (\psi_{ii} - \eta_{ii}) \\ &\qquad -(n+1)\sum_{i=1}^{p} (\psi_{1i} - \eta_{1i})^2 - \sum_{i=2}^{p} (\psi_{ii} - \eta_{ii}) + \sum_{i=2}^{p} \psi_{1i}^2 + \psi_{11} \sum_{i=1}^{p} \psi_{ii}\Big]. \end{aligned} \tag{A.27}$$

Applying (A.20), (A.24) and (A.25) to the expectation of the square of (A.8), we get

$$\begin{aligned} E[C^2(B, V)|s, \mathbf{z}] &= \frac{u^2}{4}[2 + 4(\psi_{11} - \eta_{11}) - 2\psi_{11}^2 + (n+2)(\psi_{11} - \eta_{11})^2] \\ &\quad + un[\beta \bar{z}_1]_1 (\psi_{11} - \eta_{11}) + (\frac{n}{n_1} - \frac{n}{n_1}\eta_{11} + n[\beta \bar{z}_1]_1^2). \end{aligned} \tag{A.28}$$

Substituting (A.26), (A.27) and (A.28) into (A.11), yields Theorem 1.

References

Anderson, T. W. (1957). *An Introduction to Multivariate Statistical Analysis.* John Wiley, New York.

Anderson, T. W. (1973). An asymptotic expansion of the distribution of the studentized classification statistic W. *Ann. Statist.* **1**, 964-972.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* John Wiley, New York.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics* **7**, 179-188.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*, Third Edition. Prentice Hall, New Jersey.

Lehtonen, R. and Pahkinen, E. J. (1994). *Practical Methods for Design and Analysis of Complex Surveys.* John Wiley, Chichester.

Leu, C. H. and Tsui, K. W. (1997). Discriminant analysis of survey data. *J. Statist. Plann. Inference* **60**, 273-290.

Nathan, G. and Holt, D. (1980). The effect of survey design on regression analysis. *J. Roy. Statist. Soc. Ser. B* **41**, 1491-1499.

Okamoto, M. (1963). An asymptotic expression for the distribution of the linear discriminant function. *Ann. Math. Statist.* **34**, 1286-1301. (Correction (1968) **39**, 1358-1359.)

Rao, J. N. K. Sutradhar, B. C. and Yue, K. (1993). Generalized least squares F test in regression analysis with two-stage cluster samples. *J. Amer. Statist. Assoc.* **88**, 1388-1391.

Skinner, C. J. (1982). Multivariate analysis of sample survey data. Unpublished Ph.D. thesis, University of Southampton, Dept. of Social Statistics.

Skinner, C. J. (1986). Regression estimation and post-stratification in factor analysis. *Psychometrika* **51**, 347-356.

Skinner, C. J., Holmes, D. J. and Smith, T. M. F. (1986). The effect of sample design on principal component analysis. *J. Amer. Statist. Assoc.* **81**, 789-798.

Skinner, C. J., Holmes, D. J. and Smith, T. M. F. (1989). *Analysis of Complex Surveys.* John Wiley, New York.

Department of Statistics, University of Wisconsin, 1210 W. Dayton Street, Madison, WI 53706-1685, U.S.A.

E-mail: kwtsui@stat.wisc.edu

Department of Statistics, National Cheng-Kung University, Tainan, Taiwan.

E-mail: chleu@ibm.stat.ncku.edu.tw

(Received September 1996; accepted April 1998)