Multivariate Behrens-Fisher problem using means of auxiliary variables

Jianqi Yu, Kalimuthu Krishnamoorthy & Bin Wang

Multivariate Behrens-Fisher problem using means of auxiliary variables

Jianqi Yu[a], Kalimuthu Krishnamoorthy[b], and Bin Wang[a]

[a]College of Science, Guilin University of Technology, Guilin, Guangxi, China; [b]Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana, USA

ABSTRACT

The authors considered the problem of testing equality of two multivariate normal mean vectors when the covariance matrices are unknown and arbitrary. Given auxiliary variables with known means, the authors proposed a pivotal quantity which is similar to the Hotelling T^2 statistic and obtained a satisfying approximation to its distribution. The authors also outlined hypothesis testing and confidence estimation based on the approximate distribution. The merits of the test were studied using Monte Carlo simulation. Monte Carlo studies indicated that the test is very satisfactory even for moderately small samples. At last, the authors illustrated the proposed methods by an example.

1. Introduction

In statistical practice, auxiliary variables are very common and they are usually highly related to the study variables. For example, if the study variable is sleeping time, auxiliary variables can be age, blood pressure, gender, etc. Making full use of auxiliary variables can improve the accuracy of inference. For instance, we often use sample mean to estimate population mean, but when auxiliary variables are available, there are other estimators better than the sample mean. Cochran (1940) proposed the ratio estimation of the population mean in Simple Random Sampling (SRSWOR), and pointed out that the ratio estimation reached the best when study variables and auxiliary variables were highly positively correlated and the regression line of the research variable and auxiliary variable passes through the origin. The product estimation was first proposed by Robson (1957) and rediscovered by Murthy (1964), which is suitable for the situation where the study variables and auxiliary variables are highly negatively correlated. The regression estimation proposed by Watson (1937) is suitable for the case that the regression line does not pass through the origin. These popular methods for estimation of the population mean are actually using the known mean of an auxiliary variable. Later, scholars proposed various methods using auxiliary variables to improve the estimation of population mean in Simple Random Sampling (SRSWOR). For details, see Yan and

Tian (2010), Khan et al. (2015), Kadilar (2016), etc. In this article, we consider the Behrens-Fisher problem with auxiliary variables.

The Behrens-Fisher problem is to infer the difference between the means of two normal populations without assuming that the population variances are equal. It has been solved well in both univariate and multivariate situations. Even though some exact methods are available, they have disadvantages to some extent. For example, because the precise method of Scheffé (1943) involves the random pairing of observations from two independent samples, the order of the observations affects the results. Therefore, Scheffé (1970) himself advises against using this exact method. Bennett (1950) proposed an exact method for multivariate cases, similar to Scheffé's method for univariate cases. Similarly, due to the same reasons as the univariate case, it is generally not recommended to use this method in practical applications. Many authors have proposed approximate solutions using the idea of Welch's (1947) to approximate degrees of freedom, see James (1954), Yao (1965), and Johansen (1980).

Let $x_1, ..., x_N$ be a random sample from a p-variate normal populations with mean vector μ and covariance matrix Σ_1, i.e., $N_p(\mu, \Sigma_1)$. Let $y_1, ..., y_M$ be a random sample from $N_p(\beta, \Sigma_2)$. It is assumed that Σ_1 and Σ_2 are unknown and arbitrary positive definite matrices. Let us consider the problem of testing

$$H_0 : \mu = \beta \text{ vs. } H_a : \mu \neq \beta. \tag{1}$$

It is easy to see that $H_0 : \mu = \beta$ is equivalent to $H_0 : A\mu = A\beta$ for any non singular matrix A. Hence, a practical solution should be non singular invariant. Some authors have proposed approximate solutions, in which the solutions from James (1954), Yao (1965), and Johansen (1980) are invariant whereas the one from Nel and Van der Merwe (1986) is not. Krishnamoorthy and Yu (2004) proposed a modification to the solution of Nel and van der Merwe so that the resulting test is affine invariant. We call this test the modified Nel van der Merwe (MNV) test. Monte Carlo comparative study in Krishnamoorthy and Yu (2004) showed that in terms of power and control type I error rate, the MNV test is the best among all the invariant tests available so far. In this article, we extend the MNV test to the case in presence of auxiliary variables with known means.

To formulate the problem, let x and y be the study variables that we want to compare, w and t be the auxiliary variables of x and y Meanwhile, let

$$\begin{pmatrix} w \\ x \end{pmatrix} \sim N_{q+p} \left(\begin{pmatrix} c_w \\ \mu \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

independently of

$$\begin{pmatrix} t \\ y \end{pmatrix} \sim N_{q+p} \left(\begin{pmatrix} c_t \\ \beta \end{pmatrix}, \Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right)$$

where c_w and c_t are known constants.

Assume that there are N random samples on $\begin{pmatrix} w \\ x \end{pmatrix}$ and M random samples on $\begin{pmatrix} t \\ y \end{pmatrix}$. That is, we have data matrices as shown below:

$$\begin{matrix} \mathbf{w}_1, ..., \mathbf{w}_N & \mathbf{t}_1, ..., \mathbf{t}_M \\ \mathbf{x}_1, ..., \mathbf{x}_N & \mathbf{y}_1, ..., \mathbf{y}_M \end{matrix} \tag{2}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$'s are $p \times 1$ vectors, while $\mathbf{w}_i$ and $\mathbf{t}_i$'s are $q \times 1$ vector. We consider testing the hypotheses in (1).

The structure of this article is arranged as follows. In the following section, we present the maximum likelihood estimators of the relevant parameters, and propose a statistic for the test. In Sec. 3, we derive an approximation to the distribution of the statistics and outline the procedure for the hypothesis testing. The merits of this test are studied in Sec. 4 through Monte Carlo simulation whose results indicate that the test is very satisfactory even for moderately small samples. The proposed methods are illustrated by an example in Sec. 5.

2. Maximum likelihood estimators of μ, β

Firstly, we derive the maximum likelihood estimator of μ. To finish this, let

$$\mathbf{D} = \begin{pmatrix} \mathbf{w}_1, & \cdots & , \mathbf{w}_N \\ \mathbf{x}_1, & \cdots & , \mathbf{x}_N \end{pmatrix}_{(q+p) \times N} \tag{3}$$

Let $\bar{\mathbf{D}} = \begin{pmatrix} \bar{\mathbf{w}} \\ \bar{\mathbf{x}} \end{pmatrix}$ and $\mathbf{S} = \begin{pmatrix} \mathbf{S}_{ww} & \mathbf{S}_{wx} \\ \mathbf{S}_{xw} & \mathbf{S}_{xx} \end{pmatrix}$ denote the sample mean vector and the sum of squares and sum of products matrix respectively based on $\mathbf{D}$.

Consider the density function of this data. We note that the density of $\mathbf{w}$ and $\mathbf{x}$ can be written as the marginal density of $\mathbf{w}$ times the conditional density of $\mathbf{x}$ given $\mathbf{w}$ (we indicate the density of normal distribution by $n(\cdot)$ here), that is

$$n(\mathbf{w}, \mathbf{x}|\mathbf{c}_w, \mu, \Sigma) = n(\mathbf{w}|\mathbf{c}_w, \Sigma_{11})n(\mathbf{x}|\mu_{2.1} + \mathbf{B}_{2.1}\mathbf{w}, \Sigma_{2.1})$$

where

$$\mathbf{B}_{2.1} = \Sigma_{21}\Sigma_{11}^{-1}, \mu_{2.1} = \mu - \mathbf{B}_{2.1}\mathbf{c}_w, \Sigma_{2.1} = \Sigma_{22} - \mathbf{B}_{2.1}\Sigma_{12} \tag{4}$$

The likelihood function can be written as

$$L(\mu, \Sigma) = \prod_{i=1}^{N} n(\mathbf{w}_i|\mathbf{c}_w, \Sigma_{11}) \prod_{i=1}^{N} n(\mathbf{x}_i|\mu_{2.1} + \mathbf{B}_{2.1}\mathbf{w}_i, \Sigma_{2.1}) \tag{5}$$

The maximum likelihood estimates of $\Sigma_{11}, \mu_{2.1}, \mathbf{B}_{2.1}, \Sigma_{2.1}$ are those values that maximize (5). To maximize (5) with respect to Σ_{11}, we maximize $\prod_{i=1}^{N} n(\mathbf{w}_i|\mathbf{c}_w, \Sigma_{11})$. This procedure gives us the usual maximum likelihood estimates of the parameters of a normal distribution based on N observations, namely,

$$\hat{\Sigma}_{11} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{w}_i - \mathbf{c}_w)'(\mathbf{w}_i - \mathbf{c}_w) \tag{6}$$

To maximize (5) with respect to $\mu_{2.1}, \mathbf{B}_{2.1}$ and $\Sigma_{2.1}$, we maximize the second term of the right hand side of (5). This gives the usual estimates of regression parameters, namely,

$$\hat{\mathbf{B}}_{2.1} = \mathbf{S}_{xw}\mathbf{S}_{ww}^{-1}, \quad \hat{\boldsymbol{\mu}}_{2.1} = \bar{x} - \hat{\mathbf{B}}_{2.1}\bar{\mathbf{w}}, \quad \hat{\boldsymbol{\Sigma}}_{2.1} = (\mathbf{S}_{xx} - \hat{\mathbf{B}}_{2.1}\mathbf{S}_{wx})/N \tag{7}$$

It is easy to see that the maximum likelihood estimates of $\boldsymbol{\mu}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22}$ are obtained by solving (4), where $\boldsymbol{\mu}_{2.1} = \hat{\boldsymbol{\mu}}_{2.1}, \mathbf{B}_{2.1} = \hat{\mathbf{B}}_{2.1}$ and $\boldsymbol{\Sigma}_{2.1} = \hat{\boldsymbol{\Sigma}}_{2.1}$. Hence, we have

$$\hat{\boldsymbol{\mu}} = \bar{x} - \hat{\mathbf{B}}_{2.1}(\bar{\mathbf{w}} - \mathbf{c}_w), \hat{\boldsymbol{\Sigma}}_{2.1} = (\mathbf{S}_{xx} - \hat{\mathbf{B}}_{2.1}\mathbf{S}_{wx})/N \text{ with } \hat{\mathbf{B}}_{2.1} = \mathbf{S}_{xw}\mathbf{S}_{ww}^{-1} \tag{8}$$

Secondly, we derive the maximum likelihood estimator of $\boldsymbol{\beta}$. let $\begin{pmatrix} \bar{\mathbf{t}} \\ \bar{\mathbf{y}} \end{pmatrix}$ and $\mathbf{R} = \begin{pmatrix} \mathbf{R}_{tt} & \mathbf{R}_{ty} \\ \mathbf{R}_{yt} & \mathbf{R}_{yy} \end{pmatrix}$ denote the sample mean vector and the sum of squares and sum of products matrix respectively based on the data $\mathbf{E} = \begin{pmatrix} \mathbf{t}_1, & \cdots & , \mathbf{t}_M \\ \mathbf{y}_1, & \cdots & , \mathbf{y}_M \end{pmatrix}_{(q+p) \times M}$.

Consider the density function of this data. We note that the density of $\mathbf{t}$ and $\mathbf{y}$ can be written as the marginal density of $\mathbf{t}$ times the conditional density of $\mathbf{y}$ given $\mathbf{t}$ (we indicate the density of normal distribution by $n(\cdot)$ here), that is

$$n(\mathbf{t}, \mathbf{y}|\mathbf{c}_t, \boldsymbol{\beta}, \boldsymbol{\Delta}) = n(\mathbf{t}|\mathbf{c}_t, \boldsymbol{\Delta}_{11})n(\mathbf{y}|\boldsymbol{\beta}_{2.1} + \mathbf{C}_{2.1}\mathbf{t}, \boldsymbol{\Delta}_{2.1})$$

where

$$\mathbf{C}_{2.1} = \boldsymbol{\Delta}_{21}\boldsymbol{\Delta}_{11}^{-1}, \boldsymbol{\beta}_{2.1} = \boldsymbol{\beta} - \mathbf{C}_{2.1}\mathbf{c}_t, \boldsymbol{\Delta}_{2.1} = \boldsymbol{\Delta}_{22} - \mathbf{C}_{2.1}\boldsymbol{\Delta}_{12} \tag{9}$$

The likelihood function can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\Delta}) = \prod_{i=1}^{M} n(\mathbf{t}_i|\mathbf{c}_t, \boldsymbol{\Delta}_{11}) \prod_{i=1}^{M} n(\mathbf{y}_i|\boldsymbol{\beta}_{2.1} + \mathbf{C}_{2.1}\mathbf{t}_i, \boldsymbol{\Delta}_{2.1}) \tag{10}$$

The maximum likelihood estimates of $\boldsymbol{\Delta}_{11}, \boldsymbol{\beta}_{2.1}, \mathbf{C}_{2.1}, \boldsymbol{\Delta}_{2.1}$ are those values that maximize (10). To maximize (10) with respect to $\boldsymbol{\Delta}_{11}$, we maximize $\prod_{i=1}^{M} n(\mathbf{t}_i|\mathbf{c}_t, \boldsymbol{\Delta}_{11})$. This procedure gives us the usual maximum likelihood estimates of the parameters of a normal distribution based on M observations, namely,

$$\hat{\boldsymbol{\Delta}}_{11} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{t}_i - \mathbf{c}_t)'(\mathbf{t}_i - \mathbf{c}_t) \tag{11}$$

To maximize (10) with respect to $\boldsymbol{\beta}_{2.1}, \mathbf{C}_{2.1}$ and $\boldsymbol{\Delta}_{2.1}$, we maximize the second term of the right hand side of (10). This gives the usual estimates of regression parameters, namely,

$$\hat{\mathbf{C}}_{2.1} = \mathbf{R}_{yt}\mathbf{R}_{tt}^{-1}, \quad \hat{\boldsymbol{\beta}}_{2.1} = \bar{\mathbf{y}} - \hat{\mathbf{C}}_{2.1}\bar{\mathbf{t}}, \quad \hat{\boldsymbol{\Delta}}_{2.1} = (\mathbf{R}_{yy} - \hat{\mathbf{C}}_{2.1}\mathbf{R}_{ty})/M \tag{12}$$

It is easy to see that the maximum likelihood estimates of $\boldsymbol{\beta}, \boldsymbol{\Delta}_{12}, \boldsymbol{\Delta}_{22}$ are obtained by solving (9), where $\boldsymbol{\beta}_{2.1} = \hat{\boldsymbol{\beta}}_{2.1}, \mathbf{C}_{2.1} = \hat{\mathbf{C}}_{2.1}$ and $\boldsymbol{\Delta}_{2.1} = \hat{\boldsymbol{\Delta}}_{2.1}$. Hence, we have

$$\hat{\boldsymbol{\beta}} = \bar{\mathbf{y}} - \hat{\mathbf{C}}_{2.1}(\bar{\mathbf{t}} - \mathbf{c}_t), \hat{\boldsymbol{\Delta}}_{2.1} = (\mathbf{R}_{yy} - \hat{\mathbf{C}}_{2.1}\mathbf{R}_{ty})/M \text{ with } \hat{\mathbf{C}}_{2.1} = \mathbf{R}_{yt}\mathbf{R}_{tt}^{-1} \tag{13}$$

3. Hypothesis test and confidence region for $\mu - \beta$

In the Appendix, we derive the distribution of $\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\beta}}$ and construct the quantity used for the test

Table 1. Monte Carlo estimates of the sizes of the test; $\Sigma = \text{diag}(\lambda_1, \lambda_2)$, $\Delta = I$ $-\Sigma(p,q) = (1,1)$; $\alpha = 0.05$.

(λ_1, λ_2)	(N, M)			
	(6, 12)	(12, 13)	(15, 19)	(20, 29)
(0.1,0.1)	0.049	0.050	0.048	0.050
(0.2,0.5)	0.046	0.050	0.049	0.050
(0.2,0.7)	0.052	0.050	0.051	0.051
(0.1,0.9)	0.054	0.051	0.050	0.050
(0.5,0.5)	0.047	0.049	0.050	0.050
(0.9,0.9)	0.054	0.052	0.051	0.050

$$Q = (\hat{\mu} - \hat{\beta} - (\mu - \beta))' \left[C_x \tilde{\Sigma}_{2.1} + C_y \tilde{\Delta}_{2.1} \right]^{-1} (\hat{\mu} - \hat{\beta} - (\mu - \beta)) \tag{14}$$

where $\tilde{\Sigma}_{2.1}, \tilde{\Delta}$ are unbiased estimators of $\Sigma_{2.1}$ and $\Delta_{2.1}$ respectively which are defined in (A.2) of the Appendix, and

$$C_x = \frac{1}{N} + (\bar{w} - c_w)' S_{ww}^{-1} (\bar{w} - c_w), C_y = \frac{1}{M} + (\bar{t} - c_t)' R_{tt}^{-1} (\bar{t} - c_t)$$

Moreover, we derive the approximate distribution of Q in the Appendix:

$$Q \sim \frac{pf}{f - p + 1} F(p, f - p + 1) \text{ approximately} \tag{15}$$

where $F(p, f - p + 1)$ is a F distribution with numerator degree p and denominators degree $(f - p + 1)$.

Thus, for a given level α and an observed value Q_0 of Q, the null hypothesis that $\mu - \beta = 0$ will be rejected whenever the p-value

$$P\left(\frac{pf}{f - p + 1} F(p, f - p + 1) > Q_0 | H_0 \right) < \alpha.$$

Furthermore, an approximate $1 - \alpha$ confidence set for $\mu - \beta$ is the set of values of $\mu - \beta$ that satisfy

$$(\hat{\mu} - \hat{\beta} - (\mu - \beta))' \left[C_x \tilde{\Sigma}_{2.1} + C_y \tilde{\Delta}_{2.1} \right]^{-1} (\hat{\mu} - \hat{\beta} - (\mu - \beta)) \leq \frac{pf}{f - p + 1} F_{1-\alpha}(p, f - p + 1) \tag{16}$$

where $F_{1-\alpha}(p, f - p + 1)$ is the $(1-\alpha)$th quantile of the $F(p, f - p + 1)$ distribution.

4. Accuracy of the approximations

We have used an approximation to approximate the distribution of the sum of two Wishart matrices with different scale matrices to derive the distribution of Q. So, to understand the accuracy of the approximation, we use Monte Carlo simulation to estimate the sizes of the test for the hypothesis in (1) when the nominal level is 0.05.

Each simulation is consisting of 100,000 runs. As pointed out by Yao (1965), there exists a non singular matrix M such that $\Sigma = M\Lambda M'$, $\Delta = M(I - \Lambda)M'$ and $\Sigma + \Delta = MM'$, where $\Lambda = \text{diag}(\lambda_1, ..., \lambda_p)$, $0 < \lambda_1 \leq ... \leq \lambda_p < 1$, and λ_i's are the eigenvalues of $(\Sigma + \Delta)^{-\frac{1}{2}} \Sigma (\Sigma + \Delta)^{-\frac{1}{2}}$. Because the solutions are non singular invariant, without loss of generality, we can assume $\Sigma = \Lambda$ and $\Delta = I - \Lambda$ for comparison purpose. The sizes are

Table 2. Monte Carlo estimates of the sizes of the test; $\mathbf{\Sigma} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, $\mathbf{\Delta} = \mathbf{I} - \mathbf{\Sigma}$, $(p, q) = (2, 2)$; $\alpha = 0.05$.

	(N, M)			
$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$	(6, 12)	(12, 13)	(15, 19)	(20, 29)
(0.1,0.1,0.1,0.1)	0.048	0.051	0.050	0.051
(0.1,0.1,0.5,0.5)	0.056	0.049	0.049	0.049
(0.1,0.2,0.5,0.7)	0.060	0.048	0.049	0.049
(0.1,0.5,0.5,0.9)	0.065	0.049	0.051	0.050
(0.5,0.5,0.5,0.5)	0.056	0.047	0.049	0.050
(0.5,0.5,0.9,0.9)	0.076	0.051	0.051	0.050
(0.9,0.9,0.9,0.9)	0.076	0.051	0.051	0.050

computed for $p = q = 1$ and $p = q = 2$. The estimates of the sizes are presented in Tables 1 and 2. It is clear from Tables 1 and 2 that the coverage probabilities are very close to 0.95 for all the cases considered.

5. An illustrative example

We now illustrate the methods using the 'Fishers Iris Data' which represents measurements of the sepal length and width, and petal length and width in centimeters of fifty plants for each of three types of iris: Iris setosa, Iris versicolor and Iris virginica. The data sets are posted in many websites, say, http://javeeh.net/sasintro/intro151.html. For illustrative purposes, we use the data on setosa (x) and versicolor (y).

Assume that the population means of the sepal length and width of Iris setosa and Iris versicolor are $(5.936, 2.770)$ and $(5.006, 3.428)$ respectively, which are means of all the observations. Let $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ be means of the petal length and width of Iris setosa and Iris versicolor respectively. Suppose that we want to test

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\beta} \text{ vs. } H_a : \boldsymbol{\mu} \neq \boldsymbol{\beta}$$

In this example, we sample randomly 20 observation from the Iris data. Then, we have $p = q = 2$, $N = M = 20$. After careful calculation, we get $f = 42.29$, $Q = 7.98$. The critical value $\frac{pf}{f-p+1} F_{1-\alpha}(p, f - p + 1) = 6.60$. Since Q is larger than the critical value, we have sufficient evidence to reject H_0 at 5% significance level.

Appendix

The following lemma is needed to find approximate distribution of Q in (14). In Lemma, we give the modified version of the Nel van der Merwe (1986) Wishart approximation given in Krishnamoorthy and Yu (2004).

Lemma. *Let* $\mathbf{A}_1 \sim W_p(n_1, \mathbf{\Delta}_1)$ *independently of* $\mathbf{A}_2 \sim W_p(n_2, \mathbf{\Delta}_2)$. *Then*

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 \sim W_p\left(f, \frac{1}{f}(n_1 \mathbf{\Delta}_1 + n_2 \mathbf{\Delta}_2)\right) \text{ approximately,}$$

where

$$f = \frac{p + p^2}{\frac{1}{n_1}\left\{\mathrm{tr}\left[(\mathbf{A}_1\mathbf{A}^{-1})^2\right] + \left[\mathrm{tr}(\mathbf{A}_1\mathbf{A}^{-1})\right]^2\right\} + \frac{1}{n_2}\left\{\mathrm{tr}\left[(\mathbf{A}_2\mathbf{A}^{-1})^2\right] + \left[\mathrm{tr}(\mathbf{A}_2\mathbf{A}^{-1})\right]^2\right\}}.$$

□

We consider conditional distribution of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ at first. After some complicated calculation, we have

$$\hat{\boldsymbol{\mu}}|(\mathbf{w}_1, ..., \mathbf{w}_N) \sim N_p(\boldsymbol{\mu}, (\frac{1}{N} + (\bar{\mathbf{w}} - \mathbf{c}_w)'\mathbf{S}_{ww}^{-1}(\bar{\mathbf{w}} - \mathbf{c}_w))\boldsymbol{\Sigma}_{2.1}) = N_p(\boldsymbol{\mu}, \mathbf{C}_x\boldsymbol{\Sigma}_{2.1})$$

independent of

$$\hat{\boldsymbol{\beta}}|(\mathbf{t}_1, ..., \mathbf{t}_M) \sim N_p(\boldsymbol{\beta}, (\frac{1}{M} + (\bar{\mathbf{t}} - \mathbf{c}_t)'\mathbf{R}_{tt}^{-1}(\bar{\mathbf{t}} - \mathbf{c}_t))\boldsymbol{\Delta}_{2.1}) = N_p(\boldsymbol{\beta}, \mathbf{C}_y\boldsymbol{\Delta}_{2.1})$$

Hence,

$$(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\beta}})|(\mathbf{w}_1, ..., \mathbf{w}_N, \mathbf{t}_1, ..., \mathbf{t}_M) \sim N\left[(\boldsymbol{\mu} - \boldsymbol{\beta}), \mathbf{C}_x\boldsymbol{\Sigma}_{2.1} + \mathbf{C}_y\boldsymbol{\Delta}_{2.1}\right] \qquad (A.1)$$

Meanwhile, we change the estimator of $\boldsymbol{\Sigma}_{2.1}$ and $\boldsymbol{\Delta}_{2.1}$ in (8) and (13) into unbiased estimator respectively as follows:

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}}_{2.1} &= (\mathbf{S}_{xx} - \hat{\mathbf{B}}_{2.1}\mathbf{S}_{xw})/(N - q - 1) \\ \tilde{\boldsymbol{\Delta}}_{2.1} &= (\mathbf{R}_{yy} - \hat{\mathbf{C}}_{2.1}\mathbf{R}_{yt})/(M - q - 1) \end{aligned} \qquad (A.2)$$

then

$$\begin{aligned} (N - q - 1)\tilde{\boldsymbol{\Sigma}}_{2.1}|(\mathbf{w}_1, ..., \mathbf{w}_N) &\sim W_p(N - q - 1, \boldsymbol{\Sigma}_{2.1}) \\ (M - q - 1)\tilde{\boldsymbol{\Delta}}_{2.1}|(\mathbf{t}_1, ..., \mathbf{t}_M) &\sim W_p(M - q - 1, \boldsymbol{\Delta}_{2.1}) \end{aligned} \qquad (A.3)$$

Moreover, $\tilde{\boldsymbol{\Sigma}}_{2.1}$ and $\hat{\boldsymbol{\mu}}$ are independent conditional on $(\mathbf{w}_1, ..., \mathbf{w}_N)$, $\tilde{\boldsymbol{\Delta}}_{2.1}$ and $\hat{\boldsymbol{\beta}}$ are independent conditional on $(\mathbf{t}_1, ..., \mathbf{t}_M)$.

Let $\mathbf{L}_1 = \mathbf{C}_x\tilde{\boldsymbol{\Sigma}}_{2.1}, \mathbf{L}_2 = \mathbf{C}_y\tilde{\boldsymbol{\Delta}}_{2.1}, \mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$, Using the lemma, we have

$$\mathbf{L} = (\mathbf{C}_x\tilde{\boldsymbol{\Sigma}}_{2.1} + \mathbf{C}_y\tilde{\boldsymbol{\Delta}}_{2.1})|(\mathbf{w}_1, ..., \mathbf{w}_N, \mathbf{t}_1, ..., \mathbf{t}_M) \sim W_p(f, \left[\mathbf{C}_x\boldsymbol{\Sigma}_{2.1} + \mathbf{C}_y\boldsymbol{\Delta}_{2.1}\right]/f) \text{approximately,}$$

where

$$f = \frac{p + p^2}{\frac{1}{N-q-1}\left\{\text{tr}\left[(\mathbf{L}_1\mathbf{L}^{-1})^2\right] + \left[\text{tr}(\mathbf{L}_1\mathbf{L}^{-1})\right]^2\right\} + \frac{1}{M-q-1}\left\{\text{tr}\left[(\mathbf{L}_2\mathbf{L}^{-1})^2\right] + \left[\text{tr}(\mathbf{L}_2\mathbf{L}^{-1})\right]^2\right\}} \qquad (A.4)$$

So, Q is a Hotelling Statistics conditionally

$$Q|(\mathbf{w}_1, ..., \mathbf{w}_N, \mathbf{t}_1, ..., \mathbf{t}_M) \sim T^2(p, f) = \frac{pf}{f - p + 1}F(p, f - p + 1)$$

Since this conditional distribution is free of $(\mathbf{w}_1, ..., \mathbf{w}_N, \mathbf{t}_1, ..., \mathbf{t}_M)$, we have

$$Q \sim \frac{pf}{f - p + 1}F(p, f - p + 1) \qquad (A.5)$$

Funding

References

Bennett, B. M. 1950. Note on a solution of the generalized Behrens-Fisher problem. *Annals of the Institute of Statistical Mathematics* 2 (2):87–90. doi:10.1007/BF02919505.

Cochran, W. 1940. The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce. *The Journal of Agricultural Science* 30 (2):262–75. doi:10.1017/S0021859600048012.

James, G. S. 1954. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika* 41 (1/2):19–43. doi:10.2307/2333003.

Johansen, S. 1980. The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika* 67 (1):85–95. doi:10.1093/biomet/67.1.85.

Kadilar, G. 2016. A new exponential type estimator for the population mean in simple random sampling. *Journal of Modern Applied Statistical Methods* 15 (2):207–14. doi:10.22237/jmasm/1478002380.

Khan, S.A., H. Ali, S. Manzoor, and A. Alamgir. 2015. A class of transformed efficient ratio estimators of finite population mean. *Pakistan Journal of Statistics* 31 (4):353–62.

Krishnamoorthy, K., and J. Yu. 2004. Modified Nel and Van der Merwe test for the multivariate Behrens-Fisher problem. *Statistics & Probability Letters* 66 (2):161–9. doi:10.1016/j.spl.2003.10.012.

Murthy, M. 1964. Product method of estimation. *Sankhya* 26:294–307.

Nel, D. G., and C. A. Van der Merwe. 1986. A solution to the multivariate Behrens-Fisher problem. *Communications in Statistics - Theory and Methods* 15 (12):3719–35. doi:10.1080/03610928608829342.

Robson, D. 1957. Application of multivariate polykays to the theory of unbiased ratio type estimation. *Journal of American Statistical Association* 52:411–22.

Scheffé, H. 1943. On solutions of the Behrens-Fisher problem, based on the t-distribution. *Annals of Mathematical Statistics* 14:35–44.

Scheffé, H. 1970. Practical solutions of the Behrens-Fisher problem. *Journal of American Statistician Association* 65:1501–8.

Watson, D. 1937. The estimation of leaf area in field crops. *The Journal of Agricultural Science* 27 (3):474–83. doi:10.1017/S002185960005173X.

Welch, B. L. 1947. The generalization of Student's problem when several different population variances are involved. *Biometrika* 34 (1/2):28–35. doi:10.2307/2332510.

Yao, Y. 1965. An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika* 52:139–47.

Yan, Z., and B. Tian. 2010. Ratio method to the mean estimation using coefficient of skewness of auxiliary variable *ICICA*. *Part II* 106:103–10.