# A SEQUENTIAL APPROACH TO BEHRENS-FISHER PROBLEM

**Basil M. de Silva***
School of Mathematical and Geospatial Sciences,
RMIT University, Melbourne, Australia

**Vasant B. Waikar**
Department of Mathematics & Statistics,
Miami University, Oxford, Ohio, U.S.A.

**Abstract:** This paper reexamines the scope of a two-stage methodology for constructing a fixed-width confidence interval for the difference between two independent normal population means. We address directly the problem in a Behrens-Fisher situation, that is, when the variances of the two populations are both unknown and unequal. However, we show that the proposed methodology remains valid when the two variances are equal but unknown.

We apply a two-stage procedure to the Behrens-Fisher situation using two approaches. In the first approach, we work with the differences of each observation in the first sample with all observations in the second sample. In the second approach, we work with the differences of paired observations from samples one and two. Next, to overcome the problem of over-sampling associated with the two-stage procedures, we apply bootstrap methodology to modify the two-stage procedures. These methods, with and without bootstrapping, are compared with the help of a large-scale simulation study.

**Keywords:** Behrens-Fisher situations; bootstrapping; two-stage procedure; asymptotic analysis; moderate sample comparisons; simulations; normal distributions; mixtures of distributions.

**Subject Classifications:** 62L12; 62F40.

## 1. INTRODUCTION

The problem of estimating the difference between the means of two normal populations $\delta = \mu_1 - \mu_2$ when variances are both unknown and unequal is known as the Behrens-Fisher problem. A fixed sample solution to this problem does not exist. Chapman (1950) proposed a two-stage procedure following Stein (1945, 1949), which was improved upon by Ghosh (1975). Further, a generalization of this two-stage (or multi-stage) fixed-width confidence interval for $\delta$ was given by Mukhopadhyay and de Silva (1997). Thus in general a sequential approach to solving this problem has been somewhat successful. Lately, some others have introduced the idea of resampling (bootstrap in particular) to improve the precision by reducing oversampling

*The work has been carried out while visiting Miami University, Ohio, U.S.A.

in the second or third stage (see Babu and Padmanabhan (2002) for a solution to the nonparametric Behrens-Fisher problem). Also see Mukhopadhyay and Solanky (1994) for a review of multi-stage selection and ranking procedures related to the Behrens-Fisher problem. In this paper we have reexamined the scope of a two-stage solution in the light of bootstrapping. Some simulation results are obtained and a data analysis is also provided for a real-life example from the field of political science using the data from Friedenberg and Heyward (2003) who classified a country as democratic or non-democratic on the basis of certain variables.

Given $\{X_{11}, X_{12}, \ldots, X_{1n}\}$ and, $\{X_{21}, X_{22}, \ldots, X_{2n}\}$ are two independent random samples from two independent normal populations with unknown means $\mu_1$ and $\mu_2$ respectively such that $\mu_1 - \mu_2 = \delta$. The variances, $\sigma_1^2$ and $\sigma_2^2$ of the two populations are unknown. In this paper the sample sizes of both samples are equal and denoted by $n$.

The aim of this paper is to construct a fixed-width confidence interval for the difference $\delta$. The problem is to find the sample size $n$ such that

$$\mathcal{P}\left(\delta \in I_{(n,d)}\right) \geq 1 - \alpha \tag{1.1}$$

where $\alpha$ is the confidence coefficient, $2d$ is the width of the confidence interval and

$$I_{n,d} = \left(\overline{X}_{1n} - \overline{X}_{2n} - d, \;\; \overline{X}_{1n} - \overline{X}_{2n} + d\right). \tag{1.2}$$

Here $\overline{X}_{in} = n^{-1} \sum_{j=1}^{n} X_{ij}, i = 1, 2$. If $\sigma_1^2$ and $\sigma_2^2$ are known, then

$$I_{n,d} = \left(\overline{X}_{1n} - \overline{X}_{2n} - z_{\alpha/2}\sigma_0/\sqrt{n}, \;\; \overline{X}_{1n} - \overline{X}_{2n} + z_{\alpha/2}\sigma_0/\sqrt{n}\right) \tag{1.3}$$

where $\sigma_0^2 = \sigma_1^2 + \sigma_2^2$ and $z_{\alpha/2}$ is the $50\alpha^{\text{th}}$ upper percentile of standard normal distribution. From (1.2) and (1.3) we can prove that the sample size, $n$ required to achieve the confidence interval given in (1.1), is given by

$$n \geq z_{\alpha/2}^2 \sigma_0^2 d^{-2} = n_{opt}, \text{ say.} \tag{1.4}$$

Thus, if the variances of the two populations are known, then $n_{opt}$ given in (1.4) is the *optimal sample size* for a $100(1-\alpha)\%$ confidence interval for $\delta$ with fixed-width $2d$. This paper examines methodologies for obtaining sample size when the two variances, $\sigma_1^2$ and $\sigma_2^2$ are unknown.

## 2. METHODOLOGIES

Here we formulate two computational methodologies to combine the two sample problem into one sample problem. Method I uses the traditional approach by taking the difference of two sample means and in Method II we use the idea given in Ghosh(1975). In both cases we assumed that the population variances are unknown and are estimated by their corresponding sample variances, $S_{1n}^2$ and $S_{2n}^2$.

### 2.1. Method I

Consider the Satterthwaite's method for constructing a confidence interval for the difference of normal means when the variances are unknown and unequal.

**Theorem 2.1.** *Let $\overline{U}_n = \overline{X}_{1n} - \overline{X}_{2n}$ be the difference of the two sample means for a fixed sample size $n$ and $I_{(n,d)} = \left(\overline{U}_n - d, \overline{U}_n + d\right)$ be a fixed-width confidence interval for $\delta$ such that $\mathcal{P}\left(\delta \in I_{(n,d)}\right) \geq (1-\alpha)$ then*

$$d \geq t_\nu(\alpha/2)\left(\frac{S_{1n}^2 + S_{2n}^2}{n}\right)^{1/2} \tag{2.1}$$

*and*

$$n \geq \frac{t_\nu^2(\alpha/2)}{d^2}\left(S_{1n}^2 + S_{2n}^2\right). \tag{2.2}$$

*Here $t_\nu(\alpha/2)$ is the $50\alpha^{\text{th}}$ upper percentile of the t-distribution with*

$$\nu = \left[(n-1)\left(S_{1n}^2 + S_{2n}^2\right)^2 / \left(S_{1n}^4 + S_{2n}^4\right)\right] \tag{2.3}$$

degrees of freedom (Satterthwaite, 1946) and $[a]$ is the integer part of $a$.

**Proof:** Clearly $\overline{U}_n \sim N(\delta, \sigma_0^2/n)$, thus the above theorem follows from the results given in Satterthwaite (1946). ■

### 2.2. Method II

In this method we construct the new random variable $V_j = X_{1j} - X_{2j}$ for $j = 1, 2, \ldots, n$ as in Ghosh(1975). Then clearly $V_j$'s are iid normal random variables with mean $\delta$ and variance $\sigma_0^2$. Also $\overline{V}_n = n^{-1}\sum_{j=1}^n V_j = \overline{X}_{1n} - \overline{X}_{2n}$ is $N\left(\delta, \sigma_0^2/n\right)$.

**Theorem 2.2.** *Consider a fixed sample size $n$. Let $J_{(n,d)} = \left(\overline{V}_n - d, \overline{V}_n + d\right)$ be a fixed-width confidence interval for delta such that*

$$\mathcal{P}\left(\delta \in J_{(n,d)}\right) \geq 1 - \alpha. \tag{2.4}$$

*Then*

$$n \geq \frac{t_{n-1}^2(\alpha/2) S_n^2(v)}{d^2}. \tag{2.5}$$

*Here $t_{n-1}(\alpha/2)$ is the $50\alpha^{\text{th}}$ upper percentile of the t-distribution with $(n-1)$ degrees of freedom and, $S_n^2(v) = (n-1)^{-1}\sum_{j=1}^n \left(V_j - \overline{V}_n\right)^2$.*

**Proof:** Since $V_j$'s are iid normal random variables with mean $\delta$ and variance $\sigma_0^2$, the distribution of $(n-1)S_n^2(v)/\sigma_0^2$ is $\chi^2_{(n-1)}$ and $100(1-\alpha)\%$ confidence interval for $\delta$ is given by

$$\overline{V}_n \pm \frac{t_{n-1}(\alpha/2) S_n(v)}{\sqrt{n}}. \tag{2.6}$$

Since $2d$ is the width of the confidence interval, from (2.6) we have $n \geq t_{n-1}^2(\alpha/2)S_n^2(v)/d^2$. ■

**Remark 2.1.** The inequality for the fixed sample size, $n$ given in (2.2) and (2.6) will be used respectively to derive the two-stage sequential stopping rules for Methods I and II. Since $\left(S_{1n}^2 + S_{2n}^2\right)$ in (2.2) and the corresponding factor $S_n^2(v)$ in (2.6) are different, we will have two different stopping rules for Method I and II.

## 3. TWO-STAGE PROCEDURES

This procedure follows along the lines of Stein(1945, 1949) and Mukhopadhyay and Solanky (1994). In the two-stage procedure, first we take a pilot sample of size $m$ then use this pilot sample to estimate the final sample sample size, $N$. The pilot sample is called the Stage 1 sample and in Stage 2, we sample the difference $N - m$ provided $N > m$. If $N = m$ then no observations are taken in the Stage 2. The final sample size $N$ is computed from a *stopping rule* derived from Stage 1 sample with the predetermined confidence width $(2d)$ and the confidence coefficient $(1 - \alpha)$. In *Stein's two-stage procedure*, initial sample size $m$ is a fixed number decided by the experimenter. Similarly $d$ and $\alpha$ are chosen to achieve precision required in the experiment. In the literature $m$ can be chosen as small as 2 but in most cases we need more than 2 observations to obtain a reasonably good estimate for the final sample size. In the current work we set $m = 15$ for Stein's procedure.

**Stopping Rule for Method I**

If $m$ is the starting sample size from both populations, then from (2.2), the final sample size of the two-stage procedure is given by

$$N_1 = \max\left\{m, \left\langle t_\nu^2(\alpha/2)\left(S_{1m}^2 + S_{2m}^2\right) d^{-2}\right\rangle + 1\right\} \tag{3.1}$$

where $< y >$ stands for the largest integer less than $y$ and the degrees of freedom $\nu = [(m-1)\left(S_{1m}^2 + S_{2m}^2\right)^2 / \left(S_{1m}^4 + S_{2m}^4\right)]$. As explained above, if $N_1 = m$, we do not sample any more from either population, however if $N_1 > m$, then we sample the difference $(N_1 - m)$ from both populations in the second stage. Here, $N_1$ estimates $n_{opt}$ via double sampling. Now based on the final data $\{X_{i1}, \cdots, X_{im}, \cdots, X_{iN_1}\}, i = 1, 2$, we propose the fixed-width confidence interval $I_{(N_1,d)} = \left(\overline{U}_{N_1} - d, \overline{U}_{N_1} + d\right)$ for $\delta$ where $\overline{U}_{N_1} = \overline{X}_{1N_1} - \overline{X}_{2N_1}$. From Theorem 6.2.1, Ghosh at el. (p. 154, 1997) it is clear that $\mathcal{P}(\delta \in I_{(N_1,d)}) \geq 1 - \alpha$.

**Stopping Rule for Method II**

Let $m$ be the initial sample size for both populations. From (2.5), the final sample size for this method is given by

$$N_2 = \max\left\{m, \left\langle t_{m-1}^2(\alpha/2)S_m^2(v)d^{-2}\right\rangle + 1\right\}. \tag{3.2}$$

As in Method I, if $N_2 = m$, then stop sampling and compute the confidence interval. Otherwise sample the difference $N_2 - m$ and compute the confidence interval $J_{(N_2,d)} = \left(\overline{V}_{N_2} - d, \overline{V}_{N_2} + d\right)$ using all $N_2$ observations. Then $\mathcal{P}(\delta \in J_{(N_2,d)}) \geq 1 - \alpha$.

## 3.1. Initial Sample Size and Asymptotic Results

### Initial Sample Size, $m$

It is well known that Stein's procedure oversamples. However, a significant reduction of this oversampling problem can be achieved by using the *modified two-stage procedure* introduced in Mukhopadhyay (1980). Here $m$ is computed from the following equation as a function of $d$ and $\alpha$.

$$m = \max\left\{2, \left\langle \left(z_{\alpha/2}^2 d^{-2}\right)^{1/(1+\gamma)} \right\rangle + 1\right\} \tag{3.3}$$

where $\gamma$ is a positive constant.

Mukhopadhyay and Solanky (1994) used minimum sample size 2 in (3.3) for the theoretical computations and asymptotic analysis of the modified two-stage procedure. In practice, it is clear that we require more than two observations in the first stage to achieve a satisfactory estimation for the final sample size. Thus we use a minimum sample size of 15 in all the computations and simulations in this paper. In this project we modified the initial sample size, $m$ given in (3.3) to

$$m = \max\left\{15, \left\langle \left(z_{\alpha/2}^2 d^{-2}\right)^{1/(1+\gamma)} \right\rangle + 1\right\}. \tag{3.4}$$

For a 95% confidence interval the initial sample size given above is reduced to

$$m = \max\left\{15, \left\langle \left(3.8416 d^{-2}\right)^{1/(1+\gamma)} \right\rangle + 1\right\}. \tag{3.5}$$

### Choice of $d$ and $\gamma$ for Simulation

For the simulation analysis of Method I and Method II we have considered a 95% confidence interval with $\mu_1 = 0.8$, $\mu_2 = 0.4$, $\sigma_1^2 = 0.16$ and $\sigma_2^2 = 0.09$. Thus, $\sigma_0^2 = \sigma_1^2 + \sigma_2^2 = 0.25$ and from (1.4)

$$n_{opt} = \frac{1.96^2 \times 0.25}{d^2} = \frac{0.9604}{d^2}$$

for a given fixed $d$. Table 3.1 gives values of $d$ together with corresponding optimal sample sizes ($n_{opt}$). We choose these $d$ values in the simulation to study Method I and II for small, moderate and then large sample cases.

To choose a suitable value for $\gamma$ in the above equation we plot $m$ versus $\gamma$ for each $d$ value given in Table 3.1. Figure 3.1 shows that $m$ decreases with increasing $\gamma$ for a fixed $d$ and also decreases with increasing $d$ for a fixed $\gamma$. Our interest is in $d = 0.05, 0.10, 0.15, 0.20$ and $m$ must be less than the corresponding $n_{opt}$ given in Table 3.1. At the same time it should not be too small when compared with $n_{opt}$. Figure 3.1 shows that $\gamma = 1$ gives suitable values for $m$ for these values of $d$.

**Table 3.1.** $d$ and Corresponding $n_{opt}$ Used in the Simulation Study

| $d$ | 0.20 | 0.15 | 0.10 | 0.07 |
|---|---|---|---|---|
| $n_{opt}$ | 24.01 | 42.68 | 96.04 | 195.99 |
| $d$ | 0.05 | 0.04 | 0.03 | |
| $n_{opt}$ | 384.15 | 600.23 | 1067.07 | |

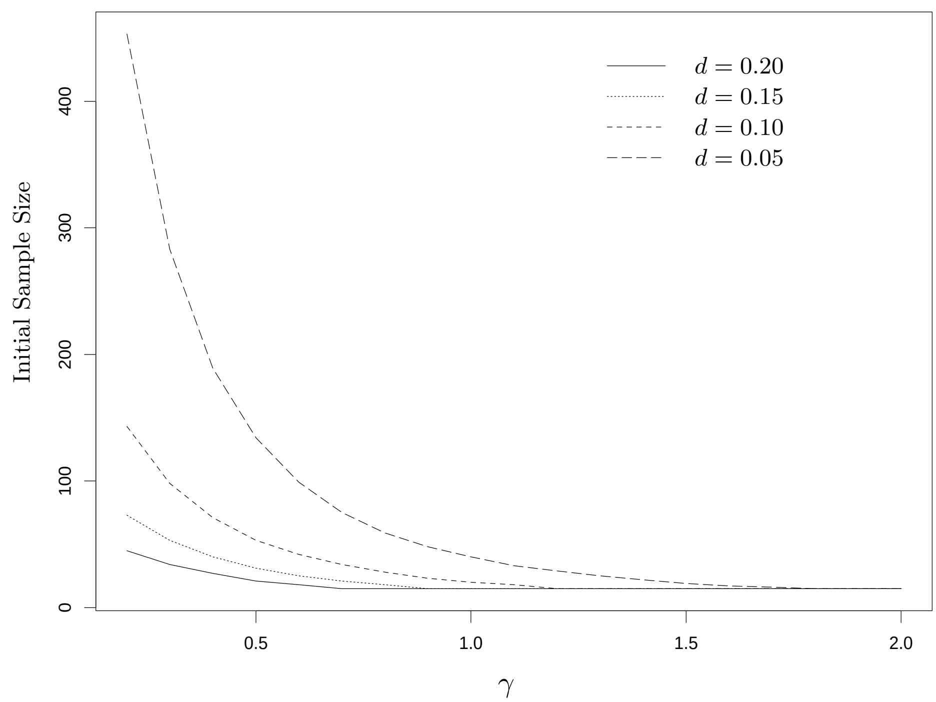


***Figure 3.1.*** Initial Sample Size, $m$ Versus $\gamma$ For Fixed $d$ and $\alpha = 0.05$

## Asymptotic Results

It has been proven that a fixed-width confidence interval obtained using the Stein's two-stage procedure is consistent (See Mukhopadhyay and Solanky, 1994). That is $\mathcal{P}(\delta \in I_{(N,d)}) \geq (1-\alpha)$. However the modified two-stage procedure where $m$ is obtained from (3.4) is only asymptotically consistent:

$$\mathcal{P}(\delta \in I_{(N_1,d)}) \rightarrow (1-\alpha) \text{ as } d \rightarrow 0. \tag{3.6}$$

Also it has been shown that the modified two-stage procedure is first order efficient, implying that

$$\mathbf{E}\left(\frac{N_1}{n_{opt}}\right) \to 1 \text{ as } d \to 0. \tag{3.7}$$

The Stein's procedure fails to satisfy the first order efficiency given in (3.7). Therefore the amount of oversampling in modified two-stage procedure is significantly less than the Stein's procedure.

## 3.2. Simulation Analysis

Consider two independent normal populations with fixed set of parameters $\mu_1 = 0.8$, $\mu_2 = 0.4$, $\sigma_1^2 = 0.16$ and $\sigma_2^2 = 0.09$. Now examine the efficiencies of the two methods described in Section 2, by generating random data from these two independent normal populations. Having fixed set of parameters given above, determine $N_1$ (denoted by $n_1$) and $N_2$ (denoted by $n_2$) respectively using (3.1) and (3.2) for a 95% confidence interval with fixed-width $2d$. Table 3.2 gives values of $d$ together with the corresponding optimal sample sizes ($n_{opt}$) used in the simulation to study effectiveness of the Methods I and II for the small, moderate and large sample cases.

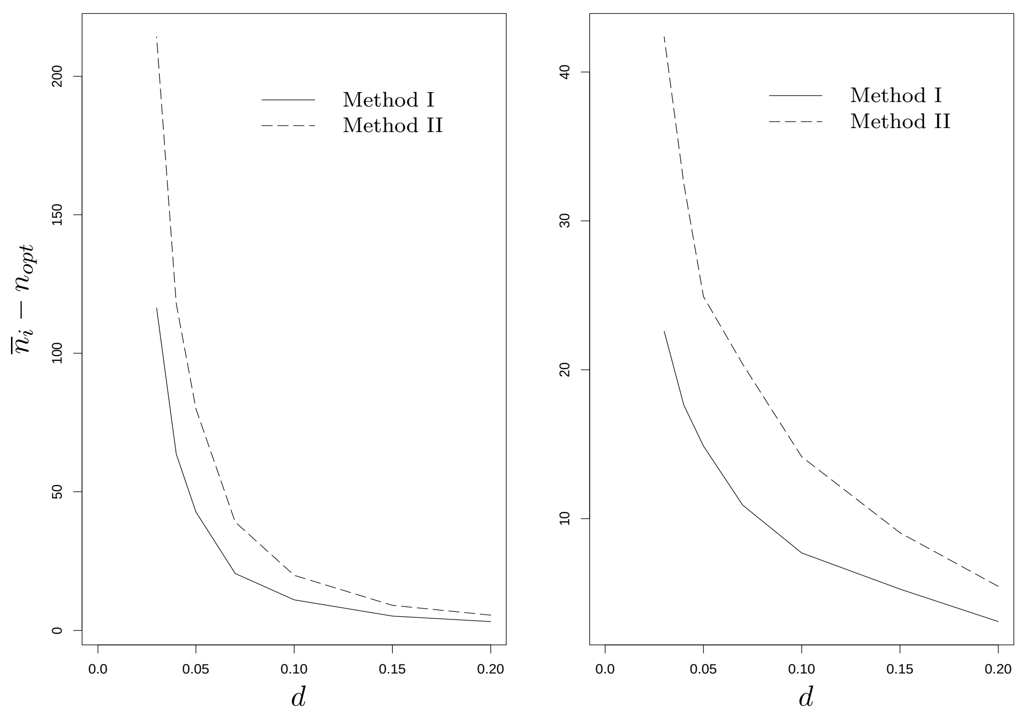

Fig 3.2(a): Stein's Two-Stage

Fig 3.2(b): Modified Two-Stage

***Figure 3.2.*** Oversampling, $\overline{n}_i - n_{opt}$ *Versus* $d$ For Method I and Method II

Simulation is replicated for 15,000 times for each $d$ value. The initial sample size is fixed at $m = 15$ for simulation of the Stein's two-stage procedure and for the modified two-stage, $m$ is obtained from (3.5) with $\gamma = 1$. Let $\overline{n}_i$ $(i = 1, 2)$ be the average value of $n_{ij}(j = 1, \ldots, 15000)$ from these 15,000 simulations. Figure 3.2 shows the amount of oversampling in these two procedures.

**Table 3.2.** Percentage Oversampling in Stein's Two-Stage and the Modified Two-Stage Procedures

| | | | Stein's Procedure | | | Modified Two-Stage | | |
|---|---|---|---|---|---|---|---|---|
| | $d$ | $n_{opt}$ | $\overline{n}$ | $s_{\overline{n}}$ | $P_{over}$ | $\overline{n}$ | $s_{\overline{n}}$ | $P_{over}$ |
| Method I | 0.20 | 24.0 | 27.2 | 0.06 | 13.4 | 27.1 | 0.06 | 12.9 |
| | 0.15 | 42.7 | 47.8 | 0.11 | 12.0 | 48.0 | 0.11 | 12.3 |
| | 0.10 | 96.0 | 107.0 | 0.25 | 11.5 | 103.7 | 0.20 | 8.0 |
| | 0.07 | 196.0 | 216.6 | 0.49 | 10.5 | 206.9 | 0.34 | 5.6 |
| | 0.05 | 384.2 | 426.9 | 0.98 | 11.1 | 399.0 | 0.55 | 3.9 |
| | 0.04 | 600.2 | 663.9 | 1.52 | 10.6 | 617.9 | 0.76 | 2.9 |
| | 0.03 | 1067.1 | 1183.4 | 2.68 | 10.9 | 1089.7 | 1.16 | 2.1 |
| Method II | 0.20 | 24.0 | 29.5 | 0.09 | 22.7 | 29.5 | 0.09 | 22.7 |
| | 0.15 | 42.7 | 51.7 | 0.16 | 21.2 | 51.7 | 0.16 | 21.2 |
| | 0.10 | 96.0 | 115.9 | 0.36 | 20.7 | 110.2 | 0.29 | 14.7 |
| | 0.07 | 196.0 | 235.1 | 0.73 | 20.0 | 216.4 | 0.48 | 10.4 |
| | 0.05 | 384.2 | 464.2 | 1.44 | 20.8 | 409.1 | 0.75 | 6.5 |
| | 0.04 | 600.2 | 718.2 | 2.21 | 19.7 | 632.7 | 1.05 | 5.4 |
| | 0.03 | 1067.1 | 1281.2 | 3.96 | 20.1 | 1109.4 | 1.59 | 4.0 |

Figure 3.2 (a) gives the plots of $\overline{n}_i - n_{opt}$ versus $d$ for the Stein's two-stage procedure. The amount of oversampling in both Method I and II, is increasing with decreasing $d$. That is, amount of oversampling appears to be directly proportional to the final sample size. Also amount of oversampling in Method II is consistently higher than that of Method I. We can make similar conclusion for the results given in Figure 3.2(b) for the modified two-stage procedure. However oversampling in the modified two-stage is significantly smaller than the Stein's procedure. This can be seen easily from Table 3.2, which gives the amount of oversampling as a percentage of the corresponding optimal sample size. Here percentage oversampling is computed from the following:

$$P_{over} = 100\left(\overline{n} - n_{opt}\right)/n_{opt} \tag{3.8}$$

where $\overline{n}$ is the average of the simulated values of the sample size. Further, Table 3.2 shows the simulated standard error $s_{\overline{n}}$ of $\overline{n}$.

Note that column 5 in Table 3.2 gives $P_{over}$ in Stein's procedure and column 8 gives $P_{over}$ in the modified two-stage procedures. Now using these results, it is apparent that as $d$ decreases from 0.20 to 0.03, for the Stein's procedure using

- Method I: $P_{over} \approx 11\%$ and
- Method II: $P_{over} \approx 21\%$.

Similarly for modified two-stage procedure using

- Method I: $P_{over}$ is decreasing from 12.9% to 2.1% and
- Method II: $P_{over}$ is decreasing from 22.7% to 4.0%.

That is, the modified two-stage gives a significant reduction in oversampling when compared with the Stein's procedure. Further it also shows that Method I performed much better than Method II in both two-stage procedures. Using these results we can claim that the modified two-stage procedure using Method I performed better than the other method and procedure considered in this analysis.

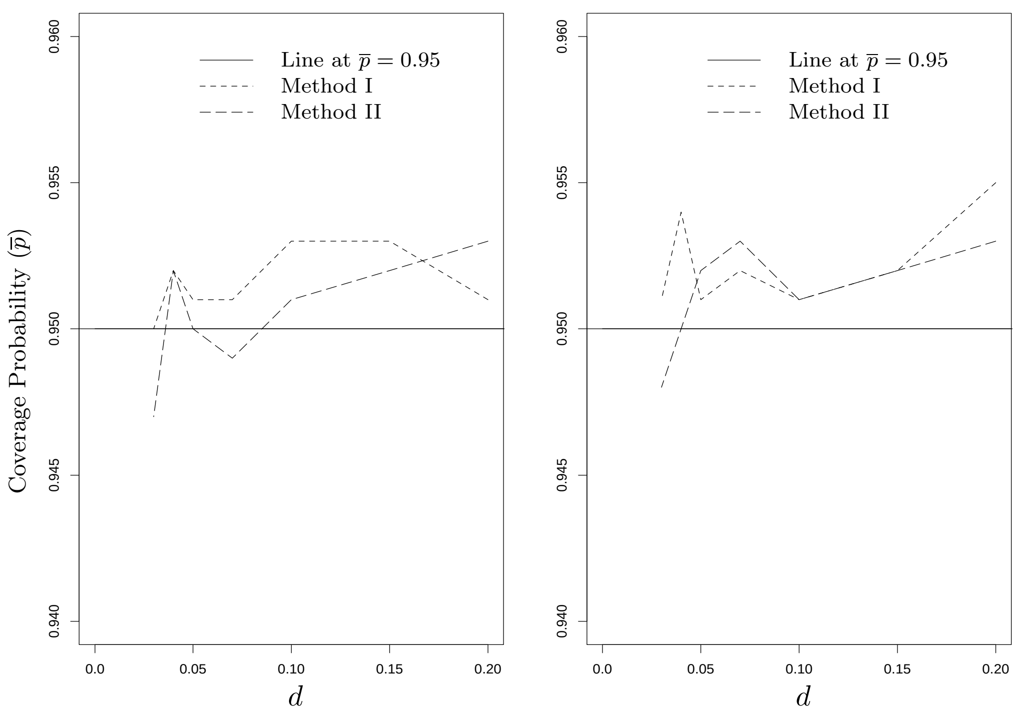


Fig 3.3(a): Stein's Two-Stage

Fig 3.3(b): Modified Two-Stage

***Figure 3.3.*** Coverage Probabilities ($\overline{p}$) Versus $d$ For Method I and Method II

Figures 3.3(a) and 3.3(b) give plots of the coverage probability, $\overline{p}$ versus $d$ respectively for Stein's and modified procedures. Here, $\overline{p}$ is the proportion of the confidence intervals that contains the theoretical difference,

$\delta = \mu_1 - \mu_2 = 5$. That is,

$$\overline{p} = n_\delta / n_{sim}$$

where $n_\delta$ is the number of confidence intervals that contain $\delta$ among $n_{sim} = 15,000$ number of simulations. These figures show that $\overline{p}$ is higher than the target value 0.95 for Method I in both procedures, thus clearly Method I is oversampling for $d$ used in the analysis. In general, the results given in these figures confirm that both procedures using Method I and Method II oversample.

# 4. APPLICATION OF BOOTSTRAPPING

From the simulation results presented in Section 3, it is clear that modified two stage procedure for Method I performed better than other procedures considered in this paper. However Figures 3.2(b) and 3.3(b) showed that this procedure is also oversampling. Among others, de Silva and Mukhopadhyay (2004) showed that a considerable reduction in oversampling in Stein's two-stage and also modified two-stage can be achieved by improving the estimate of the final sample size using bootstraps.

Consider samples of size $n$ from both normal populations. Using the notations given in Section 1, let us assume that $\{X_{11}, X_{12}, \ldots, X_{1n}\}$ is a sample from the first population and $\{X_{21}, X_{22}, \ldots, X_{2n}\}$ is a sample from the second population.

## 4.1. Bootstrapping Procedure for Method I

Let $W_k = X_{1i} - X_{2j}$ where $k = n(i-1) + j$ for $i, j = 1, 2, \ldots, n$. Clearly $k = 1, 2, \ldots, n^2$ and the sample $\left\{W_1, W_2, \ldots, W_{n^2}\right\}$ has $n^2$ observations. Note that $W_k \sim N(\delta, \sigma_0^2)$ but they are not independent random variables.

The sample mean and the variance of the above constructed sample is given by

$$\overline{W}_n = \frac{1}{n^2}\sum_{k=1}^{n^2} W_k \quad \text{and} \quad S_n^2(w) = \frac{1}{n^2-1}\sum_{k=1}^{n^2}\left(W_k - \overline{W}_n\right)^2. \tag{4.1}$$

**Theorem 4.1.**

(*i*) *For a fixed sample size, n*

$$\overline{W}_n = \overline{X}_{1n} - \overline{X}_{2n} = \overline{U}_n \tag{4.2}$$

and

$$S_n^2(w) = \frac{n}{n+1}\left(S_{1n}^2 + S_{2n}^2\right). \tag{4.3}$$

(*ii*) *The stopping rule* $N_1$ *given in (3.1) is identically equal to* the stopping rule (4.4) *given below and thus the confidence interval* $I_{(N_1,d)} = \left(\overline{W}_{N_1} - d, \overline{W}_{N_1} + d\right)$.

$$N_1 = \max\left\{m, \left\langle \frac{1}{2}\eta + \sqrt{\eta + \frac{1}{4}\eta^2} \right\rangle + 1\right\} \tag{4.4}$$

$$\text{where } \eta = \frac{t_\nu^2(\alpha/2) S_n^2(w)}{d^2} \qquad \text{and} \qquad \nu = \left[(n-1)\frac{\left(S_{1n}^2 + S_{2n}^2\right)^2}{S_{1n}^4 + S_{2n}^4}\right].$$

**Proof of Theorem 4.1**(*i*)**:** First we establish the results given in (*i*).
The result given in (4.2) follows directly from (4.1) and $W_k = X_{1i} - X_{2j}$. Next consider

$$\begin{aligned}
S_n^2(w) &= \frac{1}{n^2-1}\sum\nolimits_{k=1}^{n^2}\left(W_k - \overline{W}_n\right)^2 \\
&= \frac{1}{n^2-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{n}\left[(X_{1i} - X_{2j}) - (\overline{X}_{1n} - \overline{X}_{2n})\right]^2 \\
&= \frac{1}{n^2-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{n}\left[(X_{1i} - \overline{X}_{1n}) - (X_{2j} - \overline{X}_{2n})\right]^2 \\
&= \frac{1}{n^2-1}\left[n(n-1)S_{1n}^2 + n(n-1)S_{2n}^2\right] \\
&= \frac{n}{n+1}\left(S_{1n}^2 + S_{2n}^2\right).
\end{aligned}$$

**Proof of Theorem 4.1**(*ii*)**:**
Using (2.1) and the above results we have

$$d \geq \frac{t_\nu(\alpha/2)}{\sqrt{n}}\sqrt{S_{1n}^2 + S_{2n}^2} = \frac{t_\nu(\alpha/2)}{\sqrt{n}}\sqrt{\frac{n+1}{n}S_n^2(w)}.$$

This gives $n \geq \frac{1}{2}\eta + \sqrt{\eta + \frac{1}{4}\eta^2}$ where

$$\eta = \frac{t_\nu^2(\alpha/2) S_n^2(w)}{d^2} \qquad \text{and} \qquad \nu = \left[(n-1)\frac{\left(S_{1n}^2 + S_{2n}^2\right)^2}{S_{1n}^4 + S_{2n}^4}\right]. \tag{4.5}$$

Thus, the stopping rules $N_1$ given in (3.1) can be written as

$$N_1 = \max\left\{m, \left\langle \frac{1}{2}\eta + \sqrt{\eta + \frac{1}{4}\eta^2} \right\rangle + 1\right\} \equiv N_w. \tag{4.6}$$

Hence, confidence interval

$$I_{(N_1,d)} = \left(\overline{U}_{N_1} - d, \overline{U}_{N_1} + d\right) \equiv \left(\overline{W}_{N_1} - d, \overline{W}_{N_1} + d\right). \qquad \blacksquare$$

**Bootstrap Estimator of $N_1$**

First construct the sample $\left\{W_1, W_2, \ldots, W_{m^2}\right\}$ from the initial samples of size $m$ from two normal populations. Let $\left\{W_{1i}^*, W_{2i}^*, \ldots, W_{m^2 i}^*\right\}$ $(i = 1, 2, \ldots, B)$ be the $i^{\text{th}}$ bootstrap sample from $\left\{W_1, W_2, \ldots, W_{m^2}\right\}$.

Let $N_{1i}^*$ be the estimate of $N_1$ given in (4.6) using the $i^{\text{th}}$ bootstrap sample and

$$\overline{N}_1^* = \frac{1}{B} \sum_{i=1}^{B} N_{1i}^*.$$

The bias corrected bootstrap estimate of the final sample size (see Efron and Tibshirani, 1993) is given by

$$\tilde{N}_1^* = \max\left\{m,\ 2N_1 - \overline{N}_1^*\right\}.$$

The above $\tilde{N}_1^*$ gives the bootstrap estimate of the final sample size for Method I. Now to complete the two-stage procedure, if $\tilde{N}_1^* > m$, take a random sample of size $\tilde{N}_1^* - m$, otherwise no sampling is required in this second stage. Now using all $\tilde{N}_1^*$ observations compute the final confidence interval

$$I_{(\tilde{N}_1^*, d)} = \left(\overline{U}_{\tilde{N}_1^*} - d, \overline{U}_{\tilde{N}_1^*} + d\right)$$

where $\overline{U}_{\tilde{N}_1^*} = \overline{X}_{1\tilde{N}_1^*} - \overline{X}_{2\tilde{N}_1^*}$.

## 4.2. Bootstrapping Procedure for Method II

Here we use *critical bootstrap method* to obtain the bootstrap estimate of the final sample size using the initial sample. Consider a bootstrap sample $\left\{V_1^*, \ldots, V_m^*\right\}$ from $\left\{V_1, \ldots, V_m\right\}$ where $V_i = X_{1i} - X_{2i}$ as in Section 2.2.

Let $\overline{V}_{mi}^*$ and $S_{mi}^{*2}(v)$ be the sample mean and variance of the $i^{\text{th}}$ bootstrap sample $(i = 1, \ldots, B)$. Next compute

$$\left|\sqrt{m}(\overline{V}_{mi}^* - \overline{V}_m)/S_{mi}^*(v)\right| \quad \text{for } i = 1, 2 \ldots, B$$

and arranging them in increasing order. Let $\xi_{\alpha/2}^*$ be the $(1-\alpha)B^{\text{th}}$ largest value. Then $\xi_{\alpha/2}^*$ is the $(1-\alpha)^{\text{th}}$ bootstrap quantile of the above distribution. So, we replace $t_{m-1}(\alpha/2)$ in (3.2) by $\xi_{\alpha/2}^*$ to obtain the bootstrap stopping rule, $N_2^*$ for Method II.

$$N_2^* = \max\left\{m,\ \left\langle \xi_{\alpha/2}^{*2} \frac{S_m^2(v)}{d^2} \right\rangle + 1\right\}.$$

As in other two-stage procedures, if $N_2^* > m$ take further $N_2^* - m$ observations and compute the confidence interval using all $N_2^*$ observations.

Otherwise no further observations are required to compute the confidence interval.

Now the required confidence interval using $\tilde{N}_2^* = \max\left\{m,\ 2N_2 - \overline{N}_2^*\right\}$, the bias corrected bootstrap estimate of $N_2$ is given by

$$I_{(\tilde{N}_2^*, d)} = \left(\overline{V}_{\tilde{N}_2^*} - d, \overline{V}_{\tilde{N}_2^*} + d\right).$$

Note that $\overline{V}_{\tilde{N}_2^*} = \overline{X}_{1\tilde{N}_2^*} - \overline{X}_{2\tilde{N}_2^*}$.

## 4.3. Simulation Analysis of the Bootstrapping Procedures

Here we repeat the simulation analysis as in Section 3.2 with bootstrapping the sample $\left\{W_1, W_2, \ldots, W_{m^2}\right\}$ in Method I and sample $\left\{V_1, V_2, \ldots, V_m\right\}$ in Method II. The initial samples for constructing the samples given above are obtained by generating random variates from the $N(0.8, 0.16)$ and $N(0.4, 0.09)$ populations. Here we use Stein's procedure with $m = 15$ and the modified two-stage procedure using $m$ obtained from (3.5). The results of this simulation analysis together with the results without bootstrapping are given in Figures 4.1 - 4.3 and in Table 4.1.

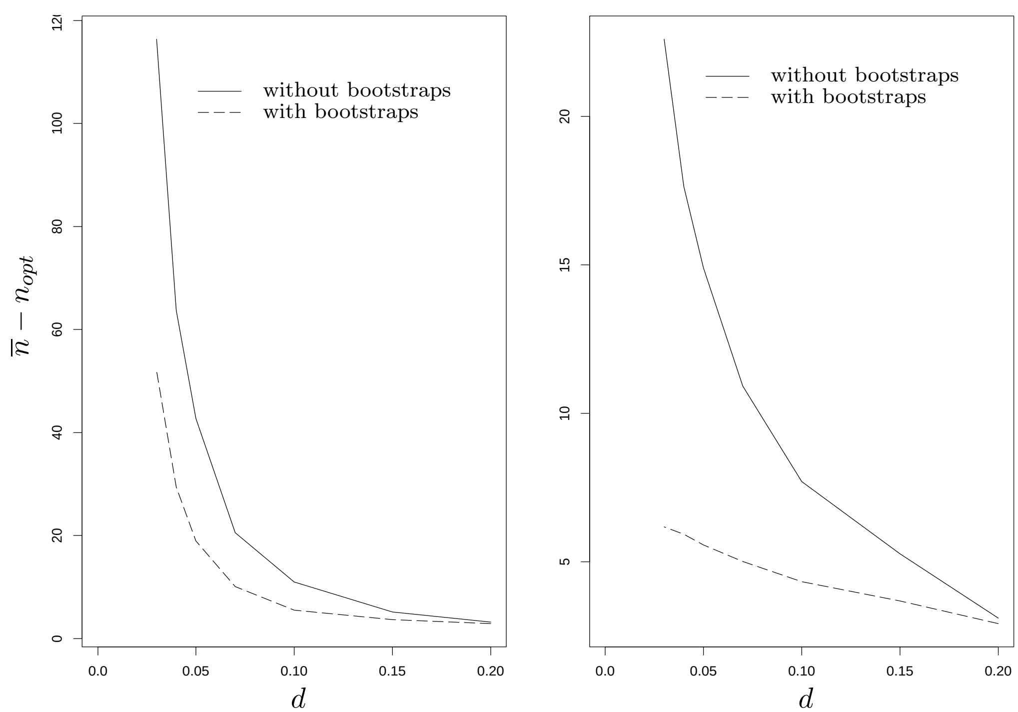


Fig 4.1(a): Stein's Two-Stage

Fig 4.1(b): Modified Two-Stage

***Figure 4.1.*** c Oversampling, $\overline{n} - n_{opt}$ Versus $d$ For Method I

Figures 4.1(a) and 4.1(b) show that the application of bootstrapping to Method I gives considerable reduction in oversampling. However from

Figure 4.2, it appears that the critical bootstrapping method does not reduce the oversampling in Method II.

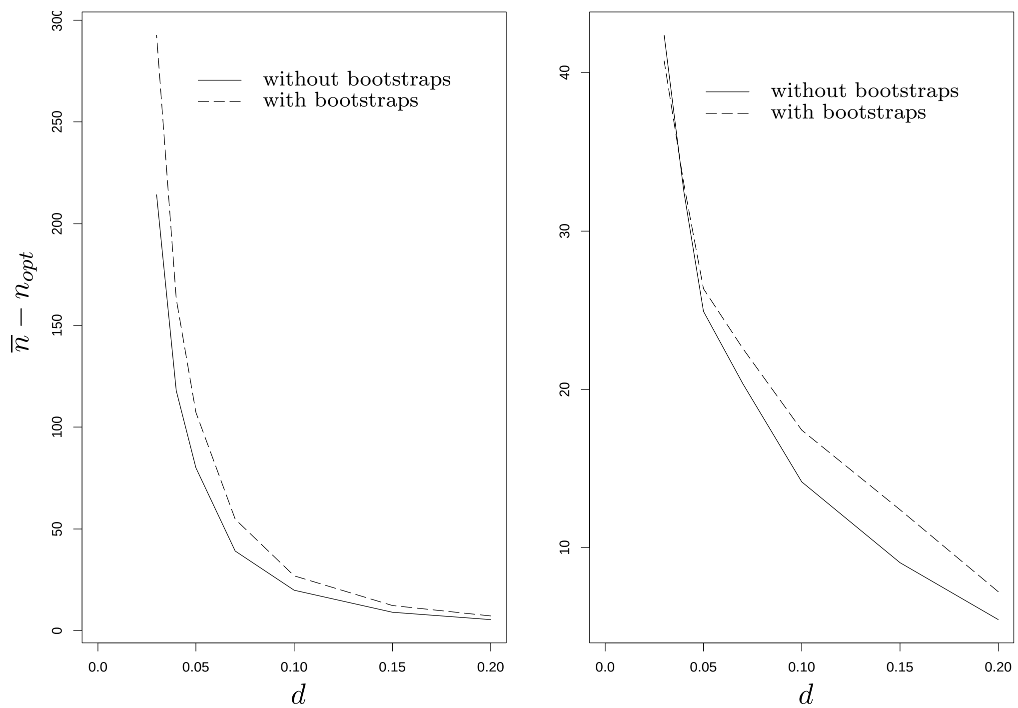


Fig 4.2(a): Stein's Two-Stage    Fig 4.2(b): Modified Two-Stage

***Figure 4.2.*** Oversampling, $\overline{n} - n_{opt}$ Versus $d$ For Method II

**Table 4.1.** Percentage Oversampling With and Without Bootstrapping

| | | Stein's Procedure | | | | Modified Two-Stage | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Method I | | Method II | | Method I | | Method II | |
| $d$ | $n_{opt}$ | $P_{over}$ | $P^*_{over}$ | $P_{over}$ | $P^*_{over}$ | $P_{over}$ | $P^*_{over}$ | $P_{over}$ | $P^*_{over}$ |
| 0.20 | 24.0 | 13.4 | 12.2 | 22.7 | 30.1 | 12.9 | -2.5 | 22.7 | 30.1 |
| 0.15 | 42.7 | 12.0 | 8.6 | 21.2 | 29.0 | 12.3 | -4.4 | 21.2 | 29.0 |
| 0.10 | 96.0 | 11.5 | 5.8 | 20.7 | 28.0 | 8.0 | -4.1 | 14.7 | 18.1 |
| 0.07 | 196.0 | 10.5 | 5.1 | 20.0 | 28.0 | 5.6 | -3.3 | 10.4 | 11.5 |
| 0.05 | 384.2 | 11.1 | 4.9 | 20.8 | 27.9 | 3.9 | -2.4 | 6.5 | 6.9 |
| 0.04 | 600.2 | 10.6 | 4.9 | 19.7 | 27.2 | 2.9 | -2.2 | 5.4 | 5.5 |
| 0.03 | 1067.1 | 10.9 | 4.9 | 20.1 | 27.4 | 2.1 | -1.7 | 4.0 | 3.8 |

Table 4.1 gives the percentage oversampling for these two methods with and without bootstrapping. As in (3.8), we define $P^*_{over}$ as the per-

centage oversampling in bootstrap stopping rule, $N^*$ and thus

$$P^*_{over} = 100(\tilde{n^*} - n_{opt})/n_{opt}$$

where $\tilde{n^*}$ is the simulated value of the bias corrected bootstrap estimate of $N^*$.

The results in Table 4.1 show that the application of bootstrapping to both the Stein's procedure and the modified two-stage procedure using Method I, gives significant reduction in final sample sizes. Corresponding results in Method II show that application of critical bootstrap increases the required final sample sizes. From this result we can conclude that the critical bootstrap method is not appropriate to both two-stage procedures in Method II. Further, it is evident that Method I performed significantly better than Method II for the problem considered in this paper.

## 4.4. Method I With and Without Bootstrapping

Figure 4.3 gives the plots of coverage probabilities of the constructed Stein's and modified two-stage procedures using Method I. These plots clearly show that the coverage probabilities obtained without using bootstraps are greater than 0.95 (target value). Thus clearly both Stein's and modified two-stage procedure oversample. Further, the application of bootstrapping shows a clear reduction in oversampling.

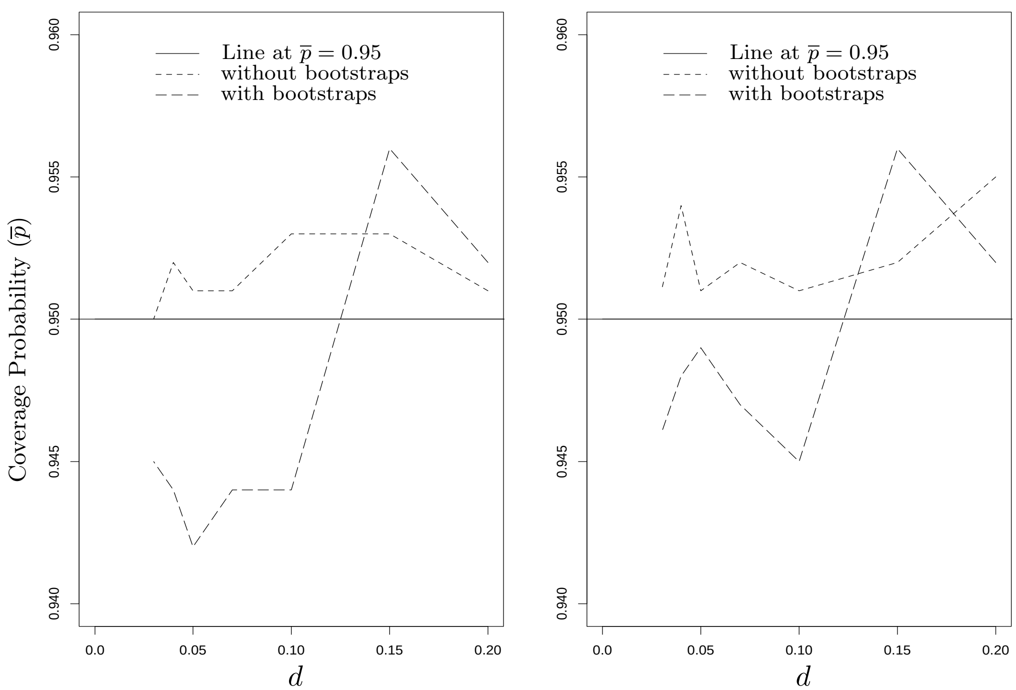


Fig 4.3(a): Stein's Two-Stage

Fig 4.3(b): Modified Two-Stage

Figure 4.3. *Coverage Probabilities* ($\overline{p}$) *Versus* $d$ *For Method I*

# 5. APPLICATION TO THE LITERACY DATA

## 5.1. Literacy Data

The methodology developed in the previous sections has been applied to compare the literacy rates of democratic and 'so called' non-democratic countries.

Friedenberg and Heyward (2003)used discriminant analysis to classify a country as a democratic or non-democratic on the basis of several variables. They obtained the data and information about the countries given in THE WORLD FACTBOOK (2003). One of the variable they used is the **literacy rate**. Using the data given in their paper we estimate the difference between the average literacy rate of these two groups of countries. To see the flavor of the data we give below the randomly selected initial sample of size $m = 15$ for the two groups. The measurement is the percentage of the population who are considered *literate* according to THE WORLD FACTBOOK (2003).

**Table 5.1.** A Random Sample of 15 Democratic and Non-democratic Countries With Their Literacy Rates

| **Democratic** | | **Non-Democratic** | |
|---|---|---|---|
| Country | Literacy(%) | Country | Literacy(%) |
| Albania | 93.0 | Algeria | 61.6 |
| Benin | 37.5 | Azerbaijan | 97.0 |
| Bolivia | 69.8 | Belarus | 98.0 |
| Central African Republic | 60.0 | Burma | 83.1 |
| Georgia | 99.0 | D.R. Congo | 77.3 |
| Greece | 95.0 | Guinea | 78.5 |
| Guatemala | 63.6 | Ethiopia | 35.5 |
| Honduras | 72.7 | Gabon | 63.2 |
| Hungary | 99.0 | Kazakhstan | 98.0 |
| India | 52.0 | Kuwait | 78.6 |
| Indonesia | 83.8 | Libya | 76.2 |
| Israel | 95.0 | Oman | 80.0 |
| Lithuania | 98.0 | Togo | 51.7 |
| Madagascar | 80.0 | Tunisia | 66.7 |
| Namibia | 38.0 | Uganda | 61.8 |

Below we give the summary results of the computation of 95% fixed-width confidence interval. Recall $\delta = \mu_1 - \mu_2 =$ Difference of the average literacy rate between democratic and non-democratic countries. Note that $d = 0.1(10\%)$

## 5.2. Computation of the Confidence Interval

Method I using modified two-stage procedure with and without bootstrapping is applied to the literacy data. For the procedures the data were randomly selected without replacement from the two populations. The following summary results were obtained from the analysis.

Initial sample size $m = 15$

For democratic country data: $\overline{x}_{1m} = 0.7465$ and $s_{1m}^2 = 0.0635$
For nondemocratic country data: $\overline{x}_{2m} = 0.7381$ and $s_{2m}^2 = 0.0268$

Degrees of freedom, $\nu = 32$

Final sample size without bootstrapping $n_1 = 38$
Final sample size with bootstrapping $\tilde{n}_1^* = 34$

Using Final sample size 34: $\overline{x}_{1\tilde{n}_1^*} = 0.7930$ and $\overline{x}_{2\tilde{n}_1^*} = 0.6948$

Therefore the 95% confidence interval for $\delta = \mu_1 - \mu_2$ is $\big(-0.0018,\ 0.1982\big)$.

## ACKNOWLEDGMENTS

We are pleased to thank Professor Nitis Mukhopadhyay (the editor) and the referees for many helpful comments and suggestions on the earlier version of this paper.

## REFERENCES

Chapman, D. G. (1950). Some Two Sample Tests, Annals of Mathematical Statistics 21: 601-606.

Babu, G. J. and Padmanabhan, A. R. (2002). Resampling Methods For the Nonparametric Behrens-Fisher Problem, Sankhya 64: 678-692.

de Silva, B. M. and Mukhopadhyay, N. (2004). Kernel density estimation of wool fibre diameter, In *Applied Sequential Methodologies*; Mukhopadhyay, N., Datta, S., Chattopadhyay, S., Eds. New York New York: Marcel Dekker Inc, 141-170.

Efron B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Friedenberg, D. and Heyward, S. (2003). A Multivariate Statistical Analysis of the Free World, Online (http://www.users.muohio.edu/porterbm/sumj/2002/TOC02.html), SUMSRI Journal - Miami University: Oxford, Ohio U.S.A.

Ghosh, B. K. (1975). A Two-stage Procedure for the Behrens-Fisher Problem, Journal of the American Statistical Association 70: 457-462.

Ghosh, M., Mukhopadhyay, N. and Sen, P.(1997). *Sequential Estimation*, New York: Wiley & Sons Inc.

Mukhopadhyay, N. (1980). A consistent and asymptotically efficient two-stage procedure to construct fixed-width confidence intervals for the mean. Metrica 27: 281-284.

Mukhopadhyay, N. and de Silva, B. M. (1997). Multistage Fixed-Width Confidence Intervals in the Two-Sample Problem: The Normal Case, Journal of Statistical Research 31: 1-20.

Mukhopadhyay, N. and Solanky, T. K. S. (1994). *Multistage Selection and Ranking Procedures, Second-order Asymptotic*, New York: Marcel Dekker Inc.

Satterthwaite, F. W. (1946). An Approximate Distribution of Estimates of Variance Components, Biometrics Bulletin 2: 110 -114.

Stein, C. (1945). A two-Sample Test for a Linear Hypothesis Whose Power Is Independent of the Variance, Annals of Mathematical Statistics 16: 243-258.

Stein, C. (1949). Some Problems in Sequential Estimation (abstract), Econometrica 17: 77-78.

*THE WORLD FACTBOOK* (2003). Website:
http://www.odci.gov/cia/publications/factbook/fields/2103.html