Future of Tuned Ratio Unbiased Mean Predictor (TRUMP) with the Unified Scrambling Approach (USA)

Sarjinder Singh and Stephen A. Sedory
Department of Mathematics
Texas A&M University-Kingsville
Kingsville, TX 78363, USA
E-mail: sarjinder@yahoo.com

Abstract
The Tuned Ratio Unbiased Mean Predictor (TRUMP) was introduced by Singh and Sedory (2017: Survey Research Methods Section, Proceedings of the American Statistical Association, pp. 1746-1759). They have shown that the proposed TRUMP when utilizing First Basic Information (FBI) about the TRUMP Care coefficient can perform better than the Best Linear Unbiased Estimator (BLUE) and, also can perform better than the Best Linear Unbiased Predictor (BLUP). Warner (1965: Journal of the American Statistical Association, pp. 63-69) introduced the idea of estimating the population proportion of a sensitive attribute by making use of randomization device. Later on, the idea was extended to estimate the population mean of a sensitive variable by making use of an approach involving additive and multiplicative scrambling variables. In this paper, we will study the future of the TRUMP with a Unified Scrambling Approach (USA) along the lines of Singh, Joarder and King (1996: Australian Journal of Statistics, pp. 201-211). Making a great adjustment (MAGA) by means of scrambling variables may help TRUMP have more precise estimates of frauds, induced abortions, illegal immigration, extramarital relations, tax returns, illegal drugs, and cheating etc. The results based on theory and simulation study will be reported.

Key Words: Population mean, Scrambled responses, Jackknifing, TRUMP Cuts, TRUMP Care coefficient, Linear model.

1. Introduction

Assume y_i and x_i, $i = 1,2,..., N$, be the values of the i[th] unit of the study variable and auxiliary variable, respectively, in a population Ω. Consider the problem of estimating the population mean

$$\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i \tag{1.1}$$

by assuming that the population mean

$$\bar{X} = N^{-1}\sum_{i=1}^{N} x_i \tag{1.2}$$

of the auxiliary variable is known.

Assume a simple random and with replacement sampling (SRSWR) s of n units is taken from the population of interest Ω. Assume (y_i, x_i), $i = 1,2,..., n$, be the values of the i[th] unit, in the sample s, of the study variable and auxiliary variable. Let the sample means for the study variable and the auxiliary variable, respectively, be given by

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1.3}$$

and

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.4}$$

Cochran (1940) defined a ratio estimator of the population mean $\bar{Y}$ given by

$$\bar{y}_{Rat} = \bar{y}_n\left(\frac{\bar{X}}{\bar{x}_n}\right) \tag{1.5}$$

In case the regression line is passing through the origin, the well-known mean predictor model is given by:

$$y_i = Rx_i + e_i \tag{1.6}$$

where

$$R = \frac{\bar{Y}}{\bar{X}} \tag{1.7}$$

is the ratio of the population mean of the study variable to that of the auxiliary variable.

$$E_m\left(e_i \mid x_i\right) = 0 \tag{1.8}$$

$$E_m\left(e_i^2 \mid x_i\right) = V_m\left(e_i \mid x_i\right) = \sigma^2 \tag{1.9}$$

and

$$E_m\left(e_i e_j \mid x_i x_j\right) = C_m\left(e_i, e_j \mid x_i, x_j\right) = 0 \tag{1.10}$$

where E_m, V_m, and C_m denote the model expectation, variance, and covariance, respectively. Singh and Sedory (2017) developed the Tuned Ratio Unbiased Mean Predictor (TRUMP) $\bar{y}_{\text{TRUMP}}$ given by:

$$\bar{y}_{\text{TRUMP}} = \frac{\sum_{j=1}^{n} \bar{y}_n(j)_{TC}\,\bar{x}_n(j)_{TC}}{\sum_{j=1}^{n}\left\{\bar{x}_n(j)_{TC}\right\}^2}\bar{X} \tag{1.11}$$

where

$$\bar{y}_n(j)_{TC} = \frac{n^g\, y_j - \bar{y}_n}{n^g - 1} \tag{1.12}$$

$$\bar{x}_n(j)_{TC} = \frac{n^g\, x_j - \bar{x}_n}{n^g - 1} \tag{1.13}$$

are called TRUMP Cuts (TC), and $\bar{w}_n(j)$ are called tuned calibrated weighs to be determined based on certain criterion.

Fig. 1.1. TRUMP Cuts versus Jackknifing

The TC is obtained by calibrating the jth sampled observation y_j by n^g, and then subtracting the sampled mean value $\bar{y}_n$. The value of $g \neq 0$ is called TRUMP Care coefficient.

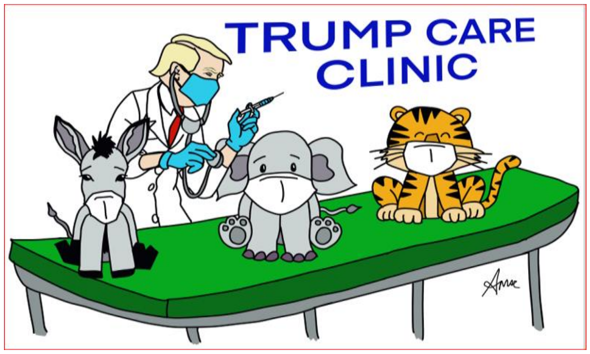

Fig. 1.2. TRUMP Care coefficient

Its value depends on the First Basic Information (FBI). For example, if $g = -1$, then

$$\bar{y}_n(j)_{TC} = \frac{y_j - n\bar{y}_n}{1-n} = \frac{n\bar{y}_n - y_j}{n-1} = \bar{y}_n(j) \tag{1.14}$$

which is the usual jackknifing due to Quenouille (1956) and was first used to estimate the variance by Tukey (1958).

In the next section, we discuss the need of randomized response techniques.

2. Need of Randomized Response Techniques

The randomized response technique is used for reducing response error problems when potentially "sensitive questions" are present in surveys of human populations. Direct questioning of the respondents about sensitive issues often results in either refusal to respond or falsification of the answers. This introduces a non-sampling error that can bias sample estimates. The problem arises when a survey asks personal questions on a sensitive issue like the illegal use of drugs, level of income, incidents of incest or acts of domestic violence. The most serious difficulty in studying these types of problems is the lack of reliable data on their incidence. Social stigma and fear of reprisals sometimes result in lying by the respondents when approached with the conventional or direct question survey method.

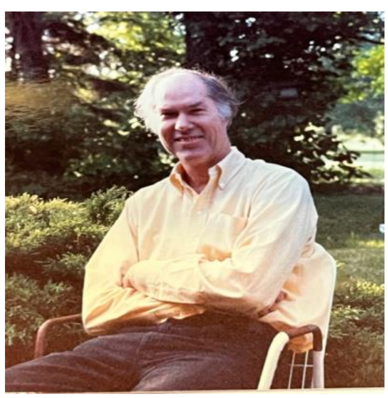

Fig. 2.1. Prof. Stanley L. Warner
(*Dec 23, 1928 – Aug 25, 1992*)
Printed with a written permission.

Warner (1965), Figure 2.1, first suggested the concept of randomized response sampling. His approach was to use a randomization device, such as a spinner, to question respondents about sensitive topics in a way that the privacy of the person being interviewed would be preserved. Instead of directly asking the respondent the confidential question, the interviewer would employ the randomization device. Utilizing a spinner, a survey respondent can use the following technique to adhere to Warner's method. There are two sections on the spinner of different sizes. One has the phrase "I belong to sensitive group A" printed on it, while the other has "I do not belong to sensitive group A." A respondent is instructed to spin the spinner in private and read the statement to themselves. Depending on their actual status, they would give the response "yes" or "no" to the interviewer. The respondent's privacy is maintained because a "yes" answer means either: "yes" - I belong to sensitive group A, or "yes" – I do not belong to sensitive group A. The same scenario exists for a "no" answer. This method shields the respondent's privacy and gives them confidence to answer the question honestly since they know the interviewer does not have any idea which group their replies are coming from. In the Warner model the true proportion of members belonging to group A is π, and the true proportion of members belonging to group A^c is $(1-\pi)$. The probability of the spinner landing on "I belong to sensitive group A" is P, and the probability of the spinner landing on "I do not belong to group A" is $(1-P)$. A pictorial representation of the Warner model is shown in Figure 2.2.

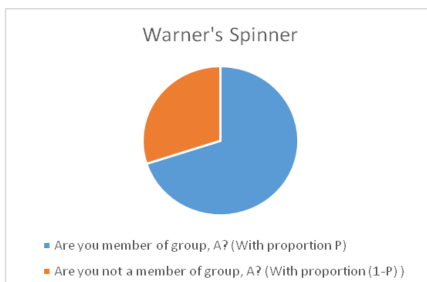

Fig. 2.2. Warner's pioneer model.

Then θ, the probability of a "yes" answer is given by:

$$\theta = P(yes) = P\ \pi + (1-P)(1-\pi) \tag{2.1}$$

Assume a SRSWR of n people is selected from the population of interest, and x is the total number of observed "yes" answers, then the maximum likelihood estimator of π given by

$$\hat{\pi}_W = \frac{\hat{\theta}-(1-P)}{2P-1}, \text{ where, } P \neq 0.5 \tag{2.2}$$

where $\hat{\theta} = x/n$ is the observed proportion of "yes" answers. The estimator $\hat{\pi}_W$ is found to be unbiased estimator of the population proportion π. The variance of estimator $\hat{\pi}_W$ is given by

$$V(\hat{\pi}_W) = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2} \tag{2.3}$$

His idea has spawned a vast literature which has been reviewed by Fox (2016), Pushadapu and Singh (2024), Pushadapu, Singh and Sedory (2024), Naatjes, Sedory and Singh (2023, 2024a 2024b), Aguirre-Hamilton, Sedory and Singh (2024), Zapata, Sedory and Singh (2022, 2023), Xu, Sedory and Singh (2021a, 2021b, 2022), Arias, Sedory and Singh (2021, 2022), Olanipekun, Zhao, Wang, Sedory, and Singh (2023), Zheng, Sedory, and Singh (2021a, 2021b), Yennum, Sedory and Singh. (2019, 2022), Murtaza, Singh, and Hussain (2020, 2021), Lee, Sedory and Singh (2013a, 2013b, 2021), Lee, Su, Mondragon, Salinas, Zamora, Sedory, and Singh (2016), Ahmed, Sedory and Singh (2018, 2020), Su, Sedory and Singh (2015, 2017, 2021), Sedory, Singh, Olanipekun, and Wark (2020), Jayaraj, Odumade, and Singh (2018a, 2018b) and Jayaraj, Sedory, Singh, and Odumade (2018), Odumade and Singh (2008, 2009a, 2009b, 2010), and Gjestvang and Singh (2006, 2007, 2009). Some modifications to the Warner's model have been suggested by Franklin (1989), Kuk (1990), Mangat (1994), Mangat and Singh (1990), Singh, Mangat, and Singh (1993), Singh (1994), Singh, Singh, Mangat and Tracy (1994) , Bansal, Singh and Singh (1994), Abdelfatah, S. and Mazloum, R. (2016), and Abdelfatah, Mazloum and Singh (2013), Arnab and Shangodoyin (2020), Batool, Shabbir and Hussain (2017), Chaudhuri (2011), Chaudhuri and Christofides (2013). Chaudhuri, Christofides, and Rao (2016) and Chen and Singh (2011) etc. Himmelfarb and Edgell (1980) introduced the idea of additive scrambled randomized response model which they used to estimate the population mean of a sensitive variable by making use of the known distribution of a scrambling variable. Eichhorn and Hayre (1983) came up with the idea of multiplicative randomized response model which could also be used to estimate the population mean of a sensitive variable. Chaudhuri and Stenger (1992) incorporated another ingenious suggestion of combining both the additive and multiplicative model together as the Unified Scrambling Approach (USA).

Fig. 2.3. Map of the USA.

In their randomization device, the i^{th} person selected in the sample s , is provided with a set of two randomization devices, say Device-I and Device-II. The Device-I consists of M_1 cards marked with numbers $\{S_{11}, S_{12}, ..., S_{1M_1}\}$ and the Device-II consists of M_2 cards marked with numbers $\{S_{21}, S_{22}, ..., S_{2M_2}\}$. The mean and variances of the numbers marked on the cards in the Device-I and Device -II are assumed to be known, and are computed, respectively, as:

$$\mu_{S_1} = \frac{1}{M_1} \sum_{i=1}^{M_1} S_{1i} \tag{2.4}$$

$$\sigma^2_{S_1} = \frac{1}{M_1} \sum_{i=1}^{M_1} \left(S_{1i} - \mu_{S_1}\right)^2 \tag{2.5}$$

$$\mu_{S_2} = \frac{1}{M_2} \sum_{i=1}^{M_2} S_{2i} \tag{2.6}$$

$$\sigma^2_{S_2} = \frac{1}{M_2} \sum_{i=1}^{M_2} \left(S_{2i} - \mu_{S_2}\right)^2 \tag{2.7}$$

The scrambling variables S_{1i} and S_{2i} can be generated from the Normal, Weibull or any discrete distribution. Chaudhuri and Stenger (1992) suggested the Unified Scrambling Approach (USA) as:

$$Z_i = \frac{Y_i S_1 + S_2 - \theta_2}{\theta_1} \tag{2.8}$$

where S_1 and S_2 are two scrambling variables such that $E(S_1) = \theta_1$, $V(S_1) = \gamma_1^2$, $E(S_2) = \theta_2$, $V(S_2) = \gamma_2^2$ and $Cov(S_1, S_2) = \rho_{s_1 s_2} \gamma_1 \gamma_2$ are known. Most of the studies considered

$\rho_{s_1 s_2} = 0$. The benefit of use of $\rho_{s_1 s_2} \neq 0$ was first time highlighted by Murtaza, Singh, and Hussain (2020, 2021, 2024). Following Singh, Joarder ad King (1996). we consider linear model:

$$z_i = \beta x_i + \eta_i \tag{2.9}$$

where

$$z_i = \frac{y_i |S_1|^{p_1} + |S_2|^{p_2} - \theta_2}{\theta_1} \tag{2.10}$$

is a new scrambled response made by the **U**nified **S**crambling **A**pproach (USA) such that $S_1 \sim Cauchy(0, h_1)$ and $S_2 \sim Cauchy(0, h_2)$. Then for $-0.5 < p_1 < 0.5$ and $-0.5 < p_2 < 0.5$, we have

$$\theta_1 = E\left(|S_1|^{p_1}\right) = h_1^{p_1} \mathrm{Sec}\left(\frac{\pi p_1}{2}\right) \tag{2.11}$$

$$\gamma_1^2 = E\left\{\left(|S_1|^{p_1}\right)^2\right\} - \left\{E\left(|S_1|^{p_1}\right)\right\}^2 = h_1^{2p_1}\left[\mathrm{Sec}(\pi p_1) - \mathrm{Sec}^2\left(\frac{\pi p_1}{2}\right)\right] \tag{2.12}$$

$$\theta_2 = E\left(|S_2|^{p_2}\right) = h_2^{p_2} \mathrm{Sec}\left(\frac{\pi p_2}{2}\right) \tag{2.13}$$

$$\gamma_2^2 = E\left\{\left(|S_2|^{p_2}\right)^2\right\} - \left\{E\left(|S_2|^{p_2}\right)\right\}^2 = h_2^{2p_2}\left[\mathrm{Sec}(\pi p_2) - \mathrm{Sec}^2\left(\frac{\pi p_2}{2}\right)\right] \tag{2.14}$$

.and

$$Cov\left\{|S_1|^{p_1}, |S_2|^{p_2}\right\} = \rho_{|S_1|^{p_1}|S_2|^{p_2}} \gamma_1 \gamma_2 \tag{2.15}$$

for

$$-1 \leq \rho_{|S_1|^{p_1}|S_2|^{p_2}} \leq +1 \tag{2.16}$$

in such a way that most of available scrambling methods are special case of it, and it also allows to use the scrambling random numbers from a Cauchy distribution.

3. TRUMP with the USA

We consider the Tuned Ratio Unbiased Mean Predictor (TRUMP) with the Unified Scrambling Approach (USA) as:

$$\bar{y}_{\mathrm{TRUMP}}^{USA} = \frac{\sum\limits_{j=1}^{n} \bar{z}_n(j)_{TC}\, \bar{x}_n(j)_{TC}}{\sum\limits_{j=1}^{n} \left\{\bar{x}_n(j)_{TC}\right\}^2} \bar{X} \tag{3.1}$$

where

$$\bar{z}_n(j)_{TC} = \frac{n^g z_j - \bar{z}_n}{n^g - 1} \tag{3.2}$$

are TRUMP Cuts (TC) on the scrambled responses and $g \neq 0$ is called TRUMP Care coefficient. Following Singh and Sedory (2017), we have

$$\bar{z}_n(j)_{TC} = R\, \bar{x}_n(j)_{TC} + \bar{\eta}_n(j)_{TC} \tag{3.3}$$

where

$$\bar{\eta}_n(j) = \frac{n^g \eta_j - \bar{\eta}_n}{n^g - 1} \tag{3.4}$$

Let E_m and E_R be the expected values over the original model and the randomization device, respectively. We have

$$E\left\{\bar{\eta}_n(j)_{TC}\right\} = E_m E_R\left[\frac{n^g \eta_j - \bar{\eta}_n}{n^g - 1}\right] = E_m\left[\frac{n^g e_j - \bar{e}_n}{n^g - 1}\right] = 0 \tag{3.5}$$

Let V_m and V_R be the variance over the original model and the randomization device, respectively. Let E_p and V_p be the expected value and variance over the sampling design p, respectively. Let C_m and C_R be the covariance over the original model and the randomization device, respectively. Let E_p and V_p be the expected value and variance over the sampling design p, respectively. Then we have

$$V\left\{\bar{\eta}_n(j)_{TC}\right\} = V_m E_R\left[\frac{n^g \eta_j - \bar{\eta}_n}{n^g - 1}\right] + E_m V_R\left[\frac{n^g \eta_j - \bar{\eta}_n}{n^g - 1}\right] = \frac{\left(n^{2g+1} + 1 - 2n^g\right)\sigma^2}{n\left(n^g - 1\right)^2} + \Psi_j \tag{3.6}$$

where

$$\begin{aligned}\Psi_j = \frac{1}{n\left(n^g-1\right)^2}\Bigg[&\left\{n^{2g+1} - 2n^g\right\}\left\{\left(R^2 x_j^2 + \sigma^2\right)\gamma_1^2 + \gamma_2^2 + 2\left(Rx_j\right)\rho_{|S_1|p_1|S_2|p_2}\gamma_1\gamma_2\right\} \\ &+ \frac{1}{n}\sum_{j=1}^{n} = \left\{\left(R^2 x_j^2 + \sigma^2\right)\gamma_1^2 + \gamma_2^2 + 2\left(Rx_j\right)\rho_{|S_1|p_1|S_2|p_2}\gamma_1\gamma_2\right\}\Bigg]\end{aligned} \tag{3.7}$$

Also, we have

$$\begin{aligned}Cov\left\{\bar{\eta}_n(j)_{TC}, \bar{\eta}_n(j')_{TC}\right\} &= E_m C_R\left[\frac{n^g \eta_j - \bar{\eta}_n}{n^g - 1}, \frac{n^g \eta_{j'} - \bar{\eta}_n}{n^g - 1}\right] \\ &\quad + C_m\left[E_R\left(\frac{n^g \eta_j - \bar{\eta}_n}{n^g - 1}\right), E_R\left(\frac{n^g \eta_{j'} - \bar{\eta}_n}{n^g - 1}\right)\right] \\ &= \frac{\left(1 - 2n^g\right)\sigma^2}{n\left(n^g - 1\right)^2} + \Psi_{jj'}\end{aligned} \tag{3.8}$$

.where

$$\begin{aligned}\Psi_{jj'} = \frac{1}{\theta_1^2\left(n^g-1\right)^2}\Bigg[&\frac{1}{n^2}\sum_{j=1}^{n}\left\{\left(R^2 x_j^2 + \sigma^2\right)\gamma_1^2 + \gamma_2^2 + 2\left(Rx_j\right)\rho_{|S_1|p_1|S_2|p_2}\gamma_1\gamma_2\right\} \\ &- n^{g-1}\left\{\left(R^2 x_j^2 + \sigma^2\right)\gamma_1^2 + \gamma_2^2 + 2\left(Rx_j\right)\rho_{|S_1|p_1|S_2|p_2}\gamma_1\gamma_2\right\} \\ &- n^{g-1}\left\{\left(R^2 x_{j'}^2 + \sigma^2\right)\gamma_1^2 + \gamma_2^2 + 2\left(Rx_{j'}\right)\rho_{|S_1|p_1|S_2|p_2}\gamma_1\gamma_2\right\}\Bigg]\end{aligned} \tag{3.9}$$

Now we have the following theorems:

Theorem 3.1. The proposed predictor $\bar{y}_{TRUMP}^{USA}$ is unbiased with the **U**nified **S**crambling **A**pproach (USA).

Proof. The proposed TRUM with the USA can be written as:

$$\bar{y}^{USA}_{\text{TRUMP}} = \bar{Y} + \bar{X} \sum_{j=1}^{n} \bar{\mathrm{H}}_n(j)_{TC}\, \bar{\eta}_n(j)_{TC} \tag{3.10}$$

.where

$$\bar{\mathrm{H}}_n(j)_{TC} = \frac{\bar{x}_n(j)_{TC}}{\sum_{j=1}^{n}\left\{\bar{x}_n(j)_{TC}\right\}^2} \tag{3.11}$$

Taking expected value on both sides of the Equation (3.10), we have

$$\begin{aligned}
E\left[\bar{y}^{USA}_{\text{TRUMP}}\right] &= E_p E_m E_R\left[\bar{y}^{USA}_{\text{TRUMP}}\right] \\
&= E_p E_m E_R\left[\bar{Y} + \bar{X}\sum_{j=1}^{n}\bar{\mathrm{H}}_n(j)_{TC}\,\bar{\eta}_n(j)_{TC}\right] \\
&= E_p\left[\bar{Y} + \bar{X}\sum_{j=1}^{n}\bar{\mathrm{H}}_n(j)_{TC}\, E_m E_R\left\{\bar{\eta}_n(j)_{TC}\right\}\right] \\
&= E_p\left(\bar{Y}\right) \\
&= \bar{Y}
\end{aligned} \tag{3.12}$$

which proves the theorem.

Now we have the following theorem:

Theorem 3.2. The variance of the proposed predictor $\bar{y}^{USA}_{TRUMP}$ with the **U**nified **S**crambling **A**pproach (USA) is given by:

$$\begin{aligned}
V\left(\bar{y}^{USA}_{TRUMP}\right) = \bar{X}^2 E_p\Bigg[&\sum_{j=1}^{n}\left\{\bar{\mathrm{H}}_n(j)_{TC}\right\}^2\left\{\frac{\left(n^{2g+1}+1-2n^g\right)\sigma^2}{n\left(n^g-1\right)^2}+\Psi_j\right\} \\
&+\sum_{j\neq j'=1}^{n}\bar{\mathrm{H}}_n(j)_{TC}\,\bar{\mathrm{H}}_n(j')_{TC}\left\{\frac{\left(1-2n^g\right)\sigma^2}{n\left(n^g-1\right)^2}+\Psi_{jj'}\right\}\Bigg] \\
+\bar{X}^2 E_p\Bigg[&\frac{\left(n^{2g+1}+1-2n^g\right)\sigma^2}{n\left(n^g-1\right)^2}\left\{\frac{1}{\sum_{j=1}^{n}\left(\bar{x}_n(j)_{TC}\right)^2}\right\} \\
&+\frac{\left(1-2n^g\right)\sigma^2}{n\left(n^g-1\right)^2}\sum_{j\neq j'=1}^{n}\bar{\mathrm{H}}_n(j)_{TC}\,\bar{\mathrm{H}}_n(j')_{TC}\Bigg]
\end{aligned} \tag{3.13}$$

Proof. We have

$$V\left[\bar{y}^{USA}_{TRUMP}\right] = E_p E_m V_R\left[\bar{y}^{USA}_{TRUMP}\right] + E_p V_m E_R\left[\bar{y}^{USA}_{TRUMP}\right] + V_p E_m E_R\left[\bar{y}^{USA}_{TRUMP}\right]$$

$$= E_p E_m V_R \left[\bar{Y} + \bar{X} \sum_{j=1}^{n} \bar{\mathrm{H}}_n(j)_{TC}\, \bar{\eta}_n(j)_{TC} \right]$$

$$+ E_p V_m E_R \left[\bar{Y} + \bar{X} \sum_{j=1}^{n} \bar{\mathrm{H}}_n(j)_{TC}\, \bar{\eta}_n(j)_{TC} \right]$$

$$+ V_p E_m E_R \left[\bar{Y} + \bar{X} \sum_{j=1}^{n} \bar{\mathrm{H}}_n(j)_{TC}\, \bar{\eta}_n(j)_{TC} \right]$$

$$= \bar{X}^2 E_p E_m \left[\sum_{j=1}^{n} \left\{ \bar{\mathrm{H}}_n(j)_{TC} \right\}^2 \left\{ \frac{\left(n^{2g+1} + 1 - 2n^g\right)\sigma^2}{n\left(n^g - 1\right)^2} + \Psi_j \right\} \right.$$

$$\left. + \sum_{j \neq j'=1}^{n} \bar{\mathrm{H}}_n(j)_{TC}\, \bar{\mathrm{H}}_n(j')_{TC} \left\{ \frac{\left(1 - 2n^g\right)\sigma^2}{n\left(n^g - 1\right)^2} + \Psi_{jj'} \right\} \right]$$

$$+ E_p V_m \left[\bar{Y} + \bar{X} \sum_{j=1}^{n} \bar{\mathrm{H}}_n(j)_{TC} \left\{ \bar{e}_n(j)_{TC} \right\} \right] + V_p E_m \left[\bar{Y} + \bar{X} \sum_{j=1}^{n} \bar{\mathrm{H}}_n(j)_{TC} \left\{ \bar{e}_n(j)_{TC} \right\} \right]$$

which proves the theorem.

In the next section, we are making a great adjustment (MAGA) to the choice of device parameters and the TRUMP Care coefficient so that the proposed TRUMP can perform better than its competitor.

4. MAGA-Making a Great Adjustment

Following Singh and Horn (1998), we created two scrambling variables

$$s_1 = \theta_1 + \left(s_1^* - \theta_1\right)\sqrt{1 - \rho_{s_1 s_2}^2} + \rho_{s_1 s_2} \frac{\gamma_1}{\gamma_2}\left(s_2^* - \theta_2\right) \tag{4.1}$$

and

$$s_2 = s_2^* \tag{4.2}$$

where $s_1^* \sim Cauchy(0,1)$ and $s_2^* \sim Cauchy(0,1)$ are independently generated random variables from the standard Cauchy distribution using R function rcauchy(10000).

Fig. 4.1. Crunching helpful TRUMP Care coefficients.

To compute θ_1, θ_2, γ_1 and γ_2, we fixed $p_1 = 0.2$, $h_1 = 10$, $p_2 = 0.4$, and $h_2 = 20$. We generated 10,000 scrambling variables ordered pairs (s_1, s_2) with a value of correlation coefficient $\rho_{s_1 s_2} = 0.70$. Following Singh and Sedory (2017), from the generator of data (GOD), we again borrowed data through the linear model

$$y_i = R x_i + e_i \tag{4.3}$$

where $e_i \sim N(0,1)$, $x_i \sim Gamma(2,2.5)$ and $R = 4.2$ for $n = 100, 150, 200 \text{ and } 250$. The values of e are generated from rnorm(n,0,1) and X_i are from rgamma(n, 2,2.5). The scrambled reposes Z_i are obtained by randomly sampling n ordered pairs (s_1, s_2) out of the generated 10,000 pairs over the $nitr = 50{,}000$ iterations. The choice of $-\frac{1}{2} < p_1 < \frac{1}{2}$, $-\frac{1}{2} < p_2 < \frac{1}{2}$, $h_1 > 0$, $h_2 > 0$ and the TRUMP Care coefficient (g) makes a **v**ariance **a**djustable **n**ewly **c**reated **e**quipment (VANCE) for the jeopardy defence (JD) system

Fig.4.2. A locksmith's security.

to achieve the minimum variance of the estimator (VOTE)

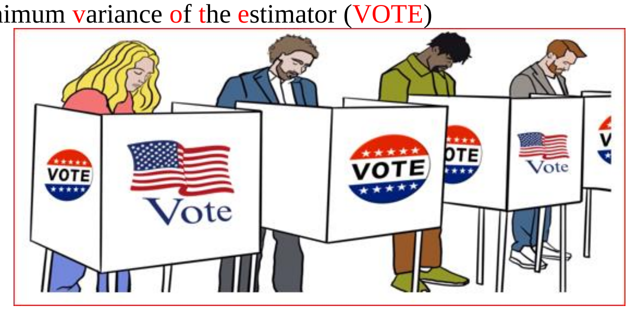

Fig. 4.3. A voting booth.

to beat the competitors.

At the same level of JD, the percent relative efficiency of the proposed TRUMP with respect to the usual ratio estimator is:

$$RE = \frac{\sum_{i=1}^{nitr} \left[\bar{y}_{rat}^{USA}(i) - \bar{Y} \right]^2}{\sum_{i=1}^{nitr} \left[\bar{y}_{TRUMP}^{USA}(i) - \bar{Y} \right]^2} \times 100\% \tag{4.4}$$

where, from Ericksson (1973),

$$\bar{y}_{rat}^{USA} = \bar{z}_n \left(\frac{\bar{X}}{\bar{x}_n} \right) \tag{4.5}$$

is the ratio estimator with the USA. By varying the value of the TRUMP Care coefficient $g \in [-0.5, 0.0) \cup (0.0, 0.5]$ with a skip of 0.02, the results are presented in graphs in Figure 4.4.

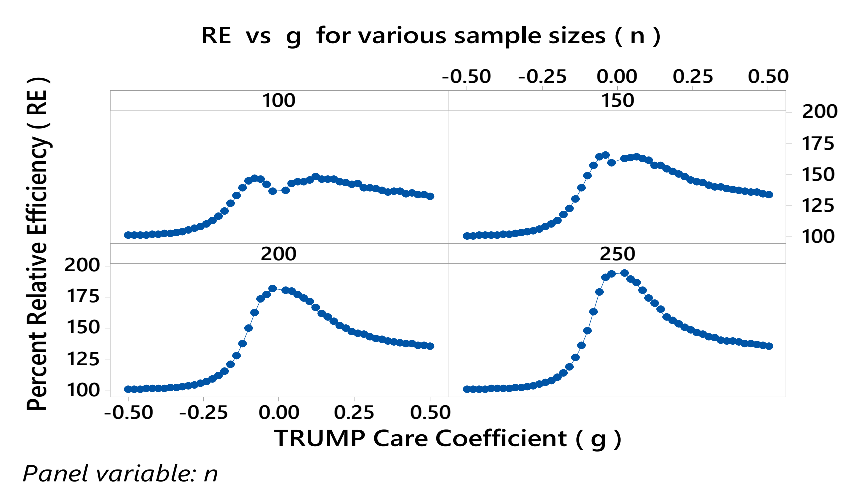

Fig. 4.4. Optimal TRUMP Care coefficients.

We noted that making a great adjustment (MAGA) of the TRUMP Care coefficient with the Unified Scrambling Approach (USA), the proposed TRUMP can perform better than the usual ratio estimator.

Fig.4.5. A golf player optimizing score.

5. Future of TRUMP with the USA

The future of the proposed **TRUMP** with the **USA** is very wide, cannot be summarized in a sentence or a paragraph, however a few directions are below:

(a) Extension to Tuned Regression Unbiased Mean Predictor (TRUMP) with the Unified Scrambling Approach (USA) on the lines of Singh and Sedory (2019) and Hansen, Hurwitz and Madow (1953).

(b) Extension to other sampling designs such as stratified random sampling, complex survey designs. (Singh and Sedory (2017a)).

(c) The use of use-phase sampling design would make it more practical to deal with situations when the population mean of the auxiliary variable is not known.

(d) Both the study variable and the auxiliary variable can be scrambled with the new USA on the lines of Diana and Perri (2011), and so on.

(e) All calibration work following Deville and Särndal (1992) and Singh (2003, 2004) is extendable to study various forms of the TRUMP methodology.

(f) Anyone who improves the TRUMP methodology, can claim I-TRUMP.

Acknowledgements

The authors are thankful to the R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. This presentation is a part of a chapter in Singh and Sedory (2017a), and R codes are also included in the chapter. All the opinions are of authors' and do not belong to any institute or organization. The use of ARTEXPLOSION 600,000 in few drawings is also duly acknowledged. Any resemblance to actual people, living or dead, or actual events is purely coincidental. The second author Prof. Sedory is now retired. Additional acknowledgements are available in Singh and Sedory (2017a).

References

Abdelfatah, S. and Mazloum, R. (2015). An improved stratified randomized response model using two decks of cards, *Model. Assist. Stat. Appl.* 10, 300–320.

Abdelfatah and, S. and Mazloum, R. (2016). An efficient two-stager and optimized response model under stratified random sampling, *Math. Popul. Stud.* 23, 222–238.

Abdelfatah, S., Mazloum, R., and Singh, S. (2013). Efficient use of a two-stage randomized response procedure. *Braz. J. Probab. Stat* .27. 608–617

Aguirre-Hamilton, C., Sedory, S.A. and Singh, S. (2024). Franklin's Randomized Response Model with Correlated Scrambled Variables. *Statistica Neerlandica*, 78(2), 302-309.

Ahmed, S., Sedory, S.A. and Singh, S. (2018). Simultaneous estimation of means of two sensitive variables. *Comm. Statist. Theory Methods* 47(2), 324–343.

Ahmed, S., Sedory, S.A. and Singh, S. (2020). Forcibly Re-scrambled randomized response model for simultaneous estimation of means of two sensitive variables. *Communication in Math and Statistics*, 8: 23-45.

Arias, R., Sedory, S.A. and Singh, S. (2021). Modified regression type estimator by ingeniously utilizing probabilities for more efficient results in randomized response sampling. Optimal decision making in operations research and statistics – methodologies and applications. 206-223. CRC Press, Boca, FL.

Arias, R., Sedory, S.A. and Singh, S. (2022). An unbiased regression type estimator in randomized response sampling. *Sankhya*, B, 84(1), 243-258.

Arnab, R. and Shangodoyin, D.K. (2020). Stratified randomized response model for multiple responses. J. Indian Soc. Agricultural Statist., 74, (1), 1-10.

Bansal, ML, Singh, S and Singh, S. (1994). Multi-character survey using randomized response technique. *Commun. Statist.- Theory Meth*. 23(6), 1705-1715

Batool, Fatima, Shabbir, Javid and Hussain, Zawar (2017). An improved binary randomized response model using six decks of cards. Communications in Statistics - Simulation and Computation, 46(4), 2548-2562.

Chaudhuri, A. and Stenger, H.S. (1992). *Survey Sampling*. CRC Press, Boca-Ratan.

Cochran, W.G. (1940). Some properties of estimators based on sampling scheme with varying probabilities. *Austral. J. Statist.*, 17, 22--28.

Chaudhuri, A. (2011). Randomized response and indirect questioning techniques in surveys. Statistics: Textbooks and Monographs. CRC Press, Boca Raton, FL

Chaudhuri, A. and Christofides, T.C. (2013). Indirect questioning in sample surveys. First Edition, Springer, Heidelberg

Chaudhuri, A., Christofides, T.C., and Rao, C.R. (2016). Data gathering, analysis and protection of privacy through randomized response techniques: qualitative and quantitative human traits. North Holland: Elsevier.

Chen, C.C. and Singh, S. (2011). Pseudo-Bayes and pseudo-empirical Bayes estimators in randomized response sampling. J. Statist Comp. and Simu., 81(6), 779-793.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc*., 87, 376-382.

Diana, G. and Perri, P.F. (2011). A class of estimators for quantitative sensitive data. *Statistical Papers,* 52, 633-650.

Eriksson, S.A. (1973). A new model for Randomized Response. *Int. Stat. Rev.*, 41(1), 101-113.
Franklin, L.A. (1989). A comparison of estimators for randomized response sampling with continuous distributions from a dichotomous population. *Communications in Statistics -Theory and Methods*, 18(2), 489-505.
Fox, J.A. (2016). *Randomized Response and Related Methods*, 2nd Ed.. SAGE, Los Angeles, CA.
Gjestvang, C.R and Singh, S. (2006). A new randomized response model. *Journal of the Royal Statistical Society*, B, 68,523-530.
Gjestvang, C. and Singh, S. (2007). Forced Quantitative Randomized Response Model: A new device. *Metrika*, 66, 2, 243-257.
Gjestvang, C. and Singh, S. (2009). An improved randomized response model: Estimation of mean. *Journal of Applied Statistics*,36(12),1361–1367.
Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods and Theory.* New York, John Wiley and Sons, 456--464.
Jayaraj, A., Sedory, S.A., Singh, S., and Odumade, O. (2018a). A new estimator of proportion with linear function using data from two-decks randomized response model. *Commu. in Statistics –Theory and Methods,* 47(6), 1475-1490.
Jayaraj A, Odumade O, Sedory S, Singh S (2018b) Weighted squared distance for two-deck randomized response model. *Behaviormetrika* 45(1). 91-109.
Jayaraj, A., Odumade, O. and Singh, S. (2018). A new quasi empirical Bayes estimate in randomized response sampling. *Comm. in Stat: Simul and Comp.*, 47 (7), 1879-1889.
Kuk, A.Y.C. (1990). Asking sensitive questions indirectly. *Biometrika*, 77(2), (1990), 436-438.
Lee, Cheon-Sig, Sedory, S.A. and Singh, S. (2013a). Simulated minimum sample sizes for various randomized response models. *Comm. in Stat.: Simu. and Comp.*, 42 (4), 771-789.
Lee, Cheon-Sig, Sedory, S.A. and Singh, S. (2013b). Estimating at least seven measures for\ qualitative variables using randomized response sampling. *Statistics and Probability Letters*, 83,399-409.
Lee, Cheon-Sig, Su, Ching-Shu, Mondragon, K., Salinas, V.I., Zamora, M.L, Sedory. S.A. and Singh, S. (2016). Comparison of Cramer-Rao lower bounds of variances for at least equal protection of respondents. *Statistica Neerlandica*, 70(2), 80-99.
Lee, Cheon-Sig, Sedory, S.A. and Singh, S. (2021). Estimation of odds ratio, attributable risk, correlation coefficient and other parameters using randomized response techniques. *Behaviormetrika*, **48**, 371–392.
Mangat, N.S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society*, Series B, 56(1), 93-95.
Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, 77(2), 439-442.
Murtaza, M., Singh, S. and Hussain, Z. (2020). An innovative optimal randomized response model using correlated scrambling variables. *Journal of Statistical Computation and Simulation*, 2823-2839.
Murtaza, M., Singh, S. and Hussain, Z. (2021). Use of correlated scrambling variables in quantitative randomized response technique. *Biom. J.*, 63(1), 134-147.
Naatjes, D., Sedory, S.A. and Singh, S. (2023a). New randomized response models for two sensitive characteristics: Theory and application. *Int. Stati. Review, (3), 511-534.*
Naatjes, D., Sedory, S.A. and Singh, S. (2024a). An unbiased regression type estimator of proportion in randomized response sampling by using analysis of variance mechanism. *Communications in Statistics: Theory and Methods*, 53(14), 5210-5217
Naatjes, D., Sedory, S.A. and Singh, S. (2024b). A class of infinite number of unbiased estimators using two-deck card method. *Journal of Applied Statistics* (Available online).
Odumade, O. and Singh, S. (2008). Generalized forced quantitative randomized response model: A unified approach. *Journal of the Indian Society of Agricultural Statistics*, 62(3), 244-252.
Odumade, O. and Singh, S. (2009a). Improved Bar-lev, Bobovitch and Boukai randomized response models. *Communication in Statistics-Simulation and Computation*, 38: 473–502.
Odumade, O. and Singh, S. (2009b). Efficient use of two decks of cards in randomized response sampling. *Communications in Statistics-Theory and Methods*, 38: 439–446.
Odumade, O and Singh, S. (2010). An alternative to the Bar-lev, Bobovitch and Boukai randomized response model. *Sociological Methods and Research*, 39: 206-221.

Olanipekun, O., Zhao, J. Wang, R., Sedory, S.A. and Singh, S. (2023). A theory of higher order interactions between sensitive variables: Empirical evidences and an application to a variety of smoking. *Sociological Methods and Research*, 52(2), 561–586.

Pushadapu, Kavya, Singh, S. and Sedory (2024). An optimized optional randomized response technique. International *Statistical Review* (Online available)

Pushadapu, Kavya and Singh, S. (2024). Chaudhuri and Mukerjee ORRT for two sensitive characteristics and their overlap. *Journal of Statistical Computation and Simulation, 94(5)*, 1056-1072.

Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

Sedory, S.A., Singh, S., Olanipekun, O.L., and Wark, C. (2020). Unrelated question model with two decks of cards. *Statistica Neerlandica.*74(2), 192-215.

Singh, R., Mangat, N.S. and Singh, S. (1993). A mail survey design for sensitive character without using randomization device. *Comm. Statist. Theory Methods*, 22 (9), 2661-2668.

Singh, S. (1994). Unrelated question randomized response sampling using continuous distributions. *J. Indian Soc. Agril. Stat.* 46(3), 349-361.

Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael Selected Amy*. Kluwer.

Singh, S. (2004). Golden and silver jubilee year-2003 of the linear regression estimators. ASA section on survey research methods. pp. 4382-4389.

Singh, S., Joarder, A. and King, M.L. (1996). Regression analysis using scrambled responses. *Australian Journal of Statistics*, 38(2), 201-211.

Singh, S., Singh, R., Mangat, N.S. and Tracy, D.S. (1994). An alternative device for randomized responses. *Statistica*, 54(2), 233-243.

Singh, S. and Sedory, S.A. (2017). TRUMP: Tuned Ratio Unbiased Mean Predictor. *Proceedings of the Joint Statistical Meeting-Survey Research Methods Section, Baltimore, Maryland*, 1746-1759.

Singh, S. and Sedory, S.A. (2017a). T.R.U.M.P: Tuned Ratio Unbiased Mean Predictor. *Working Monograph.*

Singh, S. and Sedory, S.A. (2019). TRUMP: Tuned Regression Unbiased Mean Predictor. *Proceedings of the Joint Statistical Meeting-Survey Research Methods Section, Denver, Colorado*, 2991-3008.

Singh, S., Sedory, S.A., Rueda, M.M, Arcos, A., and Arnab R. (2015). *A new concept for tuning design weights in survey sampling*. Elsevier.

Su, C.S., Sedory, S.A. and Singh, S. (2015). Kuk's model adjusted for protection and efficiency. *Sociological Methods and Research*, 44(3), 534-551.

Su, C.S., Sedory, S.A. and Singh, S. (2017). Adjusted Kuk's model using two non-sensitive characteristics unrelated to the sensitive characteristic. *Communications in Statistics -Theory and Methods*, 46(4), 2055-2075.

Su, C.S, Salinas, V.I., Zamora, M.L., Sedory, S.A. and Singh, S. (2021). Randomized response sampling with applications to tracking drugs for better life. *Statistica Sinica*, 31(4), 1871 -1890.

Tukey, J.W. (1958). Bias and confidence in not-quite large samples (abstract). *Ann. Math. Statist*., 29, 61-75.

Xu, T., Sedory, S.A.; Singh, S. (2021a). A new unrelated question model with two questions per card. *Optimal decision making in operations research and statistics-methodologies and applications*, 117–126. CRC Press, Boca Raton, FL

Xu, T., Sedory, S.A. and Singh, S. (2021b). Two sensitive characteristics and their overlap with two questions per card. *Biom. J.*, 63(8), 1688-1705.

Xu, T., Sedory, S.A. and Singh, S. (2022). Lowering the Cramer-Rao lower bounds of variance in randomized response sampling. *Commu.in Statist: Simul and Compu*, 51(7), 4112-4126.

Yennum, N., Sedory, S.A., and Singh, S. (2019). Improved strategy to collect sensitive data by using geometric distribution as a randomization device. *Comm. Statist. Theory Methods* 48 (23), 5777–5795.

Yennum, N., Sedory, S.A., and Singh, S. (2022). Improved strategy to collect sensitive data by using negative binomial and negative hypergeometric distribution as randomization devices. *Communications in Statistics: Theory and Methods,* 51(8), 2640-2658.

Zapata, Z., Sedory, S.A. and Singh, S. (2022). Zero truncated binomial distribution device. *Sociol. Methods Res*., 51, (2), 800-815.

Zapata, Z., Sedory, S.A. and Singh, S. (2023). An innovative improvement in Warner's randomized response device for evasive answer bias. *Journal of Statistical Computation and Simulation*, 93(2), 298-311.

Zheng, R., Sedory, S.A., and Singh, S. (2021a). Hybrid of crossed model and a new unrelated question model for two sensitive characteristics. *Optimal decision making in operations research and statistics—methodologies and applications*, 165–205. CRC Press, Boca Raton, FL.

Zheng, R., Sedory, S.A. and Singh, S. (2021b). Hybrid of simple model and a new unrelated question model for two sensitive characteristics. *Optimal decision making in operations research and statistics—methodologies and applications*, 127–164. CRC Press, Boca Raton, FL.