Improved Estimators of Mean of Sensitive Variables using Optional RRT Models

Sat Gupta
Department of Mathematics and Statistics
University of North Carolina – Greensboro
sngupta@uncg.edu

University of South Florida
March 9, 2016

Outline

- Randomized Response Models
- Why Randomize – SDB, Big Data Issues

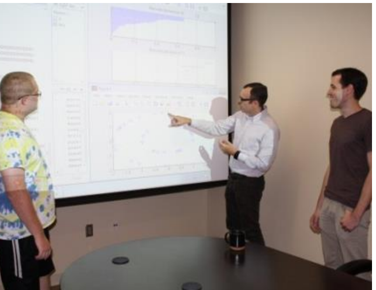
  Data Confidentiality & Respondent Privacy
- Some Applications of RRT Models
- Mean Estimation with RRT & Optional RRT Models
- Improved Ratio and Regression Estimation through Optional RRT Models
- Simulation results
- Concluding Remarks

Detour

Math & Stats at UNC Greensboro

THE UNIVERSITY *of* NORTH CAROLINA
GREENSBORO
Department *of*
Mathematics & Statistics

Department of Mathematics and Statistics



uncg.edu/mat

Computational Mathematics Ph.D. Program

Areas of research include:

- Combinatorics
- Differential Equations
- Functional Analysis
- Group Theory
- Statistics
- Mathematical Biology
- Number Theory
- Numerical Analysis
- Topology

MA in Mathematics

Concentrations in

- Mathematics
- Applied Statistics
- Actuarial Mathematics*
- College Teaching*
- Data Analytics*

* From Spring 2016 (anticipated)

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

Mathematics & Statistics Graduate Assistantships

Our Graduate Assistants usually receive:

$18,000 + tuition waivers for the Ph.D. Program in Computational Mathematics

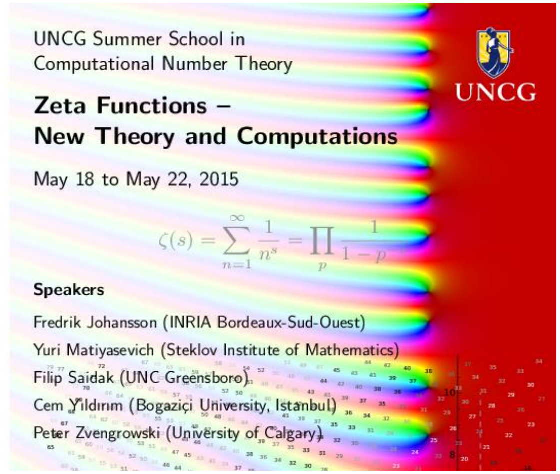
$10,800 + tuition waivers for the M.A. in Mathematics (Pure Mathematics, Applied Mathematics, Applied Statistics specialties)

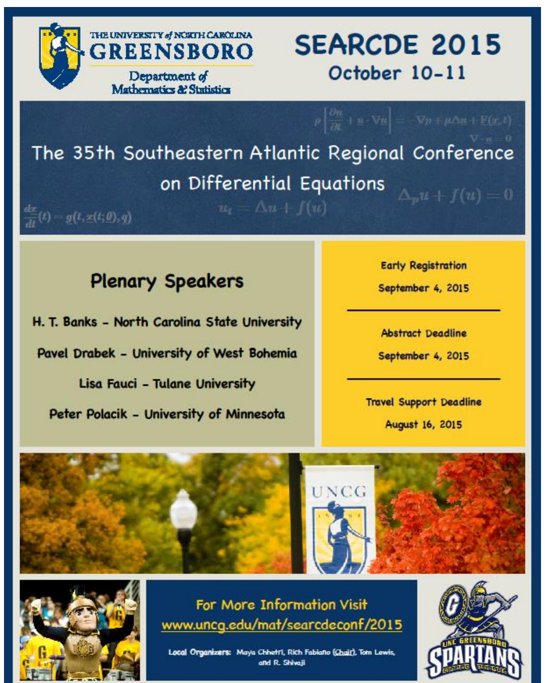
For more information, go to www.uncg.edu/mat or contact our Graduate Director, Dr. Gregory Bell at gcbell@uncg.edu.

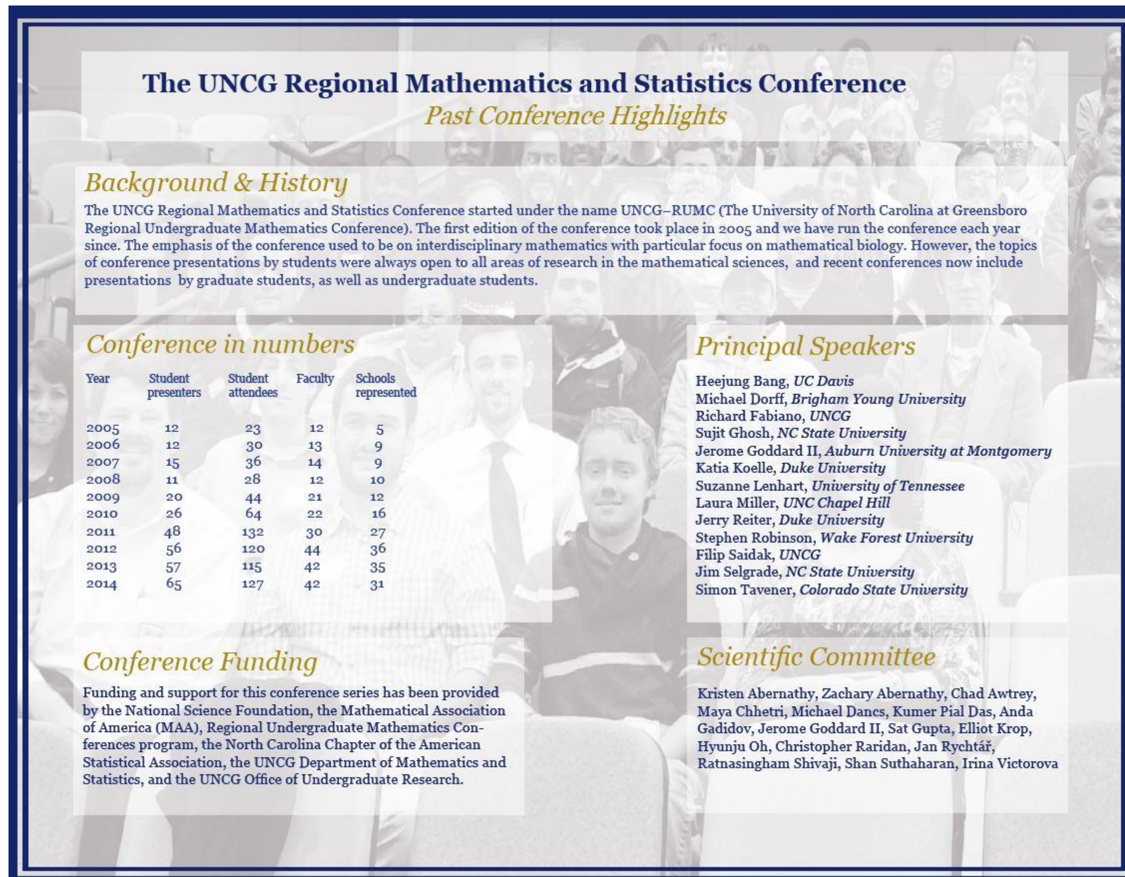
Our Graduate students are usually funded via graduate assistantships. Their duties include one or a combination of the following: teaching lower level Mathematics or Statistics courses, tutoring in the Math Help Center, or monitoring the Math Emporium Lab.

Additional summer funding is also available.

International Biennial Conference on Advances in Interdisciplinary Statistics and Combinatorics







Student Activities

Pi Mu Epsilon

Math Club

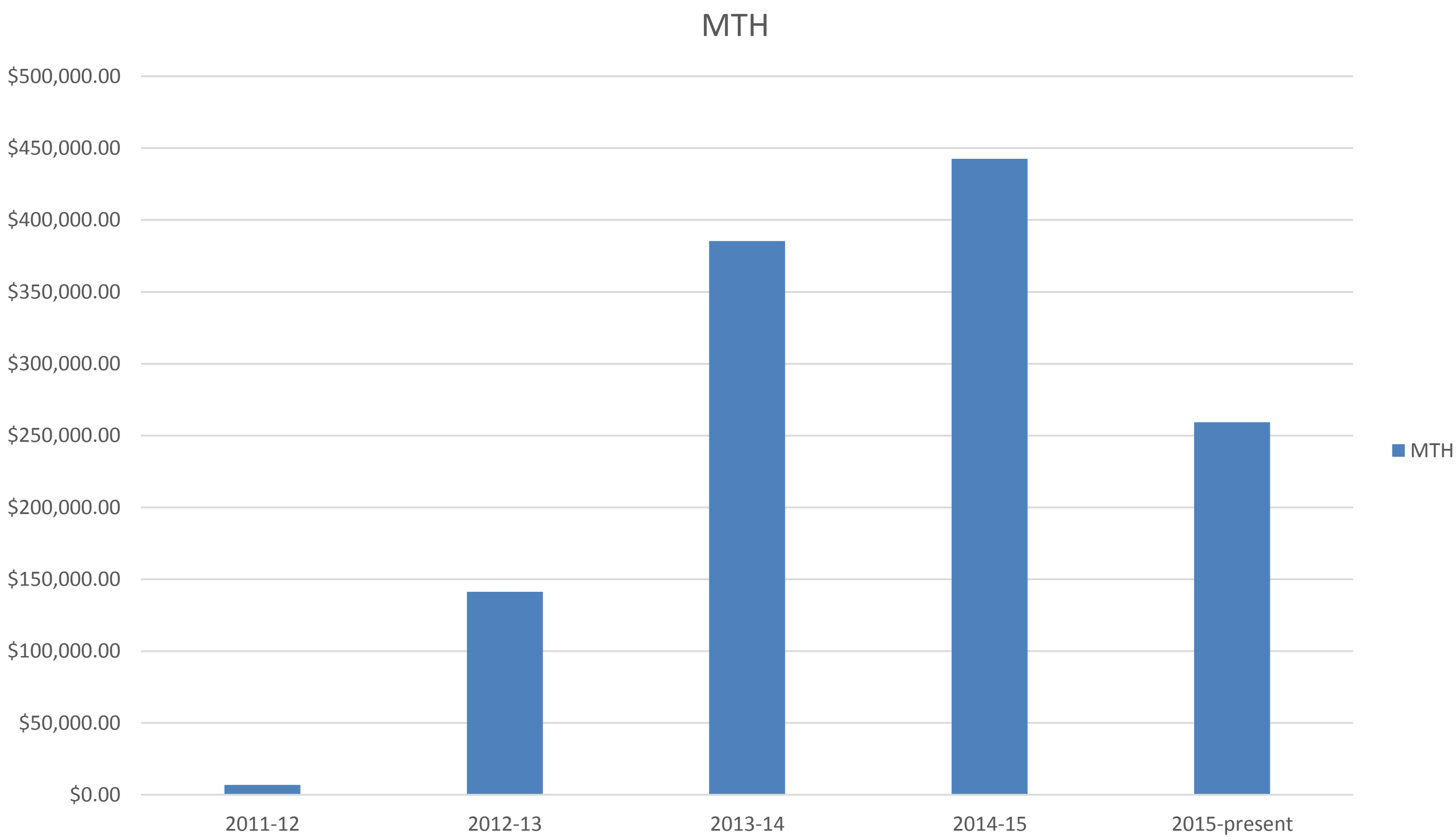
Graduate Tea

Student Publications

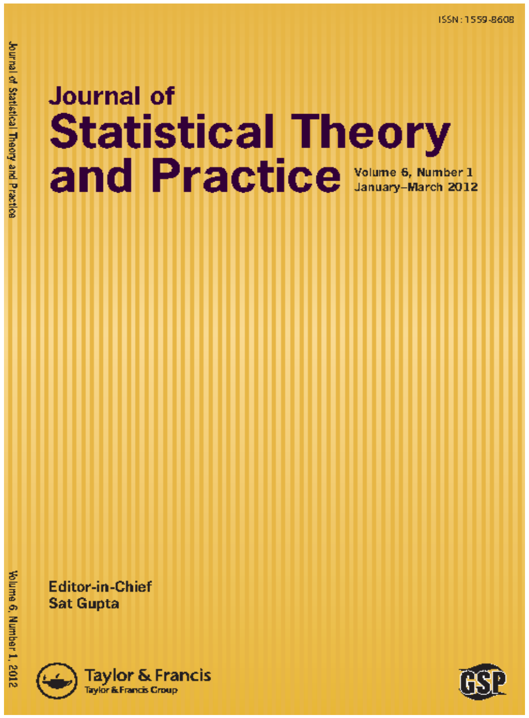
Some of the Journals that feature student publications.

Recent Highlights

Grant Awards by Academic Year

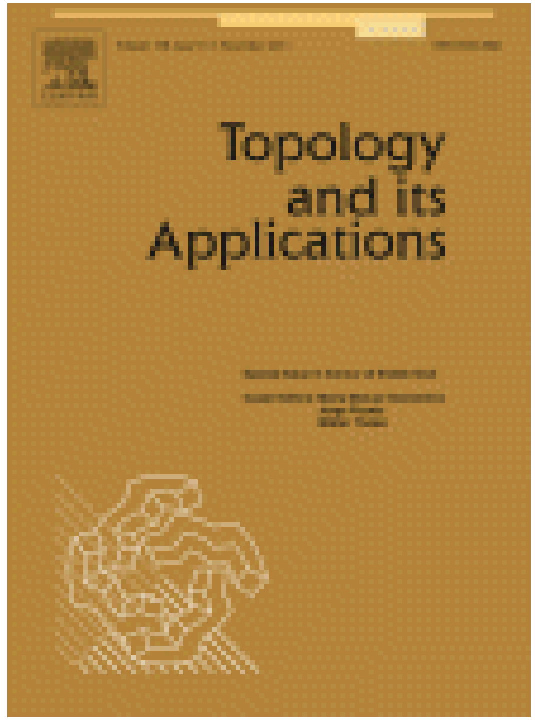


Journals Associated with the Department



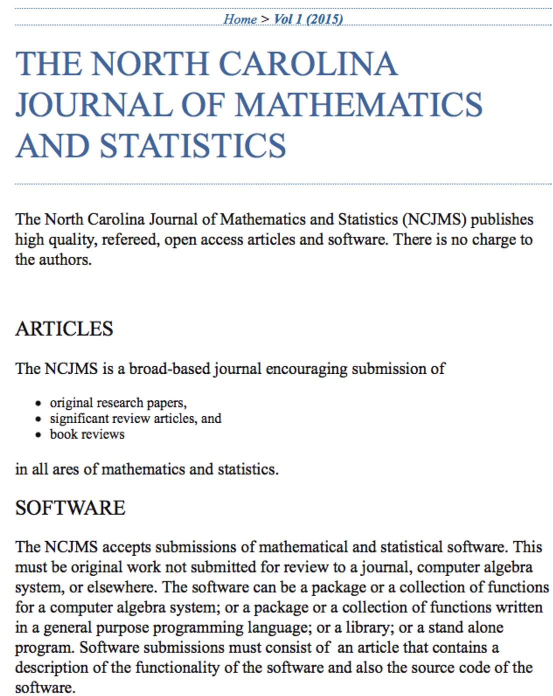
Dr. Sat Gupta serves as Editor-in-Chief

http://www.tandfonline.com/loi/UJSP20



Dr. Jerry Vaughan serves as one of two Editors-in-Chief

http://www.journals.elsevier.com/topology-and-its-applications/



Home > Vol 1 (2015)

THE NORTH CAROLINA JOURNAL OF MATHEMATICS AND STATISTICS

The North Carolina Journal of Mathematics and Statistics (NCJMS) publishes high quality, refereed, open access articles and software. There is no charge to the authors.

ARTICLES

The NCJMS is a broad-based journal encouraging submission of

- original research papers,
- significant review articles, and
- book reviews

in all ares of mathematics and statistics.

SOFTWARE

The NCJMS accepts submissions of mathematical and statistical software. This must be original work not submitted for review to a journal, computer algebra system, or elsewhere. The software can be a package or a collection of functions for a computer algebra system; or a package or a collection of functions written in a general purpose programming language; or a library; or a stand alone program. Software submissions must consist of an article that contains a description of the functionality of the software and also the source code of the software.

Dr. Jan Rychtar and Sebastian Pauli serve as Editors-in-Chief

http://ncjms.uncg.edu/

Back to Business

Randomized Response Techniques – RRT Models

- Introduced by Warner (1965) to decrease Social Desirability Bias.
- The respondents ‘randomize’ or ‘scramble’ the response to a sensitive or threatening question.
- Unscrambling can be done only at the group level, not at the individual level.
- Many Other Models – Greenberg Unrelated Question Model etc.

SDB – Social Desirability Response Bias

- Tendency in humans to look good in the face of incriminating questions
- Sensitivity may result in refusal to answer, or intentional false answers.

Getting Around SDB

- SDB Scale
- Bogus Pipeline
- RRT Models – But Privacy Protection is an Important Issue

Types of RRT Models

- Binary vs. Multi-category vs. **Quantitative**
- Full RRT vs. Partial RRT Vs. **Optional RRT**
- **Additive** vs. Multiplicative vs. Generalized

Big Data

- Number of records could be very large
  - *Large n (too many rows in the data)*
  - *Social Media Data*

- Dimension may be very high
  - *Large p (too many columns in the data)*
  - *Public Health Data*

- Big data creates additional challenges in both situations when data need to be released publicly

Why Release Data Publicly

- Advancement of science
- Student training
- Public interest
- Funding agencies may insist

Data Confidentiality – Back-End Problem

- Maintain confidentiality of record level data. Less worry at aggregate level
- Ethical/Legal Issues
- It is not enough to delete names/ subject ID's
- Respondent safety and protection

Respondent Privacy – Front-End Problem

- SDB Related Issues

- Respondent Cooperation

Data Confidentiality & Respondent Privacy

- Too much scrambling (masking) or too little scrambling

- Think of two data scrambling models for variable X

$$Y = X + S$$

$$Y = X + \theta S$$

S is a scrambling variable, θ *is a constant*

- Confidentiality is higher when θ is larger

- Data quality is better when θ is smaller

- Same dilemma as in confidence intervals ☺

Some RRT Applications

Ostapczuk, Martin, Jochen Musch, and Morten Moshagen (2009):

A randomized-response investigation of the education effect in attitudes towards foreigners, *European Journal of Social Psychology,* 39 (6)

Spears- Gill, Tracy., Tuck, Anna., Gupta, Sat., Crowe, Mary., Jennifer Figuerova (2013):

A Field Test of Optional Unrelated Question Randomized Response Models – Estimates of Risky Sexual Behaviors, *Springer Proceedings in Mathematics and Statistics*, Vol. 64, 135-146

Education Effect in Attitudes Towards Foreigners in Germany

- Under direct questioning conditions, 75% of the highly educated expressed xenophile attitudes, as opposed to only 55% of the less educated.

- Under randomized-response conditions, 53% xenophiles among the highly educated, and 24% among the less educated

Spears-Gill et al. (2013) - Field Test: Estimates of Risky Sexual Behaviors

Use of Greenberg Unrelated Question RRT Model

Sensitive question

Have you ever been told by a healthcare professional that you have a sexually transmitted disease(STD)?

Unrelated question

Were you born between January 1st and October 31st?

Estimate of STD Prevalence

Method	$\hat{\pi}_X$	95% CI
Optional RRT	0.0367	（0.0159, 0.0576）
Check Box Method	0.0900	（0.0438， 0.1362）
Face-to-face Interview	0.0200	(-0.0042， 0.0442)

Mean Estimators of Sensitive Variables

Eichhorn and Hayre (1983): Multiplicative Model *JSPI*

- *Y*: Sensitive quantitative variable of interest with unknown mean μ_Y and an unknown variance of σ_Y^2.
- *S*: Scrambling variable independent of *Y* with known mean of $\mu_S (= \theta)$ and a known variance of σ_S^2.

The reported response *Z* is given by

$$Z = \frac{YS}{\theta}$$

This suggests estimating μ_Y by $\hat{\mu}_Y$ where

$$\hat{\mu}_Y = \overline{Z}$$

Gupta et al. (2002): Optional RRT Model *JSPI*

- Multiplicative optional RRT is used to scramble the response :
  - The respondents provides a multiplicatively scrambled response for *Y* if they consider the question sensitive, and a true response otherwise.
  - The response is given by:

$$Z = S^T Y$$

where *T* is a Bernoulli random variable with parameter W and *S* is a scrambling variable with unit mean and known variance, independent of Y

Mean Estimation

- An unbiased estimator of the population mean μ_Y is given by

$$\hat{\mu}_Y = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

- Note that *W* is not involved in the unbiased estimation of μ_Y.

- The relative efficiency of Gupta et al. (2002) estimator with respect to the estimator of Eichhorn and Hayre (1983) is greater than or equal to 1.

Sensitivity Estimation

- Taking Log and then expected values on both sides of

$$Z = S^T Y$$

leads to an estimator of *W* given by

$$\hat{W} = \frac{\frac{1}{n}\sum_{i=1}^{n}\log(Z_i) - \log\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right)}{E[\log(S)]}$$

- **Asymptotics a challenge with multiplicative scrambling**
- **Split sample approach is an option**
- **Loss of anonymity**
- **Additive scrambling works better**

Additive Optional RRT Model

- The respondent is asked to provide an additively scrambled response for Y if they consider the question sensitive and a true response otherwise. Model is given by:

$$Z = Y + ST$$

where

- *T* is a Bernoulli random variable with parameter *W*
- *S* is a scrambling variable with **zero mean** and known variance independent of Y
- One equation, two unknowns

Estimation of the Mean

- An unbiased estimator of population mean is the sample mean of the reported responses

$$\hat{\mu}_{YW} = \frac{1}{n}\sum_{i=1}^{n} z_i$$

- Note that W is not involved in the unbiased estimation of μ_{YW}

Additive Optional RRT Model Using Split Sample Approach

- Total sample size n is split into two independent sub-samples of sizes n_1 and n_2

- The mean and variance respectively for Y are μ_Y and σ_Y^2.

- The mean and variance respectively for $S_i(i=1,2)$ are θ_i and $\sigma_{S_i}^2$.

- We assume that Y , and $S_i(i=1,2)$ are mutually independent.

Estimation of Mean and Sensitivity Level

- The reported response Z_i in the i^{th} sub- sample is given by

$$Z_i = \begin{cases} Y & with\ probability\ (1-W) \\ (Y+S_i) & with\ probability\ \ W \end{cases} \quad i=1,2$$

We note $E(Z_i) = \mu_Y + \theta_i W$ where $E(S_i) = \theta_i\ (i=1,2)$.
It follows

$$\mu_Y = \frac{\theta_2 E(Z_1) - \theta_1 E(Z_2)}{\theta_2 - \theta_1}, \quad \text{and} \quad W = \frac{E(Z_2) - E(Z_1)}{(\theta_2 - \theta_1)}$$

Gupta et al. (2010): Unbiased Estimators of Mean and Sensitivity Level - *JSPI*

- $\hat{\mu}_Y = \dfrac{\theta_2\bar{z}_1 - \theta_1\bar{z}_2}{\theta_2 - \theta_1}$, $\hat{W} = \dfrac{\bar{z}_2 - \bar{z}_1}{(\theta_2 - \theta_1)}$, $\theta_1 \neq \theta_2$

where $\bar{z}_1, \bar{z}_2$ respectively are the sample mean of reported responses in the two sub-samples.

- The mean square error of $\hat{\mu}_Y$ is given by

$$MSE(\hat{\mu}_Y) = \frac{1}{(\theta_2 - \theta_1)^2}\left[\theta_2^2\left(\frac{1-f_1}{n_1}\right)\sigma_{Z_1}^2 + \theta_1^2\left(\frac{1-f_2}{n_2}\right)\sigma_{Z_2}^2\right] \qquad \theta_1 \neq \theta_2$$

where

$$f_1 = \frac{n_1}{N} \quad f_2 = \frac{n_2}{N} \qquad f = \frac{n}{N} = f_1 + f_2 \qquad \sigma_{Z_2}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Z_{2_i} - \mu_Z\right)^2$$

Improvement through
Ratio and Regression Estimation

Mean Estimation with Auxiliary Information Using Non-Optional RRT Models

- Primary variable of interest *Y* is sensitive.
- Direct observation on this variable may not be possible.
- We may directly observe a highly correlated auxiliary variable *X.*
- Usual RRT mean estimators for *Y* can be improved considerably by utilizing information from the auxiliary variable X.

Sampling Framework

- Consider a finite population with N units: $U = \{U_1, U_2, U_3, ...U_N\}$

  - A sample of size n is drawn using simple random sampling without replacement (*SRSWOR*).

  - *Y* is a study variable, a sensitive variable which cannot be observed directly.

  - *X* is a non-sensitive auxiliary variable which is strongly correlated with Y .

Sousa et al. (2010): Improved Mean Estimators with Auxiliary Information - *JSTP*

Ratio Estimation Using Non-optional RRT Model

Sousa et al. (2010) proposed a **non-optional** ratio estimator for the mean of sensitive variable *Y* utilizing information from a non- sensitive auxiliary variable *X*. Their estimator is given by

$$\hat{\mu}_{AR} = \bar{z}\left(\frac{\mu_X}{\bar{x}}\right)$$

In the above expression $\bar{z}$ is the sample mean of reported responses obtained from a non- optional additive RRT model.

Gupta et al. (2012): Regression Estimator Using Non-Optional RRT Model - *CIS-TM*

- Gupta et al. (2012) proposed a **non-optional** regression estimator for the mean of sensitive variable (Y) utilizing information from a non-sensitive auxiliary variable (X). Their estimator is given by

$$\hat{\mu}_{\mathrm{Re}\,g} = \bar{z} + \hat{\beta}_{zx}\left(\bar{X} - \bar{x}\right)$$

- $\bar{z}$ is the sample mean of reported responses obtained from a non optional additive RRT model

- $\hat{\beta}_{zx}$ is the sample regression coefficient between Z and *X.*

Ratio Estimation Using Optional RRT Model

We propose the following ratio estimator for the population mean of the sensitive study variable Y using the auxiliary variable X :

$$\hat{\mu}_{AR} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1}\right)\left(\frac{\mu_X}{\bar{x}_1} + \frac{\mu_X}{\bar{x}_2}\right)\left(\frac{1}{2}\right)$$

MSE of $\hat{\mu}_{AR}$ up to first order of approximation is given by

$$MSE^{(1)}(\hat{\mu}_{AR}) \cong \left(\frac{1-f_1}{n_1}\right)\left[\left(\frac{\theta_2}{\theta_2-\theta_1}\right)^2 \sigma_{Z_1}^2 + \frac{\mu_Y^2}{4} C_x^2 - \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_2}{\theta_2-\theta_1}\right) C_x\right] +$$
$$\left(\frac{1-f_2}{n_2}\right)\left[\left(\frac{\theta_1}{\theta_2-\theta_1}\right)^2 \sigma_{Z_2}^2 + \frac{\mu_Y^2}{4} C_x^2 + \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_1}{\theta_2-\theta_1}\right) C_x\right]$$

Efficiency Comparison

- $MSE\left(\hat{\mu}_{AR}\right) < MSE\left(\hat{\mu}_{Y}\right)$ if $\rho_{yx} > \dfrac{\alpha}{4\beta}$ with $\left(C_x = C_y\right)$

  where

$$\alpha = \left(\frac{1-f_1}{n_1}\right) + \left(\frac{1-f_2}{n_2}\right)$$

$$\beta = \left(\frac{1-f_1}{n_1}\right)\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - \left(\frac{1-f_2}{n_2}\right)\left(\frac{\theta_1}{\theta_2-\theta_1}\right)$$

- **Equal sub-sample sizes:** $n_1 = n_2 = n/2$ **,** $\left(\frac{1-f_1}{n_1}\right) = \left(\frac{1-f_2}{n_2}\right)$

  and hence $\frac{\alpha}{4\beta} = \frac{1}{2}$.

  In this case $MSE\left(\hat{\mu}_{AR}\right) < MSE\left(\hat{\mu}_Y\right)$ if $\rho_{yx} > \frac{1}{2}$

- **Unequal sub-sample sizes:** $n_1 \neq n_2$

  We can choose scrambling variables and sample sizes in such a way that $\frac{\alpha}{4\beta} < \frac{1}{2}$ and hence again

  $MSE\left(\hat{\mu}_{AR}\right) < MSE\left(\hat{\mu}_Y\right)$ if $\rho_{yx} > \frac{1}{2}$

- Note that $\frac{\alpha}{4\beta} < \frac{1}{2}$ under the following parameter choices which are always possible :

- If both the scrambling variable means are strictly positive, then we associate the smaller mean with the smaller sub-sample

- If both the scrambling variable means are strictly negative, then we associate the smaller mean with the larger sub-sample.

- If the scrambling variable means are with opposite signs then we associate the one with the larger absolute value to the larger sub-sample.

- If one of the scrambling variable means is zero then we associate the smaller sub-sample size to the variable with mean zero.

- The ratio estimator $\hat{\mu}_{AR}$ is always more efficient than the ordinary additive optional mean estimator $\hat{\mu}_Y$ if

.

$$\left(C_x = C_y\right) \quad \text{and} \quad \rho_{yx} > 0.5$$

- We see $MSE\left(\hat{\mu}_{AR}\right) = MSE\left(\hat{\mu}_Y\right)$ if $\rho_{yx} = 0.5$.

Simulations

- We show the above conclusion with the following bivariate normal population:

$$N = 5000,\ \mu_X = 4,\ \mu_Y = 6,\ \sigma_X = 2,\ \sigma_Y = 3,\ \sigma_{S_1} = 2,\ \sigma_{S_2} = 1$$

$$\theta_2 = 5 > 0.5 = \theta_1,\ n = 500$$

1000 iterations.

Table 1: Estimates with Theoretical (bold) and Empirical MSE's ($\mu_Y = 6, n = 500$)

$n_2 = 300, n_1 = 200$

	$\rho_{yx} = 0.8$					$\rho_{YX} = 0.3$				
W	$\hat{W}$	$\hat{\mu}_Y$	$\hat{\mu}_{AR}$	$MSE(\hat{\mu}_Y)$	$MSE(\hat{\mu}_{AR})$	$\hat{W}$	$\hat{\mu}_Y$	$\hat{\mu}_{AR}$	$MSE(\hat{\mu}_Y)$	$MSE(\hat{\mu}_{AR})$
0.3	0.35	5.8640	5.8688	**0.0582**	**0.0400**	0.35	5.8608	5.8695	**0.0584**	**0.0623**
				0.0580	0.0388				0.0583	0.0585
0.5	0.55	5.8003	5.7979	**0.0606**	**0.0425**	0.55	5.790	5.7963	**0.0609**	**0.0648**
				0.0650	0.0474				0.0694	0.0716
0.7	0.65	5.8923	5.8981	**0.0629**	**0.0448**	0.66	5.8983	5.901	**0.0632**	**0.0671**
				0.0525	0.0366				0.0551	0.0603
0.8	0.81	5.8332	5.8340	**0.0640**	**0.0458**	0.80	5.8437	5.8502	**0.0643**	**0.0681**
				0.0616	0.0435				0.0591	0.0617
0.9	0.92	5.8211	5.8264	**0.0650**	**0.0468**	0.92	5.8451	5.8461	**0.0653**	**0.0691**
				0.0618	0.0445				0.0644	0.0660

Regression Estimation Using Optional RRT Model

- We propose the following regression estimator which modifies the ordinary optional mean estimator using split-sample approach

$$\hat{\mu}_{Areg} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1} \right) + \left(\hat{\beta}_{z_1 x_1} \left(\mu_X - \bar{x}_1 \right) + \hat{\beta}_{z_2 x_2} \left(\mu_X - \bar{x}_2 \right) \right) \left(\frac{1}{2} \right)$$

where $\beta_{z_i x_i} \left(i = 1,2 \right)$ are the sample regression coefficients between

z_i and x_i respectively, and $\bar{z}_i, \bar{x}_i \left(i = 1,2 \right)$ are the two sub-sample

means.

The mean square error, up to first order of approximation, is given by

$$MSE^{(1)}\left(\hat{\mu}_{Areg}\right)=\frac{1}{\left(\theta_2-\theta_1\right)^2}\left[\theta_2^2\left(\frac{1-f_1}{n_1}\right)\sigma_{Z_1}^2+\theta_1^2\left(\frac{1-f_2}{n_2}\right)\sigma_{Z_2}^2\right]+\frac{\rho_{YX}^2\sigma_Y^2}{4}\alpha-\rho_{YX}^2\sigma_Y^2\beta$$

where

$$\theta_2\neq\theta_1$$

$$\alpha=\left(\frac{1-f_1}{n_1}\right)+\left(\frac{1-f_2}{n_2}\right)$$

$$\beta=\left(\frac{1-f_1}{n_1}\right)\left(\frac{\theta_2}{\theta_2-\theta_1}\right)-\left(\frac{1-f_2}{n_2}\right)\left(\frac{\theta_1}{\theta_2-\theta_1}\right)$$

Efficiency Comparison

We note

- $MSE\left(\hat{\mu}_{Areg}\right) < MSE\left(\hat{\mu}_{Y}\right)$ if $\dfrac{\alpha}{4\beta} < 1$

- $MSE\left(\hat{\mu}_{Areg}\right) < MSE\left(\hat{\mu}_{AR}\right)$ if $\rho_{YX} < \dfrac{\alpha}{4\beta - \alpha}$ $\left(C_y = C_x\right)$

The above conditions can be achieved with proper choices

of sub-sample sizes and scrambling variables.

Equal Split($n_1 = n_2 = n/2$)

- We note that in this case

  $\frac{\alpha}{4\beta} = \frac{1}{2} < 1$ is always true.

  Hence $\hat{\mu}_{Areg}$ is always efficient than $\hat{\mu}_{Y}$.

- Also

  $\frac{\alpha}{4\beta - \alpha} = 1$

  Hence $MSE\left(\hat{\mu}_{Areg}\right) < MSE\left(\hat{\mu}_{AR}\right)$ if $\rho_{yx} < 1$.

  Hence $\hat{\mu}_{Areg}$ is always more efficient than $\hat{\mu}_{AR}$.

Simulations

- Consider a bivariate normal population with the following characteristics:

$$N = 5000,\ \mu_X = 4,\ \mu_Y = 6,\ \sigma_X = 2,\ \sigma_Y = 3,\ \sigma_{S_1} = 2,\ \sigma_{S_2} = 1.$$

$$\theta_1 = 5, \theta_2 = -0.5$$

1000 iterations

Table 2: Estimate with Theoretical(bold) and Empirical MSE's ($n=500,\ \rho_{yx}=0.8,\ \mu_Y=6$)

n_1	n_2	W	$\hat{W}$	$\hat{\mu}_Y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{Areg}$	$MSE(\hat{\mu}_Y)$	$MSE(\hat{\mu}_{AR})$	$MSE(\hat{\mu}_{Areg})$
200	300	0.3	0.35	5.9061	5.9084	5.9046	**0.0250**	**0.0190**	**0.0174**
							0.0231	0.0178	0.0163
200	300	0.5	0.53	5.8576	5.8590	5.8558	**0.0256**	**0.0196**	**0.0180**
							0.0257	0.0196	0.0189
200	300	0.7	0.65	5.8638	5.8679	5.8642	**0.0261**	**0.0201**	**0.0185**
							0.0253	0.0205	0.0189
200	300	0.8	0.77	5.8479	5.8549	5.8505	**0.0262**	**0.0203**	**0.0187**
							0.0277	0.0212	0.0201
200	300	0.9	0.90	5.8418	5.8435	5.8400	**0.0264**	**0.0204**	**0.0188**
							0.0292	0.0234	0.0223

Table 3: Estimate with theoretical(bold) and Empirical MSE's($n = 500,\ \rho_{yx} = 0.3,\ \mu_Y = 6$)

n_1	n_2	W	$\hat{W}$	$\hat{\mu}_Y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{Areg}$	$MSE(\hat{\mu}_Y)$	$MSE(\hat{\mu}_{AR})$	$MSE(\hat{\mu}_{Areg})$
200	300	0.3	0.3454	5.9164	5.9178	5.9155	**0.0251**	**0.0336**	**0.0240**
							0.0230	0.0333	0.0227
200	300	0.5	0.5438	5.8582	5.8618	5.8580	**0.0257**	**0.0342**	**0.0246**
							0.0276	0.0366	0.0272
200	300	0.7	0.6519	5.8746	5.8876	5.8773	**0.0262**	**0.0347**	**0.0251**
							0.0251	0.0333	0.0242
200	300	0.8	0.7718	5.8611	5.8605	5.8597	**0.0264**	**0.0349**	**0.0252**
							0.0268	0.0349	0.0261
200	300	0.9	0.9024	5.8455	5.8455	5.8442	**0.0265**	**0.0350**	**0.0254**
							0.0291	0.0365	0.0282

Conclusions

- Even for small correlation between the study variable and the auxiliary variable, the proposed regression estimator is always more efficient than both the ratio estimator and the ordinary RRT mean estimator.

- As seen in Table 2 , for $\rho_{yx} < 0.5$ the optional RRT mean estimator is more efficient than the ratio estimator. However, the proposed regression estimator is always more efficient than both the additive ratio estimator and the ordinary optional RRT mean estimator.

Conclusions

- As the sensitivity W increases, the $MSE^{'}s$ increase, highlighting the usefulness of an optional RRT model since W is highest (equal to 1) for non-optional model.

- Similar improvements possible In stratified sampling also

THANK YOU