ESTIMATION OF ODDS RATIO, ATTRIBUTABLE RISK, RELATIVE RISK, CORRELATION COEFFICIENT AND OTHER PARAMETERS USING RANDOMIZED RESPONSE TECHNIQUES

Cheon-Sig Lee[1], Stephen A. Sedory[2], Sarjinder Singh[2]
[1] Department of Mathematics, San Jacinto College - North Campus 5800 Uvalde Rd. Houston, TX 77049.
[2] Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363.
Corresponding E-mail: sarjinder@yahoo.com

ABSTRACT

In this paper, we first define odds ratio, attributable risk, relative risk, correlation coefficient, membership in at least one group, difference between two proportions, a new estimator of single proportion when the proportion of second sensitive variable is know, while considering investigating two sensitive attributes in real practice. Then we define two estimators in each one of these cases based on simple model or crossed model proposed by Lee, Sedory and Singh (2013). We derive expressions for biases and variances of the resultant estimators. We investigate the performance of estimators based on the crossed model over those based on the simple model under the same choice of parameters, as discussed in Lee *et al.* (2013). Also the values of the various statistics such as odds ratio, attributable risk, relative risk, correlation coefficient, membership in one variable and the difference between two proportions are estimated based on a real data set.

Keywords: Sensitive characteristics, estimation of proportion, crossed model, simple model.

1. INTRODUCTION

In 1965, S. L. Warner proposed the first research method in structured survey interview. Lee, Sedory and Singh (2013) introduced a new methodology for estimating the proportions of persons in a population possessing each of two sensitive characteristics, say *A* and *B*, along with the proportion possessing both, $A \cap B$, by using two different randomized response models: Simple model and Crossed model. There are many situations where their proposed models could be implemented in real practice. For example, (1) assume *A* is a group of smokers, *B* is a group of drinkers, then $A \cap B$ will be a group of both smokers and drinkers; (2) assume *A* is a group of SMAC users, *B* is a group of criminals, then $A \cap B$ will be a group of both smack users and criminally active people; and (3) assume *A* represents hidden membership in a terrorist group-I, *B* represents a hidden membership in a terrorist group-II, then $A \cap B$ will be a hidden membership in both terrorist groups. Their models also allow one to estimate several other parameters, such as correlation coefficient, conditional proportions, and relative risk, etc. A pictorial representation of such a population is shown in Figure 1.1. Let π_A, π_B and $\pi_{A\cap B}$ be the true proportions of respondents possessing the sensitive characteristics *A*, *B*, and both $A \cap B$. Also note that $\pi_{A\cap B} \le Min(\pi_A, \pi_B)$.

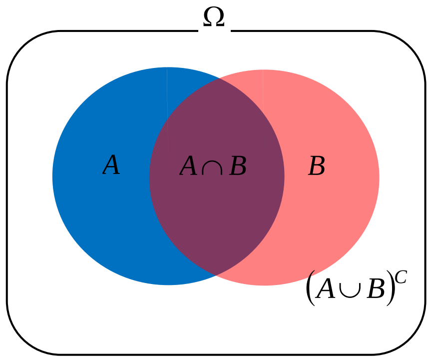

Fig.1.1. Population under study

In the following sub-sections we briefly review the two models introduced by the authors in the paper mentioned above.

1.1 Simple Model

In the simple model proposed by Lee, Sedory and Singh (2013), they suggest to using a pair of decks of cards in order: say Deck-I and Deck-II. Each respondent, selected in a simple random with replacement sample of size n, is requested to draw two cards, one from each deck and keep the responses from Deck-I and Deck-II respectively in order. Deck-I consists of cards, each bearing one of two mutually exclusive statements: "I belong to the sensitive group A", with proportion P, and "I belong to the non-sensitive group A^c", with proportion $(1-P)$. Deck-II also consists of cards, each bearing one of two mutually exclusive statements: " I belong to the sensitive group B", with proportion T, and "I belong to the non-sensitive group B^c", with proportion $(1-T)$. By following the notation of Lee, Sedory and Singh (2013) for the simple model, the probabilities of obtaining, from a given respondent, each of the following responses (*Yes*, *Yes*), (*Yes*, *No*), (*No*, *Yes*) and (*No*, *No*) are, respectively, given by:

$$\theta_{11} = (2P-1)(2T-1)\pi_{AB} + (2P-1)(1-T)\pi_A + (1-P)(2T-1)\pi_B + (1-P)(1-T), \quad (1.1)$$

$$\theta_{10} = -(2P-1)(2T-1)\pi_{AB} + (2P-1)T\pi_A - (1-P)(2T-1)\pi_B + (1-P)T, \quad (1.2)$$

$$\theta_{01} = -(2P-1)(2T-1)\pi_{AB} - (2P-1)(1-T)\pi_A + P(2T-1)\pi_B + P(1-T), \quad (1.3)$$

and

$$\theta_{00} = (2P-1)(2T-1)\pi_{AB} - T(2P-1)\pi_A - P(2T-1)\pi_B + PT. \quad (1.4)$$

Let $\hat{\theta}_{11} = n_{11}/n$, $\hat{\theta}_{10} = n_{10}/n$, $\hat{\theta}_{01} = n_{01}/n$ and $\hat{\theta}_{00} = n_{00}/n$, be the observed proportions of (*Yes*, *Yes*), (*Yes*, *No*), (*No*, *Yes*) and (*No*, *No*) responses, so that $n_{11} + n_{10} + n_{01} + n_{00} = n$. Then Lee, Sedory and Singh (2013) obtained unbiased estimators for the simple model as following:

$$\hat{\pi}_A = \frac{\hat{\theta}_{11} + \hat{\theta}_{10} - \hat{\theta}_{01} - \hat{\theta}_{00} + (2P-1)}{2(2P-1)}, \quad (1.5)$$

$$\hat{\pi}_B = \frac{\hat{\theta}_{11} - \hat{\theta}_{10} + \hat{\theta}_{01} - \hat{\theta}_{00} + (2T-1)}{2(2T-1)}, \qquad (1.6)$$

and

$$\hat{\pi}_{AB} = \frac{(P+T)\hat{\theta}_{11} + (T-P)\hat{\theta}_{10} + (P-T)\hat{\theta}_{01} + (2-P-T)\hat{\theta}_{00} - T(1-P) - P(1-T)}{2(2P-1)(2T-1)}, \quad (1.7)$$

for $P \neq 0.5$ and $T \neq 0.5$.

Let us define

$$\varepsilon_{AB} = \frac{\hat{\pi}_{AB}}{\pi_{AB}} - 1, \quad \varepsilon_A = \frac{\hat{\pi}_A}{\pi_A} - 1, \quad \varepsilon_B = \frac{\hat{\pi}_B}{\pi_B} - 1$$

so that

$$E(\varepsilon_{AB}) = E(\varepsilon_A) = E(\varepsilon_B) = 0$$

$$E(\varepsilon_{AB}^2) = \frac{V(\hat{\pi}_{AB})}{\pi_{AB}^2}, \quad E(\varepsilon_A^2) = \frac{V(\hat{\pi}_A)}{\pi_A^2}, \quad E(\varepsilon_B^2) = \frac{V(\hat{\pi}_B)}{\pi_B^2}$$

$$E(\varepsilon_{AB}\varepsilon_A) = \frac{Cov(\hat{\pi}_{AB}, \hat{\pi}_A)}{\pi_{AB}\pi_A}, \quad E(\varepsilon_{AB}\varepsilon_B) = \frac{Cov(\hat{\pi}_{AB}, \hat{\pi}_B)}{\pi_{AB}\pi_B}, \quad E(\varepsilon_A\varepsilon_B) = \frac{Cov(\hat{\pi}_A, \hat{\pi}_B)}{\pi_A\pi_B}$$

where

$$V(\hat{\pi}_A) = \frac{\pi_A(1-\pi_A)}{n} + \frac{P(1-P)}{n(2P-1)^2}, \qquad (1.8)$$

$$V(\hat{\pi}_B) = \frac{\pi_B(1-\pi_B)}{n} + \frac{T(1-T)}{n(2T-1)^2}, \qquad (1.9)$$

$$V(\hat{\pi}_{AB}) = \frac{\pi_{AB}(1-\pi_{AB})}{n} + \frac{(2P-1)^2 T(1-T)\pi_A + P(1-P)(2T-1)^2\pi_B + PT(1-P)(1-T)}{n(2P-1)^2(2T-1)^2}, \qquad (1.10)$$

$$Cov(\hat{\pi}_A, \hat{\pi}_B) = \frac{\pi_{AB} - \pi_A\pi_B}{n}, \qquad (1.11)$$

$$Cov(\hat{\pi}_{AB}, \hat{\pi}_A) = \frac{\pi_{AB}(1-\pi_A)}{n} + \frac{P(1-P)\pi_B}{n(2P-1)^2}, \qquad (1.12)$$

and

$$Cov(\hat{\pi}_{AB}, \hat{\pi}_B) = \frac{\pi_{AB}(1-\pi_B)}{n} + \frac{T(1-T)\pi_A}{n(2T-1)^2}. \qquad (1.13)$$

1.2 Crossed Model

In the crossed model, while the rest of the procedure remains the same as for the simple model, the composition of the decks is different. Deck-I consists of cards, each bearing one of two mutually exclusive statements: "I belong to the sensitive group A", with probability P and "I belong to the non-sensitive group B^c", with probability $(1-P)$ respectively. Deck-II also consists of cards, each bearing one of two mutually exclusive statements: "I belong to the sensitive group B" with probability T and "I belong to the non-sensitive group A^c "with probability $(1-T)$ respectively. By following the notation of Lee, Sedory and Singh (2013) for the crossed model, the probabilities of obtaining, from a given respondent, each of the following responses, (*Yes*, *Yes*), (*Yes*, *No*), (*No*, *Yes*) and (*No*, *No*) are, respectively, given by:

$$\theta_{11}^* = \pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A(1-P)(1-T) - \pi_B(1-P)(1-T) + (1-P)(1-T), \quad (1.14)$$

$$\theta_{10}^* = -\pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A\{(1-P)T - 1\} - \pi_B(1-P)T + (1-P)T, \quad (1.15)$$

$$\theta_{01}^* = -\pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A P(1-T) - \pi_B\{P(1-T) - 1\} + P(1-T), \quad (1.16)$$

and

$$\theta_{00}^* = \pi_{AB}\{PT + (1-P)(1-T)\} - \pi_A PT - \pi_B PT + PT. \quad (1.17)$$

Let $\hat{\theta}_{11}^* = n_{11}^*/n$, $\hat{\theta}_{10}^* = n_{10}^*/n$, $\hat{\theta}_{01}^* = n_{01}^*/n$ and $\hat{\theta}_{00}^* = n_{00}^*/n$, be the observed proportions of (*Yes*, *Yes*), (*Yes*, *No*), (*No*, *Yes*) and (*No*, *No*) responses so that $n_{11}^* + n_{10}^* + n_{01}^* + n_{00}^* = n$. Lee, Sedory and Singh (2013) obtained unbiased estimators for the crossed model as following:

$$\hat{\pi}_A^* = \frac{1}{2} + \frac{(T-P+1)(\hat{\theta}_{11}^* - \hat{\theta}_{00}^*) + (P+T-1)(\hat{\theta}_{10}^* - \hat{\theta}_{01}^*)}{2(P+T-1)}, \quad (1.18)$$

$$\hat{\pi}_B^* = \frac{1}{2} + \frac{(P-T+1)(\hat{\theta}_{11}^* - \hat{\theta}_{00}^*) + (P+T-1)(\hat{\theta}_{01}^* - \hat{\theta}_{10}^*)}{2(P+T-1)}, \quad (1.19)$$

and

$$\hat{\pi}_{AB}^* = \frac{PT\hat{\theta}_{11}^* - (1-P)(1-T)\hat{\theta}_{00}^*}{\{PT + (1-P)(1-T)\}(P+T-1)}, \quad (1.20)$$

with $P + T \neq 1$.

Define

$$\varepsilon_{AB}^* = \frac{\hat{\pi}_{AB}^*}{\pi_{AB}} - 1, \quad \varepsilon_A^* = \frac{\hat{\pi}_A^*}{\pi_A} - 1, \quad \varepsilon_B^* = \frac{\hat{\pi}_B^*}{\pi_B} - 1$$

so that

$$E(\varepsilon_{AB}^*) = E(\varepsilon_A^*) = E(\varepsilon_B^*) = 0$$

$$E(\varepsilon_{AB}^{*2}) = \frac{V(\hat{\pi}_{AB}^*)}{\pi_{AB}^2}, \quad E(\varepsilon_A^{*2}) = \frac{V(\hat{\pi}_A^*)}{\pi_A^2}, \quad E(\varepsilon_B^{*2}) = \frac{V(\hat{\pi}_B^*)}{\pi_B^2}$$

$$E(\varepsilon_{AB}^*\varepsilon_A^*) = \frac{Cov(\hat{\pi}_{AB}^*, \hat{\pi}_A^*)}{\pi_{AB}\pi_A}, \quad E(\varepsilon_{AB}^*\varepsilon_B^*) = \frac{Cov(\hat{\pi}_{AB}^*, \hat{\pi}_B^*)}{\pi_{AB}\pi_B}, \quad E(\varepsilon_A^*\varepsilon_B^*) = \frac{Cov(\hat{\pi}_A^*, \hat{\pi}_B^*)}{\pi_A\pi_B}$$

where

$$V(\hat{\pi}_A^*) = \frac{\pi_A(1-\pi_A)}{n} + \frac{(1-P)T\{PT + (1-P)(1-T)\}(1-\pi_A - \pi_B + 2\pi_{AB})}{n(P+T-1)^2}, \quad (1.21)$$

$$V(\hat{\pi}_B^*) = \frac{\pi_B(1-\pi_B)}{n} + \frac{(1-T)P\{PT+(1-P)(1-T)\}(1-\pi_A-\pi_B+2\pi_{AB})}{n(P+T-1)^2}, \quad (1.22)$$

$$V(\hat{\pi}_{AB}^*) = \frac{\pi_{AB}(1-\pi_{AB})}{n} + \frac{\pi_{AB}\left[P^2T^2+(1-P)^2(1-T)^2-\{PT+(1-P)(1-T)\}(P+T-1)^2\right]}{n\{PT+(1-P)(1-T)\}(P+T-1)^2}$$
$$+\frac{PT(1-P)(1-T)(1-\pi_A-\pi_B)}{n\{PT+(1-P)(1-T)\}(P+T-1)^2}. \quad (1.23)$$

$$Cov\left(\hat{\pi}_{AB}^*, \hat{\pi}_A^*\right) = \frac{\pi_{AB}(1-\pi_A)}{n} + \frac{\pi_{AB}T(1-P)(P-T+1)}{n(P+T-1)^2}$$
$$+\frac{PT(1-P)(1-T)(T-P+1)(1-\pi_A-\pi_B)}{n\{PT+(1-P)(1-T)\}(P+T-1)^2} \quad (1.24)$$

$$Cov\left(\hat{\pi}_{AB}^*, \hat{\pi}_B^*\right) = \frac{\pi_{AB}(1-\pi_B)}{n} + \frac{\pi_{AB}P(1-T)(T-P+1)}{n(P+T-1)^2}$$
$$+\frac{PT(1-P)(1-T)(P-T+1)(1-\pi_A-\pi_B)}{n\{PT+(1-P)(1-T)\}(P+T-1)^2} \quad (1.25)$$

and

$$Cov\left(\hat{\pi}_A^*, \hat{\pi}_B^*\right) = \frac{(\pi_{AB}-\pi_A\pi_B)}{n} + \frac{2PT(1-P)(1-T)(1+2\pi_{AB}-\pi_A-\pi_B)}{n(P+T-1)^2} \quad (1.26)$$

In this paper, we introduce two more parameters, odds ratio and attributable risk, bridging on the article by Lee, Sedory and Singh (2013) and provide a detailed study of the other parameters including the correlation coefficient, relative risk, estimation of membership to at least one group, difference between two proportions. We also introduce two new difference type estimators of the proportion of one of two sensitive variables when the proportion of the other is known. Fox (2016) has renamed the crossed model as a double decker model. To our knowledge, the problem of estimating the odds ratio and the attributable risk using randomized response sampling were first introduced and presented by Lee, Sedory and Singh (2016) at the Joint Statistical Meetings.

In the next section, we consider two estimators of odds ratio (OR); one based on the simple model and the other based on the crossed model.

2. ESTIMATION OF ODDS RATIO

In case of two sensitive characteristics A and B, the four cells of the 2×2 contingency table can be labeled as:

Attributes	B	B^c	Total
A	π_{AB}	$(\pi_A - \pi_{AB})$	π_A
A^c	$(\pi_B - \pi_{AB})$	$(1-\pi_A - \pi_B + \pi_{AB})$	$(1-\pi_A)$
Total	π_B	$(1-\pi_B)$	1

Thus, we consider a measure of odds ratio (OR) in case of two sensitive variables A and B as:

$$\text{OR} = \frac{\pi_{AB}\left(1-\pi_A - \pi_B + \pi_{AB}\right)}{\left(\pi_A - \pi_{AB}\right)\left(\pi_B - \pi_{AB}\right)} \tag{2.1}$$

In the following sub-sections, we consider estimators of the odds ratio (OR) defined in (2.1) by using the simple model and the crossed model.

2.1 ESTIMATION OF ODDS RATIO USING SIMPLE MODEL

By using the same notations for the simple model from Lee et al. (2013), we consider first estimator of the odds ratio (OR) as:

$$\hat{\text{OR}} = \frac{\hat{\pi}_{AB}\left(1-\hat{\pi}_A - \hat{\pi}_B + \hat{\pi}_{AB}\right)}{\left(\hat{\pi}_A - \hat{\pi}_{AB}\right)\left(\hat{\pi}_B - \hat{\pi}_{AB}\right)} \tag{2.2}$$

Now, we have the following theorems:

Theorem 2.1. The bias, to the first order of approximation, in the estimator $\hat{\text{OR}}$ of the odds ratio (OR) is given by:

$$\begin{aligned} B(\hat{OR}) = OR\big[&G_4 V\left(\hat{\pi}_{AB}\right) + G_5 V\left(\hat{\pi}_A\right) + G_6 V\left(\hat{\pi}_B\right) - G_7 Cov\left(\hat{\pi}_A, \hat{\pi}_{AB}\right) \\ &- G_8 Cov\left(\hat{\pi}_B, \hat{\pi}_{AB}\right) + G_9 Cov\left(\hat{\pi}_A, \hat{\pi}_B\right)\big] \end{aligned} \tag{2.3}$$

where

$$\begin{aligned} G_4 = &\frac{1}{\pi_{AB}(1-\pi_A - \pi_B + \pi_{AB})} + \frac{1}{(\pi_A - \pi_{AB})^2} + \frac{1}{(\pi_B - \pi_{AB})^2} \\ &+ \frac{1}{\pi_{AB}(\pi_A - \pi_{AB})} + \frac{1}{\pi_{AB}(\pi_B - \pi_{AB})} + \frac{1}{(\pi_A - \pi_{AB})(1-\pi_A - \pi_B + \pi_{AB})} \\ &+ \frac{1}{(\pi_B - \pi_{AB})(1-\pi_A - \pi_B + \pi_{AB})} + \frac{1}{(\pi_A - \pi_{AB})(\pi_B - \pi_{AB})}; \end{aligned}$$

$$G_5 = \frac{1}{(\pi_A - \pi_{AB})}\left\{\frac{1}{(\pi_A - \pi_{AB})} + \frac{1}{(1-\pi_A - \pi_B + \pi_{AB})}\right\};$$

$$G_6 = \frac{1}{(\pi_B - \pi_{AB})}\left\{\frac{1}{(\pi_B - \pi_{AB})} + \frac{1}{(1-\pi_A - \pi_B + \pi_{AB})}\right\};$$

$$G_7 = \frac{1}{\pi_{AB}(1-\pi_A-\pi_B+\pi_{AB})} + \frac{2}{(\pi_A-\pi_{AB})^2} + \frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}$$
$$+\frac{1}{(1-\pi_A-\pi_B+\pi_{AB})}\left\{\frac{2}{\pi_A-\pi_{AB}}+\frac{1}{\pi_B-\pi_{AB}}\right\}+\frac{1}{\pi_{AB}(\pi_A-\pi_{AB})};$$

$$G_8 = \frac{1}{\pi_{AB}(1-\pi_A-\pi_B+\pi_{AB})} + \frac{2}{(\pi_B-\pi_{AB})^2} + \frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}$$
$$+\frac{1}{(1-\pi_A-\pi_B+\pi_{AB})}\left\{\frac{1}{\pi_A-\pi_{AB}}+\frac{2}{\pi_B-\pi_{AB}}\right\}+\frac{1}{\pi_{AB}(\pi_B-\pi_{AB})};$$

and

$$G_9 = \frac{1}{(1-\pi_A-\pi_B+\pi_{AB})}\left\{\frac{1}{\pi_A-\pi_{AB}}+\frac{1}{\pi_B-\pi_{AB}}\right\}+\frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}$$

are constants.

Proof. The estimator $\hat{OR}$ of the odds ratio can be approximated as:

$$\hat{OR} = \frac{\hat{\pi}_{AB}(1-\hat{\pi}_A-\hat{\pi}_B+\hat{\pi}_{AB})}{(\hat{\pi}_A-\hat{\pi}_{AB})(\hat{\pi}_B-\hat{\pi}_{AB})}$$

$$= \frac{\pi_{AB}(1+\varepsilon_{AB})\{1-\pi_A(1+\varepsilon_A)-\pi_B(1+\varepsilon_B)+\pi_{AB}(1+\varepsilon_{AB})\}}{\{\pi_A(1+\varepsilon_A)-\pi_{AB}(1+\varepsilon_{AB})\}\{\pi_B(1+\varepsilon_B)-\pi_{AB}(1+\varepsilon_{AB})\}}$$

$$= OR(1+\varepsilon_{AB})\left[1-\frac{\pi_A\varepsilon_A+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}\right]\left[1+\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right]^{-1}\left[1+\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right]^{-1}$$

Assuming $\left|\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right|<1$ and $\left|\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right|<1$, and using the binomial expansion up to the first order of approximation, we have

$$\hat{OR} = OR(1+\varepsilon_{AB})\left[1-\frac{\pi_A\varepsilon_A+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}\right]$$
$$\times\left[1-\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}+\left(\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right)^2+\ldots\right]$$
$$\times\left[1-\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\left(\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right)^2+\ldots\right]$$

$$= OR\left[1+\varepsilon_{AB}-\frac{\pi_A\varepsilon_A+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}-\frac{\pi_A\varepsilon_A\varepsilon_{AB}+\pi_B\varepsilon_B\varepsilon_{AB}-\pi_{AB}\varepsilon_{AB}^2}{1-\pi_A-\pi_B+\pi_{AB}}\right]$$

$$\times\left[1-\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}-\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\left(\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right)^2+\left(\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right)^2\right.$$

$$\left.+\left(\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right)\left(\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right)+O(\varepsilon^2)\right]$$

$$= OR\left[1+\varepsilon_{AB}-\frac{\pi_A\varepsilon_A+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}-\frac{\pi_A\varepsilon_A\varepsilon_{AB}+\pi_B\varepsilon_B\varepsilon_{AB}-\pi_{AB}\varepsilon_{AB}^2}{1-\pi_A-\pi_B+\pi_{AB}}\right.$$

$$-\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}-\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\left(\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right)^2+\left(\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right)^2$$

$$+\left(\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right)\left(\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right)-\left(\frac{\pi_A\varepsilon_A\varepsilon_{AB}-\pi_{AB}\varepsilon_{AB}^2}{\pi_A-\pi_{AB}}\right)-\left(\frac{\pi_B\varepsilon_B\varepsilon_{AB}-\pi_{AB}\varepsilon_{AB}^2}{\pi_B-\pi_{AB}}\right)$$

$$+\left(\frac{\pi_A\varepsilon_A-\pi_{AB}\varepsilon_{AB}}{\pi_A-\pi_{AB}}\right)\left(\frac{\pi_A\varepsilon_A+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}\right)$$

$$\left.+\left(\frac{\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right)\left(\frac{\pi_A\varepsilon_A+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}\right)+O(\varepsilon^2)\right]$$

$$= OR\left[1+\left\{1+\frac{\pi_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{\pi_{AB}}{\pi_A-\pi_{AB}}+\frac{\pi_{AB}}{\pi_B-\pi_{AB}}\right\}\varepsilon_{AB}\right.$$

$$-\left\{\frac{\pi_A}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{\pi_A}{\pi_A-\pi_{AB}}\right\}\varepsilon_A-\left\{\frac{\pi_B}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{\pi_B}{\pi_B-\pi_{AB}}\right\}\varepsilon_B$$

$$+\left\{\frac{\pi_{AB}}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{\pi_{AB}^2}{(\pi_A-\pi_{AB})^2}+\frac{\pi_{AB}^2}{(\pi_B-\pi_{AB})^2}+\frac{\pi_{AB}}{\pi_A-\pi_{AB}}+\frac{\pi_{AB}}{\pi_B-\pi_{AB}}\right.$$

$$+\frac{\pi_{AB}^2}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{\pi_{AB}^2}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}$$

$$\left.+\frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}\right\}\varepsilon_{AB}^2$$

$$+\left\{\frac{\pi_A^2}{(\pi_A-\pi_{AB})^2}+\frac{\pi_A^2}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_A+\pi_{AB})}\right\}\varepsilon_A^2$$

$$+\left\{\frac{\pi_B^2}{(\pi_B-\pi_{AB})^2}+\frac{\pi_B^2}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_A+\pi_{AB})}\right\}\varepsilon_B^2$$

$$-\left\{\frac{\pi_A}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{2\pi_A\pi_{AB}}{(\pi_A-\pi_{AB})^2}+\frac{\pi_A\pi_{AB}}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}\right.$$

$$\left.+\frac{2\pi_A\pi_{AB}}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{\pi_A\pi_{AB}}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{\pi_A}{(\pi_A-\pi_{AB})}\right\}\varepsilon_A\varepsilon_{AB}$$

$$-\left\{\frac{\pi_B}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{2\pi_B\pi_{AB}}{(\pi_B-\pi_{AB})^2}+\frac{\pi_B\pi_{AB}}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}\right.$$

$$\left.+\frac{\pi_B\pi_{AB}}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{2\pi_B\pi_{AB}}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{\pi_B}{(\pi_B-\pi_{AB})}\right\}\varepsilon_B\varepsilon_{AB}$$

$$\left.+\left\{\frac{\pi_A\pi_B}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}+\frac{\pi_A\pi_B}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{\pi_A\pi_B}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}\right\}\varepsilon_A\varepsilon_B+..\right]$$

$$=OR\left[1+\left\{\frac{1}{\pi_{AB}}+\frac{1}{1-\pi_A-\pi_B+\pi_{AB}}+\frac{1}{\pi_A-\pi_{AB}}+\frac{1}{\pi_B-\pi_{AB}}\right\}\pi_{AB}\varepsilon_{AB}\right.$$

$$-\left\{\frac{1-\pi_B}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}\right\}\pi_A\varepsilon_A-\left\{\frac{1-\pi_A}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}\right\}\pi_B\varepsilon_B$$

$$+\left\{\frac{1}{\pi_{AB}(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{(\pi_A-\pi_{AB})^2}+\frac{1}{(\pi_B-\pi_{AB})^2}+\frac{1}{\pi_{AB}(\pi_A-\pi_{AB})}+\frac{1}{\pi_{AB}(\pi_B-\pi_{AB})}\right.$$

$$\left.+\frac{1}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}\right\}\pi_{AB}^2\varepsilon_{AB}^2$$

$$+\frac{1}{(\pi_A-\pi_{AB})}\left\{\frac{1}{(\pi_A-\pi_{AB})}+\frac{1}{(1-\pi_A-\pi_B+\pi_{AB})}\right\}\pi_A^2\varepsilon_A^2$$

$$+\frac{1}{(\pi_B-\pi_{AB})}\left\{\frac{1}{(\pi_B-\pi_{AB})}+\frac{1}{(1-\pi_A-\pi_B+\pi_{AB})}\right\}\pi_B^2\varepsilon_B^2$$

$$-\left\{\frac{1}{\pi_{AB}(1-\pi_A-\pi_B+\pi_{AB})}+\frac{2}{(\pi_A-\pi_{AB})^2}+\frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}\right.$$

$$\left.+\frac{2}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{\pi_{AB}(\pi_A-\pi_{AB})}\right\}\pi_A\pi_{AB}\varepsilon_A\varepsilon_{AB}$$

$$-\left\{\frac{1}{\pi_{AB}(1-\pi_A-\pi_B+\pi_{AB})}+\frac{2}{(\pi_B-\pi_{AB})^2}+\frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}\right.$$

$$\left.+\frac{1}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{2}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{\pi_{AB}(\pi_B-\pi_{AB})}\right\}\pi_B\pi_{AB}\varepsilon_B\varepsilon_{AB}$$

$$\left.+\left\{\frac{1}{(\pi_A-\pi_{AB})(\pi_B-\pi_{AB})}+\frac{1}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}+\frac{1}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}\right\}\pi_A\pi_B\varepsilon_A\varepsilon_B+..\right]$$

$$= OR\big[1 + G_1\pi_{AB}\varepsilon_{AB} - G_2\pi_A\varepsilon_A - G_3\pi_B\varepsilon_B + G_4\pi_{AB}^2\varepsilon_{AB}^2 + G_5\pi_A^2\varepsilon_A^2 + G_6\pi_B^2\varepsilon_B^2 - G_7\pi_A\pi_{AB}\varepsilon_A\varepsilon_{AB} - G_8\pi_B\pi_{AB}\varepsilon_B\varepsilon_{AB} + G_9\pi_A\pi_B\varepsilon_A\varepsilon_B + O(\varepsilon^2)\big] \quad (2.4)$$

where

$$G_1 = \left\{\frac{1}{\pi_{AB}} + \frac{1}{1-\pi_A-\pi_B+\pi_{AB}} + \frac{1}{\pi_A-\pi_{AB}} + \frac{1}{\pi_B-\pi_{AB}}\right\};$$

$$G_2 = \left\{\frac{1-\pi_B}{(\pi_A-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}\right\};$$

and

$$G_3 = \left\{\frac{1-\pi_A}{(\pi_B-\pi_{AB})(1-\pi_A-\pi_B+\pi_{AB})}\right\}$$

By the definition of bias that is

$$B\left(\hat{OR}\right) = E(\hat{OR}) - OR$$

we have the theorem.

Note that $B\left(\hat{OR}\right) \to 0$ as $n \to \infty$, thus the estimator $\hat{OR}$ is a consistent estimator of the odds ratio (OR).

Theorem 2.2. The mean squared error, to the first order of approximation, of the estimator $\hat{\mathrm{OR}}$ of the odds ratio (OR) is given by:

$$MSE(\hat{\mathrm{OR}}) = OR^2\big[G_1^2V(\hat{\pi}_{AB}) + G_2^2V(\hat{\pi}_A) + G_3^2V(\hat{\pi}_B) - 2G_1G_2Cov(\hat{\pi}_A,\hat{\pi}_{AB}) - 2G_1G_3Cov(\hat{\pi}_B,\hat{\pi}_{AB}) + 2G_2G_3Cov(\hat{\pi}_A,\hat{\pi}_B)\big] \quad (2.5)$$

Proof. By the definition of mean squared error, we have

$$MSE\left(\hat{OR}\right) = E\left(\hat{OR} - OR\right)^2$$

$$\cong OR^2E\big[G_1\pi_{AB}\varepsilon_{AB} - G_2\pi_A\varepsilon_A - G_3\pi_B\varepsilon_B\big]^2$$

$$\cong OR^2E\big[G_1^2\pi_{AB}^2\varepsilon_{AB}^2 + G_2^2\pi_A^2\varepsilon_A^2 + G_3^2\pi_B^2\varepsilon_B^2 - 2G_1G_2\pi_{AB}\pi_A\varepsilon_{AB}\varepsilon_A - 2G_1G_3\pi_{AB}\pi_B\varepsilon_{AB}\varepsilon_B + 2G_2G_3\pi_A\pi_B\varepsilon_A\varepsilon_B\big]$$

which proves the theorem.

2.2 ESTIMATION OF ODDS RATIO USING CROSSED MODEL

By using the same notations for the crossed model from Lee *et al.* (2013), we consider second estimator of the odds ratio (OR) as:

$$\hat{\text{OR}}^* = \frac{\hat{\pi}^*_{AB}\left(1-\hat{\pi}^*_A-\hat{\pi}^*_B+\hat{\pi}^*_{AB}\right)}{\left(\hat{\pi}^*_A-\hat{\pi}^*_{AB}\right)\left(\hat{\pi}^*_B-\hat{\pi}^*_{AB}\right)} \tag{2.6}$$

Now, we have the following theorems:

Theorem 2.3. The bias, to the first order of approximation, in the estimator $\hat{\text{OR}}^*$ of the odds ratio (OR) is given by:

$$B(\hat{OR}^*) = OR\Big[G_4V\left(\hat{\pi}^*_{AB}\right)+G_5V\left(\hat{\pi}^*_A\right)+G_6V\left(\hat{\pi}^*_B\right)-G_7Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_{AB}\right) \\ -G_8Cov\left(\hat{\pi}^*_B,\hat{\pi}^*_{AB}\right)+G_9Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_B\right)\Big] \tag{2.7}$$

Proof. The estimator $\hat{OR}^*$ of the odds ratio can be approximated as:

$$\hat{OR}^* = \frac{\hat{\pi}^*_{AB}\left(1-\hat{\pi}^*_A-\hat{\pi}^*_B+\hat{\pi}^*_{AB}\right)}{\left(\hat{\pi}^*_A-\hat{\pi}^*_{AB}\right)\left(\hat{\pi}^*_B-\hat{\pi}^*_{AB}\right)}$$

$$= \frac{\pi_{AB}\left(1+\varepsilon^*_{AB}\right)\left\{1-\pi_A\left(1+\varepsilon^*_A\right)-\pi_B\left(1+\varepsilon^*_B\right)+\pi_{AB}\left(1+\varepsilon^*_{AB}\right)\right\}}{\left\{\pi_A\left(1+\varepsilon^*_A\right)-\pi_{AB}\left(1+\varepsilon^*_{AB}\right)\right\}\left\{\pi_B\left(1+\varepsilon^*_B\right)-\pi_{AB}\left(1+\varepsilon^*_{AB}\right)\right\}}$$

Again expanding by the binomial distribution and by the definition of bias that is

$$B\left(\hat{OR}^*\right) = E(\hat{OR}^*)-OR$$

we have the theorem.

Note that $B\left(\hat{OR}^*\right) \to 0$ as , thus the estimator $\hat{OR}^*$ is a consistent estimator of the odds ratio (OR).

Theorem 2.4. The mean squared error, to the first order of approximation, of the estimator $\hat{\text{OR}}^*$ of the odds ratio (OR) is given by:

$$MSE(\hat{\text{OR}}^*) = OR^2\Big[G_1^2V\left(\hat{\pi}^*_{AB}\right)+G_2^2V\left(\hat{\pi}^*_A\right)+G_3^2V\left(\hat{\pi}^*_B\right)-2G_1G_2Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_{AB}\right) \\ -2G_1G_3Cov\left(\hat{\pi}^*_B,\hat{\pi}^*_{AB}\right)+2G_2G_3Cov\left(\hat{\pi}^*_A,\hat{\pi}^*_B\right)\Big] \tag{2.8}$$

Proof. By the definition of mean squared error, we have

$$MSE\left(\hat{OR}^*\right) = E\left(\hat{OR}^* - OR\right)^2$$

$$\cong OR^2 E\left[G_1 \pi_{AB} \varepsilon_{AB}^* - G_2 \pi_A \varepsilon_A^* - G_3 \pi_B \varepsilon_B^*\right]^2$$

Expanding and taking the expected value, we have the theorem.

In the next section, we consider the problem of estimation of attributable risk.

3. ESTIMATION OF ATTRIBUTABLE RISK

In order to define an attributable risk, we have the following theorem.

Theorem 3.1. The attributable risk $AR(B \mid A)$ is given by:

$$AR(B \mid A) = \frac{\pi_{AB} - \pi_A \pi_B}{\pi_B (1 - \pi_A)} \tag{3.1}$$

Proof. We know that the relative risk (RR) of being in group B given a respondent belongs to group A is defined as:

$$RR(B \mid A) = \frac{P(B \mid A)}{P\left(B \mid A^C\right)} = \frac{\frac{P(A \cap B)}{P(A)}}{\frac{P\left(B \cap A^C\right)}{P\left(A^C\right)}} = \frac{P(A \cap B)[1 - P(A)]}{P(A)[P(B) - P(A \cap B)]} = \frac{\pi_{AB}(1 - \pi_A)}{\pi_A(\pi_B - \pi_{AB})}$$

Following Rosner (2016), by the definition of attributable risk, we have

$$AR(B \mid A) = \frac{[RR(B \mid A) - 1]\pi_A}{[RR(B \mid A) - 1]\pi_A + 1} = \frac{\left[\frac{\pi_{AB}(1 - \pi_A)}{\pi_A(\pi_B - \pi_{AB})} - 1\right]\pi_A}{\left[\frac{\pi_{AB}(1 - \pi_A)}{\pi_A(\pi_B - \pi_{AB})} - 1\right]\pi_A + 1}$$

$$= \frac{[\pi_{AB}(1 - \pi_A) - \pi_A(\pi_B - \pi_{AB})]\pi_A}{[\pi_{AB}(1 - \pi_A) - \pi_A(\pi_B - \pi_{AB})]\pi_A + \pi_A(\pi_B - \pi_{AB})}$$

$$= \frac{[\pi_{AB} - \pi_{AB}\pi_A - \pi_A\pi_B + \pi_{AB}\pi_A]\pi_A}{[\pi_{AB} - \pi_{AB}\pi_A - \pi_A\pi_B + \pi_{AB}\pi_A]\pi_A + \pi_A(\pi_B - \pi_{AB})}$$

$$= \frac{[\pi_{AB} - \pi_A\pi_B]\pi_A}{[\pi_{AB} - \pi_A\pi_B]\pi_A + \pi_A(\pi_B - \pi_{AB})}$$

$$= \frac{[\pi_{AB} - \pi_A\pi_B]\pi_A}{[\pi_{AB} - \pi_A\pi_B + \pi_B - \pi_{AB}]\pi_A} = \frac{\pi_{AB} - \pi_A\pi_B}{-\pi_A\pi_B + \pi_B} = \frac{\pi_{AB} - \pi_A\pi_B}{\pi_B(1 - \pi_A)}$$

which proves the theorem.

3.1 ESTIMATION OF ATTRIBUTABLE RISK USING SIMPLE MODEL

By using the same notation for the simple model as found in Lee *et al.* (2013), we consider our first estimator of the attributable risk as:

$$\hat{AR}(B \mid A) = \frac{\hat{\pi}_{AB} - \hat{\pi}_A \hat{\pi}_B}{\hat{\pi}_B (1 - \hat{\pi}_A)} \tag{3.2}$$

Now, we have the following theorems:

Theorem 3.2. The bias, to the first order of approximation, in the estimator $\hat{AR}(B \mid A)$ of the attributable risk $AR(B \mid A)$ is given by:

$$B(\hat{AR}(B \mid A)) = \frac{AR(B \mid A)}{(\pi_{AB} - \pi_A \pi_B)} \left[\frac{\pi_{AB} V(\hat{\pi}_B)}{\pi_B^2} - \frac{(\pi_B - \pi_{AB}) V(\hat{\pi}_A)}{(1 - \pi_A)^2} - \frac{Cov(\hat{\pi}_{AB}, \hat{\pi}_B)}{\pi_B} + \frac{Cov(\hat{\pi}_{AB}, \hat{\pi}_A)}{(1 - \pi_A)} - \frac{\pi_{AB} Cov(\hat{\pi}_A, \hat{\pi}_B)}{\pi_B (1 - \pi_A)} \right] \tag{3.3}$$

Proof. The estimator $\hat{AR}(B \mid A)$ can be approximated as.

$$\hat{AR}(B \mid A) = \frac{\hat{\pi}_{AB} - \hat{\pi}_A \hat{\pi}_B}{\hat{\pi}_B (1 - \hat{\pi}_A)} = \frac{\pi_{AB}(1 + \varepsilon_{AB}) - \pi_A (1 + \varepsilon_A) \pi_B (1 + \varepsilon_B)}{\pi_B (1 + \varepsilon_B) [1 - \pi_A (1 + \varepsilon_A)]}$$

$$= \frac{(\pi_{AB} - \pi_A \pi_B) + \pi_{AB} \varepsilon_{AB} - \pi_A \pi_B (\varepsilon_A + \varepsilon_B + \varepsilon_A \varepsilon_B)}{\pi_B (1 - \pi_A) + \pi_B (1 - \pi_A) \varepsilon_B - \pi_A \pi_B \varepsilon_A - \pi_A \pi_B \varepsilon_A \varepsilon_B}$$

$$= \frac{(\pi_{AB} - \pi_A \pi_B) \left[1 + \dfrac{\pi_{AB} \varepsilon_{AB} - \pi_A \pi_B (\varepsilon_A + \varepsilon_B + \varepsilon_A \varepsilon_B)}{\pi_{AB} - \pi_A \pi_B} \right]}{\pi_B (1 - \pi_A) \left[1 + \dfrac{\pi_B (1 - \pi_A) \varepsilon_B - \pi_A \pi_B \varepsilon_A - \pi_A \pi_B \varepsilon_A \varepsilon_B}{\pi_B (1 - \pi_A)} \right]}$$

$$= AR(B \mid A) \left[1 + \frac{\pi_{AB} \varepsilon_{AB} - \pi_A \pi_B (\varepsilon_A + \varepsilon_B + \varepsilon_A \varepsilon_B)}{\pi_{AB} - \pi_A \pi_B} \right] \left[1 + \frac{\pi_B (1 - \pi_A) \varepsilon_B - \pi_A \pi_B \varepsilon_A - \pi_A \pi_B \varepsilon_A \varepsilon_B}{\pi_B (1 - \pi_A)} \right]^{-1}$$

$$= AR(B \mid A) \left[1 + \frac{\pi_{AB} \varepsilon_{AB} - \pi_A \pi_B (\varepsilon_A + \varepsilon_B + \varepsilon_A \varepsilon_B)}{\pi_{AB} - \pi_A \pi_B} \right] \left[1 + \varepsilon_B - \frac{\pi_A \varepsilon_A + \pi_A \varepsilon_A \varepsilon_B}{(1 - \pi_A)} \right]^{-1}$$

Let us assume $\left| \varepsilon_B - \dfrac{\pi_A \varepsilon_A + \pi_A \varepsilon_A \varepsilon_B}{1 - \pi_A} \right| < 1$, and by the binomial expansion, to the first order of approximation, we have

$$\hat{AR}(B \mid A) = AR(B \mid A) \left[1 + \frac{\pi_{AB} \varepsilon_{AB} - \pi_A \pi_B (\varepsilon_A + \varepsilon_B + \varepsilon_A \varepsilon_B)}{\pi_{AB} - \pi_A \pi_B} \right] \times \left[1 + \left\{ \varepsilon_B - \frac{\pi_A \varepsilon_A + \pi_A \varepsilon_A \varepsilon_B}{(1 - \pi_A)} \right\} \right]^{-1}$$

$$= AR(B\,|\,A)\left[1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right]$$

$$\times\left[1-\left\{\varepsilon_B-\frac{\pi_A\varepsilon_A+\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)}\right\}+\left\{\varepsilon_B-\frac{\pi_A\varepsilon_A+\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)}\right\}^2+\ldots\right]$$

$$= AR(B\,|\,A)\left[1+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_A}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_A\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}\right]$$

$$\times\left[1-\varepsilon_B+\frac{\pi_A\varepsilon_A}{(1-\pi_A)}+\frac{\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)}+\varepsilon_B^2+\frac{\pi_A^2\varepsilon_A^2}{(1-\pi_A)^2}-\frac{2\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)}+..\right]$$

$$= AR(B\,|\,A)\left[1+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_A}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_A\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}\right]$$

$$\times\left[1-\varepsilon_B+\frac{\pi_A\varepsilon_A}{(1-\pi_A)}+\varepsilon_B^2+\frac{\pi_A^2\varepsilon_A^2}{(1-\pi_A)^2}-\frac{\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)}+..\right]$$

$$= AR(B\,|\,A)\left[1+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_A}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A\pi_B\varepsilon_A\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}\right.$$

$$-\varepsilon_B-\frac{\pi_{AB}\varepsilon_{AB}\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}+\frac{\pi_A\pi_B\varepsilon_A\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}+\frac{\pi_A\pi_B\varepsilon_B^2}{\pi_{AB}-\pi_A\pi_B}+\frac{\pi_A}{1-\pi_A}\varepsilon_A$$

$$+\frac{\pi_{AB}\pi_A}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}\varepsilon_{AB}\varepsilon_A-\frac{\pi_A^2\pi_B\varepsilon_A^2}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}-\frac{\pi_A^2\pi_B\varepsilon_A\varepsilon_B}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}$$

$$\left.+\varepsilon_B^2+\frac{\pi_A^2\varepsilon_A^2}{(1-\pi_A)^2}-\frac{\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)}+\ldots\right]$$

$$= AR(B\,|\,A)\left[1+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_{AB}-\pi_A\pi_B}-\left\{\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}+1\right\}\varepsilon_B-\left\{\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A}{1-\pi_A}\right\}\varepsilon_A\right.$$

$$+\left\{\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}+1\right\}\varepsilon_B^2-\left\{\frac{\pi_A^2\pi_B}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}-\frac{\pi_A^2}{(1-\pi_A)^2}\right\}\varepsilon_A^2$$

$$+\frac{\pi_{AB}\pi_A\varepsilon_{AB}\varepsilon_A}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}-\frac{\pi_{AB}\varepsilon_{AB}\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}$$

$$\left.-\left\{\frac{\pi_A^2\pi_B}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}+\frac{\pi_A}{(1-\pi_A)}\right\}\varepsilon_A\varepsilon_B+..\right]$$

$$= AR(B|A)\left[1+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_{AB}\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A(\pi_B-\pi_{AB})\varepsilon_A}{(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\right.$$

$$+\frac{\pi_{AB}\varepsilon_B^2}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A^2(\pi_B-\pi_{AB})\varepsilon_A^2}{(1-\pi_A)^2(\pi_{AB}-\pi_A\pi_B)}-\frac{\pi_{AB}\varepsilon_{AB}\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}$$

$$\left.+\frac{\pi_{AB}\pi_A\varepsilon_{AB}\varepsilon_A}{(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}-\frac{\pi_{AB}\pi_A\varepsilon_A\varepsilon_B}{(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}+O\left(\varepsilon^2\right)\right]$$

By the definition of bias, we have

$$B\left(\hat{AR}(B\mid A)\right)=E\left(\hat{AR}(B\mid A)\right)-AR(B\mid A)$$

$$=AR(B\mid A)\left[\frac{\pi_{AB}E\left(\varepsilon_B^2\right)}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_A^2(\pi_B-\pi_{AB})E\left(\varepsilon_A^2\right)}{(1-\pi_A)^2(\pi_{AB}-\pi_A\pi_B)}-\frac{\pi_{AB}E(\varepsilon_{AB}\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right.$$

$$\left.+\frac{\pi_{AB}\pi_A E(\varepsilon_{AB}\varepsilon_A)}{(\pi_{AB}-\pi_A\pi_B)(1-\pi_A)}-\frac{\pi_{AB}\pi_A E(\varepsilon_A\varepsilon_B)}{(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\right]$$

$$=AR(B\mid A)\left[\frac{\pi_{AB}V(\hat{\pi}_B)}{\pi_B^2(\pi_{AB}-\pi_A\pi_B)}-\frac{(\pi_B-\pi_{AB})V(\hat{\pi}_A)}{(1-\pi_A)^2(\pi_{AB}-\pi_A\pi_B)}-\frac{Cov(\hat{\pi}_{AB},\hat{\pi}_B)}{\pi_B(\pi_{AB}-\pi_A\pi_B)}\right.$$

$$\left.+\frac{Cov(\hat{\pi}_{AB},\hat{\pi}_A)}{(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}-\frac{\pi_{AB}Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_B(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\right]$$

which proves the theorem.

Note that $B\left\{\hat{AR}(B\mid A)\right\}\to 0$ as $n\to\infty$, thus the estimator $\hat{AR}(B\mid A)$ is a consistent estimator of the attributable risk $AR(B\mid A)$.

Theorem 3.3. The mean squared error, to the first order of approximation, of the estimator $\hat{AR}(B\mid A)$ of the attributable risk $AR(B\mid A)$ is given by:

$$MSE\left(\hat{AR}(B\mid A)\right)=\frac{\{AR(B\mid A)\}^2}{(\pi_{AB}-\pi_A\pi_B)^2}\left[V(\hat{\pi}_{AB})+\frac{\pi_{AB}^2V(\hat{\pi}_B)}{\pi_B^2}+\frac{(\pi_B-\pi_{AB})^2V(\hat{\pi}_A)}{(1-\pi_A)^2}-\frac{2\pi_{AB}Cov(\hat{\pi}_{AB},\hat{\pi}_B)}{\pi_B}\right.$$

$$\left.-\frac{2(\pi_B-\pi_{AB})Cov(\hat{\pi}_{AB},\hat{\pi}_A)}{(1-\pi_A)}+\frac{2\pi_{AB}(\pi_B-\pi_{AB})Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_B(1-\pi_A)}\right] \quad (3.4)$$

Proof. By the definition of mean square error, we have

$$MSE\left(\hat{AR}(B\mid A)\right)=E\left(\hat{AR}(B\mid A)-AR(B\mid A)\right)^2$$

$$\cong\{AR(B\mid A)\}^2E\left(\frac{\pi_{AB}\varepsilon_{AB}}{\pi_{AB}-\pi_A\pi_B}-\frac{\pi_{AB}\varepsilon_B}{\pi_{AB}-\pi_A\pi_B}-\frac{(\pi_B-\pi_{AB})\pi_A\varepsilon_A}{(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\right)^2$$

$$=\frac{\{AR(B\mid A)\}^2}{(\pi_{AB}-\pi_A\pi_B)^2}E\left[\pi_{AB}\varepsilon_{AB}-\pi_{AB}\varepsilon_B-\frac{(\pi_B-\pi_{AB})\pi_A\varepsilon_A}{(1-\pi_A)}\right]^2$$

$$= \frac{\{AR(B|A)\}^2}{(\pi_{AB} - \pi_A \pi_B)^2} E\Big[\pi_{AB}^2 \varepsilon_{AB}^2 + \pi_{AB}^2 \varepsilon_B^2 + \frac{(\pi_B - \pi_{AB})^2 \pi_A^2}{(1-\pi_A)^2} \varepsilon_A^2$$

$$- 2\pi_{AB}^2 \varepsilon_{AB} \varepsilon_B - 2\frac{\pi_{AB}(\pi_B - \pi_{AB})\pi_A}{(1-\pi_A)} \varepsilon_{AB}\varepsilon_A + 2\frac{\pi_{AB}(\pi_B - \pi_{AB})\pi_A}{(1-\pi_A)} \varepsilon_A \varepsilon_B \Big]$$

$$= \frac{\{AR(B|A)\}^2}{(\pi_{AB} - \pi_A \pi_B)^2} \Big[\pi_{AB}^2 E(\varepsilon_{AB}^2) + \pi_{AB}^2 E(\varepsilon_B^2) + \frac{(\pi_B - \pi_{AB})^2 \pi_A^2}{(1-\pi_A)^2} E(\varepsilon_A^2)$$

$$- 2\pi_{AB}^2 E(\varepsilon_{AB}\varepsilon_B) - 2\frac{\pi_{AB}(\pi_B - \pi_{AB})\pi_A}{(1-\pi_A)} E(\varepsilon_{AB}\varepsilon_A) + 2\frac{\pi_{AB}(\pi_B - \pi_{AB})\pi_A}{(1-\pi_A)} E(\varepsilon_A \varepsilon_B) \Big]$$

$$= \frac{\{AR(B|A)\}^2}{(\pi_{AB} - \pi_A \pi_B)^2} \Big[V(\hat{\pi}_{AB}) + \pi_{AB}^2 \frac{V(\hat{\pi}_B)}{\pi_B^2} + \frac{(\pi_B - \pi_{AB})^2 V(\hat{\pi}_A)}{(1-\pi_A)^2}$$

$$- \frac{2\pi_{AB} Cov(\hat{\pi}_{AB}, \hat{\pi}_B)}{\pi_B} - \frac{2(\pi_B - \pi_{AB})}{(1-\pi_A)} Cov(\hat{\pi}_{AB}, \hat{\pi}_A) + \frac{2\pi_{AB}(\pi_B - \pi_{AB})}{\pi_B(1-\pi_A)} Cov(\hat{\pi}_A, \hat{\pi}_B) \Big]$$

which proves the theorem.

3.2 ESTIMATION OF ATTRIBUTABLE RISK USING CROSSED MODEL

By using the same notation for the crossed model as found in Lee *et al.* (2013), we consider our second estimator of the attributable risk as:

$$\hat{AR}^*(B|A) = \frac{\hat{\pi}_{AB}^* - \hat{\pi}_A^* \hat{\pi}_B^*}{\hat{\pi}_B^*(1-\hat{\pi}_A^*)} \tag{3.5}$$

Now, we have the following lemma:

Lemma 3.1. The bias, to the first order of approximation, in the estimator $\hat{AR}^*(B|A)$ of the attributable risk $AR(A|B)$ is given by:

$$B(\hat{AR}^*(B|A)) = \frac{AR(B|A)}{(\pi_{AB} - \pi_A \pi_B)} \Bigg[\frac{\pi_{AB} V(\hat{\pi}_B^*)}{\pi_B^2} - \frac{(\pi_B - \pi_{AB}) V(\hat{\pi}_A^*)}{(1-\pi_A)^2} - \frac{Cov(\hat{\pi}_{AB}^*, \hat{\pi}_B^*)}{\pi_B}$$

$$+ \frac{Cov(\hat{\pi}_{AB}^*, \hat{\pi}_A^*)}{(1-\pi_A)} - \frac{\pi_{AB} Cov(\hat{\pi}_A^*, \hat{\pi}_B^*)}{\pi_B(1-\pi_A)} \Bigg] \tag{3.6}$$

Proof. It follows from the previous section.

Note that $B\left\{\hat{AR}^*(B|A)\right\} \to 0$ as $n \to \infty$, thus the estimator $\hat{AR}^*(B|A)$ is a consistent estimator of the attributable risk $AR(B|A)$

Lemma 3.2. The mean squared error, to the first order of approximation, of the estimator $\hat{AR}^*(B \mid A)$ of the attributable risk $AR(B \mid A)$ is given by:

$$MSE\left(\hat{AR}^*(B \mid A)\right) = \frac{\{AR(B \mid A)\}^2}{(\pi_{AB} - \pi_A \pi_B)^2}\left[V\left(\hat{\pi}^*_{AB}\right) + \frac{\pi^2_{AB} V\left(\hat{\pi}^*_B\right)}{\pi^2_B} + \frac{(\pi_B - \pi_{AB})^2 V\left(\hat{\pi}^*_A\right)}{(1-\pi_A)^2} - \frac{2\pi_{AB} Cov\left(\hat{\pi}^*_{AB}, \hat{\pi}^*_B\right)}{\pi_B} - \frac{2(\pi_B - \pi_{AB}) Cov\left(\hat{\pi}^*_{AB}, \hat{\pi}^*_A\right)}{(1-\pi_A)} + \frac{2\pi_{AB}(\pi_B - \pi_{AB}) Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_B\right)}{\pi_B(1-\pi_A)}\right] \tag{3.7}$$

Proof. It follows from the previous section.

4. ESTIMATION OF RELATIVE RISK

Here we consider the problem of estimating the relative risk of a respondent belonging to group B given that the respondent belongs to group A. For example, it could be used to estimate the relative risk of involving a terrorist (group B) given that an accident (say A) happened (say, any type of domestic violence). The relative risk of event B given that the event A occurred is defined as;

$$RR(B|A) = \frac{P(B|A)}{P(B|A^C)} = \frac{\frac{P(A\cap B)}{P(A)}}{\frac{P(B\cap A^C)}{P(A^C)}} = \frac{\frac{P(A\cap B)}{P(A)}}{\frac{P(B)-P(A\cap B)}{1-P(A)}} = \frac{P(A\cap B)[1-P(A)]}{P(A)[P(B)-P(A\cap B)]}$$

Thus we consider the problem of estimating relative risk (RR) defined as:

$$RR(B|A) = \frac{\pi_{AB}(1-\pi_A)}{\pi_A(\pi_B - \pi_{AB})} \tag{4.1}$$

4.1 ESTIMATION OF RELATIVE RISK USING SIMPLE MODEL

We define an estimator of the relative risk as:

$$\hat{RR}(B|A) = \frac{\hat{\pi}_{AB}(1-\hat{\pi}_A)}{\hat{\pi}_A(\hat{\pi}_B - \hat{\pi}_{AB})} \tag{4.2}$$

Now we have the following theorems:

Theorem 4.1. The bias in the estimator $\hat{RR}(B\,|\,A)$, to the first order of approximation, is given by

$$B\left(\hat{RR}(B\,|\,A)\right) = RR(B\,|\,A)\left[\frac{V(\hat{\pi}_A)}{\pi_A^2(1-\pi_A)} + \frac{V(\hat{\pi}_B)}{(\pi_B-\pi_{AB})^2} + \frac{\pi_B V(\hat{\pi}_{AB})}{\pi_{AB}(\pi_B-\pi_{AB})^2} + \frac{Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A(1-\pi_A)(\pi_B-\pi_{AB})} - \frac{\pi_B Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}(1-\pi_A)(\pi_B-\pi_{AB})} - \frac{(\pi_B+\pi_{AB})Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_{AB}(\pi_B-\pi_{AB})^2}\right] \quad (4.3)$$

Proof. The estimator $\hat{RR}(B\,|\,A)$ can be approximated, in terms of ε_A, ε_B and ε_{AB} as

$$\hat{RR}(B|A) = \frac{\pi_{AB}(1+\varepsilon_{AB})[1-\pi_A(1+\varepsilon_A)]}{\pi_A(1+\varepsilon_A)[\pi_B(1+\varepsilon_B)-\pi_{AB}(1+\varepsilon_{AB})]}$$

$$= \frac{(\pi_{AB}+\pi_{AB}\varepsilon_{AB})(1-\pi_A-\pi_A\varepsilon_A)}{(\pi_A+\pi_A\varepsilon_A)(\pi_B+\pi_B\varepsilon_B-\pi_{AB}-\pi_{AB}\varepsilon_{AB})}$$

$$= \frac{(\pi_{AB}+\pi_{AB}\varepsilon_{AB})[(1-\pi_A)-\pi_A\varepsilon_A]}{(\pi_A+\pi_A\varepsilon_A)[(\pi_B-\pi_{AB})+\pi_B\varepsilon_B-\pi_{AB}\varepsilon_{AB}]}$$

$$= \frac{\pi_{AB}(1-\pi_A)+\pi_{AB}(1-\pi_A)\varepsilon_{AB}-\pi_{AB}\pi_A\varepsilon_A-\pi_{AB}\pi_A\varepsilon_A\varepsilon_{AB}}{\pi_A(\pi_B-\pi_{AB})+\pi_A(\pi_B-\pi_{AB})\varepsilon_A+\pi_A\pi_B\varepsilon_B+\pi_A\pi_B\varepsilon_A\varepsilon_B-\pi_A\pi_{AB}\varepsilon_{AB}-\pi_A\pi_{AB}\varepsilon_A\varepsilon_{AB}}$$

$$= \frac{\pi_{AB}(1-\pi_A)\left[1+\frac{\pi_{AB}(1-\pi_A)\varepsilon_{AB}}{\pi_{AB}(1-\pi_A)}-\frac{\pi_{AB}\pi_A\varepsilon_A}{\pi_{AB}(1-\pi_A)}-\frac{\pi_{AB}\pi_A\varepsilon_A\varepsilon_{AB}}{\pi_{AB}(1-\pi_A)}\right]}{\pi_A(\pi_B-\pi_{AB})\left[1+\varepsilon_A+\frac{\pi_B\varepsilon_B}{(\pi_B-\pi_{AB})}+\frac{\pi_B\varepsilon_A\varepsilon_B}{(\pi_B-\pi_{AB})}-\frac{\pi_{AB}\varepsilon_{AB}}{(\pi_B-\pi_{AB})}-\frac{\pi_{AB}\varepsilon_A\varepsilon_{AB}}{(\pi_B-\pi_{AB})}\right]}$$

$$= RR(B\,|\,A)\left[1+\varepsilon_{AB}-\frac{\pi_A\varepsilon_A}{1-\pi_A}-\frac{\pi_A\varepsilon_A\varepsilon_{AB}}{1-\pi_A}\right]\left[1+\varepsilon_A+\frac{\pi_B\varepsilon_B}{\pi_B-\pi_{AB}}+\frac{\pi_B\varepsilon_A\varepsilon_B}{\pi_B-\pi_{AB}}-\frac{\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}-\frac{\pi_{AB}\varepsilon_A\varepsilon_{AB}}{\pi_B-\pi_{AB}}\right]^{-1}$$

$$= RR(B\,|\,A)\left[1+\varepsilon_{AB}-\frac{\pi_A\varepsilon_A}{1-\pi_A}-\frac{\pi_A\varepsilon_A\varepsilon_{AB}}{1-\pi_A}\right]$$
$$\times\left[1-\varepsilon_A-\frac{\pi_B\varepsilon_B}{\pi_B-\pi_{AB}}-\frac{\pi_B\varepsilon_A\varepsilon_B}{\pi_B-\pi_{AB}}+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\frac{\pi_{AB}\varepsilon_A\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\varepsilon_A^2+\frac{\pi_B^2\varepsilon_B^2}{(\pi_B-\pi_{AB})^2}+\frac{\pi_{AB}^2\varepsilon_{AB}^2}{(\pi_B-\pi_{AB})^2}+\frac{2\pi_B\varepsilon_A\varepsilon_B}{\pi_B-\pi_{AB}}-\frac{2\pi_{AB}\varepsilon_A\varepsilon_{AB}}{\pi_B-\pi_{AB}}-\frac{2\pi_B\pi_{AB}\varepsilon_B\varepsilon_{AB}}{(\pi_B-\pi_{AB})^2}+O(\varepsilon^2)\right]$$

$$
\begin{aligned}
&= RR(B\mid A)\Bigg[1+\varepsilon_{AB}-\frac{\pi_A\varepsilon_A}{1-\pi_A}-\frac{\pi_A\varepsilon_A\varepsilon_{AB}}{1-\pi_A}-\varepsilon_A-\varepsilon_A\varepsilon_{AB}+\frac{\pi_A\varepsilon_A^2}{1-\pi_A}-\frac{\pi_B\varepsilon_B}{\pi_B-\pi_{AB}}-\frac{\pi_B\varepsilon_B\varepsilon_{AB}}{\pi_B-\pi_{AB}} \\
&\quad +\frac{\pi_A\pi_B\varepsilon_A\varepsilon_B}{(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{\pi_B\varepsilon_A\varepsilon_B}{\pi_B-\pi_{AB}}+\frac{\pi_{AB}\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\frac{\pi_{AB}\varepsilon_{AB}^2}{\pi_B-\pi_{AB}}-\frac{\pi_A\pi_{AB}\varepsilon_A\varepsilon_{AB}}{(1-\pi_A)(\pi_B-\pi_{AB})} \\
&\quad +\frac{\pi_{AB}\varepsilon_A\varepsilon_{AB}}{\pi_B-\pi_{AB}}+\varepsilon_A^2+\frac{\pi_B^2\varepsilon_B^2}{(\pi_B-\pi_{AB})^2}+\frac{\pi_{AB}^2\varepsilon_{AB}^2}{(\pi_B-\pi_{AB})^2} \\
&\quad +\frac{2\pi_B\varepsilon_A\varepsilon_B}{\pi_B-\pi_{AB}}-\frac{2\pi_{AB}\varepsilon_A\varepsilon_{AB}}{\pi_B-\pi_{AB}}-\frac{2\pi_B\pi_{AB}\varepsilon_B\varepsilon_{AB}}{(\pi_B-\pi_{AB})^2}+O(\varepsilon^2)\Bigg] \\
&= RR(B\mid A)\Bigg[1+\left(1+\frac{\pi_{AB}}{\pi_B-\pi_{AB}}\right)\varepsilon_{AB}-\left(1+\frac{\pi_A}{1-\pi_A}\right)\varepsilon_A-\frac{\pi_B\varepsilon_B}{\pi_B-\pi_{AB}}+\left(1+\frac{\pi_A}{1-\pi_A}\right)\varepsilon_A^2+\frac{\pi_B^2\varepsilon_B^2}{(\pi_B-\pi_{AB})^2} \\
&\quad +\frac{\pi_{AB}}{\pi_B-\pi_{AB}}\left(1+\frac{\pi_{AB}}{\pi_B-\pi_{AB}}\right)\varepsilon_{AB}^2+\left(\frac{\pi_A\pi_B}{(1-\pi_A)(\pi_B-\pi_{AB})}+\frac{2\pi_B}{\pi_B-\pi_{AB}}-\frac{\pi_B}{\pi_B-\pi_{AB}}\right)\varepsilon_A\varepsilon_B \\
&\quad +\left(\frac{\pi_{AB}}{\pi_B-\pi_{AB}}-\frac{2\pi_{AB}}{\pi_B-\pi_{AB}}-\frac{\pi_A\pi_{AB}}{(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{\pi_A}{1-\pi_A}-1\right)\varepsilon_A\varepsilon_{AB} \\
&\quad -\left(\frac{2\pi_B\pi_{AB}}{(\pi_B-\pi_{AB})^2}+\frac{\pi_B}{\pi_B-\pi_{AB}}\right)\varepsilon_B\varepsilon_{AB}+O(\varepsilon^2)\Bigg] \\
&= RR(B\mid A)\Bigg[1+\frac{\pi_B\varepsilon_{AB}}{\pi_B-\pi_{AB}}-\frac{\varepsilon_A}{1-\pi_A}-\frac{\pi_B\varepsilon_B}{\pi_B-\pi_{AB}}+\frac{\varepsilon_A^2}{1-\pi_A}+\frac{\pi_B^2\varepsilon_B^2}{(\pi_B-\pi_{AB})^2}+\frac{\pi_B\pi_{AB}\varepsilon_{AB}^2}{(\pi_B-\pi_{AB})^2} \\
&\quad +\frac{\pi_B\varepsilon_A\varepsilon_B}{(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{\pi_B\varepsilon_A\varepsilon_{AB}}{(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{\pi_B(\pi_B+\pi_{AB})}{(\pi_B-\pi_{AB})^2}\varepsilon_B\varepsilon_{AB}+O(\varepsilon^2)\Bigg]
\end{aligned}
$$

By the definition of bias, we have

$$
\begin{aligned}
B\left(\hat{RR}(B\mid A)\right) &= E\left(\hat{RR}(B\mid A)\right)-RR(B\mid A) \\
&= RR(B\mid A)\Bigg[\frac{E(\varepsilon_A^2)}{1-\pi_A}+\frac{\pi_B^2}{(\pi_B-\pi_{AB})^2}E(\varepsilon_B^2)+\frac{\pi_B\pi_{AB}}{(\pi_B-\pi_{AB})^2}E(\varepsilon_{AB}^2) \\
&\qquad +\frac{\pi_B E(\varepsilon_A\varepsilon_B)}{(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{\pi_B E(\varepsilon_A\varepsilon_{AB})}{(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{\pi_B(\pi_B+\pi_{AB})E(\varepsilon_B\varepsilon_{AB})}{(\pi_B-\pi_{AB})^2}+O(n^{-1})\Bigg] \\
&= RR(B\mid A)\Bigg[\frac{V(\hat{\pi}_A)}{\pi_A^2(1-\pi_A)}+\frac{V(\hat{\pi}_B)}{(\pi_B-\pi_{AB})^2}+\frac{\pi_B V(\hat{\pi}_{AB})}{\pi_{AB}(\pi_B-\pi_{AB})^2}+\frac{Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A(1-\pi_A)(\pi_B-\pi_{AB})}
\end{aligned}
$$

$$-\frac{\pi_B Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{(\pi_B+\pi_{AB})Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_{AB}(\pi_B-\pi_{AB})^2}\Bigg]$$

which proves the theorem.

Note that $B\left\{\hat{RR}(B\mid A)\right\}\to 0$ as $n\to\infty$, thus the estimator $\hat{RR}(B\mid A)$ is a consistent estimator of the relative risk $RR(B\mid A)$

Theorem 4.2. The mean squared error of the estimator $\hat{RR}(B\mid A)$, to the first order of approximation, is given by

$$MSE\left(\hat{RR}(B\mid A)\right)$$

$$=\{RR(B\mid A)\}^2\left[\frac{\pi_B^2 V(\hat{\pi}_{AB})}{\pi_{AB}^2(\pi_B-\pi_{AB})^2}+\frac{V(\hat{\pi}_A)}{\pi_A^2(1-\pi_A)^2}+\frac{V(\hat{\pi}_B)}{(\pi_B-\pi_{AB})^2}-\frac{2\pi_B Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}(1-\pi_A)(\pi_B-\pi_{AB})}\right.$$

$$\left.-\frac{2\pi_B Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_{AB}(\pi_B-\pi_{AB})^2}+\frac{2Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A(1-\pi_A)(\pi_B-\pi_{AB})}\right] \tag{4.4}$$

Proof. By the definition of mean squared error, we have

$$MSE\left(\hat{RR}(B\mid A)\right)=E\left(\hat{RR}(B\mid A)-RR(B\mid A)\right)^2$$

$$=\{RR(B\mid A)\}^2 E\left[\frac{\pi_B\varepsilon_{AB}}{\pi_B-\pi_{AB}}-\frac{\varepsilon_A}{1-\pi_A}-\frac{\pi_B\varepsilon_B}{\pi_B-\pi_{AB}}\right]^2$$

$$=\{RR(B\mid A)\}^2 E\left[\frac{\pi_B^2\varepsilon_{AB}^2}{(\pi_B-\pi_{AB})^2}+\frac{\varepsilon_A^2}{(1-\pi_A)}+\frac{\pi_B^2\varepsilon_B^2}{(\pi_B-\pi_{AB})^2}-\frac{2\pi_B\varepsilon_A\varepsilon_{AB}}{(1-\pi_A)(\pi_B-\pi_{AB})}\right.$$

$$\left.-\frac{2\pi_B^2\varepsilon_B\varepsilon_{AB}}{(\pi_B-\pi_{AB})^2}+\frac{2\pi_B\varepsilon_A\varepsilon_B}{(1-\pi_A)(\pi_B-\pi_{AB})}+O(\varepsilon^2)\right]$$

$$=\{RR(B\mid A)\}^2\left[\frac{\pi_B^2 E(\varepsilon_{AB}^2)}{(\pi_B-\pi_{AB})^2}+\frac{E(\varepsilon_A^2)}{(1-\pi_A)^2}+\frac{\pi_B^2 E(\varepsilon_B^2)}{(\pi_B-\pi_{AB})^2}-\frac{2\pi_B E(\varepsilon_A\varepsilon_{AB})}{(1-\pi_A)(\pi_B-\pi_{AB})}\right.$$

$$- \frac{2\pi_B^2 E(\varepsilon_B \varepsilon_{AB})}{(\pi_B - \pi_{AB})^2} + \frac{2\pi_B E(\varepsilon_A \varepsilon_B)}{(1-\pi_A)(\pi_B - \pi_{AB})} + O(n^{-1}) \Bigg]$$

$$= \{RR(B \mid A)\}^2 \Bigg[\frac{\pi_B^2 V(\hat{\pi}_{AB})}{\pi_{AB}^2 (\pi_B - \pi_{AB})^2} + \frac{V(\hat{\pi}_A)}{\pi_A^2 (1-\pi_A)^2} + \frac{V(\hat{\pi}_B)}{(\pi_B - \pi_{AB})^2}$$

$$- \frac{2\pi_B Cov(\hat{\pi}_A, \hat{\pi}_{AB})}{\pi_A \pi_{AB} (1-\pi_A)(\pi_B - \pi_{AB})} - \frac{2\pi_B Cov(\hat{\pi}_B, \hat{\pi}_{AB})}{\pi_{AB} (\pi_B - \pi_{AB})^2} + \frac{2Cov(\hat{\pi}_A, \hat{\pi}_B)}{\pi_A (1-\pi_A)(\pi_B - \pi_{AB})} \Bigg]$$

which proves the theorem.

4.2 ESTIMATION OF RELATIVE RISK USING CROSSED MODEL

We define an estimator of the relative risk as:

$$\hat{RR}^*(B|A) = \frac{\hat{\pi}_{AB}^* (1 - \hat{\pi}_A^*)}{\hat{\pi}_A^* (\hat{\pi}_B^* - \hat{\pi}_{AB}^*)} \tag{4.5}$$

Now we have the following theorems:

Theorem 4.3. The bias in the estimator $\hat{RR}^*(B \mid A)$, to the first order of approximation, is given by

$$B\left(\hat{RR}^*(B \mid A) \right) = RR(B \mid A) \Bigg[\frac{V(\hat{\pi}_A^*)}{\pi_A^2 (1-\pi_A)} + \frac{V(\hat{\pi}_B^*)}{(\pi_B - \pi_{AB})^2} + \frac{\pi_B V(\hat{\pi}_{AB}^*)}{\pi_{AB} (\pi_B - \pi_{AB})^2}$$

$$+ \frac{Cov(\hat{\pi}_A^*, \hat{\pi}_B^*)}{\pi_A (1-\pi_A)(\pi_B - \pi_{AB})} - \frac{\pi_B Cov(\hat{\pi}_A^*, \hat{\pi}_{AB}^*)}{\pi_A \pi_{AB} (1-\pi_A)(\pi_B - \pi_{AB})} - \frac{(\pi_B + \pi_{AB}) Cov(\hat{\pi}_B^*, \hat{\pi}_{AB}^*)}{\pi_{AB} (\pi_B - \pi_{AB})^2} \Bigg] \tag{4.6}$$

Proof. It follows from previous section.

Note that $B\left\{ \hat{RR}^*(B \mid A) \right\} \to 0$ as $n \to \infty$, thus the estimator $\hat{RR}^*(B \mid A)$ is a consistent estimator of the relative risk $RR(B \mid A)$.

Theorem 4.4. The mean squared error of the estimator $\hat{RR}^*(B \mid A)$, to the first order of approximation, is given by

$$MSE\left(\hat{RR}^*(B \mid A) \right) = \{RR(B \mid A)\}^2 \Bigg[\frac{\pi_B^2 V(\hat{\pi}_{AB}^*)}{\pi_{AB}^2 (\pi_B - \pi_{AB})^2} + \frac{V(\hat{\pi}_A^*)}{\pi_A^2 (1-\pi_A)^2} + \frac{V(\hat{\pi}_B^*)}{(\pi_B - \pi_{AB})^2}$$

$$-\frac{2\pi_B Cov\left(\hat{\pi}_A^*, \hat{\pi}_{AB}^*\right)}{\pi_A\pi_{AB}(1-\pi_A)(\pi_B-\pi_{AB})}-\frac{2\pi_B Cov\left(\hat{\pi}_B^*, \hat{\pi}_{AB}^*\right)}{\pi_{AB}(\pi_B-\pi_{AB})^2}+\frac{2Cov\left(\hat{\pi}_A^*, \hat{\pi}_B^*\right)}{\pi_A(1-\pi_A)(\pi_B-\pi_{AB})}\Bigg] \qquad (4.7)$$

Proof. The proof proceeds the same as in the previous section.

5 ESTIMATION OF CORRELATION COEFFICIENT BETWEEN TWO SENSITIVE CHARACTERISTICS

In this section, we consider the problem of estimating the correlation coefficient between the two sensitive characteristics A and B defined as:

$$\rho_{AB}=\frac{\pi_{AB}-\pi_A\pi_B}{\sqrt{\pi_A(1-\pi_A)}\sqrt{\pi_B(1-\pi_B)}} \qquad (5.1)$$

5.1. ESTIMATION OF CORRELATION COEFFICIENT WITH THE SIMPLE MODEL

We consider an estimator of the correlation coefficient ρ_{AB} as:

$$\hat{\rho}_{AB}=\frac{\hat{\pi}_{AB}-\hat{\pi}_A\hat{\pi}_B}{\sqrt{\hat{\pi}_A(1-\hat{\pi}_A)}\sqrt{\hat{\pi}_B(1-\hat{\pi}_B)}} \qquad (5.2)$$

The estimator $\hat{\rho}_{AB}$ in terms of ε_A, ε_B and ε_{AB} can be approximated as:

$$\hat{\rho}_{AB}=\frac{\pi_{AB}(1+\varepsilon_{AB})-\pi_A(1+\varepsilon_A)\pi_B(1+\varepsilon_B)}{\sqrt{\pi_A(1+\varepsilon_A)((1-\pi_A(1+\varepsilon_A))}\sqrt{\pi_B(1+\varepsilon_B)((1-\pi_B(1+\varepsilon_B))}}$$

$$=\frac{\pi_{AB}+\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(1+\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\sqrt{(\pi_A+\pi_A\varepsilon_A)((1-\pi_A)-\pi_A\varepsilon_A))}\sqrt{(\pi_B+\pi_B\varepsilon_B)((1-\pi_B)-\pi_B\varepsilon_B))}}$$

$$=\frac{(\pi_{AB}-\pi_A\pi_B)+\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\sqrt{\pi_A(1-\pi_A)+\pi_A(1-\pi_A)\varepsilon_A-\pi_A^2\varepsilon_A-\pi_A^2\varepsilon_A^2}\sqrt{\pi_B(1-\pi_B)+\pi_B(1-\pi_B)\varepsilon_B-\pi_B^2\varepsilon_B-\pi_B^2\varepsilon_B^2}}$$

$$=\frac{(\pi_{AB}-\pi_A\pi_B)\left[1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right]}{\sqrt{\pi_A(1-\pi_A)\left\{1+\varepsilon_A-\frac{\pi_A^2\varepsilon_A+\pi_A^2\varepsilon_A^2}{\pi_A(1-\pi_A)}\right\}}\sqrt{\pi_B(1-\pi_B)\left\{1+\varepsilon_B-\frac{\pi_B^2\varepsilon_B+\pi_B^2\varepsilon_B^2}{\pi_B(1-\pi_B)}\right\}}}$$

$$=\frac{(\pi_{AB}-\pi_A\pi_B)}{\sqrt{\pi_A(1-\pi_A)}\sqrt{\pi_B(1-\pi_B)}}\left[\left\{1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right\}\right.$$

$$\left. \times\left\{1+\varepsilon_A-\frac{\pi_A^2\varepsilon_A+\pi_A^2\varepsilon_A^2}{\pi_A(1-\pi_A)}\right\}^{-\frac{1}{2}}\left\{1+\varepsilon_B-\frac{\pi_B^2\varepsilon_B+\pi_B^2\varepsilon_B^2}{\pi_B(1-\pi_B)}\right\}^{-\frac{1}{2}}\right]$$

$$=\rho_{AB}\left[1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right]$$

$$\times\left[1-\frac{\varepsilon_A}{2}+\frac{\pi_A\varepsilon_A+\pi_A\varepsilon_A^2}{2(1-\pi_A)}+\frac{3}{8}\left(\varepsilon_A-\frac{\pi_A\varepsilon_A+\pi_A\varepsilon_A^2}{(1-\pi_A)}\right)^2+\cdots\right]$$
$$\times\left[1-\frac{\varepsilon_B}{2}+\frac{\pi_B\varepsilon_B+\pi_B\varepsilon_B^2}{2(1-\pi_B)}+\frac{3}{8}\left(\varepsilon_B-\frac{\pi_B\varepsilon_B+\pi_B\varepsilon_B^2}{(1-\pi_B)}\right)^2+\cdots\right]$$

where we used binomial expansion $(1+x)^\alpha=1+\alpha x+\frac{\alpha(\alpha-1)}{2}x^2+.....$ by assuming $|x|<1$ that is $\left|\varepsilon_A-\frac{\pi_A^2\varepsilon_A+\pi_A^2\varepsilon_A^2}{\pi_A(1-\pi_A)}\right|<1$ and $\left|\varepsilon_B-\frac{\pi_B^2\varepsilon_B+\pi_B^2\varepsilon_B^2}{\pi_B(1-\pi_B)}\right|<1$, so we have

$$\hat{\rho}_{AB}\approx\rho_{AB}\left[1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right]$$
$$\times\left[1-\frac{\varepsilon_A}{2}+\frac{\pi_A\varepsilon_A+\pi_A\varepsilon_A^2}{2(1-\pi_A)}+\frac{3}{8}\frac{(1-2\pi_A)^2\varepsilon_A^2}{(1-\pi_A)^2}+O(\varepsilon^2)\right]$$
$$\times\left[1-\frac{\varepsilon_B}{2}+\frac{\pi_B\varepsilon_B+\pi_B\varepsilon_B^2}{2(1-\pi_B)}+\frac{3}{8}\frac{(1-2\pi_B)^2\varepsilon_B^2}{(1-\pi_B)^2}+O(\varepsilon^2)\right]$$

$$=\rho_{AB}\left[1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right]$$
$$\times\left[1-\frac{\varepsilon_A}{2}+\frac{\pi_A\varepsilon_A+\pi_A\varepsilon_A^2}{2(1-\pi_A)}+\frac{3}{8}\frac{(1-2\pi_A)^2\varepsilon_A^2}{(1-\pi_A)^2}-\frac{\varepsilon_B}{2}+\frac{\varepsilon_A\varepsilon_B}{4}-\frac{\pi_A\varepsilon_A\varepsilon_B}{4(1-\pi_A)}\right.$$
$$\left.+\frac{\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)}+\frac{\pi_B\varepsilon_B+\pi_B\varepsilon_B^2}{2(1-\pi_B)}-\frac{\pi_B\varepsilon_A\varepsilon_B}{4(1-\pi_B)}+\frac{3}{8}\frac{(1-2\pi_B)^2\varepsilon_B^2}{(1-\pi_B)^2}+O(\varepsilon^2)\right]$$

$$=\rho_{AB}\left[1+\frac{\pi_{AB}\varepsilon_{AB}-\pi_A\pi_B(\varepsilon_A+\varepsilon_B+\varepsilon_A\varepsilon_B)}{\pi_{AB}-\pi_A\pi_B}\right]$$
$$\times\left[1-\frac{(1-2\pi_A)}{2(1-\pi_A)}\varepsilon_A-\frac{(1-2\pi_B)}{2(1-\pi_B)}\varepsilon_B+\left\{\frac{\pi_A}{2(1-\pi_A)}+\frac{3(1-2\pi_A)^2}{8(1-\pi_A)^2}\right\}\varepsilon_A^2\right.$$

$$+\left\{\frac{\pi_B}{2(1-\pi_B)}+\frac{3(1-2\pi_B)^2}{8(1-\pi_B)^2}\right\}\varepsilon_B^2+\left\{\frac{1}{4}-\frac{\pi_A}{4(1-\pi_A)}-\frac{\pi_B}{4(1-\pi_B)}+\frac{\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)}\right\}\varepsilon_A\varepsilon_B+O(\varepsilon^2)\Bigg]$$

$$=\rho_{AB}\left[1+\frac{\pi_{AB}}{\pi_{AB}-\pi_A\pi_B}\varepsilon_{AB}-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}\varepsilon_A-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}\varepsilon_B-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}\varepsilon_A\varepsilon_B\right]$$
$$\times\Bigg[1-\frac{(1-2\pi_A)}{2(1-\pi_A)}\varepsilon_A-\frac{(1-2\pi_B)}{2(1-\pi_B)}\varepsilon_B+\left\{\frac{\pi_A}{2(1-\pi_A)}+\frac{3(1-2\pi_A)^2}{8(1-\pi_A)^2}\right\}\varepsilon_A^2$$
$$+\left\{\frac{\pi_B}{2(1-\pi_B)}+\frac{3(1-2\pi_B)^2}{8(1-\pi_B)^2}\right\}\varepsilon_B^2+\left\{\frac{1-2\pi_A-2\pi_B+4\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)}\right\}\varepsilon_A\varepsilon_B+O(\varepsilon^2)\Bigg]$$

where $O(\varepsilon^2)$ denote higher order terms of ε.

$$\hat{\rho}_{AB}\approx\rho_{AB}\Bigg[1+\frac{\pi_{AB}}{\pi_{AB}-\pi_A\pi_B}\varepsilon_{AB}-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}\varepsilon_A-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}\varepsilon_B-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}\varepsilon_A\varepsilon_B$$
$$-\frac{(1-2\pi_A)}{2(1-\pi_A)}\varepsilon_A-\frac{\pi_{AB}(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_A\varepsilon_{AB}+\frac{\pi_A\pi_B(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_A^2$$
$$+\frac{\pi_A\pi_B(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_A\varepsilon_B-\frac{(1-2\pi_B)}{2(1-\pi_B)}\varepsilon_B-\frac{\pi_{AB}(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_B\varepsilon_{AB}$$
$$+\frac{\pi_A\pi_B(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_A\varepsilon_B+\frac{\pi_A\pi_B(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_B^2$$
$$+\left\{\frac{\pi_A}{2(1-\pi_A)}+\frac{3(1-2\pi_A)^2}{8(1-\pi_A)^2}\right\}\varepsilon_A^2+\left\{\frac{\pi_B}{2(1-\pi_B)}+\frac{3(1-2\pi_B)^2}{8(1-\pi_B)^2}\right\}\varepsilon_B^2$$
$$+\left\{\frac{1-2\pi_A-2\pi_B+4\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)}\right\}\varepsilon_A\varepsilon_B+O(\varepsilon^2)\Bigg]$$

$$=\rho_{AB}\Bigg[1+\frac{\pi_{AB}}{\pi_{AB}-\pi_A\pi_B}\varepsilon_{AB}-\left\{\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}+\frac{(1-2\pi_A)}{2(1-\pi_A)}\right\}\varepsilon_A-\left\{\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}+\frac{(1-2\pi_B)}{2(1-\pi_B)}\right\}\varepsilon_B$$
$$+\left\{\frac{\pi_A\pi_B(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}+\frac{\pi_A}{2(1-\pi_A)}+\frac{3(1-2\pi_A)^2}{8(1-\pi_A)^2}\right\}\varepsilon_A^2$$
$$+\left\{\frac{\pi_A\pi_B(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}+\frac{\pi_B}{2(1-\pi_B)}+\frac{3(1-2\pi_B)^2}{8(1-\pi_B)^2}\right\}\varepsilon_B^2$$
$$+\left\{\frac{1-2\pi_A-2\pi_B+4\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)}-\frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B}+\frac{\pi_A\pi_B}{2(\pi_{AB}-\pi_A\pi_B)}\left(\frac{1-2\pi_B}{1-\pi_B}+\frac{1-2\pi_A}{1-\pi_A}\right)\right\}\varepsilon_A\varepsilon_B$$
$$-\frac{\pi_{AB}(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_A\varepsilon_{AB}-\frac{\pi_{AB}(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}\varepsilon_B\varepsilon_{AB}+O(\varepsilon^2)\Bigg]$$

or, equivalently

$$\hat{\rho}_{AB} \approx \rho_{AB}\left[1 + F_1\varepsilon_{AB} - F_2\varepsilon_A - F_3\varepsilon_B + F_4\varepsilon_A^2 + F_5\varepsilon_B^2 + F_6\varepsilon_A\varepsilon_B - F_7\varepsilon_A\varepsilon_{AB} - F_8\varepsilon_B\varepsilon_{AB} + O(\varepsilon^2)\right] \quad (5.3)$$

where

$$F_1 = \frac{\pi_{AB}}{\pi_{AB} - \pi_A\pi_B}, \quad F_2 = \frac{\pi_A\pi_B}{\pi_{AB} - \pi_A\pi_B} + \frac{(1-2\pi_A)}{2(1-\pi_A)}, \quad F_3 = \frac{\pi_A\pi_B}{\pi_{AB} - \pi_A\pi_B} + \frac{(1-2\pi_B)}{2(1-\pi_B)},$$

$$F_4 = \frac{\pi_A\pi_B(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB} - \pi_A\pi_B)} + \frac{\pi_A}{2(1-\pi_A)} + \frac{3(1-2\pi_A)^2}{8(1-\pi_A)^2},$$

$$F_5 = \frac{\pi_A\pi_B(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB} - \pi_A\pi_B)} + \frac{\pi_B}{2(1-\pi_B)} + \frac{3(1-2\pi_B)^2}{8(1-\pi_B)^2},$$

$$F_6 = \frac{1-2\pi_A-2\pi_B+4\pi_A\pi_B}{4(1-\pi_A)(1-\pi_B)} - \frac{\pi_A\pi_B}{\pi_{AB}-\pi_A\pi_B} + \frac{\pi_A\pi_B}{2(\pi_{AB}-\pi_A\pi_B)}\left\{\frac{1-2\pi_B}{1-\pi_B} + \frac{1-2\pi_A}{1-\pi_A}\right\},$$

$$F_7 = \frac{\pi_{AB}(1-2\pi_A)}{2(1-\pi_A)(\pi_{AB}-\pi_A\pi_B)}, \quad \text{and} \quad F_8 = \frac{\pi_{AB}(1-2\pi_B)}{2(1-\pi_B)(\pi_{AB}-\pi_A\pi_B)}$$

are constants.

Now we have the following theorems:

Theorem 5.1. The bias in the estimator $\hat{\rho}_{AB}$, to the first order of approximation, is:

$$B(\hat{\rho}_{AB}) = \rho_{AB}\left[F_4\frac{V(\hat{\pi}_A)}{\pi_A^2} + F_5\frac{V(\hat{\pi}_B)}{\pi_B^2} + F_6\frac{Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A\pi_B} - F_7\frac{Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}} - F_8\frac{Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_B\pi_{AB}}\right] \quad (5.4)$$

Proof. By the definition of bias and using the approximation in (5.3), we have

$$B(\hat{\rho}_{AB}) = E(\hat{\rho}_{AB}) - \rho_{AB}$$

$$= \rho_{AB}\left[F_4E(\varepsilon_A^2) + F_5E(\varepsilon_B^2) + F_6E(\varepsilon_A\varepsilon_B) - F_7E(\varepsilon_A\varepsilon_{AB}) - F_8E(\varepsilon_B\varepsilon_{AB})\right]$$

$$= \rho_{AB}\left[F_4\frac{V(\hat{\pi}_A)}{\pi_A^2} + F_5\frac{V(\hat{\pi}_B)}{\pi_B^2} + F_6\frac{Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A\pi_B} - F_7\frac{Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}} - F_8\frac{Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_B\pi_{AB}}\right]$$

which proves the theorem.

Note that $B\{\hat{\rho}_{AB}\} \to 0$ as $n \to \infty$, thus the estimator $\hat{\rho}_{AB}$ is a consistent estimator of the correlation coefficient ρ_{AB}.

Theorem 5.2. The mean squared error of the estimator $\hat{\rho}_{AB}$, to the first order of approximation, is given by

$$MSE(\hat{\rho}_{AB}) = \rho_{AB}^2\left[F_1^2\frac{V(\hat{\pi}_{AB})}{\pi_{AB}^2} + F_2^2\frac{V(\hat{\pi}_A)}{\pi_A^2} + F_3^2\frac{V(\hat{\pi}_B)}{\pi_B^2} - \frac{2F_1F_2Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}} - \frac{2F_1F_3Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_B\pi_{AB}} + \frac{2F_2F_3Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A\pi_B}\right] \tag{5.5}$$

Proof. By the definition of mean squared error and using the approximation (5.3), we have

$$\begin{aligned}
MSE(\hat{\rho}_{AB}) &= E[\hat{\rho}_{AB} - \rho_{AB}]^2 \\
&\approx \rho_{AB}^2 E[F_1\varepsilon_{AB} - F_2\varepsilon_A - F_3\varepsilon_B + O(\varepsilon)]^2 \\
&= \rho_{AB}^2 E\left[F_1^2\varepsilon_{AB}^2 + F_2^2\varepsilon_A^2 + F_3^2\varepsilon_B^2 - 2F_1F_2\varepsilon_A\varepsilon_{AB} - 2F_1F_3\varepsilon_B\varepsilon_{AB} + 2F_2F_3\varepsilon_A\varepsilon_B\right] \\
&= \rho_{AB}^2\left[F_1^2E(\varepsilon_{AB}^2) + F_2^2E(\varepsilon_A^2) + F_3^2E(\varepsilon_B^2) - 2F_1F_2E(\varepsilon_A\varepsilon_{AB}) - 2F_1F_3E(\varepsilon_B\varepsilon_{AB}) + 2F_2F_3E(\varepsilon_A\varepsilon_B)\right] \\
&= \rho_{AB}^2\left[F_1^2\frac{V(\hat{\pi}_{AB})}{\pi_{AB}^2} + F_2^2\frac{V(\hat{\pi}_A)}{\pi_A^2} + F_3^2\frac{V(\hat{\pi}_B)}{\pi_B^2} - \frac{2F_1F_2Cov(\hat{\pi}_A,\hat{\pi}_{AB})}{\pi_A\pi_{AB}} - \frac{2F_1F_3Cov(\hat{\pi}_B,\hat{\pi}_{AB})}{\pi_B\pi_{AB}} + \frac{2F_2F_3Cov(\hat{\pi}_A,\hat{\pi}_B)}{\pi_A\pi_B}\right]
\end{aligned}$$

which proves the theorem.

5.2. ESTIMATION OF CORRELATION COEFFICIENT WITH THE CROSSED MODEL

We consider an estimator of the correlation coefficient ρ_{AB} as:

$$\hat{\rho}_{AB}^* = \frac{\hat{\pi}_{AB}^* - \hat{\pi}_A^*\hat{\pi}_B^*}{\sqrt{\hat{\pi}_A^*(1-\hat{\pi}_A^*)}\sqrt{\hat{\pi}_B^*(1-\hat{\pi}_B^*)}} \tag{5.6}$$

Now we have the following theorems:

Theorem 5.3. The bias in the estimator $\hat{\rho}_{AB}^*$, to the first order of approximation, is:

$$B(\hat{\rho}_{AB}^*) = \rho_{AB}\left[F_4\frac{V(\hat{\pi}_A^*)}{\pi_A^2} + F_5\frac{V(\hat{\pi}_B^*)}{\pi_B^2} + F_6\frac{Cov(\hat{\pi}_A^*,\hat{\pi}_B^*)}{\pi_A\pi_B} - F_7\frac{Cov(\hat{\pi}_A^*,\hat{\pi}_{AB}^*)}{\pi_A\pi_{AB}} - F_8\frac{Cov(\hat{\pi}_B^*,\hat{\pi}_{AB}^*)}{\pi_B\pi_{AB}}\right] \tag{5.7}$$

Proof. Obvious from the previous section.

Note that the $B\left\{\hat{\rho}^*_{AB}\right\} \to 0$ as $n \to \infty$, thus the estimator $\hat{\rho}^*_{AB}$ is a consistent estimator of the correlation coefficient ρ_{AB}.

Theorem 5.3. The mean squared error of the estimator $\hat{\rho}^*_{AB}$, to the first order of approximation, is given by

$$MSE(\hat{\rho}^*_{AB}) = \rho^2_{AB}\left[F_1^2 \frac{V\left(\hat{\pi}^*_{AB}\right)}{\pi^2_{AB}} + F_2^2 \frac{V\left(\hat{\pi}^*_{A}\right)}{\pi^2_{A}} + F_3^2 \frac{V\left(\hat{\pi}^*_{B}\right)}{\pi^2_{B}} - \frac{2F_1F_2Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_{AB}\right)}{\pi_A\pi_{AB}} - \frac{2F_1F_3Cov\left(\hat{\pi}^*_B, \hat{\pi}^*_{AB}\right)}{\pi_B\pi_{AB}} + \frac{2F_2F_3Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_{B}\right)}{\pi_A\pi_{B}}\right] \tag{5.8}$$

Proof. Obvious from the previous section.

6 ESTIMATION OF PROPORTION OF PERSONS POSSESSING AT LEAST ONE OF THE CHARACTERISTICS

Here we consider the problem of estimation of proportion of those persons in the population who possess at least one of the characteristics A or B defined as:

$$\pi_{A\cup B} = \pi_A + \pi_B - \pi_{AB} \tag{6.1}$$

6.1. AT LEAST ONE CHARACTERISTICS WITH SIMPLE MODEL

Then we have the following theorem:

Theorem 6.1. An unbiased estimator of $\pi_{A\cup B}$ is given by

$$\hat{\pi}_{A\cup B} = \hat{\pi}_A + \hat{\pi}_B - \hat{\pi}_{AB} \tag{6.2}$$

Proof. By taking expected value on both sides of (6.2), we have

$$E\left(\hat{\pi}_{A\cup B}\right) = E\left(\hat{\pi}_A + \hat{\pi}_B - \hat{\pi}_{AB}\right) = E\left(\hat{\pi}_A\right) + E\left(\hat{\pi}_B\right) - E\left(\hat{\pi}_{AB}\right) = \pi_A + \pi_B - \pi_{AB} = \pi_{A\cup B}$$

which proves the unbiased property of the estimator.

Theorem 6.2. The variance of the estimator of $\hat{\pi}_{A\cup B}$ is given by;

$$V\left(\hat{\pi}_{A\cup B}\right) = V\left(\hat{\pi}_A\right) + V\left(\hat{\pi}_B\right) + V\left(\hat{\pi}_{AB}\right) + 2Cov\left(\hat{\pi}_A, \hat{\pi}_B\right) - 2Cov\left(\hat{\pi}_A, \hat{\pi}_{AB}\right) - 2Cov\left(\hat{\pi}_B, \hat{\pi}_{AB}\right) \tag{6.3}$$

Proof. By the definition of variance, we have

$$V\left(\hat{\pi}_{A\cup B}\right) = V\left(\hat{\pi}_A + \hat{\pi}_B - \hat{\pi}_{AB}\right)$$

$$= V(\hat{\pi}_A) + V(\hat{\pi}_B) + V(\hat{\pi}_{AB}) + 2Cov(\hat{\pi}_A, \hat{\pi}_B) - 2Cov(\hat{\pi}_A, \hat{\pi}_{AB}) - 2Cov(\hat{\pi}_B, \hat{\pi}_{AB})$$

which proves the theorem.

6.2. AT LEAST ONE CHARACTERISTICS WITH CROSSED MODEL

Then we have the following theorem:

Theorem 6.3. An unbiased estimator of $\pi_{A\cup B}$ is given by

$$\hat{\pi}^*_{A\cup B} = \hat{\pi}^*_A + \hat{\pi}^*_B - \hat{\pi}^*_{AB} \tag{6.4}$$

Proof. By taking expected value on both sides of (6.4), we have

$$E(\hat{\pi}^*_{A\cup B}) = E(\hat{\pi}^*_A) + E(\hat{\pi}^*_B) - E(\hat{\pi}^*_{AB}) = \pi_A + \pi_B - \pi_{AB} = \pi_{A\cup B}$$

which proves the unbiased property of the estimator.

Theorem 6.4. The variance of the estimator of $\hat{\pi}^*_{A\cup B}$ is given by;

$$V(\hat{\pi}^*_{A\cup B}) = V(\hat{\pi}^*_A) + V(\hat{\pi}^*_B) + V(\hat{\pi}^*_{AB}) + 2Cov(\hat{\pi}^*_A, \hat{\pi}^*_B) - 2Cov(\hat{\pi}^*_A, \hat{\pi}^*_{AB}) - 2Cov(\hat{\pi}^*_B, \hat{\pi}^*_{AB}) \tag{6.5}$$

Proof. By the definition of variance, we have

$$V(\hat{\pi}^*_{A\cup B}) = V(\hat{\pi}^*_A + \hat{\pi}^*_B - \hat{\pi}^*_{AB})$$

$$= V(\hat{\pi}^*_A) + V(\hat{\pi}^*_B) + V(\hat{\pi}^*_{AB}) + 2Cov(\hat{\pi}^*_A, \hat{\pi}^*_B) - 2Cov(\hat{\pi}^*_A, \hat{\pi}^*_{AB}) - 2Cov(\hat{\pi}^*_B, \hat{\pi}^*_{AB})$$

which proves the theorem.

7. DIFFERENCE BETWEEN TWO PROPORTIONS

We consider the problem of estimation of the difference between two proportions defined as:

$$\pi_d = \pi_A - \pi_B \tag{7.1}$$

7.1 DIFFERENCE WITH SIMPLE MODEL

We consider an unbiased estimator π_d as:

$$\hat{\pi}_d = \hat{\pi}_A - \hat{\pi}_B \tag{7.2}$$

Then we have the following theorem:

Theorem 7.3. The variance of the estimator $\hat{\pi}_d$ is given by

$$V(\hat{\pi}_d) = V(\hat{\pi}_A) + V(\hat{\pi}_B) - 2Cov(\hat{\pi}_A, \hat{\pi}_B) \tag{7.3}$$

Proof. It follows from the definition of variance.

7.2 DIFFERENCE WITH CROSSED MODEL

We consider an unbiased estimator π_d as:

$$\hat{\pi}_d^* = \hat{\pi}_A^* - \hat{\pi}_B^* \tag{7.4}$$

Then we have the following theorem:

Theorem 7.4. The variance of the estimator $\hat{\pi}_d^*$ is given by

$$V(\hat{\pi}_d^*) = V(\hat{\pi}_A^*) + V(\hat{\pi}_B^*) - 2Cov(\hat{\pi}_A^*, \hat{\pi}_B^*) \tag{7.5}$$

Proof. It follows from the definition of variance.

8 REGRESSION TYPE ESTIMATOR FOR SINGLE PROPORTION

In this section we suggest a new estimator for estimating the proportion of one sensitive variable when the proportion of the second sensitive variable is known.

8.1 REGRESSION TYPE ESTIMATOR FOR SINGLE PROPORTION WITH SIMPLE MODEL

Here we first define a new difference estimator of the population proportion π_A by assuming the population proportion π_B is known as follows:

$$\hat{\pi}_{A(d)} = \hat{\pi}_A + \beta(\pi_B - \hat{\pi}_B) \tag{8.1}$$

where β is a known constant. Then we have the following theorems:

Theorem 8.1. The difference estimator $\hat{\pi}_{A(d)}$ is an unbiased estimator of π_A.
Proof. Taking expected value on both sides, we have

$$E\left[\hat{\pi}_{A(d)}\right] = E\left[\hat{\pi}_A + \beta(\pi_B - \hat{\pi}_B)\right] = E(\hat{\pi}_A) + \beta\left[\pi_B - E(\hat{\pi}_B)\right] = \pi_A + \beta(\pi_B - \pi_B) = \pi_A$$

which proves the theorem.

Theorem 8.2. The minimum variance of the difference estimator $\hat{\pi}_{A(d)}$ is given by

$$V\left[\hat{\pi}_{A(d)}\right] = V(\hat{\pi}_A)(1 - \rho_{AB}^2) \tag{8.2}$$

Proof. By the definition of variance, we have

$$V\left[\hat{\pi}_{A(d)}\right] = V\left[\hat{\pi}_A + \beta(\pi_B - \hat{\pi}_B)\right] = V(\hat{\pi}_A) + \beta^2 V(\hat{\pi}_B) - 2\beta Cov(\hat{\pi}_A, \hat{\pi}_B) \tag{8.3}$$

The variance will be minimum if

$$\frac{\partial V\left[\hat{\pi}_{A(d)}\right]}{\partial \beta} = 0 \quad \text{or} \quad 2\beta V(\hat{\pi}_B) - 2Cov(\hat{\pi}_A, \hat{\pi}_B) = 0$$

or if

$$\beta = \frac{Cov(\hat{\pi}_A, \hat{\pi}_B)}{V(\hat{\pi}_B)} \tag{8.4}$$

On substituting the value of β from (8.4) into (8.3), the minimum variance of the difference estimator is given by

$$\min V[\hat{\pi}_{A(d)}] = V(\hat{\pi}_A) + \frac{\{Cov(\hat{\pi}_A, \hat{\pi}_B)\}^2}{\{V(\hat{\pi}_B)\}^2} \cdot V(\hat{\pi}_B) - \frac{2\{Cov(\hat{\pi}_A, \hat{\pi}_B)\}^2}{V(\hat{\pi}_B)}$$

$$= V(\hat{\pi}_A) + \frac{\{Cov(\hat{\pi}_A, \hat{\pi}_B)\}^2}{V(\hat{\pi}_B)} - \frac{2\{Cov(\hat{\pi}_A, \hat{\pi}_B)\}^2}{V(\hat{\pi}_B)}$$

$$= V(\hat{\pi}_A)\left[1 - \frac{\{Cov(\hat{\pi}_A, \hat{\pi}_B)\}^2}{V(\hat{\pi}_A)V(\hat{\pi}_B)}\right]$$

$$= V(\hat{\pi}_A)\left[1 - \rho_{AB}^2\right], \text{ where } 0 < \rho_{AB}^2 < 1$$

The estimator $\hat{\pi}_{A(d)}$ is more efficient than $\hat{\pi}_A$ if $V[\hat{\pi}_{A(d)}] < V[\hat{\pi}_A]$

$$V(\hat{\pi}_A)\left[1 - \rho_{AB}^2\right] < V(\hat{\pi}_A)$$

or if

$$0 < \rho_{AB}^2$$

which is always true. Thus the difference estimator $\hat{\pi}_{A(d)}$ is always more efficient than the usual estimator $\hat{\pi}_A$. Thus we conclude that although both characteristics A and B are sensitive in nature, but if the true proportion one of the sensitive character is known (or leaked by some agency) then that information can be used to improve the estimator of proportion of the second sensitive characteristic in the population.

One of the major problems with the difference estimator $\hat{\pi}_{A(d)}$ is that it depends upon the value of an unknown constant β which further depends upon the true parameters of interest.

Thus, we suggest estimating the value of the unknown constant β as

$$\hat{\beta} = \frac{\widehat{Cov}(\hat{\pi}_A, \hat{\pi}_B)}{\hat{V}(\hat{\pi}_B)} \tag{8.5}$$

Then we suggest a linear regression type estimator as

$$\hat{\pi}_{\hat{A}(LR)} = \hat{\pi}_A + \hat{\beta}(\pi_B - \hat{\pi}_B) \tag{8.6}$$

Then it is easy to show that to the first order of approximation, we have

$$MSE\left[\hat{\pi}_{\hat{A}(LR)}\right] \approx V\left[\hat{\pi}_{A(d)}\right]$$

Thus the regression type estimator has the same approximate mean squared error value as the variance of the difference estimator.

Theorem 8.2. The difference estimator $\hat{\pi}^*_{A(d)}$ is an unbiased estimator of π_A.
Proof. Obvious.

Theorem 8.3. The minimum variance of the difference estimator $\hat{\pi}^*_{A(d)}$ is given by

$$V\left[\hat{\pi}^*_{A(d)}\right] = V\left(\hat{\pi}^*_A\right)\left(1-\rho^{*2}_{AB}\right) \tag{8.7}$$

Proof. By the definition of variance, we have

$$V\left[\hat{\pi}^*_{A(d)}\right] = V\left(\hat{\pi}^*_A\right) + \beta^{*2} V\left(\hat{\pi}^*_B\right) - 2\beta^* Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_B\right) \tag{8.8}$$

The variance will be minimum if

$$\beta^* = \frac{Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_B\right)}{V\left(\hat{\pi}^*_B\right)} \tag{8.9}$$

On substituting the value of β^* from (8.9) into (8.8), the minimum variance of the difference estimator is given by

$$\min V\left[\hat{\pi}^*_{A(d)}\right] = V\left(\hat{\pi}^*_A\right) + \frac{\left\{Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_B\right)\right\}^2}{\left\{V\left(\hat{\pi}^*_B\right)\right\}^2} \cdot V\left(\hat{\pi}^*_B\right) - \frac{2\left\{Cov\left(\hat{\pi}^*_A, \hat{\pi}^*_B\right)\right\}^2}{V\left(\hat{\pi}^*_B\right)}$$

$$= V\left(\hat{\pi}^*_A\right)\left[1-\rho^{*2}_{AB}\right] \text{, where } 0 < \rho^{*2}_{AB} < 1$$

The estimator $\hat{\pi}^*_{A(d)}$ is more efficient than $\hat{\pi}^*_A$ if $V\left[\hat{\pi}^*_{A(d)}\right] < V\left[\hat{\pi}^*_A\right]$

or if $0 < \rho^{*2}_{AB}$.

which is always true. Thus the difference estimator $\hat{\pi}^*_{A(d)}$ is always more efficient than the usual estimator $\hat{\pi}^*_A$. Thus we conclude that although both characteristics A and B are sensitive in nature, but if the true proportion one of the sensitive character is known (or leaked by some agency) then that information can be used to improve the estimator of proportion of the second sensitive characteristic in the population.

One of the major problems with the difference estimator $\hat{\pi}^*_{A(d)}$ is that it depends upon the value of an unknown constant β^* which further depends upon the true parameters of interest. Thus, we suggest estimating the value of the unknown constant β^* as

$$\hat{\beta}^{*} = \frac{\hat{Cov}\left(\hat{\pi}_{A}^{*}, \hat{\pi}_{B}^{*}\right)}{\hat{V}\left(\hat{\pi}_{B}^{*}\right)} \tag{8.10}$$

Then we suggest a linear regression type estimator as

$$\hat{\pi}_{A(LR)}^{*} = \hat{\pi}_{A}^{*} + \hat{\beta}^{*}\left(\pi_{B} - \hat{\pi}_{B}^{*}\right) \tag{8.11}$$

As before, it can be shown that

$$MSE(\hat{\pi}^{*}{}_{A(LR)}) \approx V(\hat{\pi}^{*}{}_{A(d)})$$

Thus the regression type estimator has the same approximate mean squared error value as the variance of the difference estimator.

9. RELATIVE EFFICIENCY

We define the percent relative efficiency of the estimator $\hat{OR}^{*}$ with respect to the estimator $\hat{OR}$ as:

$$\text{RE}(1) = \frac{MSE(\hat{OR})}{MSE(\hat{OR}^{*})} \times 100\% \tag{9.1}$$

We define the percent relative efficiency of the estimator $\hat{AR}^{*}(A \mid B)$ with respect to the estimator $\hat{AR}(A \mid B)$ as:

$$\text{RE}(2) = \frac{MSE(\hat{AR}(B \mid A))}{MSE(\hat{AR}^{*}(B \mid A))} \times 100\% \tag{9.2}$$

We define the percent relative efficiency of the estimator $\hat{RR}^{*}(B \mid A)$ with respect to the estimator $\hat{RR}(B \mid A)$ as:

$$\text{RE}(3) = \frac{MSE(\hat{RR}(B \mid A))}{MSE(\hat{RR}^{*}(B \mid A))} \times 100\% \tag{9.3}$$

We define the percent relative efficiency of the estimator $\hat{\rho}_{AB}^{*}$ with respect to the estimator $\hat{\rho}_{AB}$ as:

$$\text{RE}(4) = \frac{MSE(\hat{\rho}_{AB})}{MSE(\hat{\rho}_{AB}^{*})} \times 100\% \tag{9.4}$$

We define the percent relative efficiency of the estimator $\hat{\pi}^*_{A\cup B}$ with respect to the estimator $\hat{\pi}_{A\cup B}$ as:

$$\text{RE}(5) = \frac{V(\hat{\pi}_{A\cup B})}{V(\hat{\pi}^*_{A\cup B})} \times 100\% \tag{9.5}$$

We define the percent relative efficiency of the estimator $\hat{\pi}^*_d$ with respect to the estimator $\hat{\pi}_d$ as:

$$\text{RE}(6) = \frac{V(\hat{\pi}_d)}{V(\hat{\pi}^*_d)} \times 100\% \tag{9.6}$$

We define the percent relative efficiency of the estimator $\hat{\pi}^*_{A(d)}$ with respect to the estimator $\hat{\pi}_{A(d)}$ as:

$$\text{RE}(7) = \frac{V(\hat{\pi}_{A(d)})}{V(\hat{\pi}^*_{A(d)})} \times 100\% \tag{9.7}$$

We wrote FORTRAN codes, given in APPENDIX, to compute the percent relative efficiency values. We used $P = T = 0.7$ which is same choice as in Lee *et al.* (2013). The percent relative efficiency values so obtained for different choices of π_{AB}, π_A and π_B where all the parameters were computable are presented in Table 9.1.

Table 9.1. Percent Relative Efficiency (RE(j), j=1,2,3,4,5,6) values.

π_{AB}	π_A	π_B	RE(1)	RE(2)	RE(3)	RE(4)	RE(5)	RE(6)	RE(7)
0.1	0.2	0.2	1434.8	1221.2	1445.9	1298.1	367.6	527.1	173.5
0.1	0.2	0.3	1466.1	1302.7	1529.9	1366.8	390.4	499.1	154.3
0.1	0.2	0.4	1571.6	1509.4	1618.8	1529.1	424.7	487.8	140.1
0.1	0.2	0.6	1991.7	1922.3	1904.6	1998.3	568.2	504.1	118.9
0.1	0.2	0.7	2393.1	2030.6	2124.8	2350.4	743.5	533.9	110.1
0.1	0.3	0.2	1466.1	1355.6	1494.4	1366.8	390.4	499.1	154.3
0.1	0.3	0.3	1511.8	1483.8	1531.4	1486.2	424.7	464.0	135.4
0.1	0.3	0.4	1661.5	1685.1	1607.0	1703.4	478.2	445.0	121.7
0.1	0.3	0.5	1914.6	1813.7	1731.3	1979.6	568.2	437.5	111.2
0.1	0.3	0.6	2333.5	1814.4	1907.8	2353.5	743.5	439.5	103.4
0.1	0.4	0.2	1571.6	1523.7	1588.5	1529.1	424.7	487.8	140.1
0.1	0.4	0.3	1661.5	1696.6	1619.5	1703.4	478.2	445.0	121.7
0.1	0.4	0.4	1894.4	1841.4	1712.5	1981.9	568.2	419.9	109.4

0.1	0.4	0.5	2313.1	1812.6	1869.2	2369.7	743.5	406.3	101.5
0.1	0.5	0.3	1914.6	1939.0	1793.9	1979.6	568.2	437.5	111.2
0.1	0.5	0.4	2313.1	1948.5	1926.6	2369.7	743.5	406.3	101.5
0.1	0.6	0.2	1991.7	2032.9	1969.9	1998.3	568.2	504.1	118.9
0.1	0.6	0.3	2333.5	2197.8	2076.5	2353.5	743.5	439.5	103.4
0.1	0.7	0.2	2393.1	2456.3	2333.7	2350.4	743.5	533.9	110.1
0.2	0.3	0.3	1369.3	1199.9	1258.2	1341.7	360.3	527.1	181.7
0.2	0.3	0.4	1406.5	1300.0	1371.9	1388.9	389.4	499.1	160.3
0.2	0.3	0.5	1492.1	1468.7	1497.1	1494.0	434.2	487.8	144.3
0.2	0.3	0.6	1635.6	1659.4	1652.1	1644.3	507.9	489.5	131.3
0.2	0.4	0.3	1406.5	1299.9	1353.8	1388.9	389.4	499.1	160.3
0.2	0.4	0.4	1475.6	1436.8	1477.5	1463.5	434.2	464.0	139.4
0.2	0.5	0.3	1492.1	1429.7	1466.9	1494.0	434.2	487.8	144.3
0.2	0.6	0.3	1635.6	1610.3	1626.2	1644.3	507.9	489.5	131.3
0.3	0.4	0.4	1322.6	1160.2	1175.6	1318.7	357.5	527.1	186.6
0.3	0.4	0.5	1362.7	1271.9	1266.5	1370.1	395.0	499.1	163.3
0.3	0.5	0.4	1362.7	1256.9	1275.4	1370.1	395.0	499.1	163.3

From the Table 9.1, one can conclude that the use of crossed model also remains more efficient than the simple model in case of estimating odds ratio and attributable risk. The results are also consistent with the results obtained by the use of crossed model while estimating other parameters, such as the relative risk, the correlation coefficient, etc. Thus, we conclude that the crossed model is better than the simple model for all situations we have investigated.

5. APPLICATION BASED ON REAL DATASET

Lee *et al.* (2013) collected real data from 75 respondents at the Joint Statistical Meeting (2011), Miami, FL by using crossed model with $P = T = 0.7$ on smoking and drinking. Let π_{AB}, π_A and π_B be the true proportions of smokers, drinkers, and smokers and drinkers, respectively. Lee *et al.* (2013) reported respective estimates as $\hat{\pi}^*_{AB} = 0.237$, $\hat{\pi}^*_A = 0.24$, and $\hat{\pi}^*_B = 0.36$. These estimates are used for estimating estimators of odds ratio and attributable risk. With the crossed model, the estimator of odds is obtained as $\hat{OR}^* = 409.13$. Estimates of the attributable risks are $\hat{AR}^*(B \mid A) = 0.5504$ and $\hat{AR}^*(A \mid B) = 0.9804$; and estimates of relative risks are $\hat{RR}^*(B \mid A) = 6.10$ and $\hat{RR}^*(A \mid B) = 140.44$. The estimate of smoker or drinker attendees is $\hat{\pi}^*_{A \cup B} = 0.363$, and an estimate of the difference between the proportions of smokers and drinkers is $\hat{\pi}_d^* = -0.12$. The high value of $\hat{OR}^*$ indicates that smoking and drinking are highly associated to each other. The estimate of correlation coefficient between smoking and droning habits is found as $\hat{\rho}_{AB}^* = 0.7346$. The original version of this paper was presented by Lee, Sedory and Singh (2020) at the Joint Statistical Meeting, American Statistical Association.

REFERENCES

Fox, J.A. (2016). *Randomized Response and Related Methods: Surveying Sensitive Data*. Second Edition, SAGE.

Lee, C.S., Sedory, S.A. and Singh, S. (2013). Estimating at least seven measures of qualitative variables from a single sample using randomized response technique. *Statistics and Probability Letters*, 83(1), 399-409.

Lee, Cheon-Sig, Sedory, S. and Singh, S. (2016). Estimation of odds ratio and attributable risk using randomized response techniques. *Presented at the Joint Statistical Meeting* 2016.

Lee, Cheon-Sig, Sedory, S. and Singh, S. (2020). Estimation of odds ratio, attributable risk, relative risk, correlation coefficient and other parameters using randomized response techniques. Vitually presented as the Joint Statistical Meeting, American Statistical Association, August 5, 2010.

Rosner, B. (2016). Fundamentals of Biostatistics. 9th Edition, Thompson, Brooks/Cole.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60:63–69.

APPENDIX

```
      IMPLICIT NONE
      REAL P,T,PIA,PIB,PIAB,SUM, TM1, G1, G2, G3, F1,F2,F3
      DOUBLE PRECISION VA,VB,VAB,CABA,
     1CABB,CAB, VAS,VBS,VABS,
     1CABSAS,CABSBS,CASBS
      REAL ORMSE1, ARMSE1, RRMSE1, CORMSE1, UNIONV1,DIFFV1,REGMSE1
      REAL CT0, CT1, CT2, CT3, CT4
      REAL ORMSE2, ARMSE2, RRMSE2, CORMSE2, UNIONV2,DIFFV2,REGMSE2
      REAL RE1, RE2, RE3, RE4, RE5, RE6, RE7

      CHARACTER*20 OUT_FILE
       WRITE(*,'(A)') 'NAME OF THE OUTPUT FILE'
```

```
       READ(*,'(A20)') OUT_FILE
       OPEN(42, FILE=OUT_FILE, STATUS='UNKNOWN')
      P = 0.70
      T = 0.70
      WRITE(42,107)P,T
107   FORMAT(2X,'P=',F6.3,2X,'T=',F6.3)
      WRITE(42,108)
108   FORMAT( 2X,'PIAB',2X,'PIA',2X,'PIAB',2X,'RE')
      DO 10 PIAB = 0.10, 0.99, 0.10
!        PIAB = 0.2
      DO 10 PIA = 0.10, 0.991, 0.10
      DO 10 PIB = 0.10, 0.991, 0.10
      SUM = PIA+PIB
      IF ( (PIA*PIB).NE.(PIAB) ) THEN
      IF((PIAB.LE.PIA).AND.(PIAB.LE.PIB).AND.(SUM.LT.0.999)) THEN
!     SIMPLE MODEL
      VA =  PIA*(1-PIA)+P*(1-P)/(2*P-1)**2
      VB =  PIB*(1-PIB)+T*(1-T)/(2*T-1)**2
      TM1=(2*P-1)**2*T*(1-T)*PIA+P*(1-P)*(2*T-1)**2*PIB+P*T*(1-P)*(1-T)
      VAB  = PIAB*(1-PIAB) + TM1/((2*P-1)**2*(2*T-1)**2)

      CABA = PIAB*(1-PIA)+P*(1-P)*PIB/(2*P-1)**2
      CABB = PIAB*(1-PIB)+T*(1-T)*PIA/(2*T-1)**2
      CAB =  PIAB-PIA*PIB
      G1 = 1/PIAB +1/(1-PIA-PIB+PIAB)+1/(PIA-PIAB)+1/(PIB-PIAB)
      G2 = (1-PIB)/((PIA-PIAB)*(1-PIA-PIB+PIAB))
      G3 = (1-PIA)/((PIB-PIAB)*(1-PIA-PIB+PIAB))
      F1 = PIAB/(PIAB-PIA*PIB)
      F2 = PIA*PIB/(PIAB-PIA*PIB) + (1-2*PIA)/(2*(1-PIA))
      F3 = PIA*PIB/(PIAB-PIA*PIB) + (1-2*PIB)/(2*(1-PIB))
      ORMSE1 = G1**2*VAB+G2**2*VA+G3**2*VB
     1 -2*G1*G2*CABA-2*G1*G3*CABB+2*G2*G3*CAB
      ARMSE1 = VAB + PIAB**2*VB/PIB**2 + (PIB-PIAB)**2*VA/(1-PIA)**2
     1      - 2*PIAB*CABB/PIB-2*(PIB-PIAB)*CABA/(1-PIA)
     1      + 2*PIAB*(PIB-PIAB)*CAB/(PIB*(1-PIA))
      RRMSE1 = PIB**2*VAB/(PIAB**2*(PIB-PIAB)**2)
     1       + VA/(PIA**2*(1-PIA)**2)+ VB/(PIB-PIAB)**2
     1       -2*PIB*CABA/(PIA*PIAB*(1-PIA)*(PIB-PIAB))
     1       -2*PIB*CABB/(PIAB*(PIB-PIAB)**2)
     1       +2*CAB/(PIA*(1-PIA)*(PIB-PIAB))
      CORMSE1 = F1**2*VAB/PIAB**2 + F2**2*VA/PIA**2
     1        + F3**2*VB/PIB**2 -2*F1*F2*CABA/(PIAB*PIA)
     1        - 2*F1*F3*CABB/(PIAB*PIB)
     1        + 2*F2*F3*CAB/(PIA*PIB)
      UNIONV1 = VA + VB + VAB +2*CAB-2*CABA-2*CABB
      DIFFV1 = VA + VB - 2*CAB
      REGMSE1 = 1-CAB**2/(VA*VB)

!      CROSSED MODEL
```

```
      CT0 =  P*T+(1-P)*(1-T)
      CT1 = (1-P)*T*CT0*(1-PIA-PIB+2*PIAB)
      VAS = PIA*(1-PIA) + CT1/(P+T-1)**2
      CT2 = (1-T)*P*CT0*(1-PIA-PIB+2*PIAB)
      VBS = PIB*(1-PIB) + CT2/(P+T-1)**2
      CT3 = P**2*T**2+(1-P)**2*(1-T)**2-CT0*(P+T-1)**2
      VABS = PIAB*(1-PIAB) + PIAB*CT3/(CT0*(P+T-1)**2)
     1     + P*T*(1-P)*(1-T)*(1-PIA-PIB)/(CT0*(P+T-1)**2)
      CABSAS = PIAB*(1-PIA)+PIAB*T*(1-P)*(P-T+1)/(P+T-1)**2
     1     +P*T*(1-P)*(1-T)*(T-P+1)*(1-PIA-PIB)/(CT0*(P+T-1)**2)
      CABSBS = PIAB*(1-PIB) + PIAB*P*(1-T)*(T-P+1)/(P+T-1)**2
     1       +P*T*(1-P)*(1-T)*(P-T+1)*(1-PIA-PIB)/(CT0*(P+T-1)**2)
      CT4 = 2*P*T*(1-P)*(1-T)*(1+2*PIAB-PIA-PIB)
      CASBS = (PIAB-PIA*PIB)+CT4/(P+T-1)**2
      ORMSE2 = G1**2*VABS+G2**2*VAS+G3**2*VBS
     1 -2*G1*G2*CABSAS-2*G1*G3*CABSBS+2*G2*G3*CASBS
      ARMSE2 = VABS + PIAB**2*VBS/PIB**2 + (PIB-PIAB)**2*VAS/(1-PIA)**2
     1      - 2*PIAB*CABSBS/PIB-2*(PIB-PIAB)*CABSAS/(1-PIA)
     1      + 2*PIAB*(PIB-PIAB)*CASBS/(PIB*(1-PIA))
      RRMSE2 = PIB**2*VABS/(PIAB**2*(PIB-PIAB)**2)
     1       + VAS/(PIA**2*(1-PIA)**2)+ VBS/(PIB-PIAB)**2
     1       -2*PIB*CABSAS/(PIA*PIAB*(1-PIA)*(PIB-PIAB))
     1       -2*PIB*CABSBS/(PIAB*(PIB-PIAB)**2)
     1       +2*CASBS/(PIA*(1-PIA)*(PIB-PIAB))
      CORMSE2 = F1**2*VABS/PIAB**2 + F2**2*VAS/PIA**2
     1         + F3**2*VBS/PIB**2 -2*F1*F2*CABSAS/(PIAB*PIA)
     1         - 2*F1*F3*CABSBS/(PIAB*PIB)
     1         + 2*F2*F3*CASBS/(PIA*PIB)
      UNIONV2 = VAS + VBS + VABS +2*CASBS-2*CABSAS-2*CABSBS
      DIFFV2 = VAS + VBS - 2*CASBS
      REGMSE2 = 1-CASBS**2/(VAS*VBS)
            RE1 = ORMSE1*100/ORMSE2
            RE2 = ARMSE1*100/ARMSE2
            RE3 = RRMSE1*100/RRMSE2
            RE4 = CORMSE1*100/CORMSE2
            RE5 = UNIONV1*100/UNIONV2
            RE6 = DIFFV1*100/DIFFV2
           RE7 = REGMSE1*100/REGMSE2
        IF ( (RE1.GT.100).AND.(RE2.GT.100).AND.(RE3.GT.100).AND.
     1(RE4.GT.100).AND.(RE5.GT.100).AND.(RE6.GT.100).AND.
     1(RE7.GT.100) ) THEN
      WRITE(42, 101)PIAB, PIA, PIB, RE1,RE2,RE3,RE4,RE5,RE6,RE7
101   FORMAT(2X,F8.4,2X,F8.4,2X,F8.4,2X,7(F9.2,1X))
      ENDIF
      ENDIF
      ENDIF
10    CONTINUE
      STOP
      END
```