Reason- A Technical Journal
(ISSN 2277-1654)
Vol. XVI, 2017

A STUDY OF VARIOUS METHODS OF RANDOM RESPONSE AGAINST A SENSITIVE CHARACTER WHEN DISHONEST RESPONSE CAN OCCUR

P. Mukhopadhyay[1], Prof. G.N. Singh[2] and Dr. A. Bandyopadhyay[3]

1 Department of Mathematics, Asansol Engineering College, Asansol-713305, India., Email:parthamukhopadhyay1967@gmail.com Corresponding Author.

2 Department of Applied Mathematics, Indian School of Mines, Dhanbad-826004, India. Email: gnsingh ism@yahoo.com

3 Department of Mathematics, Asansol Engineering College, Asansol-713305, India., Email:arnabbandyopadhyay4@gmail.com Corresponding Author.

Paper received on : 03/05/2018, accepted after revision on : 20/05/2019
DOI : 10.21843/reas/2017/10-21/184049

Abstract: Estimating the proportion of population carrying some sensitive character has always been a challenge to the survey statisticians because exposure of such character (tax evasion, carrying venereal disease or addiction to drugs etc.) may lead to social disapproval and other serious problems of an individual. *Warner's Method* (S.L.Warner, 1965) is the pioneering one to asses the proportion of population carrying sensitive character usually denoted by A . He deployed the random response technique. The method is innovative but highly imprecise for firstly, it assumes that all responses are always true which is not very pragmatic and secondly, due to loss of information since all 'yes' responses are not given against the possession of sensitive character but also against the innocuous question as well. The next improvement came following by the work of Simmons (W.R Simmons, 1967), B.G. Greenberg et al. with the introduction of unrelated question. In the present paper attempt has been taken to include the average dishonest response factor and to obtain the modified expressions for the proportion of sensitive character.

Keywords: sensitive character, random response, dishonest response, unrelated question, interpolation

1. INTRODUCTION

Let π, p_w and θ represent the (a) a *dichotomous* proportion of population carrying the sensitive character A, (b) probability that a respondent will be asked the question "*Do you carry A ?*" through any random device and (c) the total probability of "yes" response in the survey respectively .

Then from Warner's model one has:

$$\theta=\pi.p_w+(1-\pi)(1-p_w). \tag{1}$$

clearly for a direct question $p_w=1$.

Also authors rule out the possibility of false response from the proportion (1-π) when asked the question: "*Do you carry A ?*" (for example a person not having any venereal disease will never report that he/she has it.)

2.1 Including the Factor for Dishonest Response

Let p_T be the average probability of true response for a typical respondent who belongs to A. Then the above equation is modified as :

$$\theta=\pi.p_w\,p_T+(1-\pi)(1-p_w) \tag{2}$$

2.2 Formulation of the Modified Expressions of Estimation :

2.2.1 Warner's Method : Derivation and Minimization $V(\hat{\pi})$ using Binomial Distribution:

One has from (2) , $$\pi=\frac{\theta-(1-p_w)}{p_w(p_T+1)-1}\,. \tag{3}$$

The expression reduces to Warner's expression of if $p_T = 1$. Moreover here one must have $p_w(p_T+1)-1\neq 0$ or $p_w \neq 1/(p_T+1)$

This condition is better than its precursor condition $p_w \neq 1/2$ (what Warner method imposes) for here p_w can be $1/2$ unless $p_T \neq 1$(meaning all respondents will be giving true response always which is highly impractical).

Therefore, if the estimator of θ is denoted by $\hat{\theta}$ and that of π by $\hat{\pi}$ then authors have from (3)

$$\hat{\pi}=\frac{\hat{\theta}-(1-p_w)}{p_w(p_T+1)-1}\,. \tag{4}$$

Butis the probability of "yes" response from a sample of size n and if one assumes the number of "yes" response is r and it follows Binomial distribution then $r \sim B(n,\theta) \Rightarrow E(r)=n\theta$ and $V(r)=n\theta(1-\theta)$

Hence $\hat{\theta}=r/n \Rightarrow E(\hat{\theta})=E(r/n)=E(r)/n=n\theta/n=\theta$

This implies $\hat{\theta}$ is an unbiased estimator of θ and $\hat{\pi}=\dfrac{\hat{\theta}-(1-p_w)}{p_w(p_T+1)-1}=\dfrac{r/n-(1-p_w)}{p_w(p_T+1)-1}$

Similarly the basic relation of Simon's Unrelated Question Method can be modified as follows: $\theta=\pi.p_x\,p_T+\pi_y(1-p_x)$

Therefore, $$E(\hat{\pi})=\frac{E(r)-(1-p_w)n}{n[p_w(p_T+1)-1]}=\frac{\theta-(1-p_w)}{[p_w(p_T+1)-1]} \tag{5}$$

Also authors have $$V(\hat{\pi})=\frac{V(r)}{n^2[p_w(p_T+1)-1]^2}=\frac{\theta(1-\theta)}{n[p_w(p_T+1)-1]^2} \tag{6}$$

Authors again note that if $p_T=1$ (assumed in Warner's model) the above expressions reduces to the expressions derived in Warner's model.

Further, substituting θ from (6) using, (2) : Authors obtain

$$V(\hat{\pi})=\frac{[\pi p_w p_T+(1-p_w)(1-\pi)][1-\pi p_w p_T-(1-p_w)(1-\pi)]}{n[p_w(p_T+1)-1]^2} \tag{7}$$

$$=\frac{(1+p_w\delta-\pi)(\pi-p_w\delta)}{n[p_w k-1]^2} \text{ where } k=p_T+1 \text{ and } \delta=\pi k-1$$

(7) reduces to the expression of Warner's method on putting $p_T=1$ *(vide appendix A for calculation)*

Authors now proceed further to optimize V(π)ˆ w.r.t p_w (optimizing w.r.t p_w is justified because surveyor has total control over the random device, and can change that value suitably -but p_T depends on the population nature and not in the hand of surveyor)

Differentiating $V(\hat{\pi})$ w.r.t p_w partially, authors get:

$$\frac{\partial V}{\partial p_w}=\frac{(p_w k-1)^2\{\delta(\pi-p_w\delta)+(1-\pi+p_w\delta)(-\delta)\}-2k(p_w k-1)\{(1+p_w\delta-\pi)(\pi-p_w\delta)\}}{n[p_w k-1]^4}$$

$$=\frac{p_w(2\delta^2-2\pi\delta k+\delta k)-(2\pi k+2\pi\delta-2\pi^2 k-\delta)}{n[p_w k-1]^3}$$

Hence $$p_w=\frac{(2\pi k+2\pi\delta-2\pi^2 k-\delta)}{(2\delta^2-2\pi\delta k+\delta k)}$$

Putting back the values of δ and k authors get:

$$p_w=\frac{(\pi p_T-\pi+1)}{(\pi p_T+\pi-1)(p_T-1)}$$

Hence authors get the restriction $p_T\neq 1$, which is more pragmatic in real life situation.

Further

$$\frac{\partial^2 V}{\partial p_w^2}=\frac{(p_w k\text{-}1)(2\delta^2\text{-}2\pi k\delta+k\delta)\text{-}3k\left\{p_w(2\delta^2\text{-}2\pi k\delta+k\delta)\text{-}(2\pi k+2\pi\delta\text{-}2\pi^2 k\text{-}\delta)\right\}}{n[p_w k\text{-}1]^4}$$

$$=\frac{(p_w k\text{-}1)A\text{-}3k(p_w A\text{-}B)}{n[p_w k\text{-}1]^4} \text{ where } A=2\delta^2\text{-}2\pi k\delta+k\delta\text{ ; } B=2\pi k+2\pi\delta\text{-}2\pi^2 k\text{-}\delta$$

Authors can also show $\frac{\partial^2 V}{\partial p_w^2}>0$ indicating $V(\hat{\pi})$ is minimum at $p_w=\frac{(\pi p_T\text{-}\pi+1)}{(\pi p_T+\pi\text{-}1)(p_T\text{-}1)}$ (8)

Substituting (8) in (7) one can get a value for $\hat{}\ V(\pi)_{opt}$ from estimated π

If further the respondent is asked the questions (sensitive and innocuous both) directly without resorting to any remote randomdevice then one must have p_w= 1 and cosequently: $V(\hat{\pi})=\frac{\pi(1\text{-}\pi)}{n}+\frac{(\pi^2\text{-}\pi)+p_1\pi(1\text{-}p_1\pi)}{np_T^2}$

2.2.2 Estimation of π from Modified Relation of Simmon's Unrelated Question: using Binomial Distribution

The probability θ of 'yes' response in Simmon's Method is given by $\theta=\pi.p_s p_T+\pi_y(1\text{-}p_s)$ (9)

Here the population is not *dichotomous w.r.t* the innocuous character Y where π_y is the proportion of population having the innocuous character Y. Here also authors rule out any false report from π_y

Hence $\pi=\frac{\theta\text{-}(1\text{-}p_s)\pi_y}{p_s p_T}\Rightarrow\hat{\pi}=\frac{\hat{\theta}\text{-}(1\text{-}p_s)\pi_y}{p_s p_T}$ (10)

In Simmon's strategy two situations may arise *viz.*

(*i*) Authors assume π_y is known (for that one has to suitably frame the unrelated question so that one can compute π_y irrespective of survey e.g if one asks "*Did you born in July ?*" -the expected proportion is 1/12 etc.

Then $\hat{\pi}=\frac{\hat{\theta}\text{-}(1\text{-}p_s)\pi_y}{p_s p_T}=\frac{r\text{-}n(1\text{-}p_s)\pi_y}{np_s p_T}$ again assuming binomial distribution for r

Consequently :

$$E(\hat{\pi})=\frac{E(r)-n(1-p_s)\pi_y}{np_s p_T}=\frac{\theta-(1-p_s)\pi_y}{p_s p_T} \text{ and}$$

$$V(\hat{\pi})=\frac{V(r)}{[np_s p_T]^2}=\frac{n\theta(1-\theta)}{[np_s p_T]^2}$$

$$V(\hat{\pi})=\frac{(p_s\delta+\pi_y)(1-p_s\delta-\pi_y)}{np_S^2 p_T^2} \text{ where } \delta=\pi p_T-\pi_y$$

$$\frac{\partial V(\hat{\pi})}{\partial p_s}=\frac{p_S^2\left[\delta(1-p_s\delta-\pi_y)-\delta(p_s\delta+\pi_y)\right]-2p_s\left[p_s\delta-p_S^2\delta^2-2p_s\delta\pi_y+\pi_y-\pi_y^2\right]}{np_S^4 p_T^2}$$

$$=\frac{\left(-p_s\delta-2\pi_y+2p_s\delta\pi_y+2\pi_y^2\right)}{np_S^3 p_T^2}$$

$$\text{Hence } \frac{\partial V(\hat{\pi})}{\partial p_s}=0 \text{ yields } p_s=\frac{-2\pi_y}{\delta}=\frac{2\pi_y}{\pi_y-\pi p_T} \text{ putting back } \delta=\pi p_T-\pi_y$$

$$\frac{\partial^2 V(\hat{\pi})}{\partial p_s^2}=\frac{p_S^3\left[-\delta+2\delta\pi_y\right]-3p_S^2\left[-p_s\delta-2\pi_y+2\delta p_s\pi_y+2\pi_y^2\right]}{np_S^6 p_T^2}$$

$$=\frac{1}{np_S^4 p_T^2}\left(2p_s\delta(1-2\pi_y)+6\pi_y(1-\pi_y)\right)>0$$

(ii) In the event π_y is unknown one needs take two samples of comparable sizes sayn1 and n2 obtain two eliminant involving π_y and π eliminant of π_y gives the proportion as under:

$$\theta_1=\pi.p_s^1 p_T+\pi_y(1-p_s^1) \text{ and } \theta_2=\pi p_s^2 p_T+\pi_y(1-p_s^2) \tag{11}$$

Hence $\pi_y=.\frac{p_s^2\theta_s^1-p_s^1\theta_s^2}{p_s^2-p_s^1}$ and putting this value in first equation of (10) authors get:

$$\theta_1=\pi.p_s^1 p_T+(1-p_s^1).\frac{p_s^2\theta_s^1-p_s^1\theta_s^2}{p_s^2-p_s^1}\Rightarrow\pi=\frac{\theta_s^1 p_s^2-\theta_s^2 p_s^1-(\theta_s^2-\theta_s^1)}{(p_s^1-p_s^2)p_T} \tag{12}$$

$$E(\hat{\pi})=\frac{\hat{\theta}_s^1 p_s^2-\hat{\theta}_s^2 p_s^1-(\hat{\theta}_s^2-\hat{\theta}_s^1)}{(p_s^1-p_s^2)p_T} \text{ and also}$$

$$V(\hat{\pi})=\frac{V(\theta_s^1)(1-p_s^2)^2-V(\theta_s^2)(1-p_s^1)^2}{(p_s^1-p_s^2)^2 p_T^2}=\frac{\theta_s^1(1-\theta_s^1)(1-p_s^2)^2}{n_1(p_s^1-p_s^2)^2 p_T^2}+\frac{\theta_s^2(1-\theta_s^2)(1-p_s^1)^2}{n_2(p_s^1-p_s^2)^2 p_T^2}$$

using binomial distribution with the condition $n_1+n_2=n$

Optimizing on n_1 and n_2 authors get the minimum value for $V(\hat{\pi})$ is:

$$V(\hat{\pi})_{\min}=\left[\frac{\sqrt{\theta_s^1(1-\theta_s^1)}(1-p_s^2)+\sqrt{\theta_s^2(1-\theta_s^2)}(1-p_s^1)}{\sqrt{n}(p_s^1-p_s^2)p_T}\right]^2$$

2.2.3 Comparing Warner's and Simon's model for Some Chosen values for p_T, p_w, p_s and π_y $=0.5$ Taking $p_T=1$, $p_w=p_s=0.6$ and assuming $\pi_y=0.5$ (for a suitably framed unrelated question) if a survey results in $\theta_w=0.46$ (from (2) and $\theta_s=0.38$ (from (9)) which in turn yields each π_s and $\pi_w=0.3$ (from (3) & (10)) authors get then $V(\hat{\pi}_w)=6.21/n$ and $V(\hat{\pi}_s)=0.654/n$ implies Simmon's method far outperforms Warner's method (about ten times).

3. REPLACING BINOMIAL DISTRIBUTION WITH POISSON DISTRIBUTION : OBTAINING EXPRESSIONS FOR $E(\hat{\pi})$ AND $V(\hat{\pi})$

3.1 Warner's Method : Derivation and Minimization $V(\hat{\pi})$ using Poisson Distribution:

Next authors investigate the minimum variation expressions under Poisson Distribution

The expressions $E(\hat{\pi})$ and $V(\hat{\pi})$ can be further modified if one replaces Binomial distribution with Poisson which is more practical since the ballpark figure of sample size n when binomial distribution is considerable is only ≤ 30. But in almost all practical sampling situations $n >> 30$. Keeping this fact in mind authors use Poisson Distribution.

In Poisson Distribution, both mean and variance are np i.e $n\theta$ in this case. Therefore, authors have :

$$\hat{\pi}=\frac{\hat{\theta}}{p_w(p_T+1)-1} \Rightarrow \hat{\pi}=\frac{r}{n[p_w(p_T+1)-1]}$$

$$E(\hat{\pi})=\frac{E(r)}{n[p_w(p_T+1)-1]}=\frac{\theta}{p_w(p_T+1)-1}=\frac{\pi p_w p_T+(1-p_w)(1-\pi)}{p_w(p_T+1)-1}$$

$$\text{and } V(\hat{\pi})=\frac{V(r)}{n^2[p_w(p_T+1)-1]^2}=\frac{\theta}{n[p_w(p_T+1)-1]^2}=\frac{1-\pi+p_w\delta}{(p_w k-1)^2}$$

$$\frac{\partial V(\hat{\pi})}{\partial p_w} = \frac{(p_w k-1)\delta - 2k(1-\pi+p_w\delta)}{(p_w k-1)^3} = 0$$

Hence $p_w = \dfrac{2k\pi-\delta-2k}{k\delta}$ (13)

Putting back $p_T=1$ i.e $k=2$ and $\delta=2\pi-1$ authors get

$$p_w = \frac{2\pi-3}{4\pi-2}$$

$$\frac{\partial^2 V(\hat{\pi})}{\partial p_w^2} = \frac{(p_w k-1)(-k\delta)-3k(-k\delta p_w-\delta-2k+2k\pi)}{(p_w k-1)^4}$$

$$= \frac{2k^2\delta p_w + 4k\delta + 6k^2 - 6k^2\pi}{(p_w k-1)^4} > 0 \text{ ensuring minimum variation}$$

3.2 Simmon's Method : Derivation and Minimization $V(\hat{\pi})$ using Poisson Distribution:

The expression of variation is :

$$\frac{V(r)}{[np_s p_T]^2} = \frac{n\theta}{[np_s p_T]^2} = \frac{\theta}{n[p_s p_T]^2} = \frac{\pi.p_s p_T + \pi_y(1-p_s)}{n[p_s p_T]^2} \tag{14}$$

Similar calculation done in **3.1** yields optimum value of p_s for which $V(\hat{\pi})$ is minimum.

(vide Appendix B for calculation)

5. Assessing p_T from Independent Assumption:

Authors have already mentioned that p_T is an external parameter which may be fixed from nature of population or from repeated survey experience on similar population. The moot point is that there is a lack of general guideline to fix p_T

Abdul Wahab et al. [6] have argued that p_T varies with p_w have proposed the following relationships between p_T and p_w and accordingly have derived the expression of P_T as under :

1. p_T is symmetric about $p_{w/s}=1/2$
2. p_T is monotone decreasing on $0 \le p \le 1/2$ i.e when possibility of 'yes' response is minimum the probability of dishonest answer is maximum.

Therefore, they proposed $p_T = a(1-2\,p_{w/s})^2$

where a is a parameter satisfying 0 < a < 1. It is interpreted as the probability of giving dishonest answers in case the selection of questions is not randomized.

Motivated by the earlier attempt authors can alternately and legitimately propose another relationship between P_T and θ and find an expression of p_T in terms of θ bases on the following arguments:

If 'yes' response is critically low or excessively high (called under/over statement problems) then it indicates a high degree of average dishonest response. Based on that one can assign values as follows:

$$\text{at } \theta=0,\ p_T=0\ ;\ \text{at } \theta=\tfrac{1}{2},\ p_T=(p_T)_{max}$$
$$\text{when } \theta>\tfrac{1}{2}\quad p_T\to 0 \text{ and when } \theta\to 1,\ p_T\to 0$$

Therefore by using a simple interpolation one can propose a new expression for the average value of p_T obeying above conditions as $p_T=4\beta\theta(1-\theta);\ 0<\beta<1$ (15)

Taking $4\beta=1$ in (15) and substituting it (2) authors obtain for Warner's Method (*vide 2.2.1*):

$$\theta=\pi.p_w\theta(1-\theta)+(1-\pi)(1-p_w)$$
$$\text{Hence } \pi=\frac{\theta-1+p_w}{p_w\theta(1-\theta)-1+p_w}=-(\theta-1+p_w)(1-p_w[1+\theta(1-\theta)])^{-1}$$
$$=(1-\theta-p_w)(1+p_w[1+\theta(1-\theta)])\ \text{ since } p_w[1+\theta(1-\theta)<1 \text{ as the max value of } \theta(1-\theta)=.25 \text{ when } \theta=.5$$

$$\pi=(1-\theta-p_w)+(p_w+p_w\theta+p_w\theta^2)])(1-\theta-p_w)$$
$$=1-\theta-p_w^2-p_w\theta\ ;\ \text{ignoring the 3}^{rd}\text{ or higher orders of smallness} \qquad (16)$$
$$\text{hence } \hat{\pi}=1-\hat{\theta}-p_w^2-p_w\hat{\theta}=1-\frac{r}{n}-p_w^2-p_w\frac{r}{n}$$
$$\therefore\ E(\hat{\pi})=1-\frac{E(r)}{n}-p_w^2-p_w\frac{E(r)}{n}=1-\theta-p_w^2-p_w\theta\ ;\ \because E(r)=n\theta$$

Where r is the number of 'yes' response out of n samples assumed to have followed binomial distribution as before.

$$\text{Hence } V(\hat{\pi})=\frac{V(r)}{n^2}+p_w\frac{V(r)}{n^2}=\frac{\theta(1-\theta)}{n}\ ;\ \because V(r)=n\theta(1-\theta)$$
$$=\frac{1}{n}(1-p_w^2-\pi)(\pi+p_w^2)$$
$$\frac{\partial V}{\partial p_w}=\frac{1}{n}\left[-2p_w(\pi+p_w^2)+2(1-p_w^2-\pi)p_w\right]$$

$$\frac{\partial V}{\partial p_w}=0 \Rightarrow p_w=\frac{1-2\pi}{2}$$

$$\frac{\partial^2 V}{\partial p_w^2}=\frac{1}{n}[2-4\pi-12p_w^2]>0 \text{ for a suitable range of values of } \pi \text{ and } p_w$$

$$\text{Then } V(\hat{\pi})_{opt}=\frac{1}{n}\left[-2p_w(\pi+p_w^2)+2(1-p_w^2-\pi)p_w\right]=\frac{1}{n}\left(\pi+\frac{1-2\pi}{2}\right)\left(1-\pi-\frac{1-2\pi}{2}\right)=\frac{1}{4n}$$

REFERENCE

1. Greenberg, B. G., Abul-ela, A. L. S., Simmons, W. R., and Horvitz, D. G. , The unrelated question randomized response model: Theoretical framework, Journal of the American Statistical Association, Vol. 64, pp. 520-539, 1969.

2. Warner, S. L., Randomized response A survey technique for eliminating evasive answer bias, Journal of the American Statistical Association, Vol. 60, pp. 63-69, 1965.

3. Horvitz, D. G., Greenberg, B. G., Abernathy, J. R. , Randomized response: A data-gathering device for sensitive questions, International Statistical Review, Vol. 44, pp. 181-196, 1976.

4. Horvitz, D. G., Shah, B. V.,Simmons, W. R., The unrelated question randomized response model, Proceedings of the Social Statistics Section, American Statistical Association, pp. 65-72, 1967.

5. E, dgell E, Duchan, K. L. and Arb, An empirical test of the unrelated question randomized response technique, Bulletin of the Psychonomic Society, Vol, 30 (2), pp. 153-156, 1992.

6. Alhassan, A. W. Ohuchi, S. and Taguri, M. Randomized Response Designs Considering The Probability Of Dishonest Answers, J. Japanese Soc. Comp. Statist, Vol. 4, pp. 1-24, 1991.

7. Singh G. N., Suman S., Khetan M. and Paul C, Some estimation procedures of sensitive character using scrambled response techniques in successive sampling, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2017.1327073

8. Greenberg, B.G., Abul-Ela, A.A., Simmons, W.R. and Horvitz, D.G., The Unrelated Question Randomized Response: Theoretical Framework,

Journal of the American Statistical Association, Vol. 64, pp. 520-539, 1969.

9. Chaudhuri, A. and Mukerjee, R., Randomized Response: Theory and Techniques , New York, Marcel Dekker, 1988.

10. Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R. and Horvitz, D. G. The unrelated question randomized response model: Theoretical frame-work, Journal of the American Statistical Association, Vol. 64, pp. 520-539, 1969.

11. Liu, P. T. and Chow, L. P. The efficiency of the multiple trial randomized response technique. Biometrics, Vol. 32, pp. 607-618, 1976.

12. Mukhopadhyay, P. On estimation of some confidential characters from survey data, Bull. Calcutta Statistical Association, Vol. 29, pp.133-141,1980.

13. Warner, S. L., Randomized response: A survey technique for eliminating evasive answer bias, Journal of the American Statistical Association, Vol. 60, pp. 63-69, 1965.

14. Winkler, R. L. and Franklin, L. A., Warner's randomized response model: A Bayesian approach, Journal of the American Statistical Association, Vol. 74, pp. 207-214, 1979.

APPENDIX A

Reduction of eqn. (7) to Warner's Expression :

Authors have in Eg. (7) $V(\hat{\pi})$

$$= \frac{\pi(1-\pi)}{n} + \frac{p_w(1-p_w)}{n[p_w(p_T+1)-1]^2} + \frac{\pi p_w(1+p_T)(2p_w-1) - \pi^2 p_w^2(1+p_T^2) - 2\pi p_w(1-\pi) + 2\pi^2 p_w p_T(1-p_w)}{n[p_w(p_T+1)-1]^2}$$

Putting $p_T = 1$ above one gets:

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n[2p_w-1]^2} + \frac{p_w(1-p_w)}{n[2p_w-1]^2} + \frac{2\pi p_w(2p_w-1) - 2\pi^2 p_w^2 - 2\pi p_w(1-\pi) + 2\pi^2 p_w(1-p_w)}{n[2p_w-1]^2}$$

$$= \frac{p_w(1-p_w)}{n} + \frac{\pi(1-\pi) + 2\pi p_w(2p_w-1) - 2\pi^2 p_w^2 - 2\pi p_w(1-\pi) + 2\pi^2 p_w(1-p_w)}{n[2p_w-1]^2}$$

$$= \frac{p_w(1-p_w)}{n[2p_w-1]^2} + \frac{\pi(1-\pi) + 4p_w(p_w-1)\pi(1-\pi)}{n[2p_w-1]^2}$$

$$= \frac{p_w(1-p_w)}{n[2p_w-1]^2} + \frac{\pi(1-\pi)}{n}$$

Putting $p_w = 1$ above one gets: $V(\hat{\pi}) = \dfrac{\pi(1-\pi)}{n}$.

APPENDIX B

Determining the Optimum p_s for Simmon's Modified Expression using Poisson Distribution:

From (14), one has: $V(\hat{\pi})=\dfrac{\pi.p_s p_T+\pi_y(1-p_s)}{n[p_s p_T]^2}$

$$\frac{\partial V(\hat{\pi})}{\partial p_s}=\frac{p_s\{\pi p_T-\pi_y)-2\{\pi_y(1-p_s)+\pi p_s p_T\}}{(np_T^2)p_s^3}=\frac{\pi_y p_s-\pi p_s p_T-2\pi_y}{(np_T^2)p_s^3}$$

Since denominator $\neq 0$ then one has: $\pi p_s p_T+\pi_y(1-p_s)=0 \Rightarrow (p_s)_{opt.}=\dfrac{2\pi_y}{\pi_y-\pi p_T}$

$$\frac{\partial^2 V(\hat{\pi})}{\partial p_s^2}=\frac{p_s^3\{\pi_y-\pi p_T)-3p_s^2\{\pi_y p_s+\pi p_s p_T-2\pi_y\}}{(np_T^2)p_s^6}$$

Also $$=\frac{2\pi p_s p_T+6\pi_y-2\pi_y p_s}{(np_T^2)p_s^4}$$

if $\dfrac{\partial^2 V(\hat{\pi})}{\partial p_s^2}>0$ then $2p_s(\pi p_T-\pi_y)>-6\pi_y$ or $p_s<\dfrac{3\pi_y}{\pi_y-\pi p_T}$

Hence, $V(\hat{\pi})$ is minimum for $(p_s)_{opt}$.