Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 1, 2025 (New Series), pp 385–388
https://www.ssca.org.in/journal

Understanding Fellegi Scheme for Sample Size Three

Yumnam Menon Singh[1] and Opendra Salam[2]
[1] *Research Scholar, Statistics, Manipur University, Imphal, Manipur, India*
[2] *Professor of Statistics, Manipur University, Imphal, Manipur, India*

Received: 13 May 2024; Revised: 30 September 2024; Accepted: 03 October 2024

Abstract

In this short communication, we attempt to rework on Fellegi (1963) scheme for sample size 3, taking clue from Choudhry (1981) and Sinha (1973, 1974).

Key words: Sampling designs; Sampling schemes; Inclusion probabilities of first and second orders; Mixture designs.

1. Introduction

Brewer and Hanif (1983) reviewed sampling schemes with unequal probabilities without replacement and compiled several selection procedures. Among the schemes, Brewer (1963) and Fellegi (1963) schemes for $n = 2$ are described in text books such as in Hedayat and Sinha (1991) but cannot be readily extended to $n = 3$. For Fellegi scheme, Choudhry (1981) attempted to develop computational formulae using Fortran language specifically for $n = 3$ and 4. However, satisfactory techniques are not yet available. We make an attempt to extend Fellegi scheme from algebraic consideration. Our contribution in this study is essentially a follow-up of Fellegi ($n = 2$) to $n = 3$. We are able to generalize Fellegi scheme and we explain our procedure through a numerical example.

It may be noted that Choudhry (1981) made an attempt to work out a solution for $n = 3$ underlying Fellegi scheme. He did not pursue any analytical exercise to solve for the choice of $p_3(i)$ values. He used the second stage p-values ($p_2(i)$) as trial values for the third stage p-values ($p_3(i)$) and developed a Fortran programme to approximately work out stabilized third stage p-values.

2. Fellegi scheme $(N, n = 3)$

For Fellegi Scheme $(N, n = 3)$, $P(i, j, k)$ has to be chosen in such a way that at each trial, inclusion probability of i^{th} unit is p_i for all i. Hence, overall inclusion probability for i^{th} unit is $3p_i$. To achieve this, set k^{th} trial selection probability for i^{th} unit $= p_k(i)$ for

Corresponding Author: Salam Opendra Singh
Email: sopendra@manipuruniv.ac.in

$k = 1, 2, 3$; $i = 1, 2, 3, \ldots, N$ where $\sum_{i=1}^{N} p_k(i) = 1$ for each k. Then we have the expression

$$\pi_i = p_1(i) + \sum_{j(\neq i)}^{N} p_1(j)\frac{p_2(i)}{1-p_2(j)} + \sum_{k(\neq i)}^{N}\sum_{j(\neq i,k)}^{N} p_1(k)\frac{p_2(j)}{(1-p_2(k))}\frac{p_3(i)}{(1-p_3(k)-p_3(j))} = 3p_i \quad (1)$$

It may be noted that in the above, we are tacitly using the expression for $p_2(i)$ as was derived by Fellegi (1963) for the case of $n = 2$. Set $p_1(i) = p_i$ for each $i = 1, 2, \ldots, N$.
So, $p_3(i)$'s have to satisfy

$$\sum_{k(\neq i)}^{N}\sum_{j(\neq i,k)}^{N} p_1(k)\frac{p_2(j)}{(1-p_2(k))}\frac{p_3(i)}{(1-p_3(k)-p_3(j))} = p_i, \quad i = 1, 2, \ldots, N.$$

$$\Rightarrow B_i = \frac{p_i}{p_3(i)}\left[\frac{1-2p_3(i)-p_2(i)}{(1-p_2(i))(1-2p_3(i))}\right] \quad (2)$$

where $B_i = \sum_{k=1}^{N}\sum_{j=1}^{N}\frac{p_1(k)p_2(j)}{(1-p_2(k))(1-p_3(k)-p_3(j))} - \sum_{j=1}^{N}\frac{p_1(i)p_2(j)}{(1-p_2(i))(1-p_3(i)-p_3(j))}$

$$-\sum_{k=1}^{N}\frac{p_1(k)p_2(i)}{(1-p_2(k))(1-p_3(k)-p_3(i))} - \sum_{k=1}^{N}\frac{p_1(k)p_2(k)}{(1-p_2(k))(1-2p_3(k))} + \frac{2p_1(i)p_2(i)}{(1-p_2(i))(1-2p_3(i))}$$

After simplifying (2), we obtain a quadratic equation in $p_3(i)$ as

$$2B_i(1-p_2(i))p_3^2(i) - [2p_i + B_i(1-p_2(i))]p_3(i) + p_i(1-p_2(i)) = 0 \quad (3)$$

$$\text{So, } p_3(i) = \frac{(2p_i + B_i(1-p_2(i))) \pm \sqrt{(2p_i + B_i(1-p_2(i)))^2 - 8B_ip_i(1-p_2(i))^2}}{4B_i(1-p_2(i))} \quad (4)$$

Remark 1: It must be noted that the expressions in (2) and (4) basically refer to only one relation involving B_i and $p_3(i)$. A judicial choice of B_i for evaluation of $p_3(i)$ has, so far, eluded us. Therefore, we have taken up an alternative approach that refers to a choice of $p_3(i)$ as a function of p_i and $p_2(i)$ with the solo objective: Choice of $p_3(i)$ must lead to the 3rd stage $\pi_i = 3p_i$ to best possible approximation.

Remark 2: At this stage it is pertinent to note that we will be using the concept of mixture designs of the type $pD_1 + qD_2$, $0 < p,\ q < 1,\ p + q = 1$. We recall that Sinha (1973, 1974) made a similar study with the provision that one of p and q would be negative, however, satisfying the necessary condition that $pD_1(s) + qD_2(s) > 0$ for every sample 's' in the underlying design. In our study below we will follow Sinha's approach to come up with a solution.

Remark 3: This problem is simply stated and theoretical solutions are quite hard to obtain. We make attempts to minimize the gap between π_i and $3p_i$ by making suitable choice of $p_3(i)$'s. Similar problem was encountered by Sinha (1973, 1974) who had developed a mixture solution of the type: $p_3(i) = ap_i + bp_2(i)$ with choices of a and b subject to $a + b = 1$, by admitting the solutions with negative values of a or b ! Of course, the mixture has to yield all positive fractions. Our attempt is illustrated in the following example.

Example 1: $N = 6,\ p_1 = 0.25,\ p_2 = p_3 = 0.20,\ p_4 = p_5 = 0.15,\ p_6 = 0.05.$
With reference to Fellegi (1963), for the case of $n = 2$,

(i) Solve for A from the equation: $N - 2 = \sum_{i=1}^{N} \sqrt{1 - \frac{4p_i}{A}}$, where $A = \sum_{t=1}^{N} \frac{p_t}{1 - p_2(t)}$.

(ii) Solve for $p_2(i)$ from the equation: $p_2(i) = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{p_i}{A}}$.

Newton's method is used to obtain: $A = 1.24727$, and then values for $p_2(i)$ are deduced as given below in Table 1.

Table 1: Calculation of $p_2(i)$

i	1	2	3	4	5	6	Sum
p_i	0.25	0.2	0.2	0.15	0.15	0.05	1
$p_2(i)$	0.27737	0.20058	0.20058	0.13981	0.13981	0.04184	1

Keeping the possibility of one of a and b being negative, after some trial and error, we ended up with $a = 2.55$ and $b = -1.55$. The end-result is shown below.

Table 2: Computation of π_i

i	1	2	3	4	5	6	Total
π_i	0.73560 ≈0.74	0.60475 ≈0.60	0.60475 ≈0.60	0.45445 ≈0.45	0.45445 ≈0.45	0.14598 ≈0.15	2.99998 ≈3
$3p_i$	0.75	0.6	0.6	0.45	0.45	0.15	3

Table 3: Computation of $\pi_{ij} = \sum_{s \ni (i,j)} P(s)$

	π_{ij}					
i j	1	2	3	4	5	6
1		0.40506	0.40506	0.28831	0.28831	0.08445
2	0.40506		0.31175	0.21547	0.21547	0.06174
3	0.40506	0.31175		0.21547	0.21547	0.06174
4	0.28831	0.21547	0.21547		0.14762	0.04202
5	0.28831	0.21547	0.21547	0.14762		0.04202
6	0.08445	0.06174	0.06174	0.04202	0.04202	

Remark 4: We can readily verify numerically for $n = 3$ that $\pi_{ik} > \pi_{jk}$ whenever $p_i > p_j$ for all $i \neq j \neq k$ and $\pi_i \pi_j > \pi_{ij}$ for all $i \neq j$.

3. Conclusion

From the above illustration it can be seen that if one can express $p_3(i)$ as a linear combination of p_i and $p_2(i)$ that is $p_3(i) = wp_i + (1 - w)p_2(i)$, with a suitable choice of w, the Fellegi scheme for $n = 3$ can be constructed in a simple way. Further research is needed to find an appropriate value of w, assuming that it can take negative values as well.

Acknowledgements

We are indeed very grateful to the Chief Editor of S&A for making a number of valuable suggestions towards revision on an earlier draft of the manuscript. His insightful comments have been extremely helpful in revising the manuscript. We also thank Professor Bikas Sinha, Retired Professor of ISI, Kolkata for his interest in this investigation.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Brewer, K. R. W. (1963). A model of systematic sampling with unequal probabilities. *Australian Journal of Statistics*, **5**, 5–13.

Brewer, K. R. W. and Hanif, M. (1983). *Sampling with Unequal Probabilities*, volume 15 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Choudhry, G. H. (1981). Construction of working probabilities and joint selection probabilities for Fellegi's PPS sampling scheme. *Survey Methodology*, **7**, 93–108.

Fellegi, I. (1963). Sampling with varying probabilities without replacement: rotating and non-rotating samples. *Journal of the American Statistical Association*, **58**, 183–201.

Hedayat, A. and Sinha, B. K. (1991). *Design and Inference in Finite Population Sampling.* Wiley New York.

Sinha, B. K. (1973). On sampling schemes to realize pre-assigned sets of inclusion probabilities of first two orders. *Calcutta Statistical Association Bulletin*, **22**, 89–110.

Sinha, B. K. (1974). On sampling schemes to realize invariant pre-assigned sets of inclusion probabilities of first two orders. *Calcutta Statistical Association Bulletin*, **23**, 45–72.