IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 20, Issue 4 Ser. 2 (July. – August. 2023), PP 20-22
www.iosrjournals.org

On Averaging Over Distinct Units…Revisited

Dr. Y. P. Sabharwal1
Dr. Debasree Goswami2
1 Consulting Actuary, 27 Jaina Building Roshan Ara Road Delhi, INDIA,
2 Associate Professor, Department of Statistics, Hindu College, University of Delhi, INDIA,

Abstract
This paper introduces an alternative approach to a well-established theorem in statistical theory, pertaining to the evaluation of the expected value of the reciprocal of the number of distinct units in multiple samples drawn independently and without replacement from a population. Designed with pedagogical considerations in mind, this new approach aims to make the theorem more accessible for teaching purposes. The paper emphasizes the theoretical underpinnings of the concept and its potential for assessing the diversity of any characteristic of interest in a sample, contributing to a broader understanding of this fundamental aspect of statistical theory.

Key Word: *Survey Sampling, Distinct Units, Expectation of the Reciprocal of Distinct Units, Pedagogical Considerations*

Date of Submission: 25-07-2023 Date of Acceptance: 05-08-2023

I. Introduction

The concept of evaluating the expectation of the reciprocal of the number of distinct units in a sample is a cornerstone in statistical theory, with broad applications across various fields such as biology, finance, and social sciences. This concept was elegantly addressed by Agrawal [1].

In this paper, we introduce an alternative proof of Theorem 3.1 from the cited article, which provides a method for evaluating the expected value of the reciprocal of the number of distinct units in k random samples drawn independently and without replacement from a population of size N. Offering a fresh perspective, our proof is particularly beneficial for pedagogical purposes. Our aim is to provide educators and researchers with an additional tool to further their understanding and teaching of this concept, enhancing the accessibility of the result.

II. Distinct Unit and The Main Result

In the context of the theorem mentioned above, a "distinct unit" refers to a unique entity or observation within a population, such as a specific combination of insurance attributes like policy type and coverage level. The concept of distinct units is used to quantify the average diversity within samples, providing insights into how representative the samples are of the overall insurance portfolio or customer base.

Theorem: *If ν is the number of distinct units in k random samples drawn independently and without replacement from a population of size N, i.e., each sample is replaced to the population before drawing the next, then*

$$E\left(\frac{1}{\nu}\right)=\sum_{r=0}^{S_k}\frac{1}{N-r}\prod_{j=1}^{k}\left\{\binom{N-r}{l_j}\Big/\binom{N}{l_j}\right\}$$

where l_j is the size of the jth sample ($j=1,2,\dots,k$) and $S_k=\min_{1\le j\le k}(N-l_j)$.

Proof: Define for the individual units of the population,

$$X_i=\begin{cases}1, & \text{if the } i^{th} \text{ unite of the population is not included}\\ & \text{in any of the } k \text{ samples.}\\ 0, & \text{otherwise.}\end{cases}$$

Then

$\nu = N - \sum_{i=1}^{N} X_i$ and $N - n \le \sum_{i=1}^{N} X_i \le S_k$ with $n = min\left(\sum_{j=1}^{k} l_j, N\right)$.

$$\frac{1}{\nu} = \frac{1}{N} + \sum_{I=1}^{\infty} \frac{1}{N^{I+1}} \left(\sum_{i=1}^{N} X_i \right)^I \tag{1}$$

Here

$$\begin{aligned} \left(\sum_{i=1}^{N} X_i \right)^I &= \sum_{\left\{\sum_{i=1}^{N} I_i = I, I_i \ge 0\right\}} \frac{I!}{\prod_{i=1}^{N} I_i!} \prod X_i^{I_i} \\ &= \sum_{\left\{\sum_{i=1}^{N} I_i = I, I_i \ge 0\right\}} \frac{I!}{\prod_{i=1}^{N} I_i!} \prod_{i=1}^{N} X_i^{r_i} \end{aligned} \tag{2}$$

where

$$r_i = \begin{cases} 1, & \text{if } I_i > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Next with $r = \sum_{i=1}^{N} r_i$

$$\begin{aligned} E\left[\prod_{i=1}^{N} X_i^{r_i} \right] &= \prod_{j=1}^{k} \left\{ \binom{N-r}{l_j} \Big/ \binom{N}{l_j} \right\} \\ &= p(r) \quad \text{(say)}. \end{aligned} \tag{4}$$

and

$$\begin{aligned} \sum_{\left\{\sum_{i=1}^{N} I_i = I, \sum_{i=1}^{N} r_i = r\right\}} \frac{I!}{\prod_{i=1}^{N} I_i!} &= \binom{N}{r} \sum_{\left\{\sum_{i=1}^{r} I_i = I, I_i > 0\right\}} \frac{I!}{\prod_{i=1}^{r} I_i!} \\ &= \binom{N}{r} \sum_{J=0}^{r} (-1)^J \binom{r}{J} (r-J)^I. \end{aligned} \tag{5}$$

Using (2) to (5), we get from (1)

$$\begin{aligned} E\left(\frac{1}{\nu}\right) &= \frac{1}{N} + \sum_{I=1}^{\infty} \frac{1}{N^{I+1}} \sum_{r=1}^{I} \binom{N}{r} \sum_{J=0}^{r} (-1)^J \binom{r}{J} (r-J)^I p(r) \\ &= \frac{1}{N} + \sum_{r=1}^{S_k} \binom{N}{r} p(r) \sum_{J=0}^{r} (-1)^J \binom{r}{J} \sum_{I=r}^{\infty} \frac{1}{N^{I+1}} (r-J)^I \\ &= \frac{1}{N} + \sum_{r=1}^{S_k} \binom{N}{r} p(r) \frac{1}{N^{r+1}} \sum_{J=0}^{r} (-1)^J \binom{r}{J} \frac{(r-J)^r}{N-r+J}. \end{aligned} \tag{6}$$

Finally,

$$\sum_{J=0}^{r}(-1)^J\binom{r}{J}\frac{(r-J)^r}{N-r+J} = \Delta^r\frac{(x-r)^r}{N-x+r}\Big\}_{x=r}$$
$$= \Delta^r\frac{X^r}{N-X}\Big\}_{X=0}$$
$$= \Delta'^r\frac{(N-Y)^r}{Y}\Big\}_{Y=N} \qquad (7)$$
$$= \frac{N^r r!}{N(N-1)\cdots(N-r)}.$$

where, Δ' operated at an interval of differencing of -1.
The desired result then follows on substituting from (7) in (6).

III. Conclusion

This paper has introduced an alternative proof of a significant theorem related to the expectation of the reciprocal of the number of distinct units in random samples. The fresh perspective provided by this proof enhances its accessibility, particularly for educational purposes owing to its simplicity.

References

[1]. M. C. Agrawal, "On Averaging Over Distinct Units In Replicated Samples," Statistics: A Journal Of Theoretical And Applied Statistics, Vol. 13, No. 1, Pp. 91-97, 1982.