SERVICE BASED ENUMERATION ESTIMATION

[1]Felipe Kohn and Richard Griffin
Felipe Kohn, Bureau of the Census, Washington DC 20233

Key Word: Multiplicity Estimator

I. Introduction

The Census Bureau established the Service Based Enumeration (SBE) program as the statistical program designed to include persons without usual residence that use service facilities (shelter, soup kitchen or mobile food vans). Those persons are not covered by regular Census Bureau procedures for households or persons in group quarters.

The proposed methodology for the SBE estimation for the 2000 Census is the Multiplicity estimator that is based on the number of times the respondent uses the service facilities.

In this paper we present several multiplicity estimators based on the usage question for service facilities.

II. Estimators - All estimators assume the unduplication of questionnaires has been completed.

A. Shelter Only Estimator

The enumeration is done only at shelters. The respondents are asked "Including today, how many days in the past seven days have you stayed in a shelter?" Thus, the estimator is obtained as follows:

$$\hat{X}_{sh} = \sum_{i=1}^{n} 7 / A_i \text{, where}$$

A_i = shelter usage response for the i - th respondent and n is the number of respondents

B. Soup Kitchen Estimator

The enumeration is done only at soup kitchens. The respondents are asked "Including today how many days in the past seven days did you eat in a soup kitchen?" Thus, the estimator is obtained as follows:

$$\hat{X}_{sk} = \sum_{i=1}^{m} 7 / B_i \text{ where,}$$

B_i is the number of days the i-th respondent ate a meal in a soup kitchen and m is the number of respondents.

C. Combined Estimator

The enumeration is done in both service facilities. The estimator is based on the average of the shelter estimator and soup kitchen estimator.

Thus, this estimator is obtained as follows:

$$\hat{X}_{comb} = (\hat{X}sh + \hat{X}sk) / 2$$

D. Optimal Estimator

This estimator is the optimal combination (minimum variance) of the shelter and soup kitchen estimators. It is obtained by the following expression:

$$\hat{X} = \hat{W}_{opt}\hat{X}_{sh} + (1 - \hat{W}_{opt})\hat{X}_{sk} \text{ where,}$$

$$\hat{W}_{opt} = \frac{Var(\hat{X}_{sk})}{Var(\hat{X}_{sk}) + Var(\hat{X}_{sh})}$$

E. 1995 Type Estimator

The 1995 Census Test used a SBE questionnaire that asked the following question: "How many days during the past week did you use a shelter, soup kitchen or mobile food van?" These questions were asked of all persons at the shelters on the selected day and all persons enumerated in soup kitchens or mobile food vans on the next day. The multiplicity estimator is as follows:

---

[1]Felipe Kohn and Richard Griffin are mathematical statisticians in the Decennial Statistical Studies Division of the U.S. Census Bureau. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

$\hat{Y}_{1995} = \sum_{i=1}^{n} 7 / Z_i$ where, n is the total number of persons enumerated in a shelter on the selected day and/or a soup kitchen or mobile food van on the next day and Z_i is the number of days a week person i uses a shelter, soup kitchen, or mobile food van.

F. Census Dress Rehearsal Estimator

All persons at a shelter on the selected day are asked "How many days a week including today did you use a shelter?". All persons using a soup kitchen or mobile food van on the next day are asked the same question about soup kitchen or mobile food van usage. They are asked "How many days during the past week did you receive a meal from a soup kitchen or mobile food van?". A person enumerated on both occasions is counted once based on his/her shelter usage. The multiplicity estimator is as follows:

$$\hat{X}_{DR} = \sum_{i=1}^{n} 7 / A_i + \sum_{i=1}^{m} 7 / B_i$$

where n is the number of persons enumerated at a shelter on the selected day , m is the number of persons enumerated in a soup kitchen or mobile food van the next day who did not use a shelter during the past week, A_i is the number of days in the past seven days person i uses a shelter and B_i is the number of days in the past week (for those who did not use a shelter) person i received a meal from a soup kitchen or mobile food van.

III. Hypothetical Populations

The persons without usual residence are very transient, by definition. Weather conditions, disposable income, number of service facilities in the area among other factors make it difficult to include some of them in any type of enumeration. In order to take this factor into consideration two pseudo-populations are considered for this study:

For the first four estimators we assume a population of persons without usual residence of 2,500 persons with at least ten percent of them not using services (and thus never enumerated). For the last two estimators we consider a pseudo-population of 3,000 persons with at least 250 never found in a shelter or a soup kitchen and an additional 500 who are never in a shelter but do eat meals in soup kitchens.

IV. Capture Probabilities

Persons without usual residence use service facilities (shelters and soup kitchens) depending on factors such as weather (in rough weather the population in the shelters increases while in mild weather the population in the shelters decreases), income (for example a veteran of the armed forces will not eat in a soup kitchen while he still has money from the VA and will eat in the soup kitchen when his money runs out). Thus, the enumeration of persons without usual residence at a given time varies greatly by the type of service facility (shelter or soup kitchen). To take this factor in consideration, we consider three sets of capture probabilities. For each set the pseudo-populations (3,000 persons and 2,500 persons) are randomly generated using these probabilities (except for persons designated to never use a type of service) assuming independence over days and persons; thus, for each person in the pseudo-population we generate a random number (between 0 and 1) for each day in a week for the shelter usage and a random number for each day in a week for the soup kitchen usage. If the generated random number on a given day is less than the set capture probability the person is designated as using the service facility that day. Three sets of capture probabilities are considered (a new set of random numbers is generated for each set of capture probabilities):

A. 50% probability of using a shelter and 50% probability of using a soup kitchen.
B. 50% probability of using a shelter and 25% probability of using a soup kitchen.
C. 25% probability of using a shelter and 50% probability of using a soup kitchen.

V. Statistical Criteria

A. The selected statistic to compare the estimates is the relative root mean square error defined by the following expression:

$$RRMSE = \frac{\sqrt{MSE(\hat{X})}}{X}$$

with X equal to the true population size and

$$MSE(\hat{X}) = Bias(\hat{X})^2 + Var(\hat{X})$$

B. Variance Estimation

The methodology for determining the variance for all the multiplicity estimates presented in this paper was derived by Pat Cantwell of the Census Bureau and is as follows: Divide the target population into eight groups: $G_{0,\ldots,}G_7$ of sizes $N_0,\ldots,N_7$, where the N_i's are the number of persons who use a shelter (soup kitchen) i days in a week. Assume that the days a person uses a shelter (soup kitchen) are selected randomly given the number of days a week he/she uses a shelter.

N is the total target population. On the selected enumeration day n_i out of N_i use a shelter for i=1,...,7 with each n_i distributed as a binomial random variable with parameters N_i and i/7.

The multiplicity estimator can be written as

$$\hat{N} = \sum_{i=1}^{7} \frac{7}{i} n_i$$

$E(\hat{N} | \{N_0, \ldots, N_7\}) = N - N_0$ under our assumptions and the variance is as follows:

$$Var(\hat{N}) = \sum_{i=1}^{7} (7/i)^2 (N_i)(i/7)(1-i/7)$$

$$= \sum_{i=1}^{7} N_i \left(\frac{7-i}{i}\right)$$

This binomial model allows for variation in which persons are at a shelter on the enumeration day and fixes the partition of the population into $N_1, N_2,\ldots,N_7$. For this model the variance can easily be calculated allowing the N_i values to vary. A more detailed description of this less restricted variance can be read in the paper presented at this session by Roger Shores of the Census Bureau.

VI. Results

A. Tables 1-3 below shows the results of our simulations for the estimators based on a pseudo-population of 2,500.

Table 1: Estimators with Capture Probabilities of P=.5 for Shelters and Q=.5 for Soup Kitchens

Estimator	Estimate	Variance	C.V	RRMSE
Shelter Only	2235	3008.7	2.45%	10.80%
Soup Kitchen	2237	3194.9	2.52%	10.76%
Combined	2236	1550.9	1.76%	10.67%
Optimal	2236.28	1549.51	1.76%	10.64%

Table 2: Estimators with Capture Probabilities of P=.5 for Shelters and Q=.25 for Soup Kitchens

Estimator	Estimate	Variance	C.V	RRMSE
Shelter Only	2223	3109	2.50%	11.30%
Soup Kitchen	1935	6360	4.12%	22.82%
Combined	2079	2367.42	2.34%	16.95%
Optimal	2205.81	2088.51	2.07%	11.9%

Table 3: Estimators with Capture Probabilities of P=.25 for Shelters and Q=.5 for Soup Kitchens

Estimator	Estimate	Variance	C.V.	RRMSE
Shelter Only	2250	980.38	1.39%	10.09%
Soup Kitchen	2234	3127.21	2.5%	10.87%
Combined	2242	1026.9	1.42%	10.39%
Optimal	2246.28	746.38	1.21%	10.21%

B. Table 4 below, shows the statistics of the simulation for the second pseudo-population.

Table 4: Statistics for the 95 Test and 2000 Dress Rehearsal Estimators

Estim.	Prob	Est	Var	C.V.	RRMSE
2000 Dress Reher.	P=.5 Q=.5	2747	3949.75	2.28%	8.70%
	P=.5 Q=.25	2664	4812	2.60%	11.43%
	P=.25 Q=.5	2747	1726.58	3.06%	8.88%
95 Census Test	P=.5 Q=.5	2747	1685.32	1.49%	8.54%
	P=.5 Q=.25	2664	3246.55	2.13%	11.51%
	P=.25 Q=.5	2747	1126.58	1.22%	8.61%

C. In table 5 we present results of the SBE estimation from the 2000 Census Dress Rehearsal in Sacramento and in Columbia , S.C. It is important to note that in Columbia we did not do any estimation but since all the questionnaires and field procedures were identical to their counterparts in Sacramento, we applied the estimation procedures for research purposes only.

Table 5: Results from the 2000 Census Dress Rehearsal

Site	Persons enumerated		Estimate
Sacramento	Shelter	584	2443
	Soup Kitchen	500	
Columbia	Shelter	254	833
	Soup Kitchen	187	

VII. Analysis

In all the estimators the relative mean square error is relatively large (in comparison with the coefficient of variation). The reason for this is the considerable number of persons without usual residence to whom we gave a zero probability of using service facilities.

The shelter only estimator or the soup kitchen only estimator worked fine provided that most of the persons without usual residence use that type of facility. However, based on the results of the 2000 Census Dress Rehearsal, many persons may use soup kitchens but not shelters or shelters but not soup kitchens. A large number of the persons enumerated in soup kitchens answered that they did not use a shelter in the past week (for example 80% of those enumerated in soup kitchens in Sacramento answered that they did not use shelters in the past week).

Comparing the numbers between the 95 test estimator and the 2000 Census Dress rehearsal estimator, the Cvs and the RRMSEs from the 95 Test estimator produce slightly smaller values than their counterpart of the Dress Rehearsal estimator . The main difference between these two estimators is likely the response bias (not considered in this paper) in the 95 test estimator. For example, if a respondent has eaten three days in a soup kitchen and slept in a shelter the same three days, the correct answer to the 95 test question is three days but he or she may answer six. Considering all these estimators from a total error point of view, we feel that the 2000 Dress Rehearsal estimator is best and it is proposed for Census 2000.