Measure of Size in Large Scale Sampling

Yeanok Namgung Yoon  
National Statistical Office  
920 Dunsan-dong So-gu  
Taejon, Republic of Korea  
Yoon@nso.go.kr

Sang Eun Lee  
Kyonggi University  
San 94-6, Yiui-dong Paldal-gu Suwon-si  
Kyonggi-do, Republic of Korea  
Sanglee@stat.kyonggi.ac.kr

1. Introduction

Sampling designs are usually based on PPS(Probability Proportional to Size) considering measure of size(MOS) which gives self-weighting sampling. However in large scale sampling, sometimes PPS sampling may cause some difficulties in a practical field survey such as after census, when size of enumeration district(ED) is changed, allocating the ED to the Field Operator fairly and evenly was always a problem. Here we use the 1995 Household Survey(HS) sample design to see how the measure of size plays a roll on large scale survey. And that gives the solutions to allocate the ED’s to the Field Operator fairly.

ED consist of approximately 50~60 households. Each ED is separated several Sections and Section consists of 8 households. Then ED is composed of 6~8 Sections depending on the number of households. Very rarely 5 or 9 Sections are in the ED. In HS sample design, EDs were selected by systematic sampling assuming that each ED has the same number of Sections, that is, ignoring measure of size. And from the selected ED, 3 Sections are selected. Every month about 30,000 households are surveyed in HS. Survey items are above 15 years-old population, economically active population, employed and unemployed person etc.

2. Comparison of estimated value

To see how measure of size affects HS sampling, we compared the estimated values. The estimation formula with and without applying MOS are as follows.

(1) With MOS

\[ \hat{y} = \frac{X}{\lambda'} y' \]

(2) Without MOS

\[ \hat{y} = \frac{X}{\lambda''} y'' \]
Where $X$ is the estimated above 15 years-old total population  
$\chi^\prime$ is the examined above 15 years-old population  
$y^\prime$ is the examined value with characteristic $Y$  
e$_j$ is the number of sections in the $j$th ED  

$$y^\prime = \sum k e_j y_j$$
$$\chi^\prime = \sum k e_j \chi_j$$

$k = N/ (3n)$

Based on one month data of HS, three item’s summary data are given in Table 1. F-test for difference of the variance with and without MOS shows no significant differences in each item with less than 0.005 significant level. Next T-test for difference of estimated value for each item also shows no significant differences with less than 0.005 significant level.

<table>
<thead>
<tr>
<th>Item</th>
<th>Estimated value</th>
<th>Standard Error</th>
<th>F-value</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With</td>
<td>Without</td>
<td>With</td>
<td>Without</td>
</tr>
<tr>
<td>A</td>
<td>21,219,274</td>
<td>21,186,007</td>
<td>95,310</td>
<td>92,253</td>
</tr>
<tr>
<td>B</td>
<td>19,548,183</td>
<td>19,520,888</td>
<td>92,120</td>
<td>89,890</td>
</tr>
<tr>
<td>C</td>
<td>1,671,091</td>
<td>1,665,118</td>
<td>45,066</td>
<td>44,319</td>
</tr>
</tbody>
</table>

3. Conclusion

The size of HS is approximately 30,000 households. The estimated values without applying MOS are not different from those with applying MOS. Thus in large scale sample design, when the size sampling unit is not quite different, measure of size does not affect in the estimated value and standard error. Therefore specially in a large scale sampling design, MOS can be ignored and it will be better to managing the ED in more years.

FRENCH RÉSUMÉ

Cette thèse a relation à l'utilisation de Mesure de Taille au modèle de la grandeur large. Généralement, dans ce modèle design, Mesure de Taille est utilisé au modèle pour devenir PPS. Mais, au modèle de la grandeur large, parfois Mesure de Taille n’est pas utilisé. Alors, Cette thèse est parti pour étudier le rôle de Mesure de Taille dans ce modèle de la grandeur large. Il a été trouvé que Mesure de Taille n’a rien d’impact à l’égard du modèle enquête de la grandeur large quand la taille dela unité modèle est similaire.