Using New IT in Area Sampling: An Experience in Korea

Young-je Woo

Sun-Woong Kim

Dongguk University

Outline

- Concept of Area Sampling
- Advantage of Area Sampling
- Area Sampling in Korea
- Area Sampling Using New IT
- Application
- Results

Concept of Area Sampling

- Used when a adequate sampling frame or reference is not available
- The area to be covered is subdivided into a number of smaller sub-areas which are selected at random and then subsampled or fully surveyed
- Rather than lists or registers, maps serve as the sampling frame
- Basically multistage sampling

Advantage of Area Sampling

• Allows to sample household units in equal probability by providing proper sampling frames

Recent Area Sampling in Korea

- **It has been hard to conduct the surveys using area sampling.**
  - Lack of information on dwellings
    - The list of enumeration districts in Statistic Korea is not open to the public
  - Difficulties for listing dwellings
    - Dwelling without identification
    - Complicated building structures

Area Sampling Using New IT

- **New IT enables researchers to solve the existing problems in conducting area sampling in Korea.**
  - The information we can obtain through New IT
    - The number of dwellings on certain enumeration districts from Statistic Korea to the public
    - Location and specific address of buildings
    - Type of buildings

Application

- **Pilot study for Seoul Economy and Health Survey**

  - Target Population: Households in Jung-Gu, Seoul
  - Survey Mode: Face to face with CAPI or PAPI
  - Sampling Method: Four-Stage Area Sampling
  - Sample Size: 120 households
  - Survey Questions: 36 Total number of questions
    - Categories: residential and living environment, job condition, economic condition diseases.

Sampling Process

- **Choosing the Proper Sampling Units for Each Stage**

	First stage	Second stage	Third stage	Fourth stage
	Select 3 Dong's	Selecting 5 ED's from each Dong	Select 2 Chunk's from each ED	Select a Segment from selected part
f_h	$\frac{3Mos_{h\alpha}}{\sum Mos_{h\alpha}}$	$\frac{5Mos_{h\alpha\beta}}{\sum Mos_{h\alpha\beta}}$	$\frac{2Mos_{h\alpha\beta\gamma}}{\sum Mos_{h\alpha\beta\gamma}}$	$\frac{1}{\sum Mos_{h\alpha\beta\gamma}/4}$

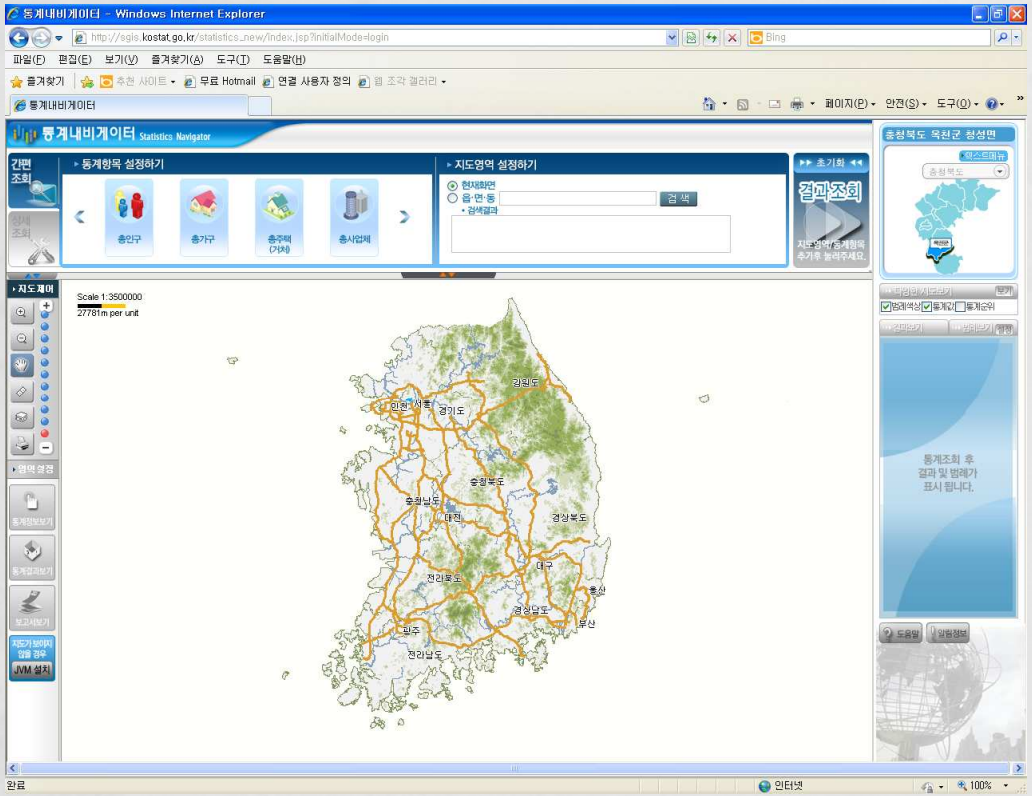
- Dong: Administrative unit
- ED: Enumerate district
- Chunk: A set of 24 dwellings
- Segment: A set of 4 dwellings

- **Selection Equation**

$$f_h = \frac{120}{N_h}$$

First Stage: Selecting Dong

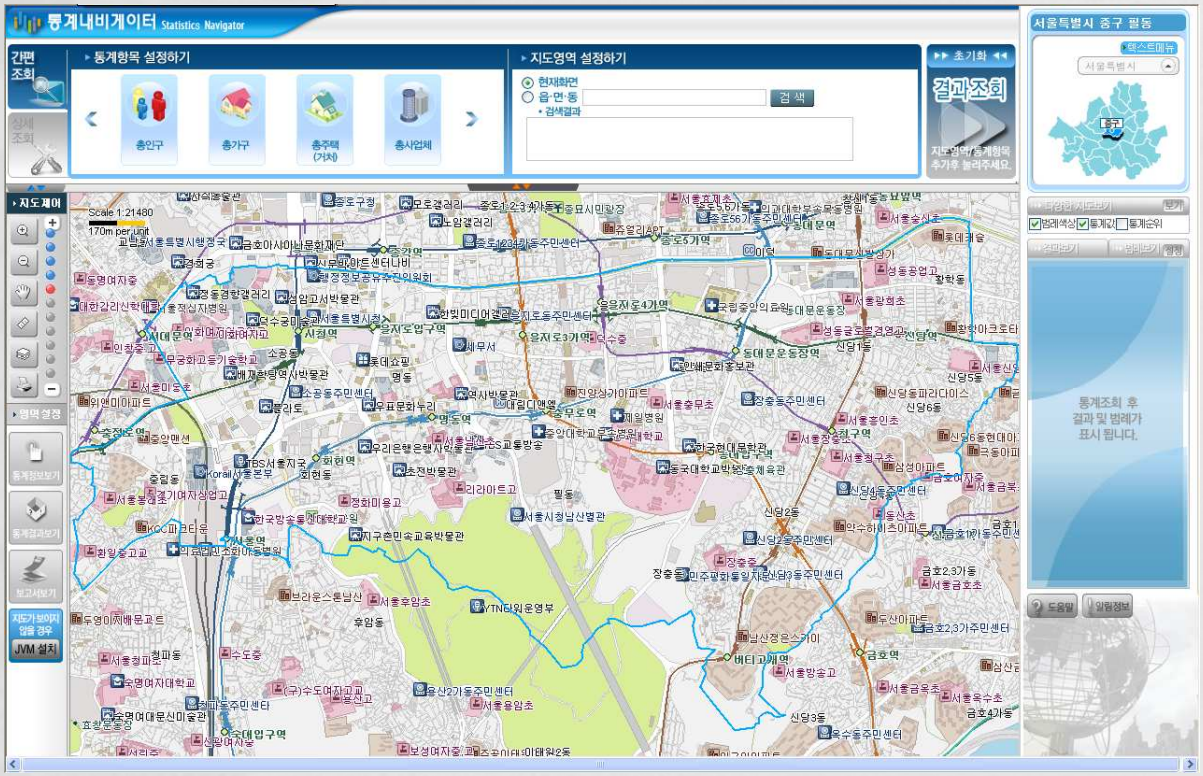
- **πPS sampling**
  - Using census data from Statistic Korea offered via the internet



<Figure.1> Census data

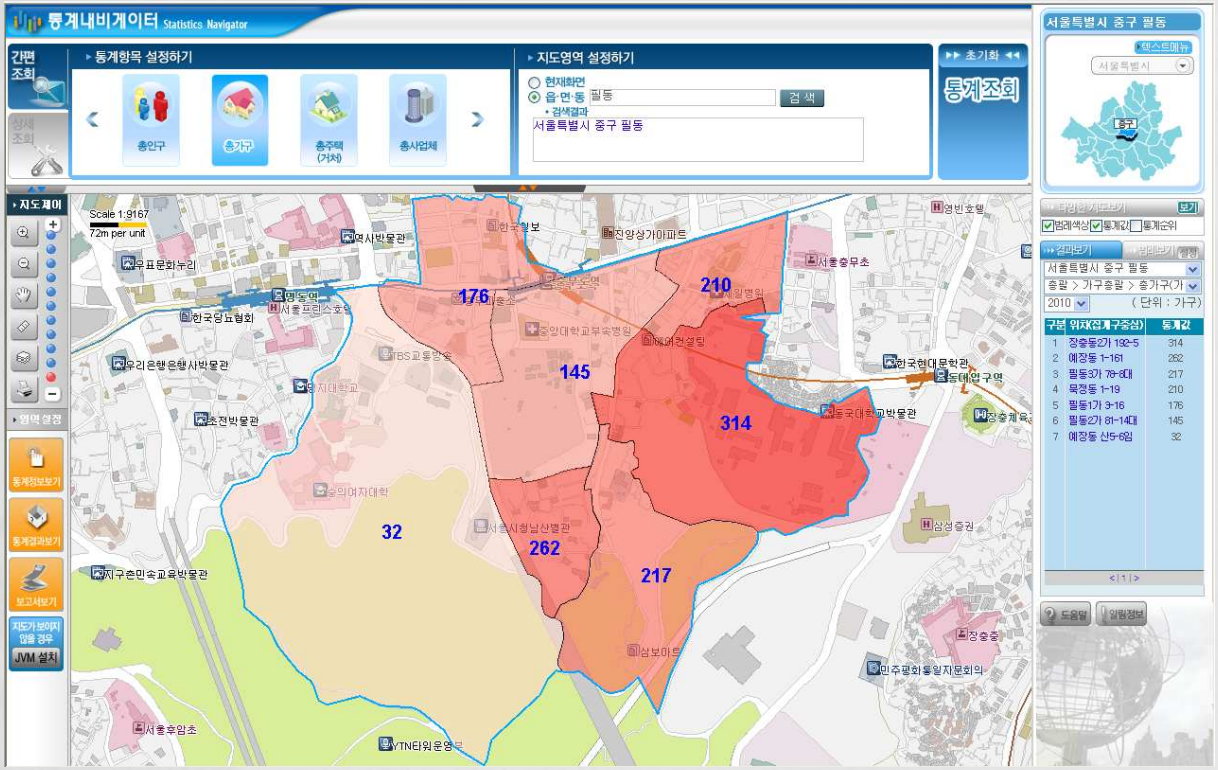
First Stage: Selecting Dong(Cont.)

- **πPS sampling**
  - From the census data, make a list of the number of dwellings for all of Dong within Jung Gu



First Stage: Selecting Dong(Cont.)

- **πPS sampling**
  - From the census data, make a list of the number of dwellings for all of Dong within Jung Gu



First Stage: Selecting Dong(Cont.)

- **πPS sampling**
  - From the census data, make a list of the number of dwellings for all of Dong within Jung Gu

Table.1 The list of the number of dwellings all of Dong within Jung Gu

ID	Dong	# of dwelling	ID	Dong	# of dwelling
1	Sogong	309	9	Sindang 2	5,590
2	Hoehyeon	2,681	10	Sindang 3	6,861
3	Myeong	1,166	11	Sindang 4	5,402
4	Pil	1,737	12	Sindang 5	3,783
5	Jangchung	2,433	13	Sindang 6	3,489
6	Gwanghui	2,043	14	Hwanghak	2,691
7	Euljiro	671	15	Jungnim	4,596
8	Sindang 1	3,110	Total		46,562

First Stage: Selecting Dong(Cont.)

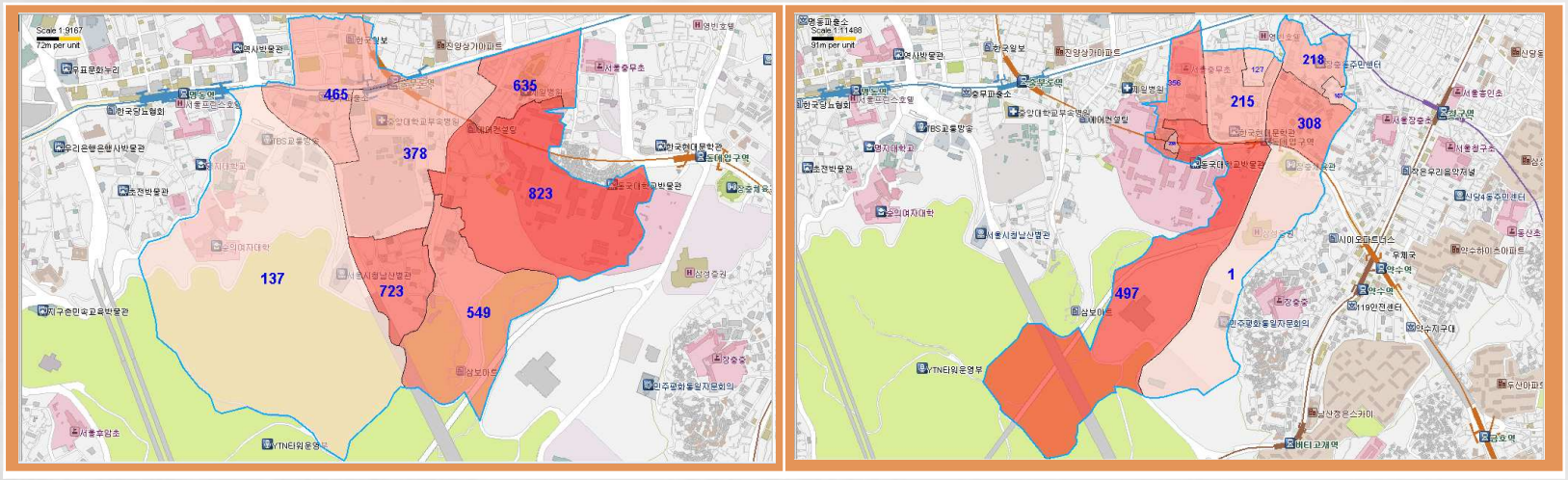
- **πPS sampling**
  - The sampled Dong's are shown Table.2

Table.2 The list of Dong's sampled with πPS

	Dong	# of dwelling	Selection Probability	Sampling Weight
5	Jangchung	2,433	0.1568	6.3792
10	Sindang 3	6,861	0.4421	2.2622
12	Sindang 5	3,783	0.2437	4.1027

Second Stage: Selecting ED(Cont.)

- **πPS sampling**
  - Statistic Korea provides the data of the Dong divided into ED
  - Make lists of the number of dwellings for all ED from each sampled Dong's



Second Stage: Selecting ED(Cont.)

- **πPS sampling**
  - Statistic Korea provides the data of the Dong divided into ED
  - Make lists of the number of dwellings for all ED from each sampled Dong's

Jangchung		Sindang 5				Sindang 3					
ED	# of dwellings	ED	# of dwellings	ED	# of dwellings	ED	# of dwellings	ED	# of dwellings	ED	# of dwellings
1	483	1	304	11	187	1	351	11	248	21	193
2	345	2	276	12	175	2	341	12	248	22	182
3	328	3	260	13	166	3	337	13	240	23	173
4	270	4	240	14	146	4	336	14	234	24	147
5	216	5	240	15	144	5	324	15	230	25	144
6	198	6	239	16	142	6	320	16	215	26	142
7	192	7	232	17	139	7	315	17	211	27	126
8	155	8	231	18	135	8	299	18	199	28	112
9	140	9	205	19	120	9	298	19	199	29	110
10	106	10	202			10	286	20	197	30	104

Second Stage: Selecting ED(Cont.)

- **πPS sampling**
  - The sampled district's are shown Table.4

Table.4 The list of sampled ED

Jangchung		Sindang 3		Sindang 5	
ED	# of dwellings	ED	# of dwellings	ED	# of dwellings
1	483	3	337	2	276
2	345	11	248	5	218
4	270	13	240	11	187
5	216	17	211	17	139
6	198	26	142	18	135

Third Stage: Selecting Chunk(Cont.)

- It is hard to make up a chunk because of lack of information
  - the list of dwellings is not open to the public
  - there is no information on how many dwellings are in a certain building
- The information we can obtain
  - the information about location and address of every buildings via the internet map service
  - the number of dwellings on a certain district

Third Stage: Selecting Chunk(Cont.)

- Make up a chunk approximately by using
  - the number of dwellings on the district
  - the number of buildings
- Algorithm for making up a chunk consists of two phase
  - decision algorithm for the number of chunk
  - decision algorithm for the number of buildings in each chunk

Third Stage: Selecting Chunk(Cont.)

- **Decision Algorithm for the number of chunk**

Step1) Divide *'the number of dwellings'* for each ED into 24*(# of dwelling per chunk)* that is, '# of dwellings ÷ 24'

Step2) If the value is an integer, using the value as the number of chunk. If not decide at random as followings

Step3) *Int(#of CH)*: the largest integer not greater than '# of dwellings ÷ 24' ,
Criteria:'# of dwellings ÷ 24' - Int(# of CH)

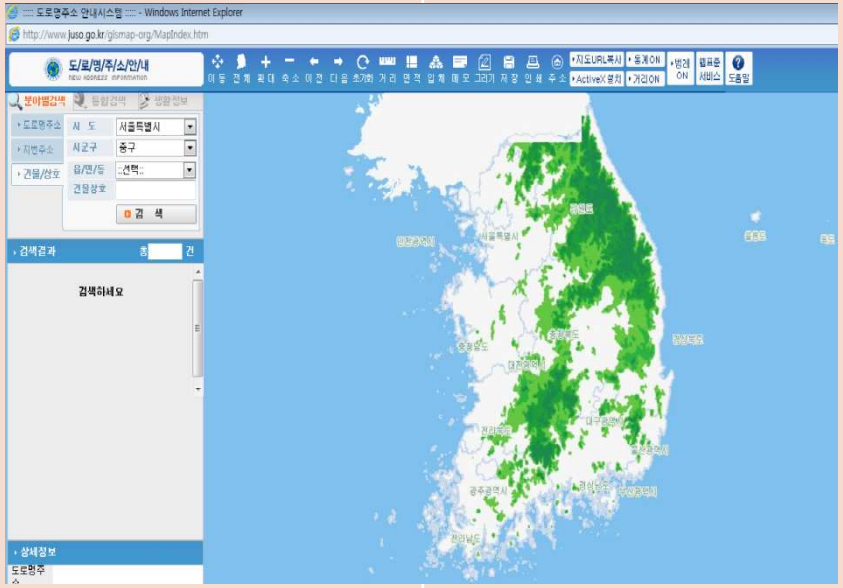
Step4) Generating a uniform(0,1) random variable (RN)

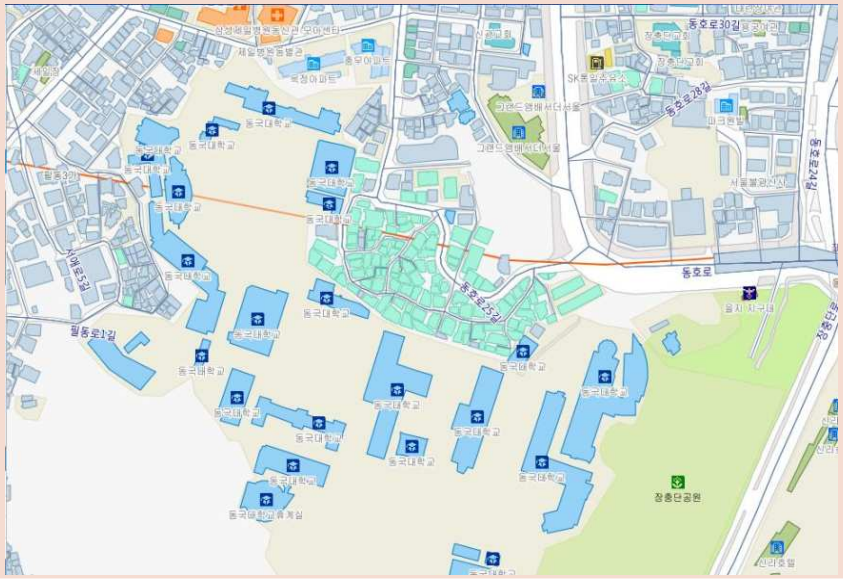
Step5) If Criteria > RN, the number of chunk is 'Int(# of CH)+1' , If not 'Int(# of CH)'

Third Stage: Selecting Chunk(Cont.)

- **Decision Algorithm for the Number of Buildings to Each Chunk**
  - Make a list of build address for each district
    - Ministry of Public Administration and Security provides New Address Information through the internet map service
    - the list of dwellings is not open to the public

Internet map service provided Ministry of Public Administration and Security





Third Stage: Selecting Chunk(Cont.)

- **Decision Algorithm for the Number of Buildings in Each Chunk**

Step.1) Divide ‘the Number of Buildings’ on sampled ED into its ‘the number of chunk’

Step.2) If the value is an integer, using the value as the number of buildings per chunk. If not, decide at random as followings

Step.3) (# of BD) is ‘the number of buildings ÷ the number of chunk’, Int(# of BD) is the largest integer not greater than ‘# of BD’, Diff: (# of BD) - Int(# of BD)

Step.4) Generating a uniform(0,1) random variable (RN_B)

Step.5) If Diff > RN_B then the number of buildings per chunk is ‘Int(# of BD)+1’. If not, ‘Int(# of BD)’

Step.6) make up the count of the number of buildings by adding or subtracting a building on chunks selected at random

Third Stage: Selecting Chunk(Cont.)

- **Example of Jangchung Dong, district 2**
  - **Decision for the number of chunk**
    - There are 345 dwellings in district 2

    1) Divide *'the number of dwellings'* into 24 *(# of welling per chunk)*, then '345 ÷ 24 =14.375'
    2) Generating a uniform random variable (RN =0.207)
    3) Criteria = 15-14.375=0.625. CR is greater than RN. Hence, the number of chunk is 15

Third Stage: Selecting Chunk(Cont.)

- **Decision for the Number of Buildings in Each Chunk**
  - Make a list of buildings for the Jangchung ED 2

<table > A list of building for the Jangchung ED 2

Jangchung					
ID	Adress	ID	Adress	ID	Adress
1	51	11	31	21	37-36
2	49	12	31-1	22	40-5
3	47	13	29	23	40
4	45-1	14	29-1	24	37-38
5	45	15	27	25	40-6
6	43	16	25	26	37-54
7	41	17	21-2	27	37-56
8	39	18	48	28	36
9	37	19	46		
10	35	20	42		

Third Stage: Selecting Chunk(Cont.)

- **Decision for the Number of Buildings in Each Chunk**
  - There are 28 buildings in district 2

    1) Devide the number of buildings into the number of chunk , that is '28 ÷ 15 =1.867'.

    2) Generating a uniform random variable(RN_B =0.708)

    3) Diff = 2-1.867=0.133. RN_B is greater than Diff. Hence, the number of build to per chunk is 2.

    4) But there are only 28 buildings, so we are short of 2 buildings. Hence select 2 chunks randomly and subtracting a building to make up the count

Third Stage: Selecting Chunk(Cont.)

- **Decision for the Number of Buildings in Each Chunk**
  - Selecting 2 chunks from the results of allocated buildings to jangchung ED 2 at random

<Table > allocated buildings to jangchung district2

ID	Adress	Allocated Chunk_id	ID	Adress	Allocated Chunk_id	ID	Adress	Allocated Chunk_id
1	51	1	11	31	6	21	37-36	11
2	49	1	12	31-1	7	22	40-5	12
3	47	2	13	29	7	23	40	12
4	45-1	2	14	29-1	8	24	37-38	13
5	45	3	15	27	8	25	40-6	14
6	43	4	16	25	9	26	37-54	14
7	41	4	17	21-2	9	27	37-56	15
8	39	5	18	48	10	28	36	15
9	37	5	19	46	10			
10	35	6	20	42	11			

Fourth Stage: Selecting Segment

- **Segment: A set of 4 dwellings**
  - Segments are formed heterogeneously using systematic selection(k=6)
  - A chunk consists of 6 segments
  - A segment is selected randomly from each chunk

<table> Composition of segment

	First	Second	third	fourth
segment1	1	7	13	19
segment2	2	8	14	20
segment3	3	9	15	21
segment4	4	10	16	22
segment5	5	11	17	23
segment6	6	12	18	24

Conclusion remarks

- This study shows how to conduct area sampling by using commercial maps, street view service and new address information map service via the Internet
- Using new IT, we can easily and correctly conduct area sampling procedure
- This methodology would lead to obtaining more reliable estimates