Using New IT for Area Sampling in a Metropolitan Household survey

Young-Je Woo

Sun-Woong Kim

So-Hyung Park

Sang-Eun Lee

Boyoon- Choi

Dongguk University
Survey & Health Policy Research Center(SHPRC)

OUTLINE

I. Area Sampling Using New IT
II. Description of the Metropolitan Household Survey
III. Study Design
IV. How to Conduct Area Sampling via New IT
V. Survey Results
VI. Concluding Remarks

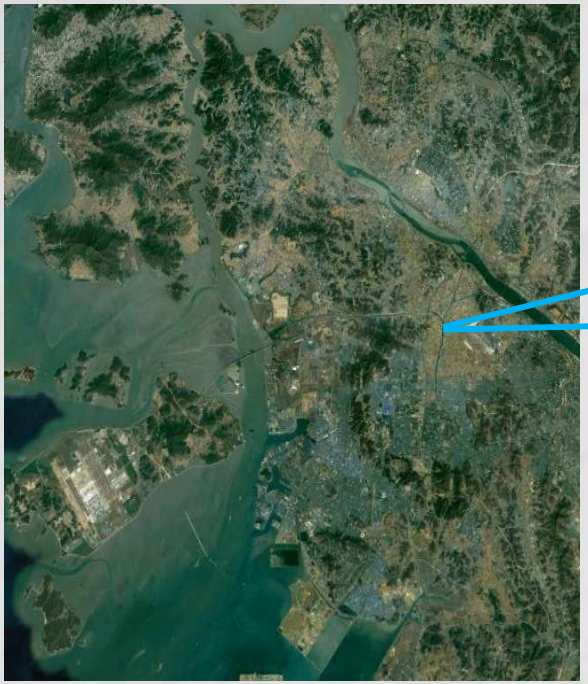
Using New IT for Area Sampling?

- Developed from Survey & Health Policy Research Center(SHPRC) in the 2011 based on the theory of sample design(2012 JSM, San diego)
- Using Information Technology(IT) such as….. enables researchers to conduct Area Sampling more efficient with respect to its own weakness(time, effort and cost)

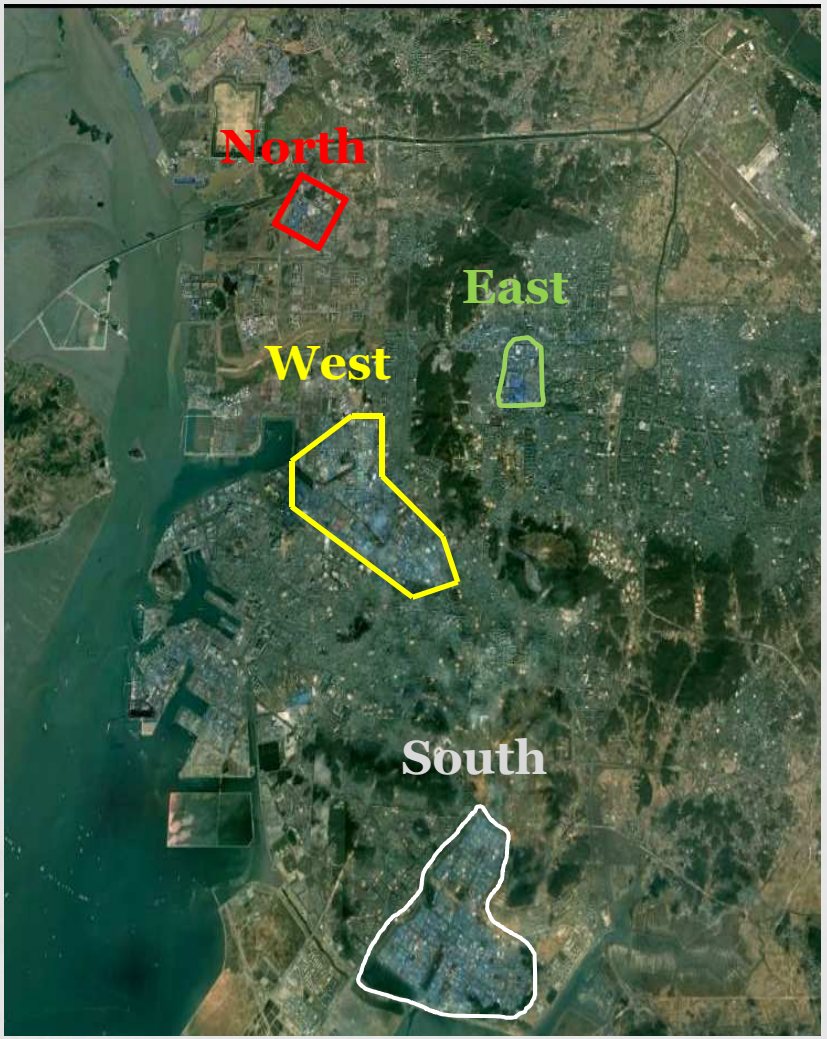
Description of the Metropolitan Household Survey

- Purpose: To investigate health problems among residents from communities surrounding industrial complexes in Incheon, South Korea
- Target Population: Residents in Incheon
- Survey Population: Residents living near the industrial complexes in Incheon
- Survey mode: CAPI(Computer Assisted Personal Interviewing)
- Sampling method: Four-stage area sampling
- Sample size: 900 households

Geographical Location of Industrial Complexes in Incheon

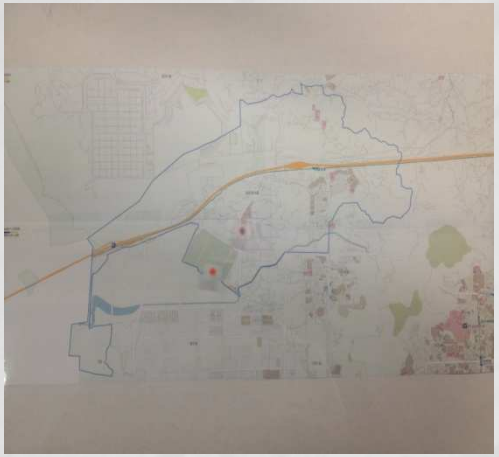


Incheon



Target population

- The population distribution of communities surrounding industrial complexes
  - Population size: 573,796(2010 Census)
  - Number of households: 199,328(2010 Census)
  - Four Industrial Complexes located



North Industrial Complex

East and West Industrial Complexes

South Industrial Complex

Summary of Survey Area

Area	Industrial complex	층1*	층2*	합계
		해당지역 (인구수 / 가구수)	**해당지역** (인구수 / 가구수)	(인구수 / 가구수)
1	East	**갈산1동** (17,728 / 6,265) **작전2동** (28,312 / 10,314) **부평1동** (34,327 / 12,512) **산곡4동** (20,174 / 6,391) **청천2동** (37,151 / 12,480)	**갈산2동** (22,983 / 7,964) **산곡2동** (32,328 / 9,656) **효성2동** (34,322 / 11,002) **청천1동** (15,583 / 6,306)	(242,908 / 82,890)
2	West	**가좌3동** (20,276 / 7,115) **도화 2·3동** (20,686 / 7,821) **가좌1동** (13,295 / 4,999) **가좌2동** (22,885 / 7,156) **십정1동** (20,952 / 7,721) **송림4동** (5,378 / 1,873)	**주안5동** (23,142 / 8,654) **송림6동** (7,601 / 2,662) **송현3동** (4,735 / 1,923) **가좌4동** (12,054 / 4,474) **간석4동** (29,548 / 10,833)	(180,552 / 65,231)

Summary of Survey Area(Cont.)

3	North	**검암경서동**** (5,841 / 1,750)			(5,841 / 1,750)
4	South	**논현고잔동**** (33,546 / 11,346) **동춘3동** (19,748 / 5,577) **남촌도림동** (23,827 / 8,073) **선학동** (21,685 / 8,590)	**동춘2동** (21,793 / 6,743) **연수2동** (23,896 / 9,128)		(144,495 / 49,457)

How to Conduct Area Sampling Using New IT

▶ Selection Procedure of Area Sampling

	First stage	*Second stage*	*Third stage*	*Fourth stage*
Sampling Unit	City	Enumeration district	Chunk	Segment
	Primary selection of city	Select EDs from sel ected cities	Select 2 chunks from each selected ED	Select a segment from selected chunk
f_h	$\frac{2\,Mos_{h\alpha}}{\sum_{\alpha} Mos_{h\alpha}}$	$\frac{d_{h\alpha} Mos_{h\alpha\beta}}{\sum_{\beta} Mos_{h\alpha\beta}}$	$\frac{2 Mos_{h\alpha\beta\gamma}}{\sum_{\gamma} Mos_{h\alpha\beta\gamma}}$	$\frac{Mos_{h\alpha\beta\gamma\delta}}{\sum_{\delta} Mos_{h\alpha\beta\gamma\delta}}$

Chunk: a set of 24 HU's
Segment: a set of 4 HU‘s

How to Conduct Area Sampling Using New IT(Cont.)

▶ Selection Probability

$$f = \frac{n}{N}$$

$$f_h = \frac{2\,Mos_{h\alpha}}{\sum_{\alpha} Mos_{h\alpha}} \times \frac{d_{h\alpha}\,Mos_{h\alpha\beta}}{\sum_{\beta} Mos_{h\alpha\beta}} \times \frac{2\,Mos_{h\alpha\beta\gamma}}{\sum_{\gamma} Mos_{h\alpha\beta\gamma}} \times \frac{Mos_{h\alpha\beta\gamma\delta}}{\sum_{\delta} Mos_{h\alpha\beta\gamma\delta}}$$

$$= \frac{2\,Mos_{h\alpha}}{\sum_{\alpha} Mos_{h\alpha}} \times \frac{(d_{h\alpha} \times Mos^{*}_{h\alpha\beta} \times 24)}{Mos_{h\alpha}} \times \frac{2}{Mos^{*}_{h\alpha\beta}} \times \frac{4}{24}$$

$$= \frac{2\,Mos_{h\alpha}}{\sum_{\alpha} Mos_{h\alpha}} \times \frac{(d_{h\alpha} \times 24)}{Mos_{h\alpha}} \times 2 \times \frac{1}{6}$$

$$= f$$

First Stage: Primary Selection of City

- Using πPS sampling select 2 cities from each stratum

Area	Industrial Complex	Stratum1	Stratum2
		City (# of population/# of household)	City (# of population/# of household)
1	East	Gal-san1 (17,728/6,265) Jak-jun2(28,312/10,314)	Hyo-sung2 (34,322/11,002) Chung-chun1 (15,583/6,306)
2	West	Ga-ja3 (20,276/7,115) Do-hwa 2·3 (20,686/7,821)	Ga-ja4 (12,054/4,474) Gan-suk4 (29,548/10,833)
3	Nort	Kyung-seo(5,841/1,750)	
4	South	Go-jan(33,546/11,346) Dong-chun3 (19,748/5,577)	Dong-chun2(21,793/6,743) Yun-su2(23,896/9,128)

▶ Equal Probability Selection for each stratum

$$d_{h\alpha} = \left(\frac{2Mos_{h\alpha}}{\sum_{\alpha} Mos_{h\alpha}} \right)^{-1} \times \left(\frac{24}{Mos_{h\alpha}} \right)^{-1} \times 3f$$

City	# of Enumeration District	# of chunk	# of Household	Total Households
Gal-san1	6	2	4	48
Jak-jun2	6	2	4	48
Ga-ja3	14	2	4	112
Do-hwa 2·3	14	2	4	112
Kyung-seo	2	2	4	16
Go-jan	12	2	4	96
Dong-chun3	12	2	4	96
Chung-chun1	6	2	4	48
Hyo-sung2	7	2	4	56
Ga-ja4	11	2	4	88
Gan-suk4	11	2	4	88
Dong-chun2	6	2	4	48
Yun-su2	6	2	4	48

Second Stage: Select Enumeration districts

- **πPS sampling**
  - Statistic Korea provides the data of the city divided into ED
  - Make the list of the number of dwellings for all ED from selected Dong

The list of sampled ED

Gal-san1		Jak-jun2		Jak-jun2	
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: Select Chunks

- **Making up the Chunk**
  - Divide each block into ‘chunks’ with approximately 24 occupied housing units
  - $\dfrac{\text{The number of HU's}}{24} = \text{the number of chunks}$
  - $\dfrac{\text{The number of buildings}}{\text{The number of chunks}} = \text{The number of buildings per chunk}$

Fourth Stage: Select Segment

- **Selecting segment**
  - Select one chunk at random from each selected block
  - Listing all dwelling units on selected chunk
  - Divide chunks into compact segments of size 4 occupied HU's each
  - Select one compact segment from each chunk

Survey Results

▶ Population and Sample Distribution

▪Gender

Gender	Population	Sample
Male	50.1%	48.0%
Female	49.9%	52.0%
total	100%	100%

▪Age

Age	Population	Sample
Under 12	18.0%	15.6%
20~64	73.5%	68.8%
Over 65	8.5%	15.6%
Total	100.0	100.0

Concluding Remarks

- Area sampling using New IT can be an alternative for saving time and effort
- It is possible to select sample in equal probability, even though there is no proper sampling frame for household.
- This approach would be applicable to national surveys

Thank you