

# 집계구 표본설계 평가

지침버전: v 1.4

작성일자: 2017.04.24

수정내용: 1. 3.7절 추가

## 1. 개요

□ 집계구를 이용한 가구조사 표본설계의 효율성 연구

## 2. 집계구 파일

### 2.1 집계구 표본틀

#### □ 파일명

전국 집계구 정보.xlsx

#### □ 탭

<u>탭(TAB)</u>	<u>내용</u>
data	집계구 정보
변수설명	변수정보

### 2.2 관심변수

#### □ 설계변수

<u>변수명</u>	<u>값(비고)</u>
sido	서울/부산/.../제주
division	수도권/충청권/호남권/대경권/동남권/강원권
urban	동부/읍면부
jgg	집계구번호 (=code)
jggapt	아파트 가구포함여부 <sup>1)</sup> (=1 아파트,=0 아님) (집계구내)
pop	총인구수 (집계구내)
hh	총가구수 (집계구내)
du	총거처수 (집계구내)
es	총사업체수 (집계구내)

#### □ 특성변수

<u>변수명</u>	<u>값(비고)</u>
in_xxx_xxx	인구관련변수
ga_xx_xxx	가구관련변수
ho_xx_xxx	주택관련변수
cp_xxx_xxx	사업체관련변수

---

1) 집계구내 한 채의 아파트가 있어도 아파트로 구분

### 3. 분석

#### 3.1 개요

□ Valliant *et al.*, (2015) 2장 참고

#### 3.2 0-1 특성

□ 특성지시자 (예, 1인 가구 여부)

$$y_k = 1 (\text{특성}), = 0 (o.w.) \quad (k=\text{단위})$$

$$A_i = \text{특성비율(proportion)} \quad (i=\text{집계구})$$

□ 집계구 총합, 모총합, 집계구 분산

$$t_i = N_i A_i$$

$$t_U = \sum_{i \in U} t_i = \sum_{i \in U} N_i A_i,$$

$$A_U = t_U / N (= \bar{y}_U)$$

$$S_{U2i}^2 = \frac{N_i}{N_i - 1} A_i (1 - A_i)$$

□ 특성분산

$$S_U^2 = \frac{N}{N-1} A_U (1 - A_U)$$

#### 3.3 Sampling Strategy I

##### 3.3.1 표본설계 및 설계가중치

□ 표본층: 해당무

□ PWR/SRS 추출: 집락크기비례 및 집락당 동일크기 표본, 즉,

$$m = 94 \text{ (집계구)}, \quad p_i = N_i / N, \quad n_i = \bar{n} \text{ (= 20 가구)}$$

□ 표본(설계)가중치: 자체가중

$$w_{ik} = w_i w_{k|i} = \frac{1}{mp_i} \frac{N_i}{n_i} = \frac{N}{mn}$$

### 3.3.2 설계요소평가

□ 상대분산분해

－ 분석변수목록

[인구특성]

in\_age\_10\_19 (10-19세 인구수)

in\_age\_65\_00 (65세 이상 인구수)

in\_univ (대졸학력 이상 인구수)

[가구특성]

ga\_own (자가 가구수)

ga\_shh (1인 가구수)

－ Valliant et al. (2015) 식 (9) 이용

변수	$A_U$	$B_*^2$	$W_*^2$	$\tilde{V}$	$k_*$	$\delta_*$	$RV$	$def f$	$keff$

여기서

$$RV = \frac{V(\hat{t}_{pur})}{t_U^2}$$

$$def f = 1 + \delta_*(\bar{n} - 1)$$

$$kdef f = k_* \times def f$$

□ 시각화분석 (추후)

### 3.4 Sampling Strategy II

#### 3.4.1 표본설계 및 설계가중치

□ 표본층: 해당무

□ SRS/SRS 추출: 단순임의집락추출 및 집락당 동일크기 단순임의추출, 즉,

$$m = 94 \text{ (집계구)}, \quad n_i = \bar{n} \text{ (= 20 가구)}$$

□ 표본(설계)가중치: 자체가중

$$w_{ik} = w_i w_{k|i} = \frac{M}{m} \frac{N_i}{n_i} = \frac{MN_i}{mn}$$

#### 3.4.2 설계요소평가

□ 상대분산분해

- 분석변수목록 (3.3.2절 참조)
- Valliant et al. (2015) 식 (7) 이용

변수	$A_U$	$B^2$	$W^2$	$\tilde{V}$	$\tilde{k}$	$\tilde{\delta}$	$RV$	$deff$	$keff$

여기서

$$RV = \frac{V(\hat{t}_\pi)}{t_U^2}$$

$$deff = 1 + \tilde{\delta}(\bar{n} - 1)$$

$$keff = \tilde{k} \times deff$$

□ 시각화분석 (추후)

### 3.5 Sampling Strategy III

#### 3.5.1 표본설계 및 설계가중치

□ 표본층: 6대권역

－ 2.2절 division 변수 사용

수도권: 서울, 인천, 경기  
충청권: 대전, 충북, 충남  
호남권: 광주, 전북, 전남, 제주  
대경권: 대구, 경북  
동남권: 부산, 울산, 경남  
강원권: 강원

□ 층별 PWR/SRS 추출: 집락크기비례 및 집락당 동일크기 표본, 즉,

－ 전국기준 전체 표본집계구  $m=94$ 를 층별로 다음과 같이 할당

수도권 29  
충청권 12  
호남권 16  
대경권 14  
동남권 15  
강원권 8

－ 층별 크기비례확률 및 집락당 표본크기

$$p_{hi} = N_{hi} / N_h, \quad n_{hi} = \bar{n} \quad (= 20 \text{ 가구})$$

여기서  $N_{hi}$ 는  $h$ 층  $i$ 번째 집락내 가구수,  $N_h = \sum_{i=1}^{M_h} N_{hi}$ 는  $h$ 층 (모집단내) 가구총수,  $M_h$ 는  $h$ 층 집락 총수

□ 표본(설계)가중치: 자체가중

$$w_{hik} = w_{hi} w_{k|hi} = \frac{1}{m_h p_{hi}} \frac{N_{hi}}{n_{hi}} = \frac{N_h}{m_h \bar{n}}$$

### 3.5.2 설계요소평가

#### □ 상대분산분해

- 분석변수목록 (3.3.2절 참조)
- 변수별로 Valliant et al. (2015)의 식 (9)을 이용하여 층 및 전체에 적용

층	$N_h$	$M_h$	$N_h/M_h$	$A_h$	$B_{*h}^2$	$W_{*h}^2$	$\tilde{V}_h$	$k_{*h}$	$\delta_{*h}$	$RV_h$	$deff_h$	$kdeff_h$
수도권												
:												
강원권												
전국					—	—		—	—		—	—
비교*												

단, \* 비교는 3.3.2절의 값으로 채울 것

$$RV_h = \frac{V(\hat{t}_{h,pwr})}{t_{U_h}^2};$$

$$deff_h = 1 + \delta_{*h}(\bar{n}_h - 1) = 1 + \delta_{*h}(\bar{n} - 1); \quad kdeff_h = k_{*h} \times deff_h$$

$$t_U = \sum_{\mathcal{J}} N_h A_h; \quad A_U = \frac{t_U}{N} = \sum_{h \in \mathcal{J}} \left( \frac{N_h}{N} \right) A_h$$

$$\tilde{V}_U = \frac{S_U^2}{A_U^2}; \quad RV_U = \frac{\sum_{h=1}^H V(\hat{t}_{h,pwr})}{t_U^2}$$

#### □ 시각화분석 (추후)

### 3.6 Sampling Strategy IV

#### 3.6.1 표본설계 및 설계가중치

□ 표본층: 6대 권역 x 동/읍면

－ 층화변수 1: 2.2절 division 변수 사용

수도권: 서울, 인천, 경기  
 충청권: 대전, 충북, 충남  
 호남권: 광주, 전북, 전남, 제주  
 대경권: 대구, 경북  
 동남권: 부산, 울산, 경남  
 강원권: 강원

－ 층화변수 2: 2.2절 urban 변수 사용

동부  
 읍면부

□ 층별 PWR/SRS 추출: 집락크기비례 및 집락당 동일크기 표본, 즉,

－ 전국기준 전체 표본집계구  $m=94$ 를 층별로 다음과 같이 할당

권역	동읍면 구분		합계
	동부	읍면부	
수도권	28	1	29
충청권	8	4	12
호남권	11	5	16
대경권	10	4	14
동남권	12	3	15
강원권	5	3	8
전국	74	20	94

－ 층별 크기비례확률 및 집락당 표본크기

$$p_{hi} = N_{hi} / N_h, \quad n_{hi} = \bar{n} \quad (= 20 \text{ 가구})$$

여기서  $N_{hi}$ 는  $h$ 층  $i$ 번째 집락내 가구수,  $N_h = \sum_{i=1}^{M_h} N_{hi}$ 는  $h$ 층 (모집단내) 가구총수,  $M_h$ 는  $h$ 층 집락 총수



□ 표본(설계)가중치: 자체가중

$$w_{hik} = w_{hi} w_{k|hi} = \frac{1}{m_h p_{hi}} \frac{N_{hi}}{n_{hi}} = \frac{N_h}{m_h \bar{n}}$$

### 3.6.2 설계요소평가

□ 상대분산분해

- 분석변수목록 (3.3.2절 참조)
- 변수별로 Valliant et al. (2015)의 식 (9)을 이용하여 층 및 전체에 적용

층		$N_h$	$M_h$	$N_h/M_h$	$A_h$	$B_{*h}^2$	$W_{*h}^2$	$\tilde{V}_h$	$k_{*h}$	$\delta_{*h}$	$RV_h$	$deff_h$	$kdeff_h$
수도권	동부												
수도권	읍부												
:													
강원권	동부												
	읍부												
전국						—	—		—	—		—	—
비교*													

단, \* 비교는 3.3.2절의 값으로 채울 것

$$RV_h = \frac{V(\hat{t}_{h,pwr})}{t_{U_h}^2};$$

$$deff_h = 1 + \delta_{*h}(\bar{n}_h - 1) = 1 + \delta_{*h}(\bar{n} - 1); \quad kdeff_h = k_{*h} \times deff_h$$

$$t_U = \sum_j N_h A_h; \quad A_U = \frac{t_U}{N} = \sum_{h \in \mathcal{J}} \left( \frac{N_h}{N} \right) A_h$$

$$\tilde{V}_U = \frac{S_U^2}{A_U^2}; \quad RV_U = \frac{\sum_{h=1}^H V(\hat{t}_{h,pwr})}{t_U^2}$$

□ 시각화분석 (추후)

### 3.7 Sampling Strategy V

#### 3.7.1 표본설계 및 설계가중치

□ 표본층: 16개 특별시, 광역시, 도 x 동/읍면

- 층화변수 1: 2.2절 sido 변수 사용
- 층화변수 2: 2.2절 urban 변수 사용

동부  
읍면부

□ 층별 PWR/SRS 추출: 집락크기비례 및 집락당 동일크기 표본, 즉,

- 전국기준 전체 표본집계구  $m=94$ 를 층별로 다음과 같이 할당

시도	동읍면 구분		합계
	동부	읍면부	
서울	12	0	12
부산	5	2	7
인천	4	0	4
대구	6	0	6
광주	4	0	4
대전	4	0	4
울산	1	1	2
경기	10	2	12
강원	6	2	8
충북	3	1	4
충남	3	3	6
전북	3	2	5
전남	2	2	4
경북	4	3	7
경남	5	2	7
제주	1	1	2
전국	73	21	94

- 층별 크기비례확률 및 집락당 표본크기

$$p_{hi} = N_{hi} / N_h, \quad n_{hi} = \bar{n} \quad (= 20 \text{ 가구})$$

여기서  $N_{hi}$ 는  $h$ 층  $i$ 번째 집락내 가구수,  $N_h = \sum_{i=1}^{M_h} N_{hi}$ 는  $h$ 층 (모집단내) 가구총수,  $M_h$ 는  $h$ 층 집락 총수

□ 표본(설계)가중치: 자체가중

$$w_{hik} = w_{hi} w_{k|hi} = \frac{1}{m_h p_{hi}} \frac{N_{hi}}{n_{hi}} = \frac{N_h}{m_h \bar{n}}$$

### 3.7.2 설계요소평가

□ 상대분산분해

- 분석변수목록 (3.3.2절 참조)
- 변수별로 Valliant et al. (2015)의 식 (9)을 이용하여 층 및 전체에 적용

층		$N_h$	$M_h$	$N_h/M_h$	$A_h$	$B_{*h}^2$	$W_{*h}^2$	$\tilde{V}_h$	$k_{*h}$	$\delta_{*h}$	$RV_h$	$deff_h$	$kdeff$
서울	동부												
부산	동부												
:													
강원	동부												
	읍부												
전국						—	—		—	—		—	—
비교*													

단, \* 비교는 3.3.2절의 값으로 채울 것

$$RV_h = \frac{V(\hat{t}_{h,pwr})}{t_{U_h}^2};$$

$$deff_h = 1 + \delta_{*h}(\bar{n}_h - 1) = 1 + \delta_{*h}(\bar{n} - 1); \quad kdeff_h = k_{*h} \times deff_h$$

$$t_U = \sum_{\mathcal{J}} N_h A_h; \quad A_U = \frac{t_U}{N} = \sum_{h \in \mathcal{J}} \left( \frac{N_h}{N} \right) A_h$$

$$\tilde{V}_U = \frac{S_U^2}{A_U^2}; \quad RV_U = \frac{\sum_{h=1}^H V(\hat{t}_{h,pwr})}{t_U^2}$$

□ 시각화분석 (추후)