



한국복지패널 가중치 조정 과정

신재동 · 오미애 (한국보건사회연구원)



1. 한국복지패널 가중치 부여체계



한국복지패널 표본변화

-
- A vertical timeline showing the evolution of the Korean Welfare Panel from 2005 to 2018. The timeline is marked with years on the left and specific survey events on the right. The events are represented by blue and dark blue horizontal bars pointing to the right. The bars alternate in color: 2005 (dark blue), 2006 (blue), 2011 (dark blue), 2012 (blue), 2013 (dark blue), 2014 (blue), 2015 (dark blue), 2016 (blue), 2017 (dark blue), 2018 (blue).
- 2005년 ● 3개 패널 통합(빈곤패널, 자활패널, 복지패널) 논의
 - 표본추출방식 논의(통합표본), 2005년 국민생활실태조사 자료 이용 결정
 - 일반 및 저소득 계층별로 3,500 가구씩 표본추출
 - 2006년 ● 1차 조사 7,072 가구의 패널을 구축 - 아동부가조사
 - 2011년 ● 6차 조사 5,735 가구 조사 완료
 - 2012년 ● 7차 조사 5,732가구 조사 완료 + 신규표본 1,800가구 조사
 - 2013년 ● 8차 조사 7,312 가구 조사 완료
 - 2014년 ● 9차 조사 7,048 가구 조사 완료
 - 2015년 ● 10차 조사 6,914 가구 조사 완료
 - 2016년 ● 11차 조사 6,723 가구 조사 완료
 - 2017년 ● 12차 조사 6,581 가구 조사 완료
 - 2018년 ● 13차 조사 6,474 가구 조사 완료

모집단

N : 10



표본

N : 6



모수 가중치

[남 : 1.5]

[여 : 2]

[총합 : 10]

표본 가중치

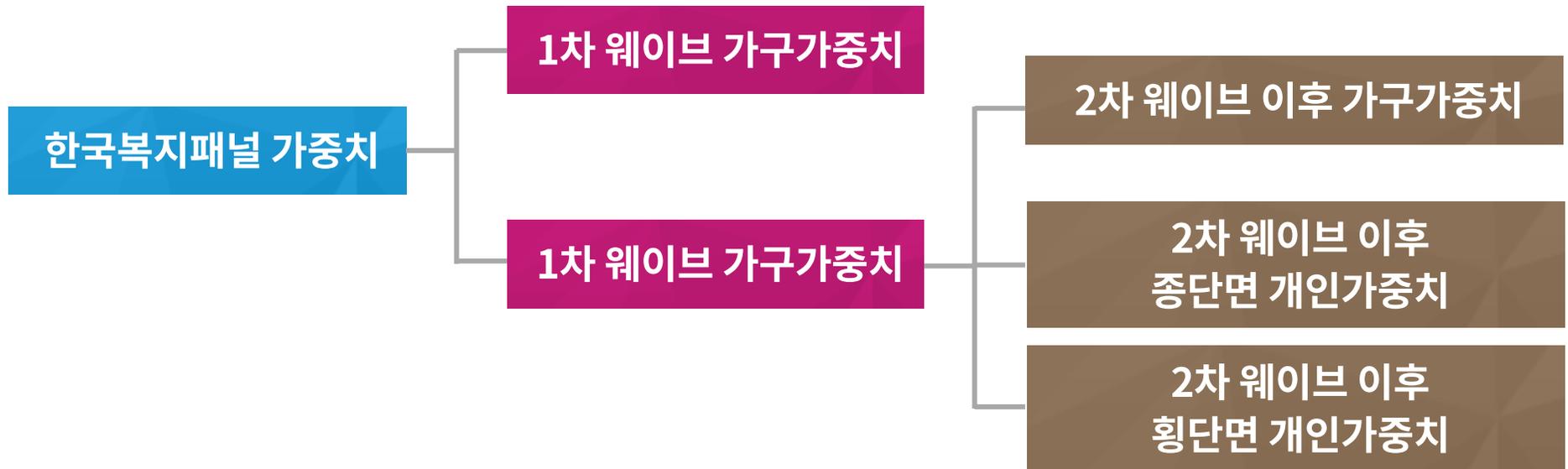
[남 : 0.9]

[여 : 1.2]

[총합 : 6]



한국복지패널 가중치 부여체계



가중치 부여 절차

- 가구가중치 (household weight)는 횡단면과 종단면 가중치를 구분하지 않고 단일가중치를 부여하며, 개인가중치 (person weight) 만 횡단면 가중치와 종단면 가중치로 구분하여 부여
- 1차 웨이브에 부여된 개인가중치를 바탕으로 개인조사표에 응답한 가구원에 대해 가중치 조정을 수행
- 종단면 가중치는 로지스틱 회귀분석을 통해 각 개인의 응답확률을 추정하여 적용

가구가중치 산출 과정

- t차 웨이브 가구가중치는 t차 웨이브 개인 종단면 가중치의 가구 내 평균값을 계산하여 산출
- 웨이브별로 새로 생성된 신규 가구의 경우, 가구 내 원표본 가구원의 가중치 평균을 적용하고, 원표본 가구원만으로 구성된 가구는 이들의 평균으로 가구가중치를 산출
- 조사 기준연도 인구추계 가구수를 기준으로 지역 및 가구규모별로 사후조정

개인가중치 산출 과정

- 2차년도 이후에는 t 차년도 종단면 가중치를 기본 가중치로 고려하여 $t+1$ 차년도 원표본 해당여부 판정
- 무응답 및 개인별 변동 상황을 반영하여 각 년도 종단면 기본 가중치 보정
- 사후조정

개인 종단가중치 산출 과정

- 패널 탈락으로 인한 무응답 보정을 위해, t차 웨이브 종단면 응답여부 변수와 t-1차 웨이브 변수들의 관계를 로지스틱 회귀모형을 이용하여 응답확률을 추정
- 로지스틱 회귀모형의 설명변수로 응답자의 성별, 연령, 지역, 경제활동상태 사용
- t차 웨이브 종단면 응답여부 변수는 t차와 t-1차 웨이브 모두 응답한 경우 1, 그렇지 않은 경우는 0을 부여
- t차 웨이브의 개인 종단면 기본가중치에 로지스틱 회귀모형으로 추정된 응답확률의 역수를 곱하여, t차 웨이브의 개인 종단면 가중치를 조정

개인 종단가중치 산출 과정

- 예를 들어, 13차 웨이브의 기본가중치를 $w_{\text{기본13차}}$, 로지스틱 회귀분석을 통해 예측된 추정 응답확률을 $\hat{p}_{13\text{차}}$ 라고 하면 무응답 보정을 통해 얻어지는 가중치는

$$w_{\text{기본13차}} = w_{\text{기본13차}} \times \hat{p}_{13\text{차}}^{-1}$$

- 개인별 변동 상황에 따라 t차년도 종단면 가중치를 조정
(신규 가구원의 경우 개인별 종단면 가중치 0값 존재)
- 지역 및 응답자의 인구학적 특성에 따라 t차년도 조사 기준연도 인구 추계 값을 이용하여 사후조정을 실시

개인 횡단가중치 산출 과정

- t차 웨이브의 개인 종단면 기본가중치에 로지스틱 회귀모형으로 추정된 응답확률의 역수를 곱하여 t차 웨이브의 개인 종단면 가중치를 조정
- 변동 상황에 따라 따라 t차년도 종단면 가중치를 조정
- t차년도 종단면 가중치 값이 0인 가중치에 대해 가구별 평균가중치를 적용
- 지역 및 응답자의 인구학적 특성(성별, 연령)에 따라 조사 기준연도 인구 추계 값을 기준으로 사후조정

2. 신규표본 추가에 따른 가중치 조정



신규표본 가중치 조정 필요성

-
- A vertical timeline showing the progression of survey waves and sample adjustments from 2005 to 2018. The timeline is marked with years on the left and corresponding survey details in blue and dark blue boxes on the right. The survey waves are numbered from 1st to 13th.
- 2005년 ● 3개 패널 통합(빈곤패널, 자활패널, 복지패널) 논의
 - 표본추출방식 논의(통합표본), 2005년 국민생활실태조사 자료 이용 결정
 - 일반 및 저소득 계층별로 3,500 가구씩 표본추출
 - 2006년 ● 1차 조사 7,072 가구의 패널을 구축 - 아동부가조사
 - 2011년 ● 6차 조사 5,735 가구 조사 완료
 - 2012년 ● 7차 조사 5,732가구 조사 완료 + 신규표본 1,800가구 조사
 - 2013년 ● 8차 조사 7,312 가구 조사 완료
 - 2014년 ● 9차 조사 7,048 가구 조사 완료
 - 2015년 ● 10차 조사 6,914 가구 조사 완료
 - 2016년 ● 11차 조사 6,723 가구 조사 완료
 - 2017년 ● 12차 조사 6,581 가구 조사 완료
 - 2018년 ● 13차 조사 6,474 가구 조사 완료



7차 웨이브 가중치 산출과정

- 7차 웨이브의 경우 기존 1~6차 웨이브에서 구축된 패널 원표본과 7차 웨이브에서 새롭게 추가된 신규표본을 병합하여 새롭게 가중치를 조정할 필요가 있음
- 1차 웨이브 이후 7차 웨이브까지 조사된 원표본은 일반적인 가중치 부여체계 적용 즉, 6차 웨이브의 종단면 가중치에 기초하여 7차 웨이브 개인별 종단면 가중치 조정
- 7차 웨이브에 추가된 신규표본의 종단면 가중치는 패널 표본으로 새롭게 진입하였으므로 모두 “0”으로 부여



7차 웨이브 가중치 산출과정

- 7차 웨이브에 추가된 신규표본의 횡단면 가중치는 표집 틀인 “2010년 차상위 실태조사”의 최종가중치를 설계가중치로 부여한 후, 추출확률의 역수를 곱하여 신규표본의 기본가중치로 설정
- 7차 웨이브의 최종 횡단면 가중치는 원표본과 신규표본을 병합하여 사후조정

8, 9, 10차 웨이브 가중치 산출과정

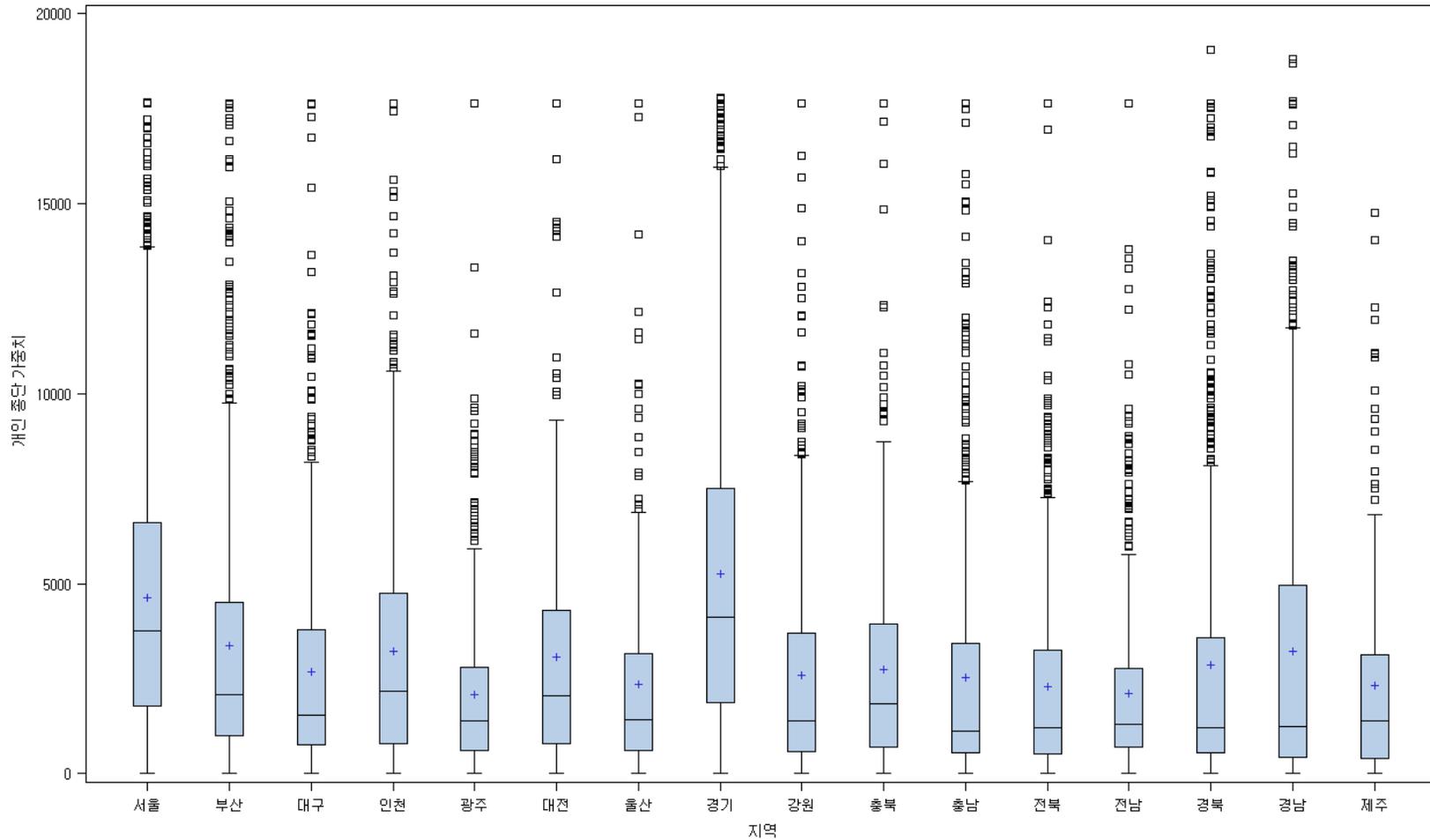
- 8차 웨이브 개인 가중치는 7차 웨이브의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여
- 단, 7차 웨이브에 새롭게 추가된 신규표본 1,800 가구의 가구원은 7차 웨이브 개인 종단면 가중치가 0이므로, 7차 웨이브 개인 횡단면 가중치를 기본가중치로 부여한 후 원표본 가중치와 병합하여 8차년도의 종단면 가중치와 횡단면 가중치를 산정
- 9차 웨이브 개인 가중치는 8차 웨이브의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여
- 10차 웨이브 개인 가중치는 9차 웨이브의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여

극단값 조정

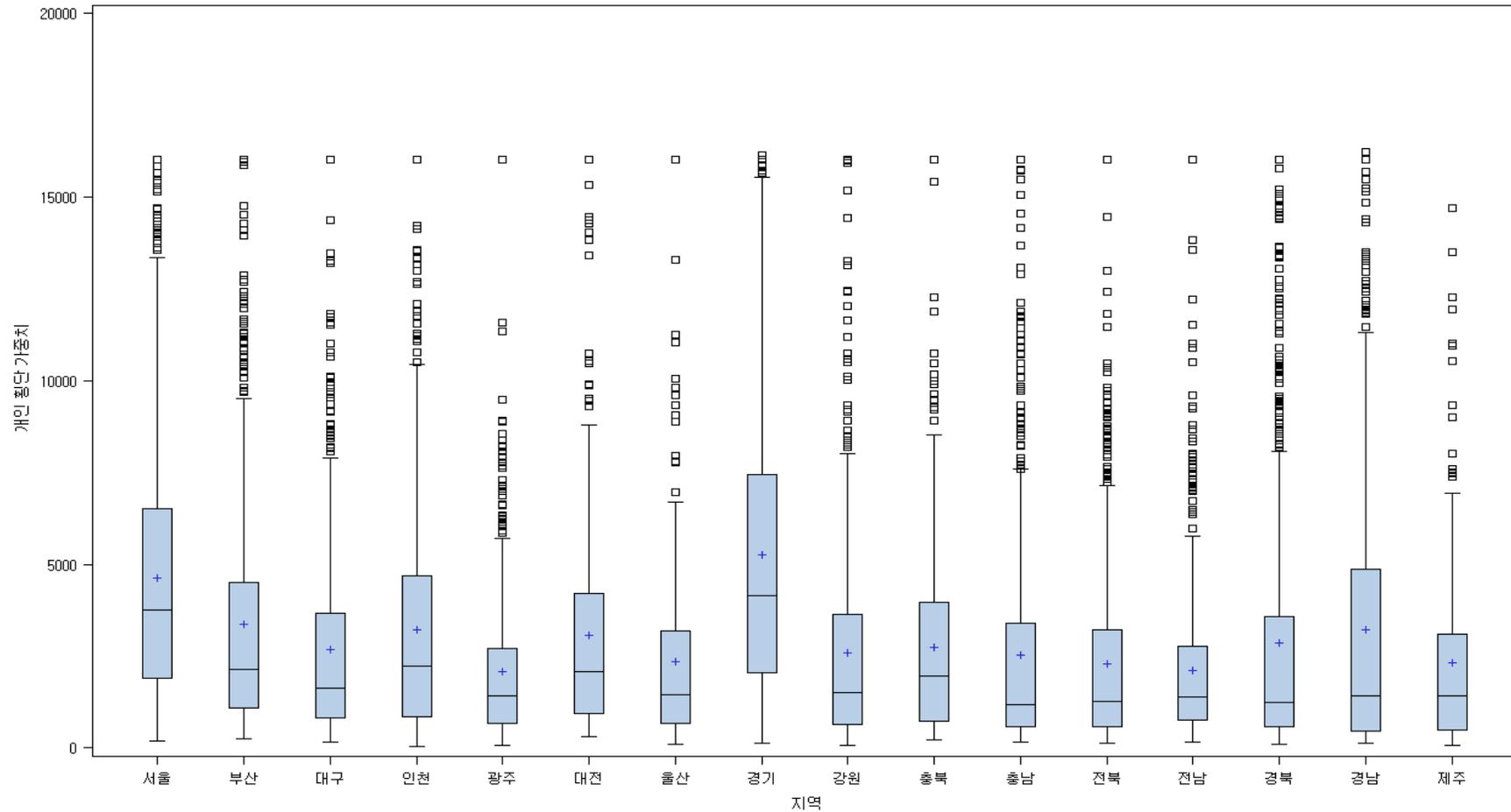
- 극단 가중치의 경우, 추정량의 분산을 크게 만들어 추정의 정확도에 영향을 미치므로 조정이 필요함
- 조사 기준년도 인구 추계값을 기준으로 가중치 조정 후, 상위 1% 극단 가중치를 찾아 상위 1% 경계값으로 절단(trimming)
- 절단값을 제외한 나머지 값을 극단 관측치가 속한 지역*가구원수(지역*성별*연령) 내의 관측치들에게 동일하게 배분
- 13차년도 가구(최대 14,645), 개인횡단(최대 16,216), 개인종단(최대 20,181) 가중치의 값을 제한

신규표본 추가에 따른 가중치 조정

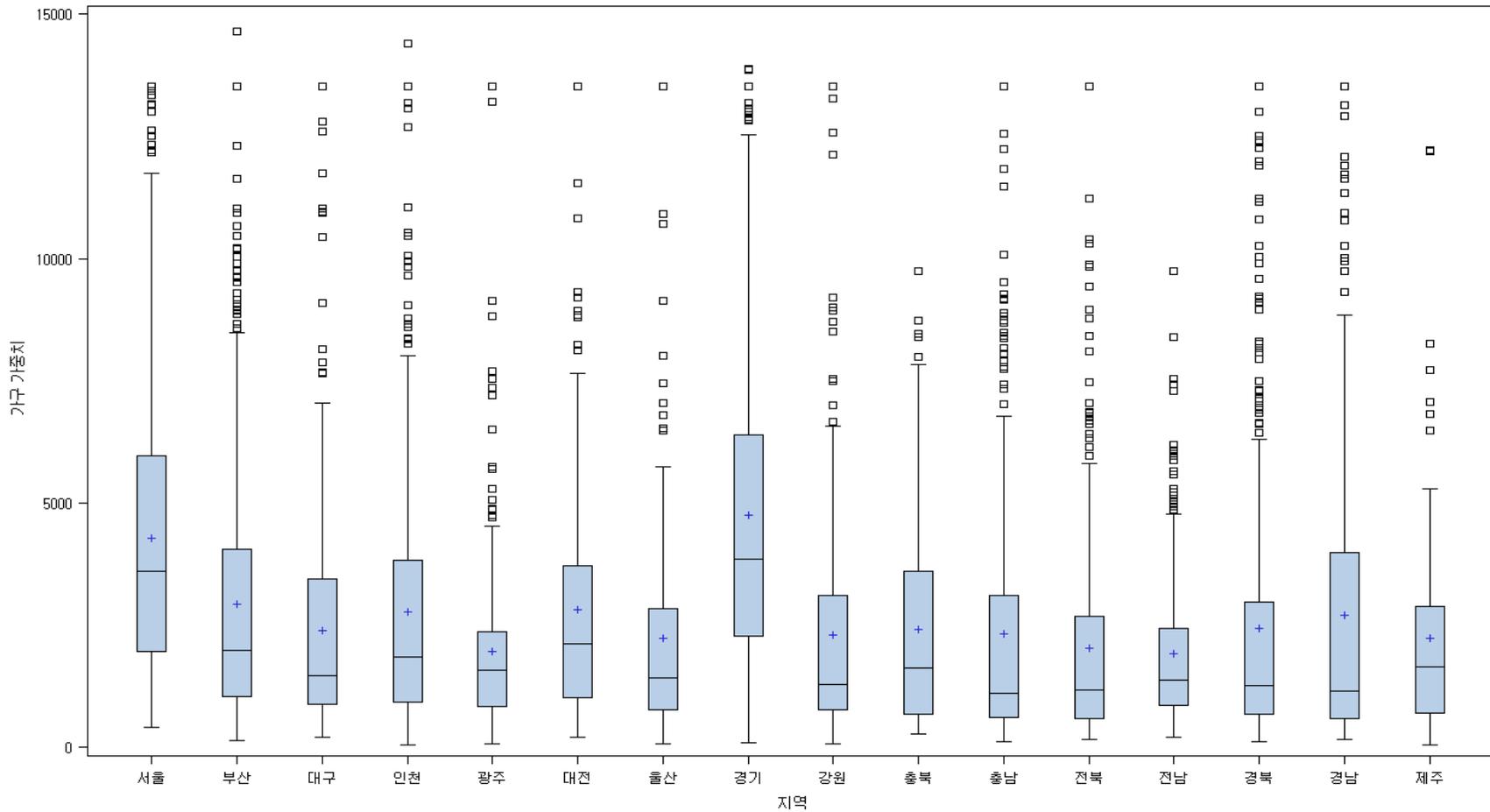
13차 지역별 개인 종단면 가중치 상자그림



13차 지역별 개인 횡단면 가중치 상자그림



13차 지역별 가구 가중치 상자그림



3. Q & A





통계 분석을 할 때 가중치를 꼭 써야하나요?

표준가중치와 일반가중치는 각각 어떠한 경우에 쓰나요?

머지데이터의 가중치는 우리나라 전체인구를 대표하는 것인가요?