

가중치 산출 및 추정

1) 가중치 작성

본 조사의 가중치는 표본설계 과정의 추출률의 차이를 보정하는 설계 가중치, 실사 과정에서 나타나는 표본사업체의 무응답을 조정하는 무응답 조정, 최신 모집단 정보를 이용하는 사후총화 조정의 과정을 통해서 산출한다. 데이터 분석 단계에서 복합표본설계를 통해서 얻은 조사데이터를 단순 임의추출법으로 추출한 것으로 가정하여 가중치를 반영하지 않고 분석한 추정결과는 심한 편향(bias)이 나타날 수 있다.

본 조사에서 1상 표본은 층화변수로 산업대분류(17)와 사업체 규모(5) 구분을 이용하는 층화계통추출법을 적용하여 표본사업체를 추출하였다. 이에 따라 각 층에서 표본사업체의 추출확률은 같은 값을 갖게 된다. 각 표본사업체의 설계 가중치는 Valliant et al. (2013, p.311)에 제시된 방식을 따라 다음 식과 같이 추출확률의 역수로 산출하였다.

$$\text{추출확률} = (\text{해당 층의 표본 사업체 수}) / (\text{해당 층의 모집단 사업체 수}) = n_h / N_h$$

$$w_{hi}^{(d)} = 1 / \text{추출확률} = N_h / n_h$$

사업체 조사에서 단위 무응답(unit nonresponse)은 표본으로 추출된 사업체가 접촉이 불가능하거나 응답을 거부하여 조사가 불가능한 경우에 발생하게 된다. 표본조사에서 단위 무응답이 발생하면 무응답 조정을 통해서 무응답 성향 차이로 인해서 나타날 수 있는 추정 편향(bias)을 줄인다. 무응답 조정을 위해서는 우선 무응답 조정셀(cell)을 정의해야 한다. 일반적으로 사업체 대상 조사에서는 산업대분류와 사업체 규모 등이 동일하다면 주요 조사변수가 유사한 값을 나타내고, 응답률도 유사하게 나타난다. 따라서 이 조사에서 무응답 조정셀은 산업대분류 구분과 사업체 규모를 사용하였다. 각 무응답 조정셀에서 무응답 조정계수는 Valliant et al. (2013, p.316)에 제시된 무응답 조정 방식을 적용하여 다음 식에 따라 산출한다.

$$\text{무응답 조정계수(사업체)} = n_h / r_h$$

여기서, n_h 는 층 h 의 표본 사업체 수, r_h 는 층 h 의 응답 사업체 수이다.

1상 표본 사업체에 대한 가중치는 다음과 같이 산출되었다.

$$w_{hi}^{(1)} = w_{hi}^{(d)} \times \text{무응답 조정계수(사업체)}$$

다음 단계로 2상 표본에 대한 표본사업체는 마찬가지로 층화변수로 산업대분류(17)와 사업체 규모(5) 구분을 이용하는 층화계통추출법을 적용하여 표본사업체를 추출하였다. 따라서 2상 표본사업체에 대한 가중치는 추출확률의 역수로 다음 식에 따라 산출한다.

$$\text{추출확률} = \frac{(\text{1상 표본 내 해당 층의 표본 사업체 수})}{(\text{1상 표본 내 해당 층의 모집단 사업체 수})} = \frac{n_h^{(1)}}{N_h^{(1)}}$$

$$w_{hi}^{(2)} = 1 / \text{추출확률} = N_h^{(1)} / n_h^{(1)}$$

여기서, $N_h^{(1)}$ 는 총 h 의 1상 표본 사업체 수, $n_h^{(1)}$ 는 총 h 의 2상 표본 사업체 수이다.

2상 표본 사업체에서 해당 여성 근로자가 표본으로 추출될 확률은 표본 사업체 내에서 전체 과장급 이상 여성 근로자 중에서 추출될 확률의 곱으로 구할 수 있다. 여성 표본 근로자의 가중치는 다단계 표본추출 과정에서의 가중치 산출 방법(Valliant et al., 2013: p.343)을 적용하여 다음과 같이 산출한다.

$$w_{hij} = w_{hi}^{(1)} \times w_{hi}^{(2)} \times \frac{1}{\text{표본사업체 내 여성(남성)관리자 추출률}_{hi}}$$

앞선 산출된 응답자에 대한 가중치는 각 산업대분류에서 여성(남성) 관리자 현황에 대하여 사후층화 조정과정을 거쳐 최종적으로 결정된다. 여성 관리자에 대한 모집단 정보는 1상 표본조사를 통해서 각 산업대분류별 여성 관리자에 대한 현황을 추정할 수 있다. 사후층화 조정 과정에서는 이 정보와 Valliant et al. (2013, p.351)에 제시된 방식을 적용하여 사후층화 조정 과정을 거치도록 한다.

과장급 이상의 남성 근로자 표본에 대한 가중치 산출은 앞서 제시한 여성 근로자 산출 방법과 유사한 방법을 통해서 산출한다.

2) 추정 방법

여성관리자패널조사에서 추정대상은 대부분 모평균이나 모비율이다. 본 조사에서 각종 모평균이나 모비율 추정은 가중치를 이용한 다음 추정량을 이용한다.

$$\hat{p} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{w_{\dots}}$$

여기서, w_{hij} 는 각 응답 근로자에 대해 산출된 가중치이고, y_{hij} 는 조사를 통해서 얻은 응답값이다. L 은 층의 수, n_h 는 총 h 에서 표본사업체 수, m_{hi} 는 총 h 에서 i 번째 표본사업체의 응답 근로자 수이다.

$w_{\dots} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$ 은 전체 응답 응답 근로자의 가중값 합이다.

앞서 제시한 모비율 추정량에 대해서 층화와 2단 집락추출 등의 표본설계를 반영한 추정분산은 다음과 같이 계산된다.

$$var(\hat{p}) = \sum_{h=1}^L \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi.} - \bar{e}_{h..})^2$$

여기서, L 은 층의 수, n_h 는 총 h 내의 표본사업체의 수, m_{hi} 는 총 h 내 i 번째 표본사업체의 응답자 수이다. $f_h = n_h/N_h$, $e_{hi.} = \left(\sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \hat{p}) \right) / w_{\dots}$, $\bar{e}_{h..} = \left(\sum_{i=1}^{n_h} e_{hi.} \right) / n_h$ 이다. 모비율 추정의 표준오차(standard error)와 95% 신뢰수준 오차의 한계는 다음과 같다.

$$s.e(\hat{p}) = \sqrt{var(\hat{p})}, \quad \text{오차의 한계} = 1.96 \times \sqrt{var(\hat{p})}$$

<참고문헌>

Valliant R., Dever J. A., Kreuter F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York