

## 1.1

조사 구역을 추출하고 구역 내에서 25가구를 선정한 다음, 조사 전에 선정된 가구를 방문하는데 여기서 가구원이 응답을 거절하는 단위 무응답이 발생할 수 있다. 혹은 조사장소에 정해진 시간에 나타나지 않아 단위 무응답이 발생할 수 있다.

청소년의 음주 혹은 흡연, 가구의 소득과 같은 민감한 질문에 대해 응답을 거절하는 항목무응답, 아니면 거짓말을 하는 측정오차가 발생할 수 있다. 의도하지는 않았지만 응답자가 외식빈도, 음료별 섭취빈도 등 각각의 질문들에 대해 기억을 제대로 하지 못해 측정 오차가 발생할 수도 있다. 혹은 여러 검진 항목들의 측정도구에 문제가 생겨 측정 오차가 발생할 수도 있다.

조사에 응하는 가구에 인센티브를 지급하거나 여러 번의 재방문, 연락을 취해 단위 무응답을 줄일 수 있다. 설문에 응할 때 익명임을 확실히 인지시키고 솔직한 응답을 요청해 측정오차를 줄일 수 있다. 또한 조사 전 측정도구들을 다시 확인해 측정 오차를 줄일 수 있다.

## 1.2

전화여론조사의 가장 큰 문제점은 낮은 응답률이다. 중앙선거여론조사심의위원회 홈페이지에 등록된 30개 여론조사의 평균 응답률(단순평균)은 10.0%로 응답률이 상당히 낮다. 응답자와 비응답자의 체계적 차이가 존재할 수 있기에 이러한 낮은 응답률이 문제가 된다.

또한 조사 방법에 따라 비표본오차가 발생할 수 있다. ARS조사는 부담을 덜 느끼고 의사표현을 자유롭게 할 수 있는 반면에 전화면접조사에서는 소수의견을 가진 사람들이 응답을 거부하거나 거짓으로 응답을 할 수 있기 때문이다.

이외에 질문 내용과 순서, 유무선 전화비율에 따른 결과 차이 등의 비표본오차가 발생할 수 있다.

## 2.1

### **경제활동인구조사**

조사단위: 가구, 가구원(개인)

추출단위: 1단계:조사구, 2단계:가구

표집틀 인구주택총조사 확대조사구역과 신축아파트 리스트로 구성된 조사구 리스트

표본: 전국 약 32,000 가구

조사모집단: 인구주택총조사 결과 대한민국의 만 15세 이상 모든 인구

### **고용형태별근로실태조사**

조사단위: 근로자 1인 이상 사업체와 사업체에 종사하는 소속 근로자(정규직 및 비정규직근로자)

추출단위: 사업체 내 근로자

표집틀: 사업체노동실태현황

표본: 근로자 1인 이상 33,000개 표본사업체 및 표본사업체에 종사하는 소속 근로자 약 100만명 내외('17년 기준) 정규직 및 비정규직근로자

조사모집단: 사업체노동실태현황(비임금근로자를 제외하고 특고를 포함한 정규 및 비정규직 근로자)

2.2

a)

$$\begin{aligned}\pi_1 &= 0.06 + 0.07 + 0.08 + 0.09 + 0.10 + 0.10 = 0.50 \\ \pi_2 &= 0.06 + 0.07 + 0.08 + 0.11 + 0.12 + 0.13 = 0.57 \\ \pi_3 &= 0.06 + 0.09 + 0.10 + 0.11 + 0.12 + 0.14 = 0.62 \\ \pi_4 &= 0.07 + 0.09 + 0.10 + 0.11 + 0.13 + 0.14 = 0.64 \\ \pi_5 &= 0.08 + 0.10 + 0.10 + 0.12 + 0.13 + 0.14 = 0.67\end{aligned}$$

b)

$$\begin{aligned}\bar{y}_{s_1} &= \frac{23}{3}, \quad \hat{p}_{s_1} = \frac{1}{3} \\ \bar{y}_{s_2} &= 8, \quad \hat{p}_{s_2} = \frac{2}{3} \\ \bar{y}_{s_3} &= \frac{26}{3}, \hat{p}_{s_3} = \frac{2}{3} \\ \bar{y}_{s_4} &= \frac{25}{3}, \quad \hat{p}_{s_4} = \frac{2}{3} \\ \bar{y}_{s_5} &= 9, \quad \hat{p}_{s_5} = \frac{2}{3} \\ \bar{y}_{s_6} &= \frac{28}{3}, \quad \hat{p}_{s_6} = 1 \\ \bar{y}_{s_7} &= 9, \quad \hat{p}_{s_7} = \frac{1}{3} \\ \bar{y}_{s_8} &= \frac{29}{3}, \quad \hat{p}_{s_8} = \frac{1}{3} \\ \bar{y}_{s_9} &= 10, \quad \hat{p}_{s_9} = \frac{2}{3} \\ \bar{y}_{s_{10}} &= \frac{31}{3}, \quad \hat{p}_{s_{10}} = \frac{2}{3}\end{aligned}$$

(c)

$$\begin{aligned}\bar{y}_{s_1,HT} &= 8.11, \quad \hat{p}_{s_1,HT} = 0.4 \\ \bar{y}_{s_2,HT} &= 8.33, \quad \hat{p}_{s_2,HT} = 0.7125 \\ \bar{y}_{s_3,HT} &= 8.79, \quad \hat{p}_{s_3,HT} = 0.6985 \\ \bar{y}_{s_4,HT} &= 8.43, \quad \hat{p}_{s_4,HT} = 0.7125 \\ \bar{y}_{s_5,HT} &= 8.89, \quad \hat{p}_{s_5,HT} = 0.6985 \\ \bar{y}_{s_6,HT} &= 9.11, \quad \hat{p}_{s_6,HT} = 1.011 \\ \bar{y}_{s_7,HT} &= 8.84, \quad \hat{p}_{s_7,HT} = 0.3125 \\ \bar{y}_{s_8,HT} &= 9.29, \quad \hat{p}_{s_8,HT} = 0.2985 \\ \bar{y}_{s_9,HT} &= 9.51, \quad \hat{p}_{s_9,HT} = 0.6110 \\ \bar{y}_{s_{10},HT} &= 9.61, \quad \hat{p}_{s_{10},HT} = 0.6110\end{aligned}$$

d)

주어진 표본설계 하에서 표본추정량들을 살펴보면

각각의  $\bar{y}$ 와  $\bar{y}_{HT}$ 들의 차이,  $\hat{p}$  와  $\hat{p}_{HT}$ 의 차이는 작은 수준이다.

하지만 각 표본이 뽑힐 확률, 즉 가중치를 반영한 비편향추정량  $\bar{y}_{HT}$  과  $\hat{p}_{HT}$ 이 모집단의 모수를 더 편향이 적게 추정했다고 볼 수 있다.

3.1

(a)

가구원을 SRS로 추출하였기 때문에 추출단위는 가구원이고, 가구원이 속한 가구의 소득에 대해 조사하였기 때문에 조사단위는 가구이다.

(b)

$$\hat{t}_{unb,wr} = \frac{1}{10} \times \left( \frac{7000}{\frac{1}{600}} + \frac{8000}{\frac{1}{600}} + \frac{6000}{\frac{1}{600}} + \frac{5000}{\frac{1}{600}} + \frac{9000}{\frac{1}{600}} + \frac{4000}{\frac{1}{600}} + \frac{7000}{\frac{1}{600}} + \frac{8000}{\frac{1}{600}} + \frac{4000}{\frac{1}{600}} + \frac{2000}{\frac{1}{600}} \right) = 3,600,000 \text{ 이고}$$

$$\hat{y}_{unb,wr} = \frac{1}{600} \hat{t}_{unb,wr} = 6,000 \text{이다.}$$