

제5장 국가통계 응답률 산출방안

민경아 · 박은영

제5장



제1절 서론

1. 연구 배경

2012년은 대통령 선거가 있는 해이다. 어느 해 보다 더 많은 여론조사결과 보도를 접하게 된다. 여론조사 보도 말미에는 의례히 ‘이번 조사결과는 신뢰수준 〇〇% 하에 표준오차는 ± 0.0 %입니다’라는 말이 따라붙는데, 이것은 해당 여론조사에 적용된 표집방법에 따른 표본오차를 제시함으로써 조사의 신뢰성과 정확성을 확인시키고자 하는 것이다. 그러나 여론조사 품질을 담보하기 위해서는 표본오차 뿐만 아니라 「응답률(Response Rate)」도 함께 제공하는 것이 필요하다. 조사결과의 품질을 평가할 경우 고려해야할 중요한 요인 중 하나가 바로 응답률의 크기이기 때문이다.(허명희, 2008)

현재 우리나라 여론조사의 응답률은 15% 내외로 매우 낮은 것으로 알려져 있다. 가령, 성인남녀 1,500명을 대상으로 여론조사를 실시했을 경우, 약 1만 건 내외의 전화조사가 실시되었다는 것을 예상할 수 있다. 여론조사 전화를 받았을 때 응하는 사람과 거부하는 사람들은 애초에 어떤 편향이 있다고 볼 수도 있다. 때문에 적은 수의 응답자 결과만으로 예측한 결과는 실제의 결과와 상당히 다른 결과를 초래할 수도 있게 된다. 그래서 미국의 경우에는 응답률이 30% 미만인 여론조사 결과는 미공표할 것을 권고하고 있으나 우리나라는 아직까지 마땅한 가이드라인조차 없는 실정이다.(허명희, 2007)

여론조사가 갖는 이러한 문제점 때문에 2007년 9월 선거관리위원회에서는 선거관련

여론조사 결과를 발표할 경우 응답률 자료도 함께 제공하도록 관련 기준안¹⁾ 및 법적인 거²⁾를 마련하였다. 그러나 이와 같은 노력에도 불구하고 실제 선거관련 여론조사 결과 발표 시에 응답률을 함께 제시하는 사례는 찾아보기 어렵다. 국가통계의 사정도 이와 별반 다르지 않다. 2012년 기준 총 382종에 달하는 우리나라의 조사통계 중에서 응답률을 제공하고 있는 조사통계는 찾아보기 어렵다.

통계청에서는 오래전부터 조사통계실무자를 위한 「국가통계 실무지침」이나 「품질관리 매뉴얼」을 각 통계 작성기관에 배포하고 매년 국가통계 품질평가를 실시함으로써 지속적으로 조사통계 공표 응답률 공표를 권고하였다. 때문에 각 작성기관의 응답률 산출에 대한 공감대는 형성되었다고 볼 수 있으나, 실제적으로 각 통계담당자가 작성하는 데에는 여러 어려움이 있는 것이 사실이다. 응답률 산출에 대해 벤치마킹할 수 있는 선례가 없기 때문이다. 그래서 작성기관의 응답률 산출을 도와줄 일종의 가이드라인이 필요하게 되었다.

2. 연구 내용과 범위

본 연구에서는 먼저 응답률에 대한 정의를 알아보고, 미국의 국가통계 작성기관인 Census Bureau와 BLS(Bureau of Labor Statistics)의 응답률 제공 사례를 살펴보기로 한다. 이들 작성기관에서 어떻게 응답률을 산출하고 또 통계사용자들에게 제공은 어떻게 하고 있는지 살펴봄으로써 우리나라 실정에 맞게 활용가능한 점을 도출하고자 한다. 또한 국내 국가통계 작성 기관의 응답률 작성현황과 문제점을 살펴봄으로써 응답률 작성을 위해 필요한 부분이 무엇인지 알아본다. 그리고, 각 국가통계 작성기관의 응답률 산출을 위해 적용 가능한 응답유형 측정방법과 응답률 계산의 가상 예제를 수록하기로 한다.

제2절 응답률의 정의

응답률은 통계조사에서 조사단위가 응답을 한 비율로, 표본에서 적격단위 중 응답에 응한 단위의 수를 전체 적격단위 수로 나눈 비율이다. 우리가 응답률에 관심을 갖는 것은

-
- 1) 선거에 관한 여론조사결과 공표·보도 기준안(선거관리위원회, 2007.9) : 응답률은 표집 틀에서 추출한 적합 표집단위의 수를 기준으로 산출해야 하며, 전화조사의 경우 응답전화수를 수신된 전화수와 부재중인 전화 수의 합으로 나누어 산출하고, 다른 조사방법에 대한 응답률은 이에 준해서 산출하도록 함
 - 2) 공직선거법 제108조 ⑤누구든지 선거에 관한 여론조사의 결과를 공표 또는 보도하는 때에는 ... 표본의 크기(연령대별·성별 표본의 크기를 포함한다), 조사지역·일시·방법, 표본오차율, 응답률, 질문내용, ...등을 함께 공표 또는 보도하여야 하며, (이하생략)

응답률이 낮을수록 무응답에 의해 편향이 발생할 가능성도 높아지기 때문이다. 따라서 응답률이 통계조사의 품질을 파악할 수 있는 측정 가능한 지표 중 하나가 되는 것이다.

표본단위 무응답이 발생하는 이유는 다양하고, 각각의 이유가 전체 자료에 미치는 영향도 다르게 나타날 수 있다. 따라서 무응답을 유형별로 나누어 이해할 필요가 있다. 대체로 무응답 유형을 크게 네 가지로 나누면, 응답을 시작했으나 부분적으로 응답하거나 중도에 포기하는 경우, 표본가구를 접촉할 수 없는 경우(비접촉, noncontact), 표본가구가 조사에 참여하기를 거부하는 경우(거부, refusal), 건강상 또는 여러 가지 이유로 조사에 응답할 수 없는 기타 경우와 같다. 이러한 개념 정의는 유형별 비율을 계산하는데 매우 중요하며 특히 첫 번째의 부분 응답이나 중도 포기 경우를 응답률 계산에 반영할 것인가는 사전에 명확히 정의하고 명시해야 한다.

응답률, 무응답률, 응답전환율은 조사결과에 대한 정보를 보여줄 수 있는 비율 지표로 각종 문헌에서 공통으로 사용되고 있다. 경우에 따라서는 근본적으로 다른 비율을 설명하는데 같은 이름이 사용된다거나 같은 비율을 설명하는데 다른 이름들이 사용되기도 한다. 가장 공통적으로 사용되는 용어로는 응답(response), 협조(cooperation), 접촉(contact), 거부(refusal)가 있다.

본 연구에서는 가장 널리 인용되고 있는 미국 여론조사협회(American Association for Public Opinion Research, AAPOR)의 표준정의와 미국 국가통계작성기관의 각종 가이드라인을 제시하고 있는 OMB의 정의 등에 대해서도 간략히 소개하기로 한다.

1. 미국 여론조사협회(AAPOR)의 응답률 표준정의

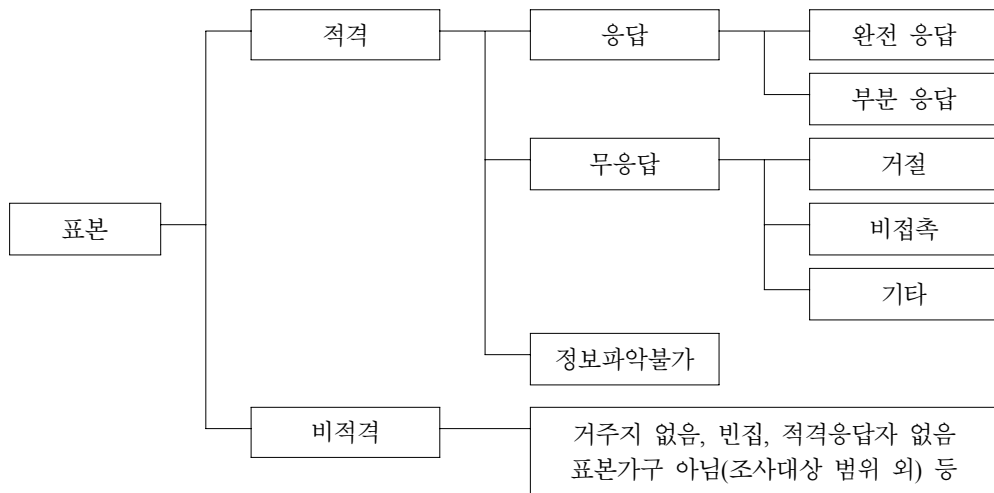
1947년 설립된 미국 여론조사협회(이하 AAPOR)는 대학, 미디어, 정부기관 및 기타 비영리조직 등의 조사연구와 관련된 전문가들로 구성된 협회이다. AAPOR는 조사방법론에서 전통적으로 중요한 영역인 조사표 설계, 조사 방법효과(mode effect), 무응답 오차, 가중치 등과 관련된 이론발전에 주요한 기여를 하고 있는데, 특히 1998년 처음 발표하고 2011년 7차 개정된 “Standard Definitions”를 통해 조사통계의 응답률과 관련된 이론적 배경을 제공하고 있다. 미국을 비롯한 여러 나라의 통계작성기관과 리서치 업체 등에서도 응답률과 관련하여 AAPOR의 기준을 준용하고 있다.

AAPOR는 표본 정보의 이용가능 여부에 따라 크게 표본을 적격(eligible)과 비적격(not eligible)로 나눈다. 또, 적격표본(eligible sample)에 대한 응답 유형을 크게 응답, 무응답, 정보파악불가 세 가지로 나누고 있다. 또한, 응답은 완전응답과 부분응답으로 나눌 수 있는데, 부분 응답의 경우에는 사전에 이를 응답으로 간주할 것인가에 대한 정의를 명확히 해



야 한다. 무응답은 여러 가지 형태로 나누고 있지만 일반적으로 크게 비접촉, 거절, 기타로 구분하고 있다. 비접촉은 표본에 대해 접촉을 시도하였으나 어떤 이유에서든 최종적으로 표본을 접촉하지 못해서 응답을 얻지 못한 경우를 말한다. 거절은 표본을 접촉하였으나 표본이 조사질문에 대해 응답하기를 거부하여 최종적으로 응답을 얻지 못한 경우를 말한다. 기타는 조사대상자가 표본으로 선택되어 표본임을 알고 있으나, 조사기간 동안 외출, 건강, 언어, 사망 등의 문제로 응답을 못한 경우를 말한다.(김서영·안다영, 2010)

비적격표본은 표본으로서 적합하지 않은 경우로, 가구의 경우 모집단 명부로부터 표본가구로 추출되었지만 조사목적에 부합하지 않아 표본자체를 이용할 수 없는 경우로서, 대상표본이 아니거나(out of sample), 주거지역이 아닌(not-a-housing -unit) 회사, 정부기관, 요양소, 군대 등 시설이거나, 장기적 혹은 정기적으로 비어있는 집(트레일러, 별장 등), 또는 적격대상자가 없는(housing unit with no eligible respondents) 가구 등이 해당된다. 지금까지 설명한 AAPOR의 유형별 응답형태를 정리하면 다음 [그림 5-1]과 같다.(김서영·안다영, 2010)



[그림 5-1] AAPOR의 표본단위 응답형태 분류(김서영·안다영, 2010)

모든 표본단위들은 위의 분류체계에 따라 응답형태가 분류되고 각 분류별 케이스를 취합하여 비율을 계산함으로써 응답률을 계산할 수 있게 된다.

AAPOR(2011)는 6가지의 응답률 계산 방법을 제시하고 있는데 공통적으로 사용되는 기호는 다음과 같다.

I = 완전응답(Complete interview)
 P = 부분응답(Partial interview)
 R = 응답거부(Refusal and break-off)
 NC = 접촉불능(Non-contact)
 O = 그 밖의 모든 무응답
 U = 적격단위인지 판단이 불가능한 단위
 e = 적격단위인지 판단이 불가능한 단위들 중 적격단위의 비율

- 1) RR1 : 전체 응답건수(완전응답건수+부분응답건수)와 무응답건수(거절+조사중단+비접촉건수) 및 적격미확인건수(가구여부미정 등)의 총합 중 완전응답건수의 비율로서 이하 제시되는 응답률 중 최소값을 가짐

$$RR1 = \frac{I}{(I+P) + (R+NC+O) + U}$$

- 2) RR2 : 완전응답과 함께 부분응답(P)까지도 포함하는 응답률로서 RR1의 분자에 부분응답(P)건수가 추가된 형태

$$RR2 = \frac{I+P}{(I+P) + (R+NC+O) + U}$$

- 3) RR3 : 분모에 적격여부가 알려지지 않은 경우들의 확률을 적용한 응답률

$$RR3 = \frac{I}{(I+P) + (R+NC+O) + e(U)}$$

- 4) RR4 : 분모에 적격여부 미확인인 경우도 포함하는 RR3에서 확장된 식으로서, RR2와 같이 부분응답(P)건수를 포함하는 응답률

$$RR4 = \frac{I+P}{(I+P) + (R+NC+O) + e(U)}$$



- 5) RR5 : RR3의 식에서 $e=0$ 를 가정하는 경우, 즉, 적격여부가 확인되지 않은 케이스 중에서 적격인 케이스는 없다는 가정하에 계산되는 응답률

$$RR5 = \frac{I}{(I+P) + (R+NC+O)}$$

- 6) RR6 : RR5와 같은 가정하에 부분응답(P)도 포함된 응답으로서 6개의 응답률 중 최대값을 가짐

$$RR6 = \frac{I+P}{(I+P) + (R+NC+O)}$$

위의 응답률 계산식에서 부분 응답(P)은 응답을 정의하는 방법에 따라 응답률 계산에 사용될 수도 있고, 제외될 수도 있다. 또한 목적에 따라 6개 계산식 중 어느 하나를 이용할 수 있다.

아래는 응답률과 함께 활용 가능한 협조율·접촉률·거절률로서 RR5와 RR6의 가정하의 계산식이다.(AAPOR, 2011)

• **협조율(co-operation rate)**

: 접촉 가능한 적격단위들 중에서 조사에 응답한 단위들의 비율

$$\frac{(I+P)}{(I+P) + R + O}$$

• **접촉률(contact rate)**

: 적격단위들 중에서 조사를 위해 접촉한 단위들의 비율

$$\frac{(I+P) + R + O}{(I+P) + (R+NC+O)}$$

• **거절률(refusal rate)**

: 적격단위들 중에서 응답을 거절한 단위들의 비율

$$\frac{R}{(I+P) + (R+NC+O)}$$

2. OMB의 응답률 정의

AAPOR가 응답률의 이론적인 배경을 제공했다고 한다면, 미국 OMB(Office of Management and Budget)는 통계작성 실무 담당자의 활용을 염두에 둔 실제적 정의와 이론을 제공한다고 볼 수 있다.

OMB는 “Standards and Guidelines for Statistical Survey”를 통해, 각 조사통계 담당자들은 응답률 산출시 가중/비가중 응답률 두 가지를 모두 산출할 것을 권고하고 있다. 또한 단위응답률이 80% 미만일 경우와 항목무응답이 70% 미만일 경우에는 무응답 편향분석을 실시할 것도 명시하고 있다.

OMB에서 제안하는 응답률 식은 다음과 같다.(OMB, 2006)(손창균, 2012)

■ 비가중 단위 응답률(Unweighted unit Response Rates, RRU)

비가중 단위 응답률은 적격 조사대상 중 조사문항에 완전하게 응답하였거나 충분히 부분적으로 응답한 단위들의 비율로서 이는 AAPOR(2011)의 기본정의와 유사하다.

$$RRU = \frac{C}{C + R + NC + O + e(U)}$$

여기서

C : 조사문항에 완전응답하거나 핵심문항에 충분히 부분적으로 응답한 케이스 수

R : 응답을 거절한 케이스 수

NC : 적격단위 중에서 접촉하지 못한 표본단위의 수

O : 거절 이외의 이유로 응답하지 않은 적격 표본단위의 수

U : 적격성을 알 수 없거나 완전하지 않은 표본단위의 수

e : 적격성을 알 수 없는 표본단위들이 적격으로 추정된 비율

■ 가중응답률(Weighted unit Response Rates, RRW)

가중응답률은 경제 분야 조사에서 많이 활용될 수 있는데 각 조사단위가 특정 조사단위가 전체 모집단 총량 추정에 미치는 영향이 클 경우, 해당 조사단위의 응답률에 가중치를 부여하는 방법이다. 예를 들어 사업체조사에서 대기업의 비율은 적으나 매출액 등의 중요 통계항목에 미치는 영향이 클 경우에 아래의 식을 활용할 수 있다.



$$RRW = \frac{\sum w_i C_i}{\sum w_i (C_i + R_i + NC_i + O_i + e(U_i))}$$

여기서

C_i : i번째 단위가 완전응답이거나 충분히 부분적으로 응답한 경우=1, 그 외= 0

R_i : i번째 단위가 거절한 경우 = 1, 그 외 = 0

NC_i : i번째 단위가 적격단위 중에서 접촉하지 못한 경우 =1, 그 외 0

O_i : i번째 단위가 거절 이외의 이유로 응답하지 않은 경우 =1, 그 외 0

U_i : i번째 단위가 적격성을 알 수 없거나 완전하지 않은 경우 =1, 그 외 0

e : 적격성을 알 수 없는 표본단위들이 적격으로 추정된 비율

w_i : i번째 단위의 추출가중치(즉, 추출율의 역수)

■ 패널조사 응답률(Longitudinal Response Rates)

패널조사(panel surveys)는 횡단면적인 단위의 표본(개인, 가구, 기업 등을) 추출하여 이 패널표본들을 복수년도에 걸쳐서 관찰·조사 하는 것을 말한다. 패널조사에서는 이전의 적격 무응답자³⁾를 다음번 조사에 포함하기 위하여 특별한 방법을 사용한다. 패널 분석에 사용되는 전체 단위 응답률(RRO^L)은 매회 차 패널조사(wave)에 참여한 표본 내의 모든 적격 응답자의 비율로, 자료 수집 모든 단계의 응답률을 포함한다. 패널조사의 단위 응답률(RRO^L)은 다음과 같다.

$$RRO^L = \prod_{k=1}^K \frac{I_k^L + P_k^L}{I_k^1 + R_k^1 + NC_k^1 + e_k(U_k^1)}$$

여기서,

K : 패널조사 자료 수집의 최종 회차

I_k^L, P_k^L : k차 패널조사에 완전히 또는 부분적으로 응답 하는 단위

R_k^1 : k차 패널조사에서 1차 wave 내 표본의 거절단위의 수

$I_k^1 + R_k^1 + NC_k^1 + e_k(U_k^1)$: 1차 웨이브에 들어온 표본 수

3) 적격의 조사 대상자이나 조사 시 무응답을 한 대상자

3. 기타 정의

가. 사업체조사의 응답률 정의

사업체, 교육기관 및 기타단체 등의 기관에 대해서도 역시 전화, 면접, 우편, 인터넷 조사 또는 혼합된 조사가 이뤄질 수 있다. 그러나 이들 기관조사는 일반적인 가구조사와는 차이가 있다. 사업체대상이므로 한 건에 대해 복수의 응답자가 연관될 수 있으며, 일정한 전체 기간에 대해 응답을 할 수도 있고 일부 기간에 대한 응답만 할 수도 있다. 예를 들면 3개월간 매월 매출액조사에서 마지막 달에 대한 매출액만 제출하고 처음 2개월의 매출액을 제출하지 않을 수 있는 것이다. 이러한 조사의 응답률은 응답 시간 단위를 고려한 분류를 기초로 계산하게 된다.

이러한 경기조사에 대한 응답률은 다음과 같이 계산할 수 있다.(통계청, 2008)

FC : 전체 기간에 대한 완전응답	FP : 전체 기간에 대한 부분응답
PC : 부분 기간에 대한 완전응답	PP : 부분 기간에 대한 부분응답
RNU : 응답하였으나 쓰지 못하는 경우	NR : 무응답
U : 부적격단위인지 판단이 불가능	
e : 부적격단위인지 판단이 불가능한 단위 중 적격한 비율	

- **완전응답률(full response rate)**

: 모든 기간에 완전하게 응답한 경우의 응답률이다.

$$\frac{FC}{(FC+FP+PC+PP)+RNU+NR+e(U)}$$

- **전체응답률(overall response rate)**

: 부분 기간 또는 부분 응답도 모두 응답한 것으로 간주한 응답률이다.

$$\frac{FC+FP+PC+PP}{(FC+FP+PC+PP)+RNU+NR+e(U)}$$



나. 조사항목에 대한 응답률

앞에서 논의한 응답률은 모두 단위응답률에 해당한다. 통계조사에서 주요 조사항목에 대한 응답률을 제공함으로써 통계이용 시 도움을 줄 수 있다.

항목 응답률은 해당 항목에 응답을 해야 하는 자격을 가진 단위의 수와 그 중에 응답한 단위의 수의 비율로 나타낸다.(통계청, 2008)

$$\text{항목응답률} = \frac{\text{항목에 응답한 조사단위의 수}}{\text{항목에 응답해야 하는 조사단위의 수}}$$

이 장에서 언급한 응답률은 가중값을 주지 않은 응답률이다. 그러나 서로 다른 선택 확률을 사용해서 표본을 추출할 때에는 가중된 응답률이 보다 더 적절한 추정값이 될 수 있고, 이는 평균제곱오차에서 무응답편향의 잠재적인 영향에 대한 보다 더 좋은 척도이다. 표본의 선택확률이 단위마다 다르다면 가중된 경우와 그렇지 않은 경우의 응답률은 상당히 다를 수 있다. 이에 대해 좋은 예가 바로 미국 지역사회조사 응답률이다. 이는 뒤의 제4절 해외사례에서 소개하기로 한다.

제3절 국가통계 응답률 작성 현황

2012년 말 현재 우리나라는 전체 385개 기관에서 902종의 국가통계가 작성되고 있다. 통계청은 이들 국가통계 작성기관을 대상으로 고품질의 통계 작성을 독려하기 위하여 2000년대 초부터 국가통계 품질관리사업을 실시하였다. 매년 일정비율의 작성통계를 대상으로 품질평가를 실시하고 있으며 실무자들이 해당 통계의 품질을 유지 또는 향상시키기 위해 조사기획, 자료수집, 자료분석, 자료 서비스 단계별로 준수하거나 참고해야 할 지침 등을 매뉴얼화하여 제공하고 있는데, 특별히 조사통계 공표 시 응답률도 함께 공표하도록 권고하고 있다.(통계청, 2008 및 2009)

그러나 상기 가이드라인 및 매뉴얼과 달리 현재 국가통계 중 응답률을 제공하는 조사통계는 찾아보기 힘들다. 국가통계 주무부처인 통계청에서도 공식적으로 응답률을 제공하지 않고, 내부적으로 조사관리 차원에서 주요 조사통계에 대해서 표본단위의 응답률을 파악·활용하고 있다. 이 응답률 자료는 통일된 기준에 의해서 작성되지 않을 뿐 아니라 앞서 제시한 국제적으로 표준화된 기준에 의해서 작성된 응답률이라고 보기 어

렵고 당초 설계된 표본수를 유지하기 위해 새로운 표본을 대체하여 조사가 완료된 비율로서 조사완료율에 가까운 개념이다.

가령 축산물생산비조사의 경우 표본농가가 폐업 등의 사유로 더 이상 조사가 불가능하게 되면 인근 지역에서 같은 축종의 사육 규모가 비슷한 농가로 표본대체가 이뤄지고 있는데, 응답률 100%가 되기까지 표본대체가 이뤄지는 형태이다. 따라서 이들 조사는 당초 설계된 표본수를 유지하며 계속 조사가 이뤄지므로 응답률이 아니라 조사완료율로 보는 것이 타당하다.

한편, 가구표본에서 가계부 작성을 기초로 하는 가계동향조사의 경우 통계청의 내부 전산시스템인 「가구통합시스템」에서 가구구분(적격/부적격), 가계부제출여부(정상제출/부실제출/미제출/대상외) 등을 입력하게 되어있어 타 조사에 비해 상대적으로 응답률 산출이 용이하다. 하지만, 가계동향조사의 응답률은 적격가구에서 장애·장기투병·연로로 가계부 작성이 불가능한 가구를 제외한 가구 중 가계부 회수가 얼마나 되었는지, 즉 가계부 회수율을 의미하므로 엄밀한 의미의 응답률과는 거리가 있다.

이와 같이 현재 국가통계의 응답률은 내부적으로만 활용되고 있고 또한 작성 방법이 표준적이지 않아 통계의 품질을 파악하거나 상호 비교하는데 문제점이 있는 실정이다.

제4절 해외 사례 연구

1. 미국 상시인구조사(Current Population Survey, CPS)

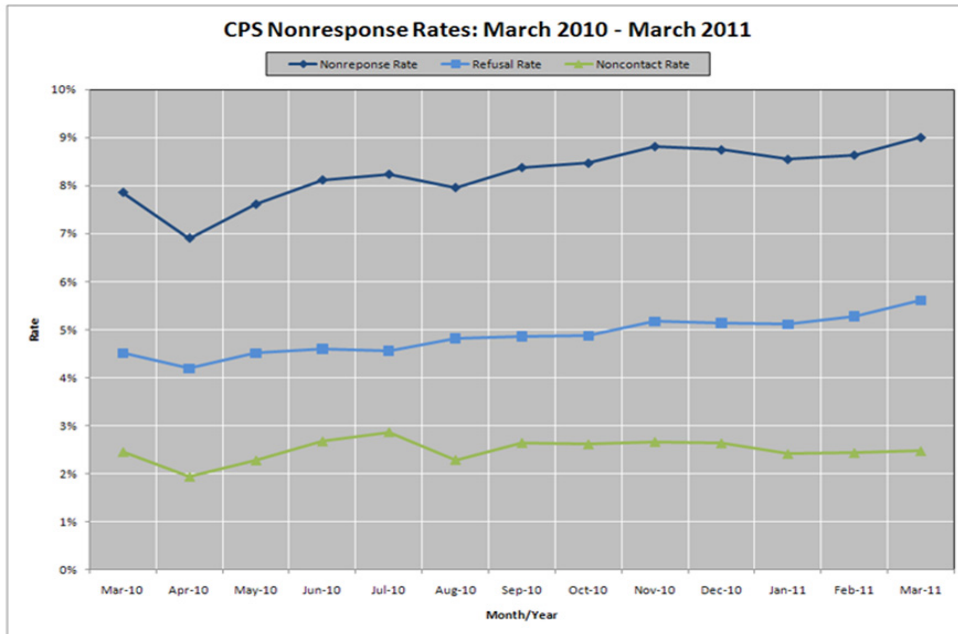
미국은 1930년대 경제 대공황을 겪게 되면서 노동시장이 불안해지고, 실업률이 크게 높아져 사회적으로 문제가 되었다. 때문에 매월 실업률 등 노동력 통계가 요구되었다. 1940년부터 실시된 상시인구조사(Current Population Survey, CPS)는 각 주(州)의 월별 실업률 등 노동력 실태파악을 목적으로 가구를 대상으로 실시하는 노동력통계조사이다. 조사수행은 센서스국(Census of Bureau)에서 실시하고 조사결과의 처리 및 분석은 노동통계국(Bureau of Labor Statistics, BLS)에서 실시한다.

상시인구조사(CPS)는 매월 약 6만여 표본가구(Housing Unit)내 16세 이상의 인구를 대상으로 고용, 비고용, 소득, 노동시간 등 노동력관련 정보와 나이, 성별, 인종, 기혼여부, 교육 등의 인구학적 특징을 조사한다. 시설조사구(Group Quater)내 살고 있는 사람은 제외되지만 국내외 직업군인이나 외국인 취업자는 포함된다.



CPS는 1953년부터 4-8-4월 연동표본제(Rotation sampling system)를 도입하였는데, 추출된 표본가구는 연속된 4개월 동안 조사에 참여하고 이후 8개월은 조사를 중단한 뒤 다음 4개월 동안 다시 조사에 참여하고 표본에서 교체된다. 즉 매월 1/4의 표본이 교체되므로 전월 중복률은 75%, 전년 동월 중복률은 50%인 표본구조를 가진다.

CPS는 웹사이트(www.census.gov/csp/methodology/nonresponse.html)를 통해 아래와 같이 최근 12개월간(2010.3~2011.3)의 무응답률과 함계, 무응답의 여러 사유 중 가장 대표적인 거절률(Refusal Rate) 및 비접촉률(Noncontact Rate)을 함께 보여 주고 있다.



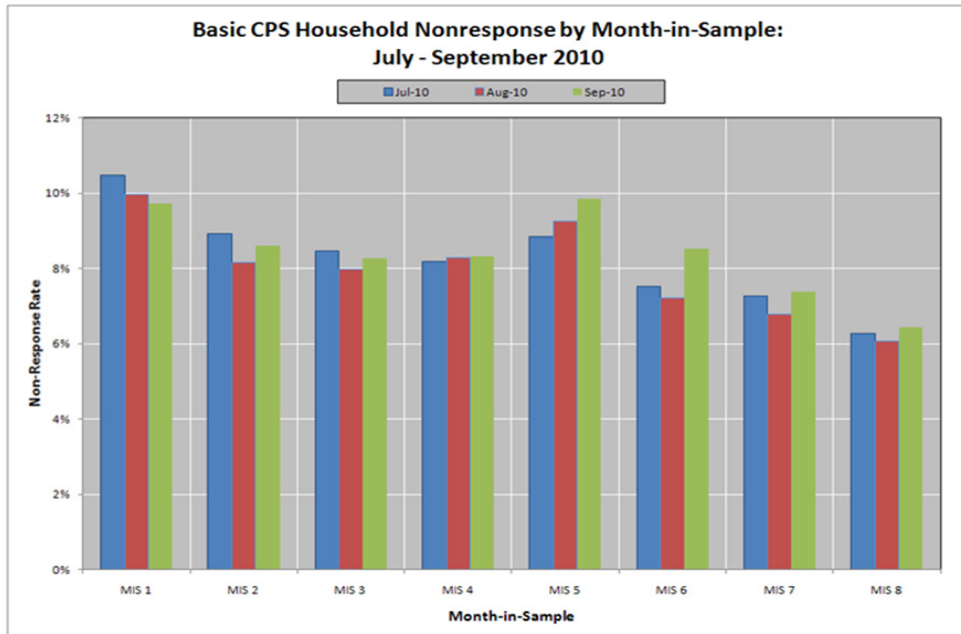
출처 : CPS web site, www.census.gov/csp/methodology

[그림 5-2] CPS의 무응답률(2010.3~2011.3)

CPS의 표본가구는 4-8-4 연동표본제이므로 총 8번 조사에 임해야하는데, 이 8번 중 첫 번째 조사에 들어간 그룹을 MIS 1(Month in Sample 1)이라하고 두 번째 조사에 임하는 그룹은 MIS 2라 한다. 따라서 여덟 번 중 마지막 8번째 조사만 마치면 조사에서 제외될 그룹은 MIS 8이 된다. CPS의 연동표본은 4개월(4번) 조사 후 8개월 쉬고 4개월 다시 조사한다. 이 조사방법 때문에 조사가 시작되는 첫 번째와 다섯 번째 즉, MIS 1과 5의

무응답률이 높게 나타나는 것이다. 협조도를 높이기 위해 센서스국에서는 첫 번째와 5 번째 달 조사 실시 전 앞으로 있을 4번의 조사에 대한 안내장이 발송한다. 또한 회를 거듭할수록 협조도는 높아지기 마련이며, 사전에 전화조사(CATI)도 실시하므로 MIS 2~5 및 MIS 6~8은 무응답률이 점차 낮아진다.

CPS 홈페이지에는 2010.7~9월 3개월간 연동그룹(MIS 1~8)별 응답률이 소개되어 있다. 조사 횟수가 많아질수록 무응답률은 점차 낮아진다. 즉, MIS 1보다는 MIS 2가, MIS 2보다는 MIS 3의 무응답률이 낮다. 그러나 4개월 조사를 마치고 8개월 간 중단한 뒤 다시 조사하게 되는 MIS 5에는 다시 무응답률이 높아졌다가 이후 다시 점차 낮아진다.



출처 : CPS web site, www.census.gov/cps/methodology

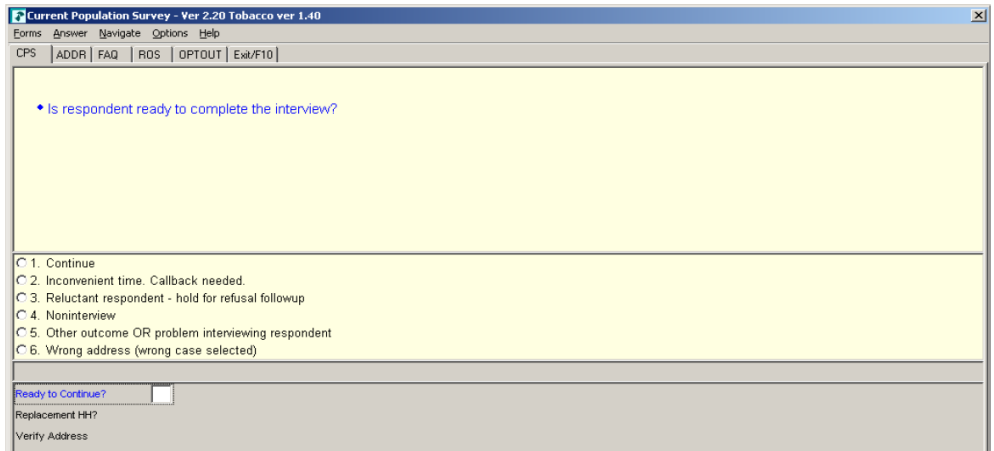
[그림 5-3] CPS MIS 1~8의 응답률 (2010.7~9월)

센서스국에서는 CPS의 응답률 산출을 위하여 표본가구의 접촉결과 상황을 구체적으로 분류, 코드화하여 각 표본가구의 접촉 결과 데이터를 수집하고 있다. 특히 무응답의 경우에는 원인에 따라 Type A, B, C 등 세분화하여 자료를 수집한다. 또한 분류 코드는 조사방법이 전화(CATI)인지 면접조사(CAPI)인지에 따라 각 Type 내 세분류 중 일부 달라진다.

Type A는 적격가구로 확인은 되었으나 조사가 이뤄지지 않은 경우를 말한다. 조사원이 방문하였을 때 조사기간 동안 일시적으로 집에 아무도 없거나, 가구원 일부는 집에 있으나 응답할 만한 사람이 부재중일 경우(CPS는 16세 이상 대상이므로), 그리고 조사를 거절한 경우 등이 이에 해당된다. Type B는 현재는 비적격이나 향후 적격가구가 될 가능성이 있는 경우를 말한다. 매매나 렌트를 위한 빈집이 매매 또는 렌트되어 새로운 가구가 이사 올 경우가 이에 해당된다. 언제든 조사적격가구가 변동 가능하므로 전월조사에서 Type B에 속했다 하더라도 매월 방문하여 조사가능 여부를 확인한다. Type C는 현재도 조사적격이 아니며 향후에도 조사적격이 될 가능성이 희박한 경우에 해당되는데, 예를 들면 영구 상업지구 또는 수용되거나 철거된 지역, 추출된 주소가 조사구 밖의 지역인 경우 등이 있다.

CPS 조사원들은 조사표내용을 현장에서 휴대용 컴퓨터(Laptop)에 기입하게 되는데, 아래의 조사화면에서 해당 표본가구의 응답여부를 체크하게 된다.

Figure C94. INTROB Screen, Noninterview Option



출처 : CPS Interviewing manual, Page C6-3

[그림 5-4] CPS 입력시스템 - 표본가구의 응답여부 체크

표본가구의 적격 응답자가 조사에 응하게 되면 위 화면에서 “1. Continue”를 체크한다. 무응답 상황이라면 “4. Noninterview”를 체크하게 되는데, 해당 무응답 상황이 Type A, B, C 중 어느 것에 해당되는지도 체크하도록 한다.

Figure C95. NONTYP, Type of Noninterview

Current Population Survey - Ver 1.30

Forms Answer Navigate Options Help

CPS ADDR FAQ Chg Resp RDS F9 Exit/F10

• What type of Noninterview do you have?

C 1. TYPE A
C 2. TYPE B
C 3. TYPE C

출처 : CPS Interviewing manual, Page C6-4

[그림 5-5] CPS 입력시스템 - 무응답 TYPE 기입

위에서 Type A를 선택했을 경우, 다음단계로 아래와 같이 Type A의 6개 사유 중 어느 것에 해당되는지 체크하도록 한다. 해당 케이스가 어느 사유에 해당되는지는 CPS 조사원들을 위한 면접조사 매뉴얼(CPS Interviewing Manual)에 잘 설명되어 있다.

Current Population Survey - Ver 1.30

Forms Answer Navigate Options Help

CPS ADDR FAQ Chg Resp RDS F9 Exit/F10

• Do Not F10 From This Screen

• Enter Type of Noninterview

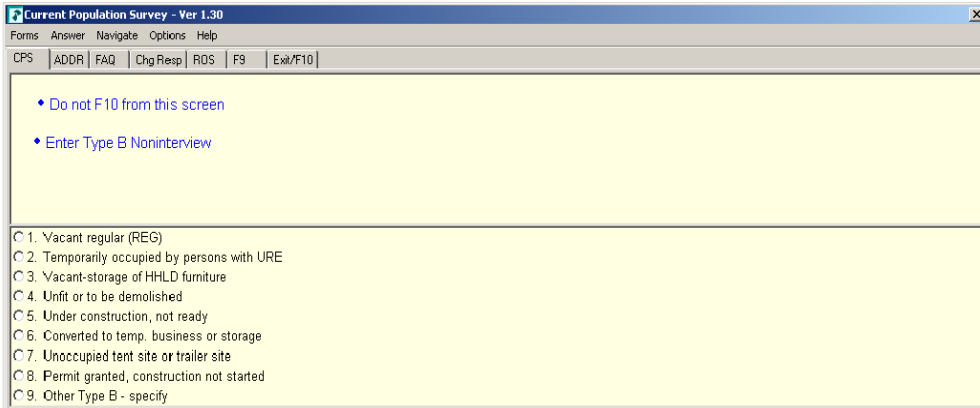
C 1. No one home (NOH)
C 2. Temporarily Absent (TA)
C 3. Refused (REF)
C 4. Language Barrier
C 5. Unable to locate
C 6. Other occupied-specify

출처 : CPS Interviewing manual, Page C6-5

[그림 5-6] CPS 입력시스템 - 무응답 TYPE A 세부 사유 기입

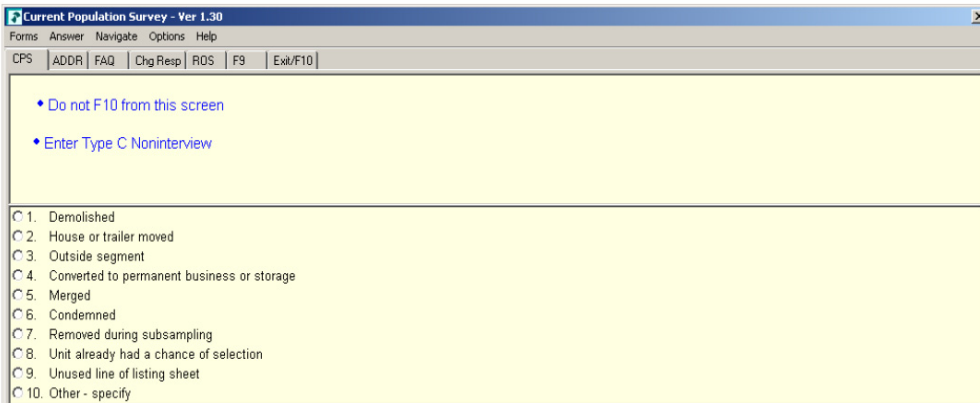


Type B와 Type C의 무응답일 경우의 조사 진행화면은 각각 다음과 같다.



출처 : CPS Interviewing manual, Page C6-8

[그림 5-7] CPS 입력시스템 - 무응답 TYPE B 세부 사유 기입



출처 : CPS Interviewing manual, Page C6-13

[그림 5-8] CPS 입력시스템 - 무응답 TYPE C 세부 사유 기입

이상의 프로세스에 의해 집계된 2004년 9월의 응답률 결과는 다음 표와 같다. 표에서 보면 최초 조사대상 가구는 71,575개이고, 이중 적격가구는 59,641개이다. 즉, 현장조사 결과 11,934개 가구는 비적격가구로 확인되어 조사에서 제외된 것이다. 적격가구 중 응답가구는 55,130개로 응답률은 92.44%이다. 그런데 이 표에서는 무응답률이 22.98%

나타나 있는데 이것은 응답률 계산에는 적격가구에 한해서 계산하고 무응답률 계산 시에는 최초 표본 수 기준으로 계산하여 비적격가구 수도 모두 분모에 포함된 결과이다.

Figure 7-9. Interviewing Results (September 2004)

Description	Result
Total HHLD	71,575
Eligible HHLD	59,641
Interviewed HHLD	55,130
Response rate	92.44%
Noninterviews	16,445
Rate	22.98%
Type A	4,511
Rate	7.56%
No one home	1,322
Temporarily absent	432
Refused	2,409
Other—specify	348
Callback needed—no progress	0
Type B	11,522
Rate	16.19%
Entire HH Armed Forces	130
Entire HH under 15	3
Temp. occupied with persons with URE	1,423
Vacant regular (REG)	7,660
Vacant HHLD furniture storage	514
Unfit, to be demolished	428
Under construction, not ready	429
Converted to temp. business or storage	167
Unoccupied tent or trailer site	354
Permit granted, construction not started	50
Other Type B	364
Type C	412
Rate	0.58%
Demolished	82
House or trailer moved	45
Outside segment	7
Converted to permanent business or storage	48
Merged	26
Condemned	9
Built after April 1, 2000	14
Unused serial no./listing Sheet Line	54
Removed during subsampling	0
Unit already had a chance of selection	4
Other Type C	123

출처 : CPS Design and Methodology, p.7-10

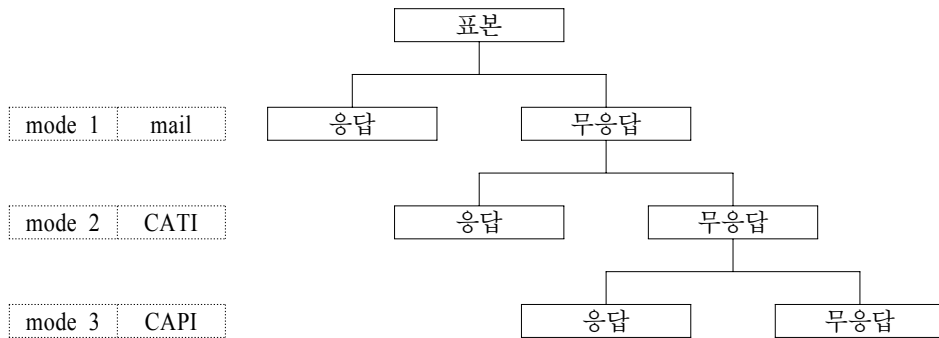
[그림 5-9] CPS의 무응답률(2004.9)

앞에서도 설명했듯이 무응답 분류 중 Type A는 ‘적격가구 중 무응답’이고 Type B와 C는 ‘비적격가구의 무응답’이다. 따라서 적격가구 중에서 무응답만을 의미하는 Type A만을 고려한 7.56%가 CPS의 최종 무응답률이다. 앞서 소개한 CPS Web Site의 무응답률 ([그림 5-2])은 바로 이와 같은 과정으로 계산된 값이다.

2. 미국 지역사회조사(American Community Survey, ACS)

미국의 지역사회조사(American Community Survey, ACS)는 기존의 10년 주기 센서스의 표본조사(short form)항목의 자료를 매년 제공하기 위하여 개발된 조사로서, 1995~1996년 몇 개의 카운티지역에서의 시험조사를 시작으로 점차 지역을 확대해 나가 2005년부터 전국을 대상으로 실시하여 지금에 이르렀다.

현재 매년 약 300만 가구를 대상으로 조사를 실시하는데, 추출된 가구를 대상으로 우편조사(mail), 전화조사(CATI) 및 방문조사(CAPI)의 3가지 자료수집 방법을 순차적으로 실행된다. 이는 일종의 Sequential mixed mode data collection으로 이해를 돕기 위해 아래 [그림 5-10]과 같이 도식화하였다.



[그림 5-10] 미국 ACS의 순차적 조사 mode

2005년부터 최근까지의 지역사회조사 표본 수 및 응답 수는 아래와 같다.

<표 5-1> 미국 지역사회조사 연도별 최초 표본 수 및 최종 응답 수

(단위 : 가구, 명)

년도	가구(Housing Unites)		시설(Group Quaters)	
	최초 표본 수	최종 응답 수	최초 표본 수	최종 응답 수
2010	2,899,676	1,917,799	197,045	144,948
2009	2,897,256	1,917,748	198,808	146,716
2008	2,894,711	1,931,955	186,862	145,974
2007	2,886,453	1,937,659	187,012	142,468
2006	2,885,384	1,968,362	189,641	145,311
2005	2,922,656	1,924,527	-	-

출처 : ACS 홈페이지 www.census.gov/acs/www/methodology

〈표 5-2〉 ACS 가구 응답률 및 사유별 무응답률

년도	응답률	사유별 무응답률						
		거절률	접근 불가	부재중	임시 부재	언어 문제	불충분 자료	기타
2010	97.5	1.1	0.0	0.6	0.1	0.0	0.2	0.4
2009	98.0	0.8	0.0	0.6	0.1	0.0	0.2	0.3
2008	97.9	0.8	0.0	0.6	0.1	0.0	0.2	0.3
2007	97.7	0.9	0.2	0.5	0.1	0.0	0.4	0.2
2006	97.5	1.0	0.3	0.5	0.1	0.0	0.4	0.3
2005	97.3	1.0	0.4	0.5	0.1	0.0	0.3	0.3

출처 : ACS 홈페이지 www.census.gov/acs/www/methodology

2006년부터 실시된 시설조사구(GQ)의 응답률 현황도 아래와 같이 소개되어있다.

〈표 5-3〉 ACS 시설조사구 응답률 및 사유별 무응답률

년도	응답률	사유별 무응답률							
		GQ 거절률	접근 불가	거주민 임시부재	언어 문제	불충분 자료	GQ Person Other	시설단위 집단거절	Whole GQ Other
2010	97.6	0.9	0.2	0.1	0.0	0.2	0.4	0.1	0.5
2009	98.0	0.9	0.1	0.1	0.0	0.1	0.4	0.1	0.3
2008	98.0	0.5	0.1	0.1	0.0	0.1	1.0	0.2	0.0
2007	97.8	0.4	0.2	0.1	0.0	0.0	1.2	0.3	0.0
2006	97.4	0.8	0.2	0.1	0.0	0.1	0.6	0.5	0.2

출처 : ACS 홈페이지 www.census.gov/acs/www/methodology

ACS는 세 가지 유형의 조사방법을 이용한 복잡한 샘플디자인에 추출률이 각기 다른 subsampling⁴⁾까지 이루어지므로, 응답률 계산 시 이 추출률을 반영하여 가중응답률(weighted response rates)을 산출한다.

4) ACS의 면접조사 대상 표본 추출은 우편조사를 할 수 없는 조사대상 가구와 우편조사와 전화조사에 미응답 가구 중에서 이뤄진다. 이들을 우편조사 가능 가구와 불가능 가구로 나누어 카운티별로 순서대로 정렬한 뒤 조건에 따라 33.3%~66.7% 비율로 계통추출한다.



앞의 <표 5-2>에서 2010년 기준 Housing Unit 즉 가구의 단순계산 응답률은 66.1% (1,917,799/2,899,676)이나, 제공된 응답률은 97.5%이다. 이 두 값의 차이는 바로 응답률 계산 시 표본추출률의 역수가 가중치로 곱해진 가중응답률 계산에 의한 것이다. 다시 말해, ACS는 세 조사방법(우편조사, CATI, CAPI)별 표본추출률이 각기 다르므로 각 응답 케이스에도 각각의 조사방법별 추출률의 역수로 가중치가 부여되기 때문이다. 이 ACS 응답률은 조사의 'tabulation year'가 아니라 'sample year' 기준으로 계산된다.

ACS 홈페이지에 소개된 응답률 계산식은 아래와 같다.

$$RR = \left(\frac{\text{initially weighted estimate of interviews}}{\text{initially weighted estimate of cases eligible to be interviewed}} \right) \times 100$$

또한, 아래의 식에 의해 조사방법(mode)별로도 응답률을 계산할 수 있다.

$$RR_i = \frac{\sum_{NUM_i=1} WTNUM_i}{\sum_{DEN_i=1} WTDEN_i} \quad (i = \text{mail, CATI, CAPI 등})$$

단, $WTNUM_i = \text{조사방법(mode) } i \text{에서 조사단위(응답=1, 무응답=0)} \times 1 / \text{추출률}$

$WTDEN_i = \quad \quad \quad \text{조사단위} \times 1 / \text{추출률}$

3. 미국 BLS의 사업체조사

1980년대 중반 미국 BLS(Bureau of Labor Statistics)는 자료수집에 관한 TF팀을 구성하여 자료수집단계에서 각각의 표본단위들의 조사상황을 체계적으로 관리하는 방안을 모색하였다. BLS는 주로 사업체를 대상으로 하는 조사가 많았지만, 가구조사도 일부 있으므로 사업체, 가구 모두에 공통적으로 적용할 수 있는 자료수집단계의 표본단위 분류체계를 마련하게 되었다. 이 분류체계는 센서스국과 비교해서 비교적 단순하다.

〈표 5-4〉 BLS의 표본단위 분류 체계

자료 구분		코드	내용	비고
적격	응답	10	응답	응답률 계산식에 포함
	무응답	20*	거절	
		21	거절(자료부재-활용불가)	
		22	거절(협조의지 없음)	
적격확인불가		23	적격여부 확인 불가	
비적격		30*	비적격	응답률 계산식에 불포함
		31	존재하나 조사범위를 벗어남	
		32	존재하지 않음	

*는 하위분류에 포함하기 어려운 경우에 부여

모든 표본단위는 자료수집과정에서 그 결과 응답하였는지 혹은 무응답하였는지, 적격인지 비적격인지, 위 코드체계에 따라 각 표본단위의 상황에 맞는 코드를 부여받게 되고 그 결과 각 코드별 케이스를 취합하여 응답률을 계산한다.

BLS에서는 비가중응답률 및 가중 응답률의 산식을 아래와 같이 정의하고 있다.

$$\begin{aligned} \cdot \text{비가중 응답률} &= \frac{\sum \text{응답한 단위}}{\sum (\text{적격단위} + \text{적격여부 미확인 단위})} \\ \cdot \text{가중 응답률} &= \frac{\sum w_i (\text{응답한 단위})}{\sum w_i (\text{적격단위} + \text{적격여부 미확인 단위})} \quad (w_i = \frac{1}{\text{추출률}}) \end{aligned}$$

기본적으로 비가중/가중응답률 모두 산출하도록 되어있으나 BLS의 일부 통계에서는 비가중응답률만 제공되는 경우도 있다. 또, BLS는 응답률 산출식 분모에 적격단위가 확인되지 않은 단위를 모두 포함하고 있다고 하지만, 일부 조사에서는 적격단위만 포함하는 경우도 확인된다.(BLS, 2011)

BLS에서 작성하는 여러 통계 중 먼저 OES(Occupational Employment Statistics) 조사의 응답률 사례를 소개한다.

OES는 산업별 고용상황과 임금 등을 파악하기 위하여 사업체를 대상으로 실시하



는 조사로서, 사업체 수, 종업원 수(fulltime/parttime) 및 임금구간별 종업원 수 등을 반기별로 조사하고 있다. 미국의 약 670만개에 달하는 전체 사업체 중에서 약 120만개 사업체를 표본으로 선정하여 매년 2회(5월, 11월)씩, 3년간 총 6회에 걸쳐 조사한다. 즉, 120만개 표본을 약 20만개의 6개 사업체 패넬로 나눠 각 패넬그룹당 한 번씩 조사하는 3년 단위 순환조사로서, 자료수집 방법은 먼저 우편조사를 실시하고 이에 회신하지 않을 경우 즉 무응답 사업체에 대해 추가적으로 전화·방문·E-mail 조사 등을 실시한다.

OES 조사의 최근 3년(2008.11~2011.5) 전체 누적 표본 수는 다음 <표 5-5>와 같다.

<표 5-5> OES 조사의 최근 3년 전체 누적 표본 수

패넬	조사기준	사업체 수(개)
6	2011. 5.	201,275
5	2010.11.	201,553
4	2010. 5.	200,282
3	2009.11.	203,546
2	2009. 5.	203,805
1	2008.11.	201,849
누 계		1,212,310

2011년 5월 기준 OES의 경우, 최초 추출된 표본사업체 약 121만개 중에서 추출표본에서 비적격 사업체를 제외한 1,110,296개 사업체 중 858,474개의 사업체가 응답하고 251,822개의 사업체가 무응답하였다. 따라서 사업체 응답률은

$$\frac{858,474}{1,110,296} \times 100 = 77.3\%$$

로 계산된다. OES 조사의 주요항목이라 할 수 있는 ‘고용’에 대한 항목응답률은 가중응답률 73.3%, 비가중응답률 62.2%로 보고하고 있다. (BLS, 2011 : Survey Methods and Reliability Statement for the May 2011 Occupational Employment Statistics Survey) BLS의 보고서에서는 비가중응답률은 산출내역을 구체적으로 보여주고 있는 반면 가중응답률의 산출식은 구체적으로 소개하고 있지 않지만, 전체 사업체의 종업원 수 합계에서 표본단위인 각 사업체의 종업원 수 비율로 가중치로 계산하여 응답률을 산정하였을 것으로 추

측된다. 이처럼 OES 조사는 조사단위인 사업체 수 기준 응답률은 비가중응답률로 제공하고, ‘고용’ 항목 응답률에 대해서는 비가중 및 가중응답률 모두를 제공함으로써 통계 사용자의 이해를 돕고 있다. (BLS, 2011(OES Technical Notes for May 2011) 및 BLS & Census Bureau, 2004)

다음은 BLS의 주요통계 중 하나인 소비자물가지수조사, CPI(Consumer Price Indexes) 조사의 응답률 사례이다.

BLS는 매년 초에 발표하는 CPI 세부보고서(Detail Report)에서 각 조사년도 표본 수와 응답률 등을 세부 통계표와 함께 제공하고 있다. CPI의 응답률은 크게 두 가지 형태로 제공되는데, 하나는 전국단위의 응답률로서, CPI의 8개 물가 부문인 식음료·주택·의류·교통·의료·여가·교육·기타 별로 조사대상인 사업체 수(비율)와 이중 응답한 사업체 수(비율), 그리고 최종 추정에 반영된, 즉, 통계작성에 사용된 표본 사업체 수(비율)가 제공된다.

〈표 5-6〉 CPI 응답률 - 전국단위(2011)

구 분	표본 사업체	CPI 부문별									
		전체	식음료	주택	의류	교통	의료	여가	교육	기타	
적격표본 수	330,927	1,183,424	464,023	147,974	144,681	152,896	73,118	78,666	78,527	43,539	
응답 표본	개	317,110	999,863	424,256	132,683	80,697	138,531	47,790	65,007	71,489	39,410
	%	95.8	84.5	91.4	89.7	55.8	90.6	65.4	82.6	91.0	90.5
통계작성 에 사용된 표본	개	302,512	973,782	418,485	129,349	76,051	133,827	45,633	61,717	70,022	38,708
	%	91.4	82.3	90.2	87.4	52.6	87.5	62.4	78.5	89.2	88.9

출처 : Response Rates for the CPI indexes, 2011(CPI Detail Report, BLS)

다음은 CPI의 지역별 응답률이다. CPI 조사는 지역 규모에 따라 통계제공 주기를 달리 적용하고 있는데, 규모가 가장 큰 권역인 시카고, LA, 뉴욕 등 3개 지역은 매월, 다음으로 규모가 큰 아틀란타, 보스턴, 샌프란시스코 등 11개 지역은 격월, 규모가 가장 작은 그룹인 피츠버그, 세인트루이스 등 13개 지역은 반기별로 CPI 자료를 제공하고 있다. BLS는 이들 지역단위별 연간 응답률 자료를 제공하는데, 캘리포니아 주의 응답률 사례를 소개하면 아래와 같다. 아래는 캘리포니아 주에 소속된 도시 중 LA(매월), 샌프란시스코(격월), 세인트루이스(반기)의 응답률 예이다.



〈표 5-7〉 CPI 응답률 - 캘리포니아 주(州) 3개 도시(2011)

구 분		LA (매월)		샌프란시스코 (격월)		세인트루이스 (반기)	
		사업체 기준	8개 부문 총합	사업체 기준	8개 부문 총합	사업체 기준	8개 부문 총합
적격표본 수		24,209	75,478	5,263	16,550	1,316	6,084
응답	개	23,074	60,581	4,960	13,335	1,287	5,652
	%	95.3	80.3	94.2	80.6	97.8	92.9
통계작성 에 사용된 표본	개	21,999	58,955	4,650	12,900	1,266	5,595
	%	90.9	78.1	88.4	77.9	96.2	92.0

출처 : Response Rates for the CPI indexes, 2011(CPI Detail Report, BLS)

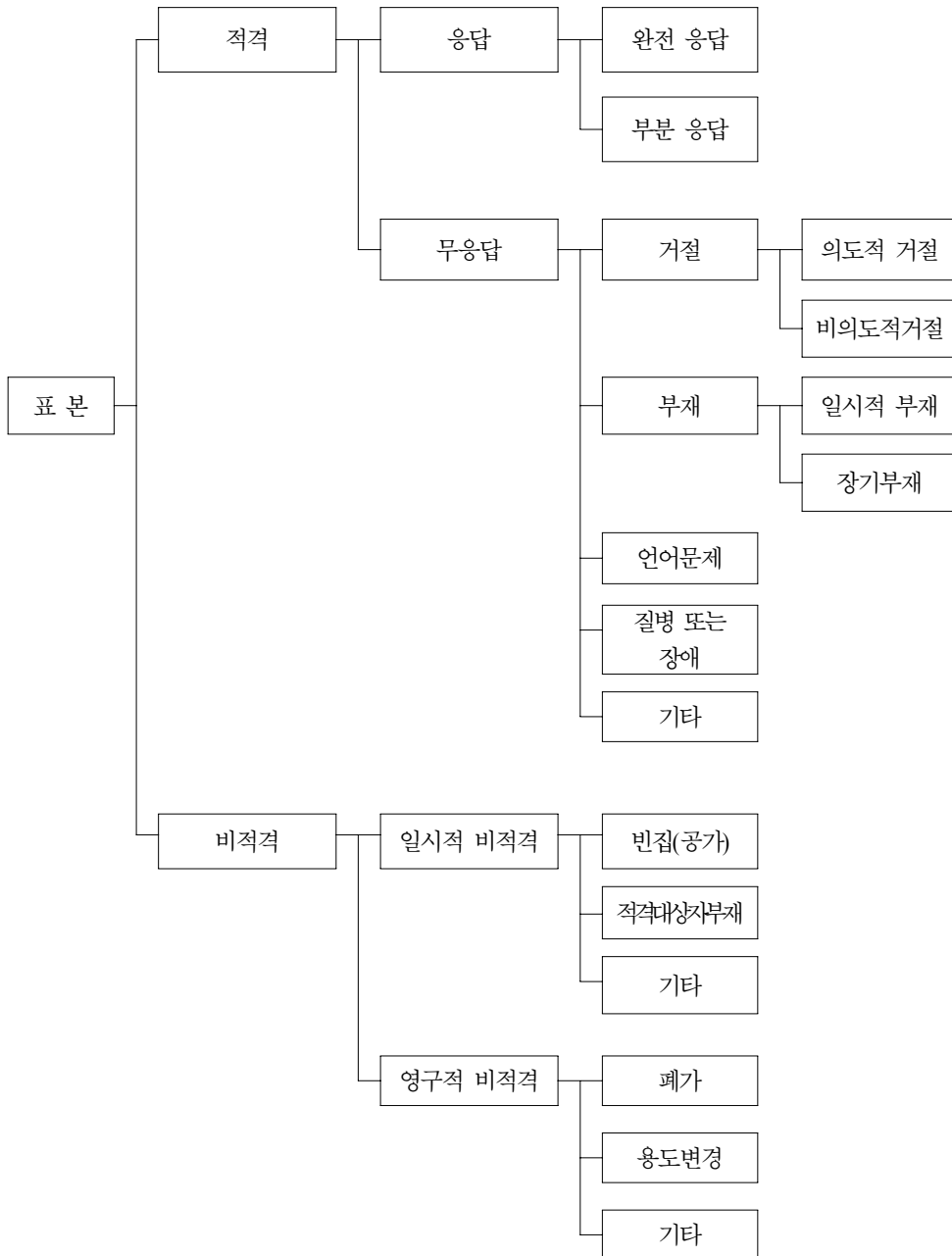
CPI 응답률의 특징은 적격표본 중 조사에 응답한 표본의 비율뿐만 아니라, 실제로 분석에 포함된 표본의 비율까지 제공한다는 점이다. 위의 LA 경우 사업체 수 기준의 응답률(95.3%)에 비해 통계작성에 사용된 사업체 비율(90.9%)은 다소 떨어짐을 알 수 있다. 즉, 수집된 자료의 약 5%는 자료가 불충분 하거나 신뢰성에 대한 문제로 집계에서 제외되었음을 보여주는 것이다. 또한 CPI의 가중응답률은 제공하지 않는 점에서 앞의 OES 조사와 다소 차이가 있다.

제5절 국가통계 응답률 산출 방법 제안

1. 응답상황 분류

응답률 산출을 위해 우선적으로 먼저 선행되어야 할 것은 바로 표본단위(가구, 사업체 등)의 응답상황을 적절한 기준에 따라 분류하는 것이다. AAPOR에서는 ‘disposition codes’로, OMB, BLS 등에서는 ‘outcome codes’라고 부르는 이 표본단위 접촉결과 분류코드는 표본단위의 적격/비적격, 응답/무응답, 무응답일 경우에는 그 사유까지 체계적으로 분류하여 관리함으로써 각 표본단위의 조사 성공 여부와 향후 조사가능성을 파악할 수 있는 중요한 자료가 된다. 또한 이들 분류코드별로 집계하면 자연스럽게 응답률 계산도 가능해진다.

우리나라 조사환경에 적합한 조사 응답상황 분류를 다음 [그림 5-11]과 같이 제안한다.



[그림 5-11] 응답 상황 분류(안)

이 분류코드를 작성하기 위해서는 첫째, 조사통계의 표본단위에 대한 정확한 정의가 있어야 한다. 해당통계의 조사대상이 무엇인지, 적격단위에 대한 명확한 정의가 명시되지 않으면 실제 현장(field)에서 확인한 표본단위가 적격인지 비적격인지 판단이 모호하게 되고, 이는 응답률산출식의 분모부분이 부정확해지는 결과를 초래한다. 이를 방지하기 위해서는 해당 조사지침서에 표본단위의 적격여부를 가능할 수 있는 다양한 예제를 들어 어떤 표본단위가 적격으로 조사대상에 포함되는지 혹은 제외되는지 구체적인 설명과 안내가 필요하다.

둘째, 응답여부의 판정기준이 마련되어야 한다. 수개에서 많게는 수십 개에 달하는 조사항목에 대하여 100% 자료수집이 안될 수도 있다. 이들 불완전 응답 케이스에 대하여 전체 조사항목의 몇 %이상 응답하였을 경우에 응답된 케이스로 간주할지, 혹은 일부 무응답하였으나 중요항목에 대해 응답한 경우 응답 케이스로 간주할지 정의되어야 한다. 참고로 AAPOR(2011)는 80% 이상 응답하였을 경우 완전응답, 50~80% 응답하였을 경우 부분응답, 50% 미만 응답하였을 경우 중도포기 혹은 무응답으로 보고 있다.

표본설계에 따라 추출된 표본단위는 조사목적에 부합하는 표본인 적격표본과 그렇지 않은 비적격표본으로 나뉜다. 적격은 다시 응답과 무응답으로 나눌 수 있는데, 다시 응답은 완전응답과 부분응답으로, 무응답은 거절, 부재, 언어문제, 질병 또는 장애 등으로 나뉘질 수 있다. 특히, 거절은 의도적 거절과 비의도적 거절로 구분가능한데, 예를 들어 사생활보호나 정부불신 등의 사유일 경우에는 의도적 거절로 볼 수 있으며, 사업체조사에서 해당되는 자료가 없어서 거절하는 경우는 비의도적 거절로 볼 수 있다. 그 외의 무응답 분류는 뒤에 나오는 분류코드 표를 참고하기 바란다.

비적격은 일시적 비적격과 영구적 비적격인 경우로 나뉘볼 수 있다. 일시적 비적격은 현재는 조사목적에 부합하지 않지만 향후 적격표본이 될 가능성이 남아있는 경우로서, 현재는 빈집이나 향후 전입자가 발생하여 적격표본이 될 수 있는 경우, 1인 이상의 가구원은 있으나 조사대상에 맞는 가구원이 없는 경우⁵⁾등이 해당된다. 영구적 비적격은 거처로서 기능을 상실한 폐가 또는 상업지역 등으로 용도변경 된 경우 등이 해당된다.

5) 예) 경제활동인구조사 표본가구로 선정되었으나, 실제 확인결과 15세 미만의 가구원으로 구성된 가구로 적격대상자가 없는 경우, 또, 다문화가족실태조사 표본가구로 선정되었으나 실제 확인 결과 다문화가정이 아닌 경우 등

본 연구에서는 특정조사에 국한되지 않고 범용으로 활용할 수 있는 수준에서 응답상황 분류코드를 설계하였다. 이 분류는 각 통계의 조사대상(가구, 사업체 등)에 따라, 혹은 조사방법(면접조사, 전화조사 등)에 따라 달라질 수 있는데, 가령 어떤 코드는 세분화되어야 하고, 또 어떤 코드는 필요 없을 수도 있을 것이다. 따라서 제안된 분류(안)은 작성기관의 필요에 따라 특정단계를 세분화 혹은 단순화 하여 사용할 수 있도록 계층적인 구조로 구성하였다. 예를 들면 가계동향조사와 같이 동일 표본을 반복해서 조사하는 경우에는 “의도적 거절(20211)”을 보다 더 세분화 하여 조사함으로써 거절 가구들에 대한 원인분석과 대응방안에 활용할 수 있을 것이다. 또, 사업체조사라면 보다 간결한 2단계까지의 분류코드로 작성하여도 무방할 것으로 보인다.

이상의 내용을 정리한 분류별 설명과 부여된 코드는 다음 <표 5-8>과 같다.

<표 5-8> 응답 상황 분류 코드(안)

분류 코드 (단계)				분류명	분류 설명
1	2	3	4		
10				최초 표본	아직 조사하지 않은 표본
20				적격	적격표본
	201			응답	표본단위 접촉결과 조사에 응답한 경우로 완전응답과 부분응답으로 나눔
		2011		완전응답	조사항목에 모두 응답한 경우
		2012		부분응답	조사항목에 일부분만 응답한 경우
	202			무응답	표본단위 접촉결과 조사에 무응답한 경우
		2021		거절	여러 차례의 협조요청(관리자 포함)에도 조사를 거절하는 경우로 의도적 거절과 비의도적거절로 나눔
		20211		의도적 거절	사생활보호, 정부불신 등을 이유로 거절한 경우
		20212		비의도적 거절	자료제출 곤란(자료 부재 등)으로 거절한 경우
		2022		부재	조사대상처(거처)에 조사대상자가 없는 경우
			20221	일시적 부재	출장, 여행, 입원 등으로 조사기간 내 일시적으로 조사대상자가 집에 없는 경우로서 다음 조사에서는 응답할 가능성이 있음
			20222	장기 부재	표본단위가 거처에 없는 경우. 여러 차례 재방문 (또는 전화)하고, 메시지를 남기거나 이웃 등에 고지하여도 조사대상자와 접촉이 되지 않는 경우로서 일시적 부재와 달리 지속적인 무응답 가능성이 있음
		2023		언어문제	한국어로 의사소통이 불가능하여 조사에 응답할 수 없는 경우
		2024		질병 또는 장애	장기투병, 중증장애, 노환 등으로 조사에 응하기 어려운 경우
		2029		기타	위의 무응답 사유 외의 사유로 무응답한 경우 (상중(喪中)인 경우, 가족전염병 발생으로 이동제한지역인 경우, 자연재해로 접근이 불가능한 경우 등)



분류 코드 (단계)				분류명	분류 설명
1	2	3	4		
30				비적격	조사목적에 맞지 않는 표본단위인 경우
	301			일시적 비적격	현재는 비적격이나 향후 적격표본이 될 가능성이 있는 경우
		3011		빈집(공가)	거처로서 기능은 유지한 채로 장기적으로 비어있는 집(매매 혹은 전월세 입주 전 거처인 경우 등)
		3012		적격대상자 부재	조사대상에 부합하는 사람이 없는 경우(예 : 15세 이상을 대상으로 하는 조사에서 15세 이상 가구원이 없는 가구)
		3019		기타	그 외의 일시적인 사유로 비적격인 경우
	302			영구적 비적격	향후에도 적격가능성이 없는 경우
		3021		폐가	오랫동안 빈집으로 방치되어 거처로서 기능을 상실한 경우
		3022		용도변경	거처가 상가, 식당 등으로 용도가 변경된 경우
		3029		기타	그 외의 사유 외의 영구적 비적격인 경우

2. 응답률 산출 예제

각 통계작성기관의 실제 적용이 용이하도록 아래 가상예제를 통하여 응답률을 산출해 보기로 한다. 동 예제에서는 단위응답률만 산출하고, 항목에 대한 가중응답률은 계산하지 않는다. 만약 항목응답률을 산출하고자 한다면 표본단위의 추출률의 역수를 가중치로 적용하여 이용가능하다.

< 1단계 : 응답상황 조사 >

먼저 전체 표본가구(20가구)를 대상으로 응답여부와 비적격여부를 파악한다. 응답 외의 무응답과 비적격의 사유를 파악하는 것은 표본의 지속적이고 체계적인 관리 차원에서 매우 중요한 사항이다. 각 가구별로 가상의 응답상황을 구성하여 앞서 제안된 해당 분류코드를 부여하였다.

〈표 5-9〉 (예) 20가구 응답상황 조사결과

가구	조사 결과	분류 코드	가구	조사 결과	분류 코드
1	완전응답	2011	11	완전응답	2011
2	완전응답	2011	12	완전응답	2011
3	부분응답	2012	13	빈집	3011
4	부분응답	2012	14	식당으로 용도변경	3022
5	거절(정부불신)	20211	15	거절(자료 없음)	20212
6	귀화가정으로 한국어미숙	2023	16	자연재해로 도로가 유실되어 접근불가	2029
7	중증장애인으로 조사곤란	2024	17	"	2029
8	완전응답	2011	18	완전응답	2011
9	완전응답	2011	19	완전응답	2011
10	부재(입원)	20221	20	부분응답	2012

< 2단계 : 응답상황 코드별 집계 >

1단계에서 조사된 각 표본가구의 응답상황을 응답상황 분류코드별로 집계하면 다음 표와 같다. 총 20개 표본 중 응답가구는 11개, 무응답가구는 7가구, 비적격가구는 2가구로 집계되었다. 그런데, 대부분 작성기관에서 무응답발생시 표본대체를 실시하므로 위의 무응답(202) 7가구 모두 다른 가구로 대체하여 응답자료를 확보하였다고 가정한다.

〈표 5-10〉 응답상황 코드별 집계

분류 코드 (단계)				코드정의	표본 수	대체
1	2	3	4			
10				최초 표본	20	
20				적격 응답	18	7
	201			완전응답	8	5
		2011		부분응답	3	3
		2012				
	202			무응답	7	
				거절	2	
			2021	의도적 거절	1	
			20211	비의도적 거절	1	
			20212			
		2022		부재	1	
			20221	일시적 부재	1	

분류 코드 (단계)				코드정의	표본 수	대체
1	2	3	4			
			2022	장기 부재	-	
		2023		언어문제	1	
		2024		질병 또는 장애	1	
		2029		기타	2	
30				비적격	2	
	301			일시적 비적격	1	
		3011		빈집(공가)	1	
		3012		적격대상자 부재	-	
		3019		기타	-	
	302			영구적 비적격	1	
		3021		폐가	-	
		3022		용도변경	1	
		3029		기타	-	

< 3단계 : 응답률 계산 >

외국의 경우에는 표본대체를 하지 않으므로 외국의 사례에서 확인되는 모든 응답률은 표본대체를 하지 않은 응답률이다. 그러나, 우리나라는 거의 모든 조사에서 표본대체가 이뤄진다. 통계청에서도 무응답이든 비적격이든 조사가 안 된 표본은 표본대체를 통해서 자료를 확보하고 있다. 실제 우리 조사현실을 볼 때 표본대체를 원칙적으로 금지하기는 어려운 문제이다. 따라서, 우리 이런 실정을 고려한다면, 표본대체 전의 응답률과 표본대체 후의 응답률 두 가지 모두를 산출하는 것이 현실적인 대안이 될 수 있다.

먼저, 표본대체 전의 응답률을 계산해보자.(단, 부분응답도 응답률계산에 포함하기로 한다.) 응답률의 정의대로 적격표본 중 응답표본의 비율로 계산하면 이래 ①식과 같다. 전체 20개 가구 중 비적격으로 확인된 2가구를 제외한 18 가구가 분모에 해당되며, 완전 응답한 8가구와 부분응답한 3가구를 합한 총 11가구가 분자에 해당되어 응답률은 ①식과 같이 계산된다.

$$\text{응답률} = \frac{\text{응답가구수}}{\text{적격가구수}} \times 100 = \frac{(8+3)}{18} \times 100 = 61.1\% \quad \dots\dots\dots ①$$

다음은 표본대체를 감안한 응답률로서 무응답(202)으로 확인된 7가구에 대해서 모두 다른 가구로 대체하여 자료를 확보하였다고 가정하였다. 따라서 분모에는 최초 적격가구인 18가구에 대체한 7가구를 합한 총 25가구가 되고 분자는 최초 응답가구인 11가구에 대체한 7가구를 합한 18가구가 되어 응답률은 아래 식 ②와 같이 계산된다.

$$\text{대체포함응답률} = \frac{\text{최초 응답가구수} + \text{대체가구수}}{\text{최초 적격가구수} + \text{대체가구수}} \times 100 = \frac{(11+7)}{(18+7)} \times 100 = 72.0\% \dots\dots ②$$

대체를 감안한 응답률 ②는 72.0%로, ①의 응답률과 비교할 때 약 11%p 높게 계산됨을 알 수 있다. 그러나 통계청 작성 통계 중 식 ②와 같이 계산하는 것이 아닌 분자에만 대체가구를 더함으로써 식 ③과 같이 계산하여 응답률이 100.0%가 나오게 되는 경우가 있다. 이는 추가된 대체가구수를 분모에는 포함하지 않고 분자에만 포함하여 계산한 것으로 응답률이 아니라 조사완료율로 보는 것이 타당할 것이다.

$$\text{조사완료율} = \frac{\text{응답가구수} + \text{대체가구수}}{\text{적격가구수}} \times 100 = \frac{11+7}{18} \times 100 = 100.0\% \dots\dots ③$$

②번식의 경우 표본대체 1차 시도에서 모두 대체가 성공한 경우를 가정하였으나, 실제 현장에서는 표본대체가 몇 차례 걸쳐 완료될 가능성이 많다. 이 경우에는 아래 ④식과 같이 대체를 위해 접촉한 모든 가구를 분모에 더해주고 최종적으로 대체에 성공한 가구 수를 분자에 더해주는 방법으로 식을 일반화 할 수 있다.

$$\text{대체포함응답률} = \frac{\text{최초 응답가구수} + \sum_{\text{대체횟수}=1}^n \text{대체 성공 가구수}}{\text{최초 적격가구수} + \sum \text{대체 접촉 가구수}} \times 100 \dots\dots ④$$

식 ①, ②, ③, ④ 중에서 응답률에 대한 국제적인 기준에 부합하는 응답률 산출식은 ①번식이다. 그러나, 현재 표본대체가 상당부분 발생하고 있는 우리 통계현실을 무시하고 ①번식만을 고집할 수는 없을 것이다. 현실이 그러하다면, 이 대체가구수를 응답률 계산식의 분자와 분모에 모두 더해줌으로써 실제 대체가 이뤄진 상황을 반영한 ④번 기준의 응답률도 함께 작성하여 제공함으로써 이용자가 상황에 맞게 응답률을 판단하도록 관련 응답률을 모두 제공해 주는 것이 필요하다.



응답률+무응답율=100%라는 식에 부합하기 위해서는 대체가 포함된 무응답률을 아래와 같이 정의할 필요가 있다.

$$\text{대체포함무응답률} = \frac{\text{최초 무응답가구수} + \sum \text{대체 실패 가구수}}{\text{최초 적격가구수} + \sum \text{대체 접촉 가구수}} \times 100 \dots\dots\dots ⑤$$

제6절 결론

각 통계작성기관에서 응답률 제공이 실행되지 않는 데에는 분명히 어떤 제약조건들이 있을 것이다. 당초 확률추출에 의해 표본설계를 하였더라도 막상 현장에서 표본을 확보할 때 간혹 일부는 할당표집이 되기도 하고 조사가 시작되기도 전에 여러 차례의 표본대체를 해야만 할 때도 있을 것이다.

응답률 산출은 응답(혹은 무응답)과 관련된 정보들을 체계적으로 수집하는 것에서부터 시작한다. 즉, 조사항목에 대한 조사이전부터 표본대상에 대한 정보를 파악하는 것이다. 따라서, 이에 수반되는 추가적인 인력과 비용의 문제가 있을 수 있으며, 실제로 응답률을 어떻게 계산해야할지, 또 응답률 산출을 위해서 특별히 어떤 자료를 수집해야 할지 고민스러운 경우도 있을 것이다. 또한 내부적으로 응답률을 파악하고 있더라도 이를 공개하기를 주저하는 경우도 있을 것이다. 많은 작성기관들이 통계의 신뢰성 시비에 노출되고 있으므로 낮은 응답률 수치를 공표하였을 때 뒤따를 부담을 염려하고 있기 때문이다. 하지만 응답률이 설사 낮은 수치라 하더라도 그 자체로 해당 조사의 특성과 한계를 보여주는 중요한 정보가 되며, 해당통계의 품질향상도 기대할 수 있다는 점을 인지할 필요가 있다.

실제 조사응답 상황코드가 작성되는 경우는 거의 없는데, 국내 조사환경을 고려할 때 조사원의 조사 부담이 부가되는 문제가 있으나, 그럼에도 불구하고 그 중요성은 간과할 수 없을 것이다. 다행히 통계청의 가구종합시스템에서 이미 표본가구의 적격여부가 조사되고, 적격가구 중에서 응답여부(가계부회수)가 확인되고 있으므로 이를 기반으로 조사하는 가구조사의 경우에는 여타 조사보다는 응답률을 산출하기 좋은 조건을 가지고 있다. 해결해야 할 것은 표본대체를 어떻게 볼 것인가이다. 표본대체를 안하는 것이 가장 좋지만, 현실적으로 대체를 해야만 한다면, 대체이전의 응답률과 대체이후의 응답률을 함께 제공하는 것이 필요하다. 대체전의 응답률로 국제기준을 충족할 수 있고, 대체

후의 응답률로는 실제 조사에 수집된 자료를 가늠할 수 있는 지표가 될 수 있다. 단, 대체 후 응답률의 경우에는 표본대체수를 분자분모에 모두 포함하도록 한다. 분자에만 포함하는 경우는 응답률을 과대 계산하게 되는 문제가 있다. 그리고, 응답률의 공개는 정책적 결정이 필요한 문제로 정책결정권자의 관심과 응답률 제공의 필요성에 대한 인식이 요구되는 부분이다.

최근 통계개발원에서는 조사과정에 발생하는 데이터, 즉 조사과정자료(Paradata)를 효율적으로 수집 관리하기 위한 연구가 진행 중에 있다. 이 조사과정자료에는 표본단위의 적격/비적격과 응답여부 항목 등이 포함되어 조사되므로 이 자료수집에 의해 응답률 계산이 가능할 수 있다. 향후 이 조사과정자료 수집이 각 작성통계에 실용화된다면 응답률 산출은 보다 용이해질 것이다. 물론 자료의 관리나 조사비용 등을 고려할 때 조사응답 상황코드를 비롯한 조사과정자료(paradata)를 수집한다는 것은 통계작성기관에게는 큰 부담일 수 있다. 하지만 각 통계 작성기관에서 응답률이 가지는 품질지표로서의 중요성을 인식하고 다소 어려움이 있더라도 응답률 산출에 동참한다면 보다 신뢰성 있는 국가통계작성에 한걸음 더 나아갈 수 있을 것이다.

참고문헌

- 통계청(2008), “조사통계의 정확성지표 품질관리 매뉴얼”
- 통계청(2009), “국가통계 실무지침 - 조사통계 실무자를 위한 가이드”
- 통계청(2009), “조사·보고통계 품질관리 안내서”
- 중앙선거관리위원회(2007), “선거에 관한 여론조사결과 공표·보도 기준안”
- 허명희(2007), “여론조사 방법론: 과제와 전망”(대한통계협회 <통계 33(1): 27-36>)
- 허명희(2008), “선거여론조사 자료의 표준적 요약과 시각화”(응용통계연구 21(5))
- 김서영·안다영(2010), “가구 면접조사에서 무응답률과 무응답 편향”(통계개발원 2010년 하반기 보고서)
- 손창균(2012), “조사응답률 제고방안”(제2회 국가통계방법론 심포지엄 발표자료)
- AAPOR(2011), “Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys”.
- U.S. Office of management and Budget(2006), “Standards and Guidelines for Statistical surveys”.
- Nancy Bates, Pat Doyle and F. Winters(2000), Survey Nonresponse : New Definitions and Measurement Methods(U.S. Census Bureau)
- Bethlehem, J., Cobben, F. and Schouten, B.(2010), “Handbook of nonresponse in household surveys”, Wiley
- U.S. Bureau of Labor Statistics(2003), “Survey Response Measurement at the Bureau of Labor Statistics”
- U.S. Bureau of Labor Statistics and Census Bureau(2004), “Response Rates and Nonresponse in Establishment Surveys-BLS and Census Bureau”
- U.S. Bureau of Labor Statistics and Census Bureau(2006), “Design and Methodology : Current Population Survey”.
- U.S. Census Bureau(2010), “Design and Methodology : American Community Survey”.
- U.S. Bureau of Labor Statistics(2011), “Detail Report-Response Rates for the Consumer Price Indexes, 2011”
- U.S. Bureau of Labor Statistics(2011), “Survey Methods and Reliability Statement for the May 2011 Occupational Employment Statistics Survey”
- U.S. Bureau of Labor Statistics(2011), “Occupational Employment Statistics Survey Technical Notes for May 2011”