

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

2021-02

정책연구

A Study on Improving Survey Methods
in the Culture and Tourism Sector According
to Changes in the Social Environment

김지학
박근화



한국문화관광연구원
Korea Culture & Tourism Institute

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

A Study on Improving Survey Methods in the Culture and
Tourism Sector According to Changes in the Social Environment

김지학·박근화



한국문화관광연구원
Korea Culture & Tourism Institute

연구책임

김지학	한국문화관광연구원 차석전문원
박근화	한국문화관광연구원 선임전문위원

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구



연구개요

1. 서론

가. 연구배경 및 목적

- 최근 첨단 IT 기술의 발전과 더불어 빅데이터가 등장하며, 생산되는 데이터를 정보로 만들어 활용하는 등 다양한 방법의 자료수집방법이 이용되고 있지만, 여전히 공공영역의 조사통계에서는 전통적인 조사방법을 고수하고 있음
- 하지만 사회의 발전과 함께, 그리고 새롭게 나타나는 사회환경의 변화는 이러한 전통적 방식의 조사환경에 큰 영향을 끼치고 있음
 - 인구구조의 변화: 1인가구 및 아파트가구의 증가로 인한 대면조사의 어려움
 - 디지털화(스마트기기 일상화): 스마트폰 보급률과 인터넷인프라를 통해 온라인 조사(웹, 모바일)를 다방면에 활용할 수 있는 환경 마련
 - 데이터 관련 법령의 개정 및 시행: 행정자료를 활용한 데이터 활성화 증대
 - 코로나19의 대유행: 비대면 조사로의 전환
- 이에 사회환경 변화에 따라 효율적이고 대표성 있는 통계값을 도출하는 조사방안을 제시하고자 함
 - 혼합조사 방법의 활용
 - 행정자료 활용을 위한 검토
 - 문화체육관광 관련 조사를 활용 방안

나. 연구범위 및 방법

- 시간적 범위는 사례연구는 최근의 현황을 살피기 위해 2010년 이후로 설정하였으며, 연구에서의 실험의 시간적 범위는 2020년 이후로 설정함
- 대상적 범위는 문화·관광 분야의 통계를 기반으로 하나, 전 분야의 조사방법에 범용적 적용이 가능하도록 설정함
- 내용적 범위는 조사방법별 비교와 혼합조사의 모드효과를 분석하고, 행정자료와 빅데이터의 활용 검토 등을 통해 조사의 효율성과 대표성을 확보하는 방안을 제시하는 것임
- 연구방법은 문헌연구, 사례분석, 전문가 자문회의를 실시하고, 혼합조사의 모드효과를 파악하고 추정하기 위한 시험조사를 설계하여 진행함

2. 사회환경 및 조사기법의 변화

가. 사회환경의 변화

■ 인구구조의 변화

- 1인가구는 1980년 4.8%, 2000년 15.5%, 2020년 31.7%로 증가추세가 지속되고 있으며, 맞벌이 부부는 2019년 기준 전체 유배우자 가구의 46%에 해당함
- 주 주거형태 역시 아파트는 1980년 4.9%, 2000년 36.8%, 2020년 54.3%로 절반이 넘게 차지하고 있음
- 이러한 변화는 실제 조사대상을 만날 수 있는 기회가 줄어들었음을 의미하며, 전통적 조사방식인 대면조사의 어려움을 보여주는 수치임

■ 디지털화(스마트기기의 일상화)

- 우리나라의 스마트폰 보급률은 약 95%, 인터넷인프라 역시 OECD 국가 중 1위로, 온라인 조사를 다방면에 활용할 수 있는 환경이 이미 마련됨
- 민간영역에서는 이러한 조사방법으로 변화하는 추세가 증가하고 있지만 국가통계의 경우 여전히 온라인조사는 일부에서만 쓰이고, 모바일조사 역시 소극적 태도를 보이고 있음
- 개인정보와 모집단 대표성 등의 문제를 해결한다면 향후 핵심적 조사방법이 될 것임

■ 데이터 관련 법령의 개정 및 시행

- 데이터3법 개정, 공공데이터법 개정, 데이터기반행정법 제정 등 데이터 활용을 위해 관련 법률이 제정 또는 개정됨으로써, 개인정보문제로 제약이 있어왔던 행정자료를 활발히 활용할 수 있는 방안이 마련됨
- 이러한 행정자료의 활용은 응답자들의 응답피로도를 낮추고, 비용을 줄일 수 있는 효과가 있음

■ 코로나19의 대유행

- 전염병의 확산은 사람들의 접촉을 꺼려지게 만드는 요인인데, 2020년부터 시작된 코로나19의 확산으로 사회적 거리두기가 시행되고 대면조사가 어려워짐
- 때문에 많은 경우 비대면 조사방식으로 일시적 변환하거나 전환되었으며, 기존

조사의 연속성을 살리려는 노력이 있어왔음

- 즉, 조사방식의 다변화 혹은 전통적 조사방법에서의 전환점이 필요함을 각인시킴

■ 인건비의 상승

- 물가상승에 따라 인건비의 상승은 비례적으로 증가해왔으나, 조사예산은 변화가 없어 조사품질을 높이기 위해 조사예산의 개선이 절대적으로 필요함

나. 조사기법의 변화

- 조사통계의 조사방법은 조사를 진행하는 수단(우편, 온라인, 면접, 전화 등)으로 구분할 수 있으며, 여러 조사 방식을 결합하는지 여부에 따라 단일조사와 혼합조사로 구분할 수 있음
- 사회와 발전과 변화는 조사환경에도 많은 변화를 가져왔는데, 2000년대 들어서며 과거의 전통적인 조사방법은 가구방문을 통한 대면조사만 거의 유지되고 있음
- 행정자료 및 빅데이터의 활용과 온라인(웹, 모바일 등)으로 조사하고자 하는 경향과, 응답률의 향상을 위해 단일조사가 아닌 혼합조사의 방식으로 하는 경향도 증가하고 있음
- 이러한 혼합조사는 응답률을 높이면서 응답시간과 비용을 줄이는 장점이 있으나, 포함률(Coverage rate)의 차이와 모드효과는 불안정성으로 작용하게 되므로 이를 파악하여 보정할 필요가 있음

3. 국내 통계 현황 및 사례검토

가. 통계 현황

1) 조사통계 현황

■ 국가승인통계

- 2021년 2월 기준 국가승인통계는 총 1,238종이며, 이 중 조사통계는 539종이며, 조사방법별로는 면접조사 79.0%, 인터넷조사 6.3%, 우편(FAX)조사 3.5%, 전화조사 1.9% 임

- 이렇듯 면접조사가 대부분인 이유는 다른 조사방법은 표본들이 부재한 경우가 많기 때문이다

■ 문화·관광통계

- 문화·관광 통계 분야의 승인통계 15종 중 조사통계는 12종임
- 조사대상이 국민인 경우는 가구방문 면접조사를, 사업체인 경우는 대부분 면접 조사를 원칙으로 하되 이메일, 팩스 등을 병행하고, 특정대상인 경우는 조사목록을 이용해 전화 컨택 후 조사대상자와 조사시기를 조정하여 진행하고 있음

2) 행정자료 현황

- 조사통계는 많은 예산과 시간이 소요되어 통계의 생산을 행정자료를 활용하는 방식으로 전환하는 추세임

■ 국가통계포털

- 통계청의 국가통계포털은 우리나라 승인통계를 한 곳에 모아 서비스하고 있음
- 현재 396개의 통계작성기관이 작성하는 1,239종의 국가승인통계를 수록함
- 하지만 통계표 위주의 수록으로, 필요한 통계값의 동일한 차원을 쓰는 용도 외에 행정자료를 결합하여 작성하는데에는 활용에 어려움이 있음

■ 공공데이터포털

- 행정안전부와 한국지능정보화사회진흥원이 운영하는 공공데이터포털은 공공기관이 생성 또는 취득하여 관리하는 공공데이터를 제공하고 있음
- 현재 총 57,034건의 공공데이터를 제공하고 있으며, 제공하고 있지 않은 공공데이터도 해당 포털에 신청서를 제출하면 심의를 거쳐 제공여부를 결정함
- 실제 분석 및 가공이 가능한 데이터 형태를 서비스하기 때문에, 그 차원변수를 활용할 수 있다면, 기존 조사와 결합하여 사용할 수 있을 것임

■ 통계데이터센터

- 통계청의 통계데이터센터는 마이크로데이터를 제공하는 것에 그치는 것이 아니라, 실제 데이터 간 연계분석이 가능하도록 직접 DB에 접속하여 작업이 가능함
- 특히 여기서 제공하는 DB 중 기업통계등록부는 조사자료와 행정자료를 상호 연계하여 구축한 통합모집단으로 행정통계자료 중 그 활용성이 매우 높음

나. 선행연구 및 사례검토

1) 조사과정자료(Paradata)

- 조사과정자료란 조사를 진행하는 과정에서 자연적으로 발생하는 파생자료임
- 조사과정자료 분석은 조사 과정에서 발생하는 다양한 오차들을 축소시킬 수 있음
- 이러한 과정은 조사의 개선으로 이어져 조사의 품질을 향상시킨다고 볼 수 있음

2) 혼합조사와 모드효과

■ 모드효과의 이해

- 혼합조사는 두 개 이상의 조사방법을 이용하여, 포함오차 및 무응답 축소 등의 장점이 있으나, 모드효과로 인한 품질의 우려가 존재함

- 모드효과는 선택효과와 측정효과로 구분하는데 이는 다음과 같음

$$\text{선택효과: } S_a(\pi_j) = P(Y_a = j \mid M = a) - P(Y_a = j \mid M = b)$$

$$\text{측정효과: } M_b(\pi_j) = P(Y_b = j \mid M = b) - P(Y_a = j \mid M = b)$$

- $P(Y_a = j \mid M = b)$ 는 실제로 관측이 불가능한 경우이므로, 아래와 같은 전확률 공식을 이용함

$$P(Y_a = j) = P(Y_a = j \mid M = a)P(M = a) + P(Y_a = j \mid M = b)P(M = b)$$

- 선택효과와 측정효과는 조사결과에 영향을 주므로, 이를 추정하여 혼합조사의 결과에 반영하는 것이 필요함

■ 모드효과 평가에 대한 이해

- 정확한 모드효과를 살펴보려면 선택효과와 측정효과를 파악하여야 하고 이를 위해 이 둘을 분리 추정하는 것이 필요함

- 이를 Vannieuwenhuyze 등(2010)이 제안한 방법으로 적용하면 다음과 같음
모드효과 = ⑥ - ①

$$\text{선택효과} = \text{㉓} - \text{㉑}$$

$$\text{측정효과} = \text{㉑} - \text{㉒}$$

■ NordChild survey의 조사방법별 차이비교

- 사회환경의 변화에 따라 조사 응답률이 낮아지자 대안 마련을 위해 4가지 방법(종이, 종이/웹, 웹, 웹/태블릿)을 비교하여 효율적인 방법을 파악하고자 함

- 조사 초기 웹/태블릿 모드의 응답률이 타 조사모드 대비 월등히 높아, 상대적으로 짧은기간의 조사에 적합함을 시사
- 첫 문항에 응답한 응답자가 마지막 문항까지 완료한 비율을 살펴보면 종이/웹 모드가 98.6%로 가장 높게 나타나, 응답자들이 선호하는 모드를 선택하는 편의성으로 인해 응답률에 긍정적 효과를 가져오는 것을 볼 수 있음
- 비용적인 측면에서도 종이 모드는 웹 모드 대비 약 2배가 높아, 한정된 조사비용 내에서 효율성 있는 조사를 위해 조사여건에 맞는 모드를 사용하는 것도 조사설계에 고려해야 할 사항임
- 병행적 혼합모드 시험조사
 - 통계개발원은 혼합조사의 모드효과 파악을 위해 시험조사를 실시하였고, 서울에 거주하는 총 1,600개 표본가구를 추출하여 조사함
 - 조사방법은 종이조사와 웹조사이며, 조사방법을 선택 할 수 있는 집단과 선택할 수 없는 집단으로 구분함
 - 모드효과를 선택효과와 측정효과를 분리 추정한 결과, 일부 항목을 제외한 대부분의 항목에서 유의미한 것으로 나타남
- 유럽사회조사
 - 응답률 감소, 면접조사 비용 등의 문제 해결을 위해 2003, 2005, 2008년에 실험연구를 실시함
 - 단일면접조사 2,674명, 혼합조사 878명을 랜덤하게 추출하고, 혼합조사 표본에 대해 CAPI, CATI, CAWI를 선택하게 함
 - 정치적 관심 항목에 대한 모드효과 분리 추정 결과, 자기기입방식 응답자는 면접방식 응답자에 비해 “꽤 있음”에 과소 응답하고, “거의 없음”에 과대 응답하는 측정효과가 관찰됨
- 미국 일반 공공 설문조사들의 혼합모드 분석
 - 웹과 우편의 혼합조사를 설계하여 항목 무응답 차이를 평가함
 - 조사자료는 Lewiston & Clarkston Quality of Life Survey’, ‘Washington Community Survey’, ‘Washington Economic Survey’를 사용
 - 응답 옵션은 ①우편전용 ②우편+웹 후속 조치 ③웹+우편 후속조치 으로 진행함
 - 각 실험결과 ②우편+웹 후속 조치 방식의 웹에서 가장 낮은 항목 무응답률을 보임

3) 행정자료와 빅데이터의 활용

■ 전통적인 설문조사와 빅데이터 서베이

- 설문조사는 현지조사의 형태를 가진 대면조사로 시작하여, 전화조사로 발전하였으며, 인터넷 조사, 모바일 조사 등 새로운 접근 방법이 활용되고 있음
- 표본의 대표성과 객관성 확보 등의 이유로 여전히 대면조사 방법론이 지속 활용되고 있으나, 급변하는 조사환경에서 새로운 조사방법에 대한 시도와 연구가 필요한 상황임
- 이에 빅데이터 분석방법론과 설문조사가 결합한 형태의 빅데이터 서베이가 대안으로 활용되기도 함

■ 빅데이터를 이용한 조사 및 활용

- 최근 들어 빅데이터의 활용이 늘고 있으나 개인정보문제로 이용이 제한적임
- 빅데이터를 보유한 기업의 정보를 활용하여 대표성 있는 표본추출을하고, 그들의 소비행태 등을 연계하는 즉 빅데이터 서베이의 방안이 있음
- 이러한 정보를 공신력있게 활용하기 위해서는 실제 조사한 결과가 우리나라 전체의 정보와 연계할 수 있는 지에 대한 부분이 필요함
- 또한 조사방법을 온라인조사로만 하고 있는데, 이러한 부분의 값이 대면조사 등과 비교할 때 어느 정도 차이가 나는지 등의 연구를 통해 정합성 등의 검증을 할 필요가 있음

■ 행정자료를 이용한 조사 및 활용

- 통계청의 인구주택총조사는 대내외적인 조사환경의 변화와 조사비용의 증대로 기존 전수조사로 실시하던 방식에서 2015년부터 등록센서스 방식으로 최초 전환 하였음
- 기업통계등록부를 이용하여 작성한 조사통계로는 중소기업실태조사, 장애인기업실태조사 등이 있으며, 가공통계로는 지자체행정통계 등이 있음
- 문화·관광 분야에서 행정자료를 활용하여 작성한 대표적 사례는 한국문화관광연구원원의 문화체육관광산업통계로, 기업통계등록부를 활용해 문화체육관광산업 영위 사업체의 경영활동 현황 등을 산출하는 가공통계임

4. 현황 분석

가. 분석개요

- 현황분석 I. 조사과정자료 분석
 - 조사과정 중에 수집되는 자료를 통한 조사수행 단계에서의 문제점 파악 및 개선 방안 도출
- 현황분석 II. 조사방식에 따른 차이 분석
 - 조사방식에 따라 결과값의 차이가 발생하는지 여부 검토
- 현황분석 III. 혼합모드 분석
 - 혼합조사에서 모드효과가 어떻게 발생하는지 파악

나. 조사과정자료 분석

- 국민여행조사
 - 코로나19로 조사를 멈출 수 밖에 없는 상황이되어, 온라인패널을 이용하여 조사를 일부 진행하는 방안을 고려하였고, 온라인 패널의 활용성 여부 판단이 필요해 시범조사를 실시함
 - 1주일간 4회 컨택, 원표본 2주간 유지에도 응답률이 저조하여, 전화 컨택으로 독려하였음에도 응답률이 6.8%에 불과함
 - 전체 패널 중 적극적 패널의 응답률은 25.5%였지만, 최근 접속하지 않은 패널의 응답률은 3.0%에 불과하였고, 6개월 이내 비접속자들이 95%에 해당하여 패널로서의 역할을 한다고 보기 어려움
 - 온라인패널은 국민 전체를 포괄한다고 보기 어려운 면이 있으며, 이를 활용할 시에는 사전 홍보방안과 더불어 패널 관리 방안이 필요함
- 국민여가활동조사 및 국민문화예술활동조사
 - 조사원의 업무량이 과다하게 배정되지 않게 하여 비표본오차를 줄이고자 하였으나, 실제 조사에서 1일 업무량의 과소는 조사성공률이 얼마나 낮은가를 짐작하게 해줌. 이는 행정자료의 활용 등을 통해 응답 피로도를 낮추고, 조사방식의 다변화를 통해 응답 편의성을 높이며, 적절한 보상을 통한 응답유도가 필요함을 시사함

- 조사구 대체 사유를 보면, ‘조사 거절’(70% 내외)이 제일 높고, 다음으로 ‘노년층 거주 비중이 높음’(13% 내외)으로 나타남. 이는 결국 젊은층을 조사하기 어려웠다는 점을 반증함
- 두 조사의 가구 접촉 현황을 보면, ‘부재’, ‘거절’, ‘성공’, ‘기타’, ‘가구없음’ 순으로 나타남. 부재로 인한 조사불가가 많은 것은 지역적인 특성이나 주택유형에 따른 주 연령층에 따라 요일이나 시간대별로 방문전략을 세울 필요가 있음을 시사함

다. 조사방식에 따른 차이 분석

1) 조사 방법(mode)별 차이

■ 문화체육관광 기업경기동향조사

- 기업경기동향조사는 문화체육관광 분야를 영위하는 사업체들을 4개산업(19개 중분류)으로 구분하여, 경영활동 실적과 전망을 파악하고 있음
- 기존에는 전화조사로 진행되어 왔으나, 응답률을 높이기 위한 방안으로 전화조사와 온라인조사를 혼합하여 수행하고 이의 결과차이를 비교함
- 전화조사와 온라인조사의 결과값에 차이는 있지만 통계적으로 유의하지는 않은 것으로 나타남

■ 국민여행조사

- 국민여행조사는 우리나라 국민들을 대상으로 여행 총량, 여행 행태, 여행 평가, 여행 제약 요인 등을 파악하고 있음
- 기존에는 대면조사로 진행되어 왔으나, 코로나로 인해 온라인패널 조사를 추가로 진행하였음
- 모든 여행 경험률에서 온라인패널 조사가 대면조사 대비 높게 나타남
- 이러한 차이는 온라인패널의 경우 적극적 패널 위주로 응답이 이루어질 가능성이 높은 것이 하나의 원인으로 판단됨

2) 조사 도구(tool)별 차이

■ 국민여가활동조사

- 2018년에 조사원에 의한 비표본오차를 줄이고 데이터 오염을 방지하기 위해,

- 기존 PAPI를 기본으로 하되 TAPI를 활용한 시범적 조사를 진행함
- 조사도구별 차이만을 보기 위해 인구특성에 맞게 쿼터를 관리하였음에도, 고연령층 응답에 어려움이 있어(TAPI의 경우 태블릿PC를 이용하기 때문에) PAPI에 비해 적은 비율로 조사됨
 - 여기비용, 휴가 일수, 지속적 여가활동 비율은 조사도구별로 유의미한 차이를 보이지 않았으나, 여가활동 개수, 동호회 활동 경험, 자원 봉사 활동 경험, 스마트기기 활용 시간, 여가생활 만족도 등의 경우는 조사도구별 유의미한 차이를 보임

라. 혼합조사의 모드효과 분석

1) 문화체육관광 소비자동향조사

■ 선택형 혼합조사

- 모바일조사의 선택성향은 연령대가 낮을수록, 학력이 높을수록, 소득이 높을수록 높게 나타남
- 현황에서는 4개 항목(‘스포츠용품 및 기구’와 ‘미디어기기 구입비’, ‘국내여행비’, ‘도서구입비’)을 제외한 나머지 12개 항목에서 조사방법별로 유의한 차이가 있는 것으로 나타났고, 전망에서는 7개 항목(‘오락시설 및 지출비용’과 ‘온라인 문화생활비’, ‘스포츠 경기 관람료’, ‘아웃도어용품 구입비’, ‘해외여행비’, ‘음악미술학원비’, ‘여행관광비’)에서 조사방법별로 유의한 차이가 있는 것으로 나타남
- ‘종합지출 현황’으로 선택형 혼합조사의 모드효과를 살펴보면 모드효과는 -12.11이며, 선택효과와 측정효과를 분리하여 순수한 모드효과는 -12.26이 존재하는 것으로 분석됨

■ 할당형 혼합조사

- ‘스포츠경기 관람료 현황’에 대한 문항을 제외하고는 모두 전화조사와 모바일조사 두 조사방법에 따라 차이가 있는 것으로 나타남
- 다른 변수들의 영향을 배제한 순수한 조사방법의 효과를 파악하기 위해 회귀분석을 실시한 결과, ‘오프라인 문화생활비’의 경우 전화조사가 온라인조사보다 12.832점이 더 높은 것으로 나타남

2) 문화체육관광 기업경기동향조사

■ 선택형 혼합조사

- 전화조사와 웹조사 중 선택하여 조사를 하도록 하였으나, 소비자동향조사와 같이 응답성향에 대한 차이가 사업체의 특성별로 발생하지는 않음
- 현황에서는 투자규모와 외부고용사정, 상품단가에서 전화조사의 동향조사 결과가 높게 나타났으며, 전망에서는 투자규모와 외부고용사정에서 전화조사의 동향조사 결과가 높게 나타남
- 측정효과는 매출규모의 경우 매우 적게 나타났으나, 현황에서는 투자규모와 외부 고용사정, 전반적 업황에서 크게 나타났고, 전망에서는 투자규모와 전반적 업황의 경우 절대값이 모드효과보다 더 크게 차이가 나타남

■ 할당형 혼합조사

- 투자규모와 외부고용사정은 현황과 전망에서 모두 유의수준 0.001 수준에서도 유의하게 나타나, 모드별 차이가 큰 것으로 나타남
- 현황에서는 채산성, 자금사정, 상품단가는 유의수준 0.05 수준에서 유의한 것으로 나타났으며, 전망에서는 투자규모, 외부 고용사정, 채산성, 자금사정이 유의수준 0.001수준에서 유의하게 나타남
- 순수한 조사방법의 효과를 파악하기 위해 회귀분석을 실시한 결과, '외부 고용사정'의 경우 웹조사방식이 전화조사방식보다 -10.165점 차이가 나는 효과가 있는 것으로 나타남

마. 소결 및 시사점

- 기존 대면조사 방식을 온라인 조사방식으로 변경할 때, 온라인 패널이 국민 전체를 대표하고, 비활성화 패널이 최소화 될 수 있도록 패널 관리가 필요함
- 가구방문조사의 경우, 표본 대체 사유를 통해 응답 대상자의 특성에 맞는 방문 전략이 필요함
- 원표본 유지율을 높이기 위해 노후화된 조사구의 업데이트가 필수적이며, 지리적 특성 뿐만 아니라 1인가구와 같은 특정 대상의 조사 성공을 위한 전략이 사전에 수립되어야 함

- 조사방법 혹은 조사도구에 따라 결과에 차이가 나는 것을 확인함. 따라서 응답률을 높이기 위해 혼합조사를 할 경우 모드효과를 파악할 필요가 있으며, 이를 종합적으로 반영한 추정방안 까지 고려되어야 함

5. 혼합조사 추정 방안

가. 혼합조사 추정방법

- 최적결합방법
 - 최적결합방법은 서로 다른 조사방법으로 인해 도출된 결과의 분산을 이용하여 최소평균제곱오차가 되도록 결과를 결합하는 방법임
 - 이는 분산이 작은 조사방법의 조사결과의 가중치를 높이고, 분산이 큰 조사방법의 조사결과의 비중을 적게 작동하도록 하는 방법임
 - M개의 조사방법을 이용한 혼합조사를 가정할 때, 최적가중값은 다음과 같음

$$w_m = \frac{\sigma_m^{-2}}{\sigma_1^{-2} + \sigma_2^{-2} + \dots + \sigma_M^{-2}}$$

- 성향점수를 이용한 추정방법
 - 성향점수보정 방법은 서로 다른 두 집단을 비교할 때, 공변량이 영향을 주는 선택편의를 보정하기 위한 방법임
 - 성향점수는 통제집단을 기준으로 실험집단의 결과를 비교하여 실험효과를 측정하며, 실험효과를 정확하게 측정하기 위해 두 집단의 공변량의 분포를 안정시킴
 - 혼합조사(ex. 대면조사+온라인조사)에서 대면조사가 참조조사의 역할을 한다고 가정할 때, 온라인조사의 추정값은 다음과 같음

$$\hat{y} = \frac{\sum_c \sum_{j \in s_c^I} d_j^{I.PSA} y_j}{\sum_c \sum_{j \in s_c^I} d_j^{I.PSA}}$$

나. 혼합조사 추정을 위한 시험조사

1) 시험조사 설계

- 문화·관광분야의 사업체 대상, 국민 대상 조사의 모드효과 분석을 위해 기존의 문화체육관광 동향조사를 기반으로 전화조사와 온라인조사를 혼합한 시험조사를 아래와 같이 추가 설계함

구분		소비자동향조사	기업경기동향조사
모집단		우리나라 전 국민	해당분기에 문화체육관광 관련업을 영위한 사업체
조사 대상		전국 17개 시도 만 19세 이상 가구주 및 가구주의 배우자	문화체육관광 표준산업구분에 해당되는 문화예술, 문화산업, 스포츠, 관광사업체 중 상시 종사자 수 5인 이상 사업체
혼합 I	표본 크기	전화조사: 1,000명 온라인(모바일)조사: 500명	전화조사: 1,700개소 온라인(이메일)조사: 500개소
	조사 방법	2가지 조사방법별 따로 설계하여, 각각 전화/온라인조사	2가지 조사방법별 따로 설계하여, 각각 전화/온라인조사
혼합 II	표본 크기	500명	500개소
	조사 방법	RDD를 통한 번호 추출하여 조사대상을 컨택한 후, 응답자가 전화조사 or 온라인(모바일)조사를 선택하여 응답진행	추출된 표본리스트를 통해 전화 컨택 후, 응답자가 전화조사 or 온라인(이메일)조사를 선택하여 응답진행

2) 혼합조사 추정

① 소비자동향조사 혼합조사 추정

- 최적결합방법에 의한 추정
 - 본문 <표5-2> 선택형 혼합조사에서 최적결합방법과 기존 방법의 추정값 제시
 - 본문 <표5-3> 할당형 혼합조사에서 최적결합방법과 기존 방법의 추정값 제시
- 성향점수를 이용한 추정
 - 기준조사는 전화조사 방법, 비교조사는 모바일조사 방법으로 설정하고, 로지스틱 회귀분석을 적용하며, 예측값을 성향점수로 활용함
 - 본문 <표5-6> 선택형 혼합조사에서 성향점수보정과 기존 추정값 비교 제시
 - 본문 <표5-9> 할당형 혼합조사에서 성향점수보정과 기존 추정값 비교 제시

② 기업경기동향조사 혼합조사 추정

- 최적결합방법에 의한 추정

- 본문 <표5-10> 선택형 혼합조사에서 최적결합방법과 기존 방법의 추정값 제시
- 본문 <표5-11> 할당형 혼합조사에서 최적결합방법과 기존 방법의 추정값 제시
- 성향점수를 이용한 추정
 - 기준조사는 전화조사 방법, 비교조사는 웹조사 방법으로 설정하고, 로지스틱 회귀분석을 적용하며, 예측값을 성향점수로 활용함
 - 본문 <표5-14> 선택형 혼합조사에서 성향점수보정과 기존 추정값 비교 제시
 - 본문 <표5-17> 할당형 혼합조사에서 성향점수보정과 기존 추정값 비교 제시
- ③ 추정방법의 비교
 - 소비자동향조사의 결과 비교
 - 선택형과 할당형에서 최적결합추정방법이 분산을 낮게 만들어 안정적인 추정값을 제시하는 것으로 나타남
 - 추정값에서는 선택형과 할당형의 결과에서 차이가 있는 것으로 나타남
 - 모드효과보다 측정효과의 값이 크게 나타나기 때문에 분산을 줄여 측정결과의 효율성을 높여주는 최적결합추정방법이 현재조사결과에서는 더 타당한 것으로 분석됨
 - 기업경기동향조사의 결과 비교
 - 선택형과 할당형에서 성향점수를 이용한 방법이 기존값과의 차이가 적은 것으로 나타났으며, 분산추정에서는 최적결합추정방법이 더 안정적으로 나타남
 - 기업경기동향조사에서도 소비자동향조사와 거의 유사한 결과가 나타났는데, 성향점수를 이용한 방법의 분산이 낮아지지 않았기 때문에 성향점수를 이용한 추정방법의 결과를 활용하기는 어려울 것으로 분석됨

라. 소결 및 시사점

- 소비자동향조사와 기업경기동향조사를 각각 선택형과 할당형으로 전화조사와 모바일조사(웹조사)를 진행한 결과를 이용하여 모드효과와 추정방법을 적용한 결과, 모드효과에서는 선택효과보다는 측정효과가 크게 나타났으며, 전화조사가 모바일조사(웹조사)보다 크게 나타남
- 추정방법에서는 모두가 최적결합방법이 적절한 것으로 나타남

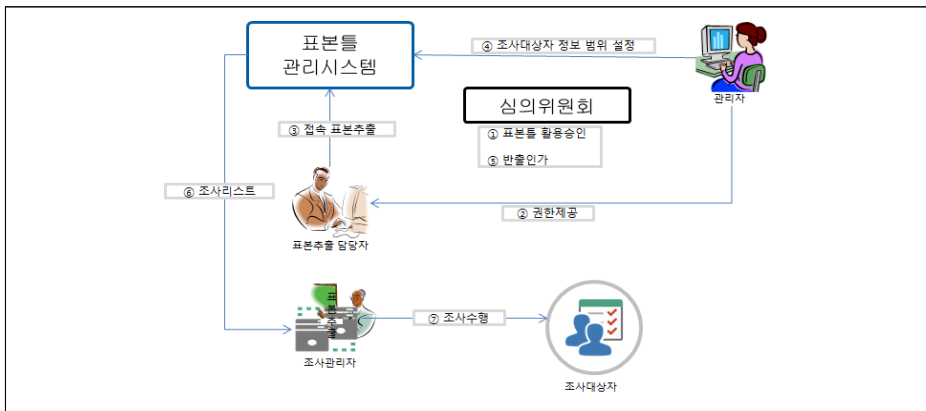
- 혼합조사로 이루어진 조사방법은 각 조사방법으로 인해 차이가 발생하는 것을 확인하였고, 따라서 이러한 문제를 해결하기 위해서는 모드효과를 파악하고 적절한 추정방법을 적용해야 하며, 조사 이전에 어떻게 조사를 진행할지에 대한 충분한 조사설계 방안을 마련할 필요가 있음

6. 결론 및 제언

가. 조사설계 개선 방안

- 혼합조사 설계 방안
 - 혼합조사 설계 시, 개인정보를 포함한 표본들의 활용이 가능해야 단순 병행조사에 머무르지 않을 수 있으며, 이를 통해 다양한 방식의 조사 진행이 가능함
 - 또 다른 방식인 조사패널의 이용은, 문화·관광 분야의 조사를 위한 대표성있는 패널 확보가 중요하며, 이는 현재 조사업체들의 온라인 패널조사와 유사한 방식임
 - 하지만 혼합조사는 모드효과가 발생하기 때문에, 이를 파악하고 최적의 추정방안을 마련하는 것이 필요함
- 문화·관광 분야의 조사 표본틀 구축 방안
 - 문화·관광 분야의 표본틀을 지속적으로 활용하기 위해서는 표본틀을 별도로 관리할 필요가 있으며, 이를 위한 시스템 구축과 함께 관리할 수 있는 제도 마련이 필요함

[그림 6-2] 표본틀 활용과정



- 또한 문화·관광 분야의 국민대상 조사패널 구축의 경우, 기존 온라인조사패널에 내재된 선택편의의 제거가 필요함
- 이를 해결하기 위해 마스터샘플(master sample)을 구축하여 조사패널로 활용하는 방안이 있음
- 통계청과 협의 하에 전국의 일정비율의 조사구를 추출하여 구축하는 방안과 현재 진행하고 있는 가구방문조사에서 패널 참여에 대한 정보를 얻어 구축하는 방안이 있음

나. 조사방식 개선 방안

■ 행정자료를 활용한 조사 방안

- 행정자료 활용을 위해 매크로데이터 이용과 마이크로데이터를 활용하는 방법이 있음
- 매크로데이터 활용은, 매크로데이터에 있는 정보는 빠고 조사를 수행하여 두 통계를 결합하는 방법이며 어렵지않게 활용할 수 있는 장점이 있음
- 마이크로데이터의 경우, 조사된 데이터와 매칭하는 방법이 있는데, 이는 연계변수의 여부에 따라 정확매칭과 통계적매칭 방법이 가능함
- 다른 하나는, 마이크로데이터를 표본으로 활용하는 방법인데, 이는 정확매칭과 동일한 수준의 정보활용이 가능한 장점이 있으나, 개인정보 문제가 직결되는 민감한 부분이 있음. 따라서 익명화 처리방법을 함께 고려해야 함
- 특히 표본틀을 이용하는 방법은 조사방법을 개선할 수 있는 중요한 방법 중 하나임
- 예를들어, 국세청의 신고자료를 활용할 수 있다면, ‘전국사업체조사’자료에 없는 1인 사업체의 정보가 있기 때문에, 프리랜서가 많은 예술분야와 콘텐츠분야의 세밀한 조사에 활용할 수 있기 때문임

■ 빅데이터와 연계한 조사 방안

- 조사대상자의 동의를 구해 지속적으로 조사하는 패널조사 방식으로 조사할 수 있다면 더욱 활용도 높은 정보로 활용이 가능함
- 빅데이터 가입자들이 우리나라 국민을 대표한다고 하기 어렵기 때문에, 이를 활용하기 위해서는 모집단의 포괄성을 분석해야 하며, 이러한 이유로 마스터샘플

을 만들어 이용하는 방법을 고려할 필요가 있음

- 이는 개인정보 문제가 발생하므로 행정자료 활용과 동일하게 마스터샘플의 정보를 관리하고, 데이터를 분석할 수 있는 시스템을 갖춘 보안센터가 있어야 할 것이며, 이들의 응답에 대한 적절한 보상도 중요한데, 이 부분은 패널조사의 패널관리와 유사하다고 할 수 있음
- 빅데이터와 조사데이터를 연계하는 방안의 하나인 빅데이터 가입자들을 대상으로 조사하는 방안은 데이터 활용에서 매우 중요한 과제가 될 것임

다. 제언

■ 조사 포괄성 확보

- 공공분야에서 사회조사는 대표성이 있어야 하며, 이를 위해서는 문화·관광 분야의 특정 대상을 조사할 수 있는 표본틀 구축과 마스터샘플 구축 방안이 있음

■ 개인정보 활용방안 마련

- 비대면조사의 경우 조사대상자들의 정보가 없다면 실제 조사대상자를 컨택하는 과정에서 대면이 불가피하기 때문에, 조사를 위해 필요한 개인정보는 충분히 활용할 수 있어야 함
- 결국 이를 위해서는 개인정보를 관리하고 분석할 수 있는 보안센터가 필요하며, 데이터 익명화 처리방안도 함께 마련되어야 함

■ 혼합조사 방안과 데이터 익명화 방안의 지속적 연구

- 문화·관광 분야에 종사하는 대상들은 상대적으로 이동량이 많아 접촉하여 조사하는데 어려움이 따르고, 이러한 저조한 응답률을 높이기 위해선 응답하기 편한 방법으로 조사할 수 있어야 함
- 따라서 혼합조사를 통해 원활한 조사를 진행한 후 이를 추정할 수 있도록 설계가 되어야 하며, 더 좋은 추정을 위한 지속적 연구가 진행되어야 함
- 세밀한 정보 활용과 분석에서 개인정보의 문제는 언제든지 발생할 수 있기 때문에, 개인의 인구통계학적 자료 외에 분석하는 자료들도 익명화 처리 방안이 필요하며, 이는 조사통계뿐만 아니라 빅데이터 등 다양한 분야에서 데이터를 활용하는데 있어 매우 중요함

■ 조사예산 확보

- 양질의 데이터 확보를 위해서는 조사원, 응답자 모두에게 충분한 사례가 반영되어야 하며, 설문지 양과 시간등을 고려한 합당한 지불 방안 마련이 필요함
- 마스터샘플 구축의 경우, 이들을 유지, 관리하는 비용과 보안센터 운영의 비용도 발생하기 때문에, 향후 조사방안 개선이 이뤄진다면 이러한 사항들을 고려한 예산확보 마련이 필요함

목차

제1장 서론	1
제1절 연구배경 및 목적	3
1. 연구의 배경	3
2. 연구 목적	4
제2절 연구범위 및 방법	6
1. 연구 범위	6
2. 연구 방법	7
제2장 사회환경 및 조사기법의 변화	11
제1절 사회환경의 변화	13
1. 인구구조의 변화	13
2. 디지털화(스마트기기의 일상화)	15
3. 데이터 관련 법령의 개정 및 시행	16
4. 코로나19의 대유행	17
5. 적정 조사단가 및 인건비의 상승	19
제2절 조사기법의 변화	22
1. 자료수집방법	22
2. 조사방법의 변화	24
3. 혼합조사	27
4. 표본추출틀(표본틀)	29
제3절 소결	31
제3장 국내 통계 현황 및 사례검토	35
제1절 통계 현황	37
1. 조사통계 현황	37
2. 행정자료 현황	41

제2절 선행연구 및 사례검토	49
1. 조사과정자료(Paradata)	49
2. 혼합조사와 모드효과	52
3. 행정자료와 빅데이터의 활용	61
제4장 현황 분석	77
제1절 분석 개요	79
제2절 조사과정자료 분석	81
1. 국민여행조사	81
2. 국민여가활동조사 및 국민문화예술활동조사	83
제3절 조사방식에 따른 차이 분석	91
1. 조사 방법(mode)별 차이	91
2. 조사 도구(tool)별 차이	97
제4절 혼합조사의 모드효과 분석	102
1. 문화체육관광 소비자동향조사	102
2. 문화체육관광 기업경기동향조사	110
제5절 소결 및 시사점	117
제5장 혼합조사 추정 방안	121
제1절 혼합조사 추정방법	123
1. 최적결합방법	123
2. 성향점수를 이용한 추정방법	124
제2절 혼합조사 추정을 위한 시험조사	128
1. 시험조사 설계	128
2. 혼합조사 추정	129
제3절 소결 및 시사점	148
제6장 결론 및 제언	151
제1절 조사설계 개선 방안	153
1. 혼합조사 설계 방안	154
2. 문화·관광 분야의 조사 표본틀 구축 방안	157

제2절 조사방식 개선 방안	162
1. 행정자료를 활용한 조사 방안	163
2. 빅데이터와 연계한 조사 방안	165
제3절 제언	168

참고문헌 /	173
--------	-----

ABSTRACT /	177
------------	-----

표 목차

〈표 1-1〉 전문가 자문회의 대상	8
〈표 2-1〉 데이터 3법의 주요 개정 내용	17
〈표 2-2〉 데이터관련 법령의 제정 목적	17
〈표 2-3〉 2020년 문화·체육·관광 관련 국가승인통계 조사의 대응 방식 사례	18
〈표 2-4〉 우리나라 조사의 표본당 평균 단가 조사 결과	20
〈표 2-5〉 2019년 주요 국가의 법정 최저 임금 대비 조사인건비	20
〈표 2-6〉 문화·관광 분야 국민대상 승인통계 표본당 단가	21
〈표 2-7〉 조사방법에 따른 장단점 비교	23
〈표 2-8〉 인구주택총조사의 등록센서스 도입 효과	26
〈표 2-9〉 사회환경 변화에 따른 조사방법의 대응 방식	31
〈표 3-1〉 국가승인통계(조사통계) 조사방법별 비율	37
〈표 3-2〉 문화·관광 분야의 주요 승인통계 현황	39
〈표 3-3〉 통계작성기관 종류별 승인통계현황	43
〈표 3-4〉 통계데이터센터 제공 자료 요약	45
〈표 3-5〉 기업통계등록부 주요 항목	46
〈표 3-6〉 행정자료 이용을 위한 절차 요약	48
〈표 3-7〉 2011년 조사과정자료 관련 사업	50
〈표 3-8〉 조사과정자료를 활용한 분석 사례	51
〈표 3-9〉 조사설계 및 모드에 따른 응답률/무응답률 비교	61
〈표 3-10〉 2020 인구주택총조사 조사항목	69
〈표 3-11〉 문화체육관광산업통계 산출 과정	74
〈표 4-1〉 현황 분석의 틀	80
〈표 4-2〉 국민여행 온라인패널 시범조사 조사진행 현황	82
〈표 4-3〉 현장조사 관리자 업무량	83
〈표 4-4〉 조사원 업무량 배정 결과	84
〈표 4-5〉 조사구 대체 현황(국민여가활동조사)	85
〈표 4-6〉 조사구 대체 현황(국민문화예술활동조사)	86

〈표 4-7〉 조사구 대체 사유	87
〈표 4-8〉 가구 접촉 현황(국민여가활동조사)	88
〈표 4-9〉 가구 접촉 현황(국민문화예술활동조사)	88
〈표 4-10〉 최종 조사 완료 표본 수 현황(국민여가활동조사)	90
〈표 4-11〉 최종 조사 완료 표본 수 현황(국민문화예술활동조사)	90
〈표 4-12〉 조사범위	91
〈표 4-13〉 조사항목	92
〈표 4-14〉 2021년 1분기 기업경기 실적(BSI)의 조사방법별 비교	93
〈표 4-15〉 2021년 2분기 기업경기 전망(BSI)의 조사방법별 비교	94
〈표 4-16〉 조사항목	95
〈표 4-17〉 국민여행 온라인 패널 시범조사 조사결과	96
〈표 4-18〉 국민여행 대면조사 조사결과	97
〈표 4-19〉 조사항목	97
〈표 4-20〉 응답자 특성 비교	99
〈표 4-21〉 지난 1년간 참여한 여가활동 개수 비교	99
〈표 4-22〉 동호회 활동 참여 경험 여부 비교	99
〈표 4-23〉 자원봉사 활동 참여 경험 여부 비교	100
〈표 4-24〉 스마트기기 활용 시간_평일 비교	100
〈표 4-25〉 스마트기기 활용 시간_휴일 비교	100
〈표 4-26〉 일과 여가생활 간 균형 비교	101
〈표 4-27〉 여가생활에 대한 전반적인 만족도 비교	101
〈표 4-28〉 소비자동향조사의 선택한 조사방법에 따른 인구통계학적 변수 비교	102
〈표 4-29〉 소비자동향조사 선택형 조사방법에 따른 조사항목의 차이분석	104
〈표 4-30〉 소비자동향조사 항목의 모드효과분해	105
〈표 4-31〉 소비자동향조사의 할당한 조사방법별 인구통계학적 변수 비교	107
〈표 4-32〉 소비자동향조사 할당형 조사방법에 따른 조사항목 차이분석	108
〈표 4-33〉 소비자동향조사에서 인구사회학적 변수를 통제한 조사방법 효과분석	109
〈표 4-34〉 기업경기동향조사의 선택한 조사방법별 사업체특성별 변수 비교	111
〈표 4-35〉 기업경기동향조사 선택형 조사방법에 따른 조사항목의 차이분석	112
〈표 4-36〉 기업경기동향조사 항목의 모드효과분해	113
〈표 4-37〉 기업경기동향조사의 할당한 조사방법별 사업체특성별 변수 비교	114
〈표 4-38〉 기업경기동향조사 선택형 조사방법에 따른 조사항목의 차이분석	115
〈표 4-39〉 기업경기동향조사에서 사업체 특성 변수를 통제한 조사방법 효과분석	115

〈표 5-1〉 조사설계 비교	129
〈표 5-2〉 소비자동향조사 선택형 혼합조사의 최적결합 추정값 비교	130
〈표 5-3〉 소비자동향조사 할당형 혼합조사의 최적결합 추정값 비교	131
〈표 5-4〉 소비자동향조사 선택형 혼합조사에서 공변량과 범주	133
〈표 5-5〉 소비자동향조사 선택형 로지스틱 회귀모형 추정	133
〈표 5-6〉 소비자동향조사 선택형 혼합조사의 성향점수 보정 추정값 비교	134
〈표 5-7〉 소비자동향조사 할당형 혼합조사에서 공변량과 범주	135
〈표 5-8〉 소비자동향조사 할당형 로지스틱 회귀모형 추정	136
〈표 5-9〉 소비자동향조사 할당형 혼합조사의 성향점수 보정 추정값 비교	137
〈표 5-10〉 소비자동향조사 선택형 혼합조사의 최적결합 추정값 비교	138
〈표 5-11〉 소비자동향조사 할당형 혼합조사의 최적결합 추정값 비교	139
〈표 5-12〉 기업경기동향조사 선택형 혼합조사에서 공변량과 범주	140
〈표 5-13〉 기업경기동향조사 선택형 로지스틱 회귀모형 추정	140
〈표 5-14〉 기업경기동향조사 선택형 혼합조사의 성향점수 보정 추정값 비교	141
〈표 5-15〉 기업경기동향조사 할당형 혼합조사에서 공변량과 범주	141
〈표 5-16〉 기업경기동향조사 할당형 로지스틱 회귀모형 추정	142
〈표 5-17〉 기업경기동향조사 할당형 혼합조사의 성향점수 보정 추정값 비교	143
〈표 5-18〉 소비자동향조사 선택형 혼합조사의 추정방법 비교	144
〈표 5-19〉 소비자동향조사 할당형 혼합조사의 추정방법 비교	145
〈표 5-20〉 기업경기동향조사 선택형 혼합조사의 추정방법 비교	146
〈표 5-21〉 기업경기동향조사 할당형 혼합조사의 추정방법 비교	147

그림 목차

[그림 1-1] 연구의 수행방법 및 과정	9
[그림 2-1] 1인가구의 비율 추이	13
[그림 2-2] 주택유형 중 아파트의 비율 추이	14
[그림 2-3] 향후 통계청의 현장조사 체계도	19
[그림 3-1] KOSIS 제공 국내통계 분야	42
[그림 3-2] 분류별 데이터 보유 비율	43
[그림 3-3] 문화, 관광 개방데이터 관계도	44
[그림 3-4] 통계데이터센터 자료 구조	47
[그림 3-5] 총조사오차의 구성 및 조사과정자료와의 관계	49
[그림 3-6] 조사과정자료의 수집과 활용	51
[그림 3-7] 조사 소요 일자별 응답률의 변화 추이	57
[그림 3-8] 빅데이터 서베이를 이용한 관광분석	67
[그림 3-9] 총조사의 불응률 및 예산 추이	70
[그림 3-10] 자료연계 상세 구조도	74
[그림 4-1] 주간 단위 실사 진행률(국민여가활동조사)	89
[그림 4-2] 주간 단위 실사 진행률(국민문화예술활동조사)	89
[그림 6-1] 혼합조사 대상 선정 방법	156
[그림 6-2] 표본틀 활용과정	160

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

제1장

서론

제1절 연구배경 및 목적

1. 연구의 배경

최근 모바일 등의 첨단 IT 환경의 발전과 더불어 데이터가 급속히 증가하고 있으며, 빅데이터(Big data)의 등장과 더불어 관련 기술의 발전으로 생산되는 데이터를 정보로 만들어 활용하고 있다. 이러한 이유로 최근에는 전통적인 조사(survey)와 빅데이터 포함하는 행정자료를 이용하여 정보를 만들어 활용하는 방법이 공존하고 있다.

빅데이터가 능동적(active)인 형태의 데이터라면, 조사데이터(survey data)는 고정된(static) 형태의 데이터이다. 따라서 빠르게 변화하는 흐름을 실시간(real time)으로 정보를 파악하기 위해서 빅데이터를 이용하려는 노력을 하고 있다. 그러나 빅데이터는 대부분이 몇 개의 결과(fact)들이 지속적으로 변화하고 있기 때문에, 특정 시점의 하나의 주제와 관련된 다양한 값들을 측정하고자 할 경우에는 고정된(static) 시점의 값을 산출하는 전통적인 조사방법을 이용하고 있다. 이러한 이유로 여전히 조사통계를 활용하고 있다. 특히 조사통계는 사람들의 인식 등 사람들의 마음을 정보로 활용할 수 있다는 점에서 여전히 매력적이다. 우리나라는 빅데이터의 중요성과 활용이 증가하고 있지만, 여전히 많은 조사가 이뤄지고 있으며, 그에 따라 통계는 물론 정보의 많은 부분을 조사통계가 담당하고 있다.

그러나 최근에 사회의 발전과 그에 따라 사회환경이 많은 변화가 있어왔다. 1인 가구, 맞벌이 부부의 증가, 개인정보보호의 인식 강화 등 사회환경 변화에 따라 현장조사의 어려움은 점차 증가하고 있다(Baek 외 1인, 2016). 이러한 현상은 조사 응답 거절률이 증가하는 것으로 나타난다. 응답거절은 직접적으로는 단위무응답(unit nonresponse)으로 나타나지만, 일부는 항목무응답(item nonresponse)으로 나타난다. 단위무응답은 표본대체를 이용한 추가조사, 항목무응답은 무응답보정을 통해 완전한 데이터(complete data)를 만들어야 하는 등, 이로 인해 시간과 예산 등의 추가비용이 발생한

다. 즉, 사회변화는 조사비용의 증가와 함께 응답자와의 접촉 자체가 어렵게 되는 등 조사효율을 떨어뜨리는 원인으로 작용되고 있다. 따라서 조사에 영향을 미치는 사회환경 변화의 유형을 우선적으로 살펴보는 것이 필요하며, 최근의 변화를 바탕으로 한 주요사항을 정리하면 다음과 같다.

- 인구구조의 변화: 1인가구 및 아파트가구의 증가로 인한 대면조사의 어려움
- 디지털화(스마트기기 일상화): 스마트폰 보급률과 인터넷인프라를 통해 온라인조사(웹, 모바일)를 다방면에 활용할 수 있는 환경 마련
- 데이터 관련 법령의 개정 및 시행: 행정자료를 활용한 데이터 활성화 증대
- 코로나19의 대유행: 비대면 조사로의 전환

이렇듯 사회가 변화하면서 다양한 이유로 기존의 조사방법을 통한 조사환경은 점차 나빠지고 있다. 따라서 다양한 조사방법을 이용하여 조사하는 방안에 대한 연구와 다른 자료와 결합(또는 매칭)하여 활용하는 방안 연구 등이 있어왔다. 그러나 여전히 조사의 대표성과 신뢰성 등을 고려할 때, 다양한 조사방법을 활용하는 것은 어려움이 있다. 또한 다른 조사 또는 행정자료의 결합은 제도적으로 초기 단계이기 때문에 많은 시행착오를 통한 개선이 필요할 것이다. 한편, 사회가 빠르게 변화하고 발전하면서, 더 빠르고 많은 정보의 요구로 인해 빅데이터 등을 활용하고 있지만, 여전히 조사로 인한 정보습득은 요구되고 있다. 따라서 사회환경의 변화에 따른 조사방법의 개선도 필요하며, 이에 대한 연구는 중요하다고 할 수 있다.

2. 연구 목적

본 연구의 목적은 사회환경 변화에 따라 조사를 효율적이면서도 대표성 있는 통계값을 도출할 수 있는 방안을 마련하여 제시하는데 있다. 이를 위해 조사방법에 대한 특성을 살펴보고, 사회환경 변화에 따른 조사의 변화과정을 살펴본 후, 최근의 사회환경과 변화에 맞춰 조사방안을 어떻게 할지를 검토할 필요가 있다.

이를 위해서 본 연구에서는 크게 세 가지 방안에 대해 고려하고자 한다.

첫 번째로는 응답자의 상황에 따라 조사를 선택하거나, 상황에 적합한 조사방법으로 조사를 하게 만드는 혼합조사방법(Mixed Mod Survey)의 활용이다. 이는 기존의 전통적 조사방식(면접조사, 전화조사, 우편조사 등)과 혼합조사 방법의 효율성과 정확성 등을 비교함으로써, 조사의 효율성을 높이면서도 대표성과 정확성을 마련할 수 있는 방안을 제시할 수 있을 것이다.

두 번째는 최근의 데이터3법 개정 및 데이터기반행정법, 공공데이터법 등 데이터활용의 확대 등에 기반하여 다른 행정자료 혹은 타 통계와 연계하는 방안이다. 행정자료를 사용하면 정확성(precision)이 높아질 것이라 생각하지만, 실제 행정자료는 조사의 상황에 맞는 데이터가 있는 경우는 거의 없다. 따라서 이를 활용하기 위한 조건과 고려사항 등을 검토하고자 한다.

세 번째는 문화체육관광 관련된 조사들을 마련하여 활용하는 방안으로, 표본조사론의 이중추출법(two-phase sampling)을 응용한 방법이라고 할 수 있다. 이는 사전에 모집단의 정보를 이용하여 조사대상을 마련하고, 조사대상에서 표본을 추출하여 조사하는 방법인데, 사전에 정보를 파악하고 있기 때문에 사회환경의 변화와 거의 무관하게 조사를 수행할 수 있다는 점에서 효율적인 방법이라 할 수 있다.

즉, 제시한 3가지 방안 중 혼합조사 방안은 사회환경의 변화에 따라, 특정한 방법의 조사를 할 수 없을 경우에 다양한 조사방법을 통해 조사를 원활하게 수행하는 방안이며, 행정자료 이용은 조사가 어려운 경우 이를 대체할 수 있는 자료나 정보가 있으면, 이를 활용함으로써 조사진행이 어려울 때 이를 전부 또는 일부를 대체함으로써 효율성을 높일 수 있는 방법을 살펴보고자 하였다. 세 번째 방법인 표본들은 대표성 있는 통계를 산출하기 위해서는 포괄성(coverage)을 충족시켜야 하는데, 현재는 통계청의 모집단을 이용해야만하기 때문에 사회변화에 대처하기 어렵다. 따라서 조사에 활용할 수 있는 표본들을 마련하고, 표본으로 추출된 대상자들의 정보를 활용하여 조사를 진행할 수 있어야 한다. 따라서 좀 더 유연하게 활용할 수 있는 표본들을 구축하는 방안을 고려할 필요가 있다.

제2절 연구범위 및 방법

1. 연구 범위

가. 시간적 범위

본 연구에서 시간적 범위는 두 가지 경우를 고려하였는데, 사례연구의 경우는 최근의 현황을 살펴보기 위해 2010년 이후로 설정하였다. 다만, 조사환경의 변화에 따른 문화·관광통계의 변화를 분석하기 위해서는 충분한 사례들을 확보하여야 하며, 이는 간접 조사실험의 방법을 적용하는 방법이기도 하다.

실제 연구에서의 실험의 시간적 범위는 2020년과 2021년으로 정하여 최근 문화와 관광 분야에서 적용한 방법들을 비교하여 분석하고자 한다.

나. 대상적 범위

대상적 범위는 문화·관광 분야의 조사통계를 기반으로 한다. 승인통계에 국한하지 않고¹⁾ 지속적으로 생산되는 통계 중에서 과거에 조사방법의 변화가 있었던 통계와 조사설계를 통해 실험하는 방법으로 구분하여 살펴보고자 한다.

다. 내용적 범위

연구의 내용적 범위는 크게 네 가지로 나눠 살펴 볼 수 있다.

첫 번째는 사례분석을 통해 전통적인 조사방법인 대면조사와 전화조사, 온라인조사 등 단일조사의 비교 연구들을 파악하고 이를 정리하였으며, 간접 조사실험 방법을 통해

1) 국가승인통계는 통계청에서 관리하는 공식적인 통계이기 때문에, 조사의 변화를 연구자가 마음대로 조정하기 어려움

다양한 사례를 확보하여 혼합조사 시 모드별 차이를 분석하여 정리하였다.

두 번째는 실제 모드별 효과분석을 위한 시험조사를 실시하였다. 이는 예산의 한계가 있기 때문에, 현재 이미 수행하고 있는 국민여행조사, 문화체육관광 동향조사를 활용해 추가조사를 실시하여 모드효과를 파악하고자 하였다.

세 번째는 현재 통계청에서 제공하는 행정자료 활용 가능 데이터를 파악하여 최근의 데이터 관련법을 고려하여 적용할 수 있는 범위를 검토하였다. 특히 행정자료의 경우 해당 통계의 포함(Coverage)정도 등에 따라 활용 가능성이 달라질 수 있기 때문에, 행정자료의 이용 절차는 물론 이를 잘 활용하기 위해 필요한 사항 등을 정리하여 제시하였다. 행정자료의 활용은 직접 조사를 하지 않고 정보를 얻을 수 있는 예산 절감 효과뿐만 아니라 유사·중복 문항들을 제거할 수 있는 조사방법 개선의 효율적 방법 중 하나이기 때문이다.

네 번째는 새로운 조사대상이 되는 집단을 마련하거나 선정하여 조사함으로써, 국민 전체 혹은 사업체 전체로 추정하는 방법이다. 이를 위해서 문화·체육·관광 관련 조사를 위한 조사대상을 선정하는 방법을 제시하고, 개인정보를 가지고 있는 조사집단을 선정하여 조사함으로써 조사의 효율성과 대표성을 확보하는 방안을 제시하고자 한다.

2. 연구 방법

가. 문헌연구 및 사례분석

조사기법의 변화, 행정자료와 빅데이터의 활용, 혼합조사의 모드효과 등에 대한 국내외 학술논문 및 연구자료 등의 문헌검토 및 사례분석을 하였다. 또한 문화체육관광부 및 한국문화관광연구원 등에서 실시했던 통계의 산출물들을 활용하여 조사과정자료(Paradata), 조사도구별 차이 검토 등 2차 분석을 실시하였다.

나. 전문가 자문회의

조사방법의 개선 방안의 모색을 위해 학계와 현장 전문가들을 통해 최근 조사방법 경향 및 조사기법, 조사환경의 변화 등에 대한 자문회의를 실시하였다.

〈표 1-1〉 전문가 자문회의 대상

구분	대상자
학계	김규성 교수(서울시립대) 박인호 교수(부경대) 변종석 교수(한신대) 서우석 교수(서울시립대) 신기일 교수(한국외대) 임종호 교수(연세대) 한근식 교수(한신대) 한 준 교수(연세대)
민간영역 (조사업계 및 빅데이터 관련)	김준호 부부장(신한카드, Data Biz팀) 김유미 팀장(코리아데이터네트워크) 김창연 상무(케이스탯리서치) 김태영 상무(글로벌리서치) 원성훈 전무(코리아리서치) 이찬복 상무(입소스코리아) 차광섭 부장(한국리서치) 하동균 이사(케이스탯리서치) 한철수 이사(한국갤럽)
공공영역 (통계 담당)	김대진 팀장(한국은행, 경제통계국 기업통계팀) 황희진 팀장(한국은행, 경제통계국 통계조사팀) 심규호 사무관(통계개발원, 통계방법연구실) 임대철 차장(한국통계진흥원) 정용찬 박사(정보통신정책연구원) 주재선 박사(한국여성정책연구원) 천영민 박사(한국고용정보원)

학계는 조사과정자료(Paradata)분석, 혼합조사에서의 유의사항과 향후 조사의 변화에 대한 의견을 받아 연구의 방향설정과 모드별 조사실험결과 활용에 맞춰 자문을 받도록 하였다. 민간영역에서의 조사업계는 코로나로 인한 조사의 변화와 최근의 조사의 어려움, 향후 조사의 변화에 대한 의견을 들어 활용하였으며, 빅데이터 분야는 빅데이터와 조사를 연계하는 현황과 향후 계획에 대한 의견을 듣도록 설계하였다. 공공영역에서는 현재 조사의 흐름과 어려움, 코로나로 인해 달라진 점, 향후 조사의 방향에 대한 의견을 청취하였다.

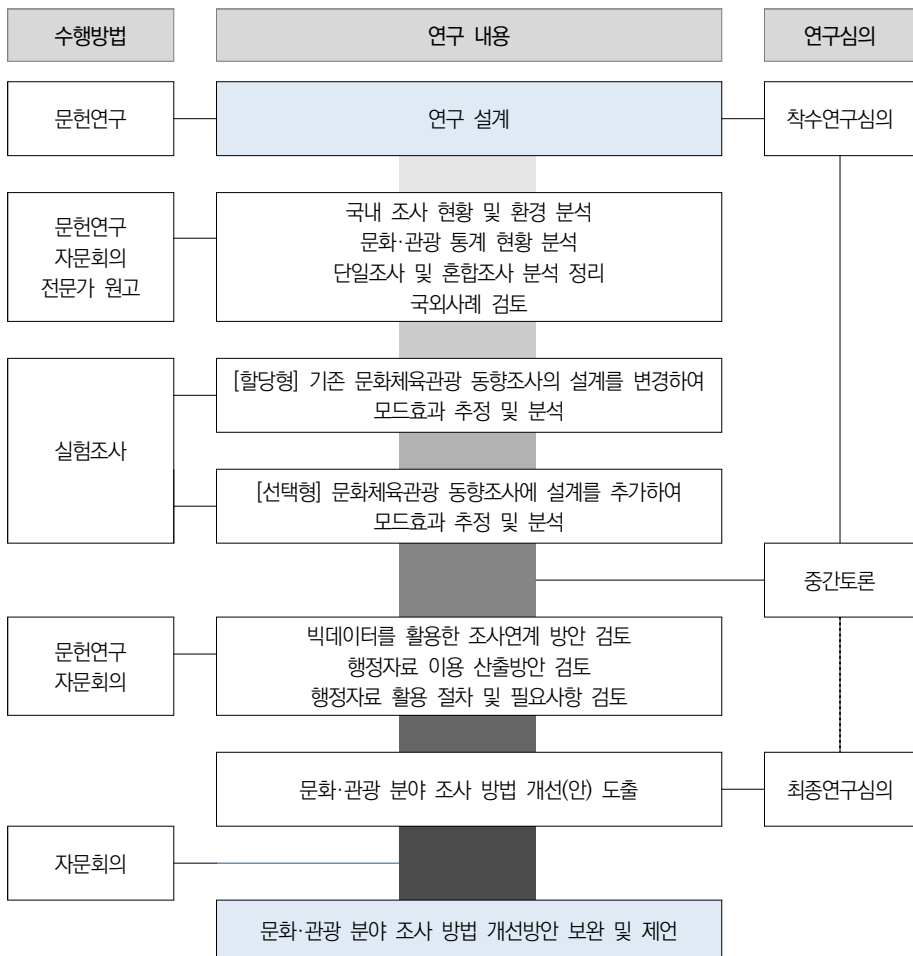
본 연구에서 조사 관련된 모든 분야 전문가들의 의견을 반영하지는 못하였지만, 공통된 의견과 최신 조사동향, 향후 조사의 흐름 등에 대한 의견은 최대한 반영할 수 있도록 하였다.

다. 시험조사 및 결과 분석

특히 혼합조사의 모드효과 파악 및 추정방안을 마련하기 위해 기존 한국문화관광연구원에서 실시하고 있는 문화체육관광동향조사를 활용해 이를 혼합조사 설계(할당형)로 재구성하여 실시하였다. 또한, 1,000표본(기업경기동향조사 500표본, 소비자동향조사 500표본)을 추가한 혼합조사(선택형)를 설계하고 동기간에 실시함으로써, 모드효과 분석을 실시하였다.

연구 전반의 수행방법과 과정은 [그림 1-1]과 같다.

[그림 1-1] 연구의 수행방법 및 과정



사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

제2장

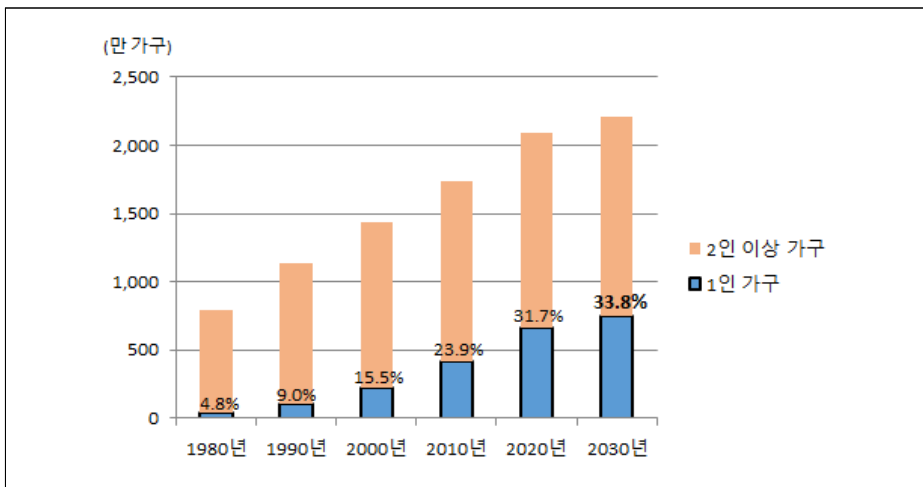
사회환경 및 조사기법의 변화

제1절 사회환경의 변화

1. 인구구조의 변화

우리나라의 인구구조 변화는 압축적으로 ‘초저출산’과 ‘인구 고령화’로 요약할 수 있으며(김현주 외, 2020), 이러한 저출산과 고령화로 인한 한국사회의 가족구성 및 가구형태의 변화는 오늘날 1인가구를 보편적인 가구형태로 인식되게 하고 있다(송정연 외, 2020). 실제 과거에서 현재에 이르기까지 일반적인 가족의 구성은 대가족(과거~농업시대)에서 핵가족(산업혁명~현대시대), 그리고 1인 가구(현재~미래시대)로 점차 그 형태가 단순화되고 있다. 이는 통계청의 인구총조사 결과를 보면 확인할 수 있는데 1인가구의 증가추세가 지속되고 있는 것은 물론, 2030년에는 전체가구의 약 1/3이 1인가구일 것으로 전망²⁾하고 있는 것이다.

[그림 2-1] 1인가구의 비율 추이

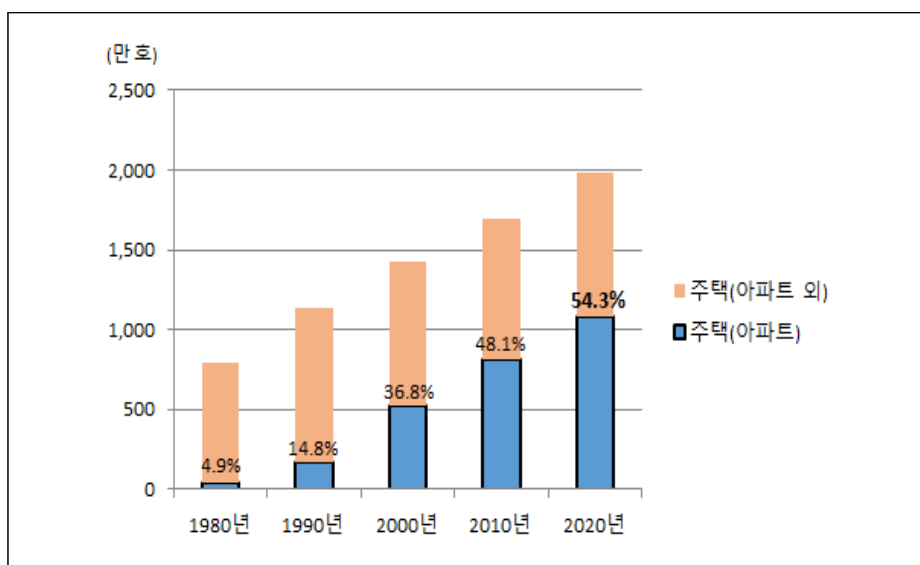


자료: 통계청 ‘인구주택총조사’, ‘장래가구추계’

2) 2030년은 통계청의 장래가구추계자료임

또한 맞벌이 부부도 2019년 기준 약 570만 가구로 전체 유배우자 가구의 46%에 해당할 정도로 증가하였다.³⁾ 주거형태 역시 큰 변화를 보이고 있는데, 단독주택은 과거 우리나라의 주 주거형태(1980년, 89.7%)였으나, 2000년 49.9%에 이어 2020년은 32%로 1/3가량을 차지하고 있다. 이에 반해 아파트는 과거 1980년 4.9%였으나, 2000년에 36.8%, 2020년은 절반이 넘는 54.3%를 차지하고 있다.

[그림 2-2] 주택유형 중 아파트의 비율 추이



자료: 통계청 '인구주택총조사'

이렇게 인구구조가 변화됨에 따라, 전통적인 설문조사방법인 대면조사의 어려움은 증가되고 있다. 1인가구와 맞벌이 부부 등의 증가로 인해, 가구를 방문해도 가구원을 만나지 못하는 경우가 많이 발생하고 있다. 이는 실제 조사대상을 만날 수 있는 기회가 줄어들었음을 의미한다. 게다가 최근 새로 생긴 아파트의 경우는 외부인의 출입을 허용하지 않는 곳이 생기고 있다. 즉, 설문조사를 할 수 있는 여건자체가 차단되는 것이다. 이러한 현상들은, 사회가 개인위주로 변화하고, 이에 따라 개인의 사생활 보호에 대한 중요성이 증가되는 방향으로 변화되고 있으며, 따라서 전통적 조사방식인 대면조사를 진행하기가 점점 어려워지고 있다.

3) 통계청 「지역별 고용조사」

2. 디지털화(스마트기기의 일상화)

디지털기술의 발달과 더불어 디지털 전환(Digital Transformation)의 시대를 맞이한 현시대에도 불구하고, 여전히 사람이 직접 조사대상을 찾아가 조사하는 대면면접 조사가 주요한 조사 방법으로 진행되고 있다. 하지만 우리나라의 경우 스마트폰 보급률은 약 95%로 전세계 1위, 인터넷인프라 역시 OECD 국가 중 1위로 나타나(OECD Digital Economy Outlook 2020), 온라인 조사(웹, 모바일 등)를 다방면에 활용할 수 있는 환경이 이미 마련되어 있다고 볼 수 있다.

오스트리아의 경우는 많은 국민대상 조사에서, 우편 등을 통해 조사대상자로 선정되었다는 것을 알려주고, 온라인으로 조사를 시도하는 방법들이 증가하고 있는데, 우리나라의 「인구주택총조사」의 경우도, 과거에는 대면조사로만 진행했지만 2005년 인터넷 조사방식을 최초 도입한 이후 점차 확대 적용하였고, 2020년부터 모바일조사와 전화조사 방식도 도입하여 실시하고 있다.

그러나 온라인조사의 경우, 온라인조사를 할 대상들의 이메일 또는 모바일 전화번호 등을 알아야 이메일 또는 문자로 조사 URL(uniform resource locator)을 제공하여 조사를 진행할 수 있다. 이러한 경우 개인정보의 문제가 있기 때문에, 법으로 정한 특별한 경우를 제외하고는 개개인이 조사에 대한 동의를 하지 않는다면 이메일이나 전화번호를 제공할 수 없다. 따라서 조사할 수 있는 표본들이 없는 경우에는 사용하기 어렵다.

박선희(2017)에 따르면, 모바일조사의 경우 역시 민간 영역에서는 활발히 활용되고 있지만, 국가통계에서는 이를 적용하는데 신중한 태도를 취하고 있다. 하지만 모바일조사를 적극 활용하고 있는 통계청 ‘사교육비조사’를 살펴보면, 모바일기기 응답비율(2014년 10.3%, 2015년 12.7%, 2016년 22.3%)이 점차 증가하고 있으며, 조사를 위한 표본들이 있는 경우, 접근성이 편리한 온라인조사를 통한 조사의 비율이 높게 나타나고 있다.

이렇듯 공식통계에서도 모바일기기를 기존의 조사방법과 같이 활용하는 비율이 점차 높아지고 있다. 민간영역에서는 많은 조사에서 온라인(웹, 모바일)을 이용한 조사방법으로 변화하는 추세가 증가하고 있는 것이다.⁴⁾ 따라서 개인정보와 모집단의 대표성 등의

4) 민간에서는 마케팅 차원에서 활용하고 있기 때문에, 공공의 사회조사처럼 대표성을 확보할 필요가 없기 때문에 더 많이 활용되고 있음

문제를 해결한다면 향후 핵심적인 조사방식이 될 것이다. 즉, 개인정보를 이용할 수 있는 대표성 있는 표본틀이 있다면, 이를 이용한 시의성 있는 다양한 조사방식을 활용할 수 있어 조사방식 역시 크게 개선될 것으로 예상할 수 있다.

3. 데이터 관련 법령의 개정 및 시행

우리나라에서는 다양한 분야에서 많은 조사들이 실시되고 있기 때문에, 국민 또는 사업체들이 설문조사에 자주 노출되고 있다. 특히 일정 규모 이상의 사업체의 경우, 경영부서의 담당자(인사 혹은 회계 등)들은 사업 또는 기업체 대상의 조사에서 많은 조사 요청을 받고 있어 응답 피로도가 높다. 많은 조사로 인해 업무에 지장을 초래하는 경우가 많아, 조사 응답에 대한 요구가 증가할수록 조사를 무성의하게 임할 가능성 또한 높아지게 된다. 이러한 이유로 이미 조사된 자료가 있거나, 행정자료(사업체조사의 매출액과 같은 정보)를 이용할 수 있다면, 이를 활용하는 것을 통계청은 권고해 왔다. 하지만 행정자료 이용 역시 개인정보문제로 인해 활용에 제약⁵⁾이 있어왔다. 그리고 원데이터(raw data)를 가지고 있는 기관에 데이터 제공에 대한 요청을 할 경우, 해당 기관의 데이터 관리 정책 등의 이유로 제공하지 않는 등 실제 행정자료를 이용하는 것도 한계가 있다.

그러나 최근 데이터의 중요성이 점차 증대하고 활용하고자 하는 수요가 확대되면서 데이터 활용을 위한 법률 등이 개정 또는 제정되었다. 데이터 3법⁶⁾ 개정('20.1월), 공공데이터의 제공 및 이용 활성화에 관한 법률(약칭: 공공데이터법) 개정('20.6월), 데이터기반행정 활성화에 관한 법률(약칭: 데이터기반행정법) 제정('20.6월) 등이 데이터 활용을 위해 개정 또는 제정된 법률들이다. 이러한 법률은 통계 작성 등을 위해 데이터를 요구했을 경우 제공받을 수 있도록 하고 있어, 데이터를 생산하지 않는 기관에서도 활용할 수 있는 방안이 마련된 예이다.

〈표 2-1〉는 데이터 3법에서 데이터 활용을 위한 주요 개정 내용을 정리한 것이다. 그리고 〈표 2-2〉는 데이터 관련 법령의 제정 목적을 제시한 것이다. 이러한 법들의 재정

5) 행정자료를 활용하기 위해서는 key변수를 이용하여 결합하여야 하는데, 이러한 key변수는 개인정보에 해당하는 경우가 대부분임

6) 「개인정보 보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」

과 개정으로 인해, 데이터 활용을 적극적으로 할 수 있는 기반이 마련되었으며, 공공과 민간 모두 데이터 활용을 통해 가치 있는 정보를 얻어 활용할 수 있도록 노력하고 있다.

〈표 2-1〉 데이터 3법의 주요 개정 내용

법령(약칭)	주요 개정 내용
개인정보보호법	<ul style="list-style-type: none"> - 개인정보 관련 개념을 개인정보, 가명정보, 익명정보로 구분한 후 가명정보를 통계 작성 연구, 공익적 기록보존 목적으로 처리할 수 있도록 허용 - 가명정보 이용 시 안전장치 및 통제 수단을 마련
정보통신망법	<ul style="list-style-type: none"> - 정보통신망법에 규정된 개인정보보호 관련 사항을 개인정보보호법으로 이관 - 온라인상 개인정보보호 관련 규제 및 감독 주체를 방송통신위원회에서 개인정보보호 위원회로 변경
신용정보법	<ul style="list-style-type: none"> - 가명조치한 개인신용정보로서 가명정보 개념을 도입해 빅데이터 분석 및 이용의 법적 근거를 명확히 마련 - 가명정보는 통계작성, 연구, 공익적 기록보존 등을 위해 신용정보 주체의 동의 없이도 이용, 제공할 수 있음

자료: 법제처 국가법령정보센터(<https://www.law.go.kr>)

〈표 2-2〉 데이터관련 법령의 제정 목적

법령(약칭)	주요 개정 내용
공공데이터법	<ul style="list-style-type: none"> - 국민의 공공데이터에 대한 이용권을 보장하고, 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로 함
데이터기반행정법	<ul style="list-style-type: none"> - 데이터를 기반으로 한 행정의 활성화에 필요한 사항을 정함으로써 객관적이고 과학적인 행정을 통하여 공공기관의 책임성, 대응성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키는 것을 목적으로 함

자료: 법제처 국가법령정보센터(<https://www.law.go.kr>)

4. 코로나19의 대유행

전염병의 확산은 사람들의 접촉을 꺼리게 만드는 요인이다. 과거 메르스(MERS; Middle East Respiratory Syndrome), 사스(SARS; Severe Acute Respiratory Syndrome)때에도 일시적으로 대면조사는 어려움을 겪었었다. 승인통계의 경우 많은 부분이 대면조사 방식으로 조사가 진행되고 있다. 그러나 2020년부터 현재에 이르기까지 코로나19의 확산으로 사회적 거리두기가 시행되고 사람들 간의 접촉을 꺼리는 상황이 발생했다. 이러한 상황으로 대면조사를 진행하기 어렵게 되었는데, 조사는 여전히 진행되어야 하기 때문에, 많은 조사담당자들은 조사 여건이 만족되는 대응방식으로 조사를

수행하며 이를 타개하고자 하였다. 그러나 비대면 조사방식으로 조사할 경우 모집단을 대표할 조사대상자를 찾을 수 있는 방안이 마련되어 있지 않은 경우가 대부분이어서 실제 대면조사가 아닌 비대면 조사방법으로 조정된 경우는 많지 않다.⁷⁾

즉 코로나19는 장기적으로 지속되며 사회경제 다방면에 영향을 끼쳤는데, 조사 환경에서도 큰 영향을 준 것이다. 이러한 이유로 통계청은 물론 각 부처와 지자체 등에서도 조사방식의 다변화를 모색하면서, 기존 조사의 연속성을 살리려는 노력이 있어왔다. 코로나19의 확산은 국가 승인통계 등에서 조사방식의 다변화 혹은 전통적인 조사방법에서의 전환점이 필요함을 각인시킨 것이다. <표 2-3>은 코로나19에 대응한 문화·체육·관광 분야의 국가승인통계 조사의 대응 방식 사례를 정리한 것이다.

<표 2-3> 2020년 문화·체육·관광 관련 국가승인통계 조사의 대응 방식 사례

대응 방식	조사명	내용
온라인조사 병행	-근로자휴가조사 -관광사업체조사 -외래관광객조사 -문화체육관광 일자리 현황조사	- 이메일, QR코드 등을 활용한 온라인(모바일) 조사 병행 - FAX, 전화 병행
유치배포조사 병행	-국민여가활동조사 -국민문화예술활동조사	- 대면 최소화를 위해 응답자의 선택에 따라 설문지를 유치·배포 후 회수하는 방식 병행
상황에 따라 일시중단 후 조사 재개	-국민여행조사	- 지역별 코로나 확산 정도에 따라 조사를 일시 중단 후, 확산 추이가 감소하면 조사 재개 - 온라인패널조사와 전화조사 방법을 시도하였으나, 기존 조사와 차이가 커 활용이 어려움

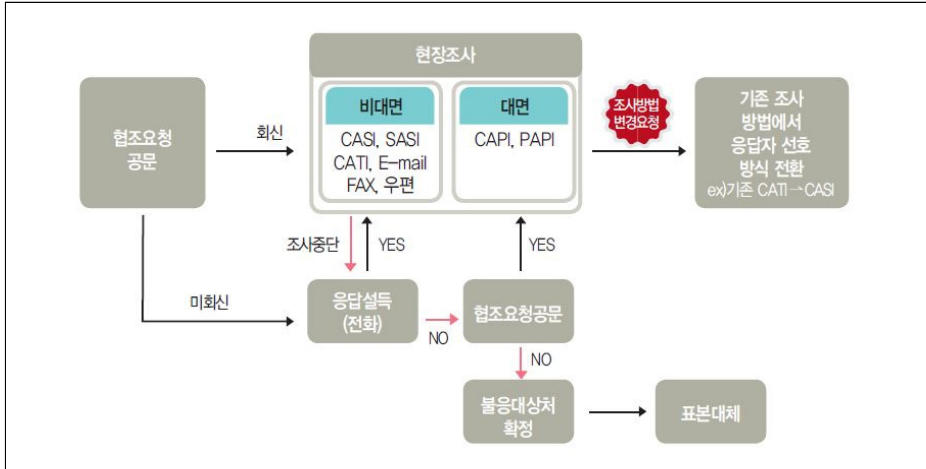
통계청 역시 비대면조사의 확대를 추진하고 있다. 2021년까지 비대면조사의 효율적 운영을 위한 비대면조사 통합관리시스템 구축 예정이고, 비대면조사 전담 기구인 원격조사센터를 신설하여 22년부터 전국을 대상으로 한 비대면조사 실시 추진 계획에 있다(박원용, 2020)

[그림 2-3]은 향후 통계청의 현장조사 체계를 제시한 것이다. 현재 개인정보 등의 문제 등으로 많은 조사에서, 특히 가구방문조사의 경우 가구방문을 통해 대면을 하지 않고 조사를 요청하거나, 조사에 대한 홍보를 할 수 있는 방안이 없었다. 그러나 체계도

7) 많은 대면조사는 대부분 가구방문조사인데, 가구를 방문하지 않고 가구원을 컨택할 수 있는 정보가 없기 때문에, 가구방문을 하지 않고 조사를 할 수 있는 방법이 현재는 없는 상황임

를 살펴보면 응답설득을 요청하는 방안이 들어 있는 것으로 볼 때, 조사를 위해 개인정보를 활용할 수 있는 방안이 마련될 것으로 보인다. 조사를 위해 전화번호나 이메일 주소를 활용할 수 있는 방안이 마련된다면, 온라인조사와 같은 비대면조사방식이 더욱 활성화될 것이다.

[그림 2-3] 향후 통계청의 현장조사 체계도



자료: 박원용(2020)

5. 적정 조사단가 및 인건비의 상승

사회가 변화하면서 다양한 분야에서 많은 변화가 있어왔다. 그 중에서도 물가상승에 따라 인건비의 상승은 비례적으로 증가하여왔다. 따라서 인건비가 많이 드는 분야의 비용은 많은 예산이 따르게 된다. 그러나 우리나라 조사예산은 오랫동안 변화가 없어왔다. 즉, 물가와 인건비는 증가하고 있지만, 조사 1부당 단가는 거의 증가가 없어 조사를 진행하는데 어려움을 겪고 있다.

〈표 2-4〉는 한국조사협회에서 회원사를 대상으로 한 설문조사결과인데, 적정 표본단가에 비해 현재 표본단가가 24~32%가량 낮은 것을 알 수 있다. 〈표 2-5〉는 한국조사협회에서 조사원 인건비를 비교한 표인데, 이를 살펴보면 우리나라 조사원 인건비는 55.76\$로 8시간 업무로 산정할 때, 최저 인건비에도 못 미치는 것으로 나타나 조사원 활용에 어려움이 있는 것으로 나타났다.

〈표 2-4〉 우리나라 조사의 표본당 평균 단가 조사 결과

조사방법		현재 표본당 평균 단가(원)	적정 표본당 평균 단가(원)	차이	
				원	%
전화 조사	CATI, RDD, 유선	15,792	19,750	3,958	25.1
	CATI, RDD, 무선	16,625	20,667	4,042	24.3
면접 조사	조사1(개인)	33,125	39,500	6,375	19.2
	조사2(가구)	59,111	77,556	18,445	31.2
온라인조사		14,500	18,083	3,583	24.7
좌담회(FGD)		3,216,667	3,733,333	516,667	16.1
기업체조사		42,909	60,909	18,000	41.9

자료: 한국조사협회, 회원사 설문조사 결과(2019년)

〈표 2-5〉 2019년 주요 국가의 법정 최저 임금 대비 조사인건비

단위: US달러

국가	조사원 인건비/일	법정 최저임금/일	최저임금 대비 조사원 인건비	1인당 GDP
한국	55.76	55.84	99.9%	31,838
호주	163.20	131.28	124.3%	54,907
미국	143.12	58.00	246.8%	65,281
독일	148.48	80.64	184.1%	46,259
영국	167.44	80.80	207.2%	42,300
일본	224.08	73.76	303.8%	40,247

자료: 한국조사협회, 1인당 GDP는 통계청 KOSIS(한국은행, The World Bank, 대만통계청) 기준임.(2019년)

문화·관광 분야의 설문조사 1부당 단가는 〈표 2-6〉에서 제시한 것처럼, 〈표 2-4〉의 표본당 단가보다 적은 것으로 나타나, 이를 개선할 필요가 있는 것으로 나타났다. 고품질의 통계자료를 생산하기 위해서는 많은 연구원과 조사원이 투입되어야 하지만, 현재의 단가로는 1명의 연구원이 많은 조사를 관리하고, 조사원도 많은 조사에 참여하여야 생활이 가능할 수밖에 없다. 따라서 좋은 조사품질의 조사통계가 생산되는데 하나의 어려움으로 작동되고 있다.

〈표 2-6〉 문화·관광 분야 국민대상 승인통계 표본당 단가

조사명	표본당 단가(원)	조사대상	조사방법
국민여가활동조사 (2020)	25,400	국민	가구방문 통한 개인조사
국민문화예술활동조사 (2020)		국민	가구방문 통한 개인조사
국민여행조사 (2021)	34,161	국민	가구방문 통한 개인조사
외래관광객조사 (2021)	31,938	한국방문한 외래관광객	출국자 대상 외래관광객조사

현재 사회환경의 변화로 인해 조사는 점점 어려워지고 있어, 조사품질을 높이기 위해서는 조사예산도 같이 증가되어야 할 것이다. 그러나 조사예산은 증가하지 않음으로 인해, 응답사례비, 조사관리 단계를 줄이는 등 조사품질에 직접적으로 관련이 있는 비용을 줄이고 있는 실정이다. 좋은 정보를 얻기 위해서는 적절한 예산을 투입하는 것은 당연할 것이다. 따라서 향후 조사예산의 개선은 절대적으로 필요하다고 할 수 있다.

제2절 조사기법의 변화

1. 자료수집방법

자료(data)란 해결하려는 문제와 직접적으로 관계가 있는 정보로, 이러한 자료를 얻어 내는 자료 수집방법은 연구의 목적이나 방법 및 대상에 따라 학자마다 다양하게 분류하고 있다(김수택 외, 2010). 이러한 자료를 얻는 방법은, 조사(survey)를 통한 데이터 수집, 실험(experiment)을 통한 데이터 수집과 IT기술의 발달로 인해 어떠한 행위에 대해 자동으로 DB(database)에 데이터가 자동으로 쌓이는 경우 등으로 구분 할 수 있다.

본 연구에서는 Bailey(1981)의 분류에 따른 구분 중 ‘조사’에 의한 자료수집방법을 다루고자하며, 최근 데이터 활용성을 높이고자 하는 정책 시행 및 관련 법령 개정 등의 흐름에 따라 행정자료에 대한 내용도 함께 다루고자 한다. 다만 조사의 효율성 측면에서 행정자료를 이용하는 방안, 즉 행정자료 자체로만 통계를 생성하는 것이 아닌 조사통계에서 조사의 어려움 또는 효율성을 높이기 위한 수단으로 행정자료를 이용하는 방안도 고려하여 제시하고자 한다.

조사통계의 조사방법은 조사를 진행하는 수단, 즉 우편조사, 온라인조사(이메일조사, 웹조사 등), 면접조사, 전화조사 등으로 구분할 수 있다. 또한 하나의 조사에 단일 조사방법을 사용하는지, 여러 가지 방법을 결합해서 사용하는지에 따라 단일조사와 혼합조사로 구분할 수 있다. 각각의 방법과 방법별 장단점을 <표 2-7>에 제시하였다.

〈표 2-7〉 조사방법에 따른 장단점 비교

구분	장점	단점
단일 조사	우편 조사 <ul style="list-style-type: none"> - 상대적으로 적은 비용 - 지리적으로 유연하여 넓은 지역을 대상으로 설문조사 실시 가능 - 면접원에 의해 발생할 수 있는 편의를 방지하고 광범위한 표본에 접근 가능 - 시간적으로 유연 - 다른 수단들보다 접근하기 어려운 대상들에게 접근 용이 - 직접적인 대면조사가 아니라 익명성을 보장하여 타당성을 높이는 대답 유도 가능 	<ul style="list-style-type: none"> - 낮은 응답률 - 응답을 하더라도 발생하는 응답오류와 특정 문항에 대한 무응답 오류
	전자 우편 조사 <ul style="list-style-type: none"> - 빠른 시간에 메시지 발송(신속성) - 저렴한 비용(반송용 봉투, 우편료 등 별도로 들지 않음) - 시간 제약이 없으며 24시간 편리하게 발송 및 수신이 가능 - 다른 조사방법보다 개방형 질문에 대한 응답이 가능하여 응답품질 더 좋음 - 무응답자에 대한 신속한 후속 작업 - 설문조사의 응답 편리 - 발송의 유연성 - 중개자 없이 수신자에게 직접 발송 가능하여 어려운 계층과 직접적으로 연결이 가능 - 설문응답의 발송 지연 없음 - 조사원의 바이어스 개입 안 됨 - 프로그램을 활용하여 필요한 표본만 조사가 가능 - 이중응답 제거 가능 	<ul style="list-style-type: none"> - 이메일 사용자로 대상이 한정되어 모집단의 편중과 대표성의 문제 - 컴퓨터 바이러스 감염의 두려움 등으로 낮은 응답률 - 응답자의 컴퓨터 숙련도에 좌우 - 응답자의 ID가 제시되어 익명성 결여 - 용량의 크기에 따라 발송이 되지 않거나 반송 및 로드될 때 속도에 문제발생 - 스팸메일 취급 - 복잡한 설문지 부적합 - 이메일 주소가 빨리 낙후 - 시스템 호환의 문제 - 응답자의 ID를 정확히 입력하지 않으면 발송이 불가능하여 정확한 입력을 요구함 - 초기 비용이 높음 - 메일 삭제 및 거절이 높음
	패널 조사 <ul style="list-style-type: none"> - 반복조사로 표본이 축적되어 분석대상의 규모가 증가하여 응답자에 대한 정확한 추정이 가능하여 생태학적 오류를 막을 수 있음 - 다양한 통계분석 모형의 적용이 가능 - 시계열자료나 횡단자료에 비해 인간행동의 복잡성을 포착하는데 유용 - 반복조사로 통계적으로 신뢰성이 높음 	<ul style="list-style-type: none"> - 시간이 오래 걸림 - 반복된 조사로 응답자들이 오염될 가능성이 있음 - 패널탈락 문제 발생: 반복된 조사로 기존에 설문에 응답한 응답자가 꾸준히 참여하기 어려움 - 다른 조사방법에 비해 패널유지로 많은 비용이 듦 - 경험과 지식을 갖춘 전문 연구인력이 필요 - 패널조사 자료를 분석하기 위해 자료축적이 필요
	면접 조사 <ul style="list-style-type: none"> - 조사과정에서 응답자의 반응을 실시간으로 확인 가능 - 자유응답형 등 다양한 형태의 설문가능 - 다른 면접 방법에 비해 얻을 수 있는 정보의 양이 많음 	<ul style="list-style-type: none"> - 시간과 비용이 많이 듦 - 비표본 오차를 통제하기 어려움 - 면접원에 대한 균일화된 면접교육이 필요함 - 면접자가 다른 응답자의 면접 내용과 비교할 수 있음

구분		장점	단점
	전화 조사	<ul style="list-style-type: none"> - 개별면접조사보다 비용이 저렴 - 신속하고 응답자에 접근하기 쉬움 	<ul style="list-style-type: none"> - 적절한 표본 확보가 어려움 - 젊은 층의 참여율이 낮음
	혼합조사	<ul style="list-style-type: none"> - 두 개 이상의 조사방법을 실시하여 각각의 장 단점이 보완 - 조사방법을 혼합하여 효율적이고 저렴한 비용으로 시행 가능 - 연령대별로 익숙한 조사방법을 사용한다면 응답률을 높일 수 있음 - 더 높은 표본을 확보할 수 있음 - 모드별로 효율적인 접촉가능 계층을 다양하게 구성할 수 있어 단일조사와 비교할 때 표본오차 축소가능 - 다양한 조사모드의 선택을 통해 무응답률과 무응답 편향 축소 가능성 증대 	<ul style="list-style-type: none"> - 혼합조사로 실제 오차가 줄어들었는지 알 수 없음 - 응답률이 높더라도 다양한 조사방법 사용으로 모드효과가 발생할 수 있음 - 단일조사에서 발생하지 않는 모드효과 발생으로 혼합형태의 편향 초래 가능성

자료: 채서일, 손영석(1993); 류진화, 김종훈(2003); 전중우(2008); 이현우(2009); 김수택 외(2010); 한신갑(2012)

2. 조사방법의 변화

사회의 발전과 변화는, 조사환경에도 많은 변화를 가져왔다. 「인구주택총조사」결과를 이용한 조사구를 표본틀⁸⁾로 활용하는 가구방문 조사, 전화번호부를 표본틀로 활용한 전화조사⁹⁾, 주민등록 현황의 주소를 이용한 우편조사가 공식통계에서 주로 사용하던 대표적인 전통적인 조사방법이다. 이 중에서 전화번호부를 이용한 조사는 개인정보 문제와 이동통신 기기의 보급률이 증가되면서 무작위전화걸기(RDD, random digit dialing) 방식으로 바뀌었다.

이 중에서 전화번호부를 이용한 조사는 개인정보 문제와 이동통신 기기의 보급률이 증가되면서 무작위전화걸기(RDD, random digit dialing) 방식으로 바뀌어 많이 활용하고 있다. 그러나 전화번호부와 같은 표본틀이 없기 때문에, 대표성 있는 공식적인 통계에서는 활용하지 못하고, 시의성이 필요한 여론조사와 같은 조사에서 주로 활용되고 있다.

8) 표본추출단 단위를 목록화한 것을 의미하며, 실제 표본추출을 하는 모든 대상을 리스트한 목록

9) 우리나라의 경우 휴대폰 보급과 개인정보유출 등의 문제로 2005년 5월부터 전화번호부 발매가 중단되었고, 이로 인해 전화번호부를 표본틀로 활용하지 않고 있음. 그러나 여전히 많은 나라에서는 전화번호부를 표본틀로 활용하고 있음

우리나라 주거형태가 일반주택에서 아파트로 변화하고¹⁰⁾, 1인가구의 증가, 맞벌이 가구 증가 등으로 인해, 가구 방문을 통한 응답대상자와의 접촉이 점차 어려워지고 있다. 이러한 이유로 최근 가구방문은 평일은 4시 이후에 조사가 진행되며, 많은 사람들이 출근을 하지 않는 주말에도 조사를 진행하고 있다. 이렇듯 응답대상자와 접촉할 기회를 높이기 위해 많은 노력을 하고 있다. 그러나 최근에 건설된 많은 아파트들은 보안을 이유로 외부인들의 출입을 차단하는 경우가 있어, 응답대상자와의 접촉은 더욱 어려워지고 있다.

최근 우편조사는 거의 사용되지 않는 조사방법이 되었다. 우편함의 많은 우편물들이 대부분 광고·홍보물인 관계로 이를 개봉하거나 열어보지 않고 버려지게 된다. 따라서 수취인이 우편물을 열어보게 하는 것이 중요한데, 대부분 본인의 이름이 적혀있을 경우에 우편물을 개봉한다. 그러나 이 또한 개인정보문제로 거주자의 이름을 제공하지 않고 있기 때문에, 우편조사방법으로 조사하는 것은 더욱 어려워지고 있다.

한편 2000년대 들어와서 IT기술이 급속도로 발전하면서 개인정보문제가 크게 대두되었다. 따라서 개인정보와 관련된 정보의 활용에 제약이 되면서 전통적인 조사방법은 가구방문을 통한 대면조사만 거의 유지되고 있다. 대신 IT기술의 발달과 함께, 행정자료(빅데이터 포함)의 활용과 이메일이나 스마트폰의 문자를 이용한 온라인조사를 이용하여 조사하고자 하는 경향이 증가하고 있다.

따라서 최근의 조사 경향은 하나의 조사방식으로 조사하는 단일조사(unit mode survey)방법보다는 두 개 이상의 조사방식으로 조사하는 혼합조사(mixed mode survey)방법으로 하고자 하는 경향이 많다. 혼합조사 방법은 어려운 조사 환경에서 조사 대상자들의 응답 부담과 조사비용의 증가 없이 고품질의 통계를 생산할 수 있는 방법들 중 하나로 주목 받고 있다(임경은 외, 2013).

통계청의 인구주택총조사의 경우에도, 인터넷조사를 적극 도입함으로써 인건비, 인쇄 및 입력 비용 등 204억원의 예산을 절감하였으며(임경은 외, 2013), 2015년부터 등록센서스 방식을 도입하며 약 1,455억원의 절감효과를 가져왔다. 이는 <표 2-8> 제시하였다.

10) 2020년 통계청의 인구주택총조사 자료에서 우리나라의 아파트 비율은 54.3%임

〈표 2-8〉 인구주택총조사의 등록센서스 도입 효과

	등록센서스 도입 전	등록센서스 도입 후	도입효과
응답부담	전 국민의 100% 참여 (약 5,000만명)	전 국민의 20% 참여 (약 1,000만명)	전 국민의 80% 응답부담 경감
소요예산	2,712억원	1,275억원	1,455억원 절감

자료: 2020 인구주택총조사(<https://www.census.go.kr>)

혼합조사는 이와 같은 비용 절감뿐만 아니라 응답률을 높이는데 매우 효과적인 것으로 알려져 있다(Daniel et al., 2012). 여러 가지 방법을 하나의 조사에 이용하면 조사 대상자들에 대한 접촉 가능성을 높이고 무응답을 줄임으로써 포함오차와 무응답오차 축소를 통한 총 조사오차(total survey error) 축소 효과를 얻을 수 있으며(de Leeuw, 2005; Dillman et al., 2009), 나아가 고품질의 조사 결과를 얻을 수 있는 것이다(Voogt and Saris, 2005).

혼합조사를 할 경우, 사전에 응답대상자(표본)에게 조사방식을 선택하게 하거나, 또는 가능한 방식으로 접촉하여 조사를 하게 하여야 한다. 그러나 국민대상 조사에서 가구방문 조사의 경우에는 통계청의 ‘인구총조사’ 자료를 표본틀로 활용하여, 조사구를 추출하기 때문에 가구를 방문하는 방법을 제외하고는 응답자와 접촉이 어렵다. 따라서 최근 코로나 19와 같은 문제로 대면조사가 어려운 경우에도 다른 조사방식을 도입하기가 쉽지 않다.

최근 국민대상의 가구방문조사에서 응답자와 접촉하지 않고 조사하려는 노력들이 다양하게 있어왔다. 예를 들면, 조사취지를 일리는 설명과 함께 설문지를 두거나 QR (Quick response)코드를 넣어 조사에 응답하도록 하는 방법이 최근에 시험적으로 사용되고 있다. 설문지를 유치하는 경우에는 우편함에 언제까지 넣어달라는 문구를 넣고 해당일에 가서 설문지를 회수하는 방법이며, QR코드 방식은 QR코드를 스캔하면 스마트폰으로 온라인조사가 연결되는 방식이다. 그러나 이러한 방법들도 방문을 통해 홍보하지 않는 경우 응답률이 매우 낮으며, 설문내용이 어려운 경우 접촉했다가 중도에 포기하는 경우가 많다. 결과적으로 응답률이 매우 낮으며, 표본의 대체율이 매우 높은 것으로 나타났다.

한편 조사방법 중 조사 도구(Tool)에 따른 변화도 있는데 대표적인 경우가 면접조사의 태블릿PC 활용이다. 과거 대면조사에서는 모두 구조화된 종이설문지를 활용하여 조사(Paper Assisted Personal Interview; PAPI)였으나, 기기의 발달로 설문지를 출력

하지 않고 전자파일을 그대로 Table PC에 저장하여 조사(Tablet Assisted Personal Interview; TAPI)하는 사례가 늘고 있다.

이러한 태블릿 기반의 조사방식은 비용이나 시간을 상당히 줄임으로써 효율성이 증가한다는 장점이 있는데(Leisher, 2014), 문화체육관광부의 국민여가활동조사와 국민문화예술활동조사에서는 이러한 장점들과 함께 비표본오차(데이터의 오입력 등)를 줄이기 위한 방안으로 전면 TAPI를 적용하여 진행(2019년 조사)한 바 있다.

3. 혼합조사

전통적인 조사방법은 하나의 조사를 하나의 조사방법으로 진행하는 것이다. 그러나 사회환경 변화에 따라 조사대상자, 즉 응답자들을 접촉하기가 점차 어려워지고 있다. 또한 개인정보 문제로 전화번호나 집주소 정보를 설문조사에 이용하는 것도 어려워졌다. 최근에 스마트폰 등의 이동통신 전화기를 국민들 대다수가 보유하면서 집전화가 점차 줄어들고 있는 것도, 전화번호를 이용하여 조사하기 어려워진 하나의 예로 볼 수 있다. 따라서 조사에 걸리는 시간과 비용은 증가할 수밖에 없지만, 조사예산은 증가하지 않아 조사의 어려움은 더욱 커져가고 있다. 이러한 다양한 이유로 최근 조사는 하나의 조사에 2개 이상의 조사방법을 적용하는 혼합조사(Mixed-mode survey) 방식을 이용하여 조사의 효율성을 높이는 노력을 하고 있다.

혼합조사는 악화일로를 견고 있는 조사환경에서 조사대상자들의 응답 부담과 조사 비용의 증가 없이 고품질의 통계를 생산할 수 있는 방법이다(임경은 외, 2013). 혼합조사는 저렴한 비용으로 각 조사방법의 약점을 서로 보완할 수 있다는 점에서 각광받고 있는데(de Leeuw, 2005), 혼합조사 방식은 응답자가 선호하는 방식 또는 조사대상자를 빠르게 접촉할 수 있는 방법으로 진행되기 때문에, 응답률을 높이면서도 응답시간과 응답 비용을 줄이는 장점이 있다. 이러한 이유로 조사업체에서 특히 선호하여 활용하고 있다. 또한 하나의 조사에 여러 가지 조사방법을 적용함으로써, 조사대상자들에 대한 접촉 가능성을 높이고 무응답률을 줄일 수 있다. 이는 포함오차와 무응답오차 축소를 통한 총 조사오차(total survey error) 축소 효과를 얻을 수 있는(de Leeuw, 2005; Dilman et al, 2009) 장점이 있다고 할 수 있다. 즉 무응답을 줄임으로써 무응답으로 인한 조사

결과의 편의(bias)를 줄일 수 있으며, 이를 보정하기 위한 무응답대체 방법을 연구로 인한 시간과 비용 등을 줄이는 효과가 있는 것이다.

그러나 혼합조사가 반드시 장점만 있는 것은 아니다. 혼합조사를 이용하여 응답률을 높이고 포함오차를 축소한다 하더라도, 서로 다른 조사방법에 따른 포함률(Coverage rate)의 차이, 다양한 조사방법 중 하나를 선택함으로써 발생하는 선택효과(selection effects) 및 조사방법에 따른 측정효과(measurement effects) 등과 같은 모드 효과(mode effects)는 조사 결과에 또 다른 불안정성으로 작용하게 된다(임경은 등, 2013). 또한 De Leeuw(1992)는 모드 간 차이를 일으키는 주요인은 면접원기입¹¹⁾과 자기기입 차이라고 하였다(박지선, 2016). 조사원의 간섭으로 인한 응답차이, 자기 응답은 비주열에 의존하는 반면 조사원 조사는 듣는 것에 의존하게 되는 것에 의한 응답차이, 조사 인센티브의 차이로 인한 무응답의 차이(박상언 등, 2012)가 그것이며, 자기 기입식은 일반적으로 더 높은 항목무응답을 가지게 된다(Dillman, 2000) 따라서 조사방법(조사모드)이 다른 것도 고려할 필요가 있지만, 기입 방식의 차이도 고려할 필요가 있다고 할 수 있다.

결국 혼합조사의 장점을 살리고 조사결과를 대표성 있게 활용하기 위해서는 조사방법이 서로 다름에 따라 나타나는 차이를 파악하여 보정할 필요가 있다. 혼합조사에서 조사방법 간에 생기는 차이를 모드효과(mode effects)라 하는데, 이를 잘 파악하여 평가하고 조정된 추정을 할 필요가 있다. 이렇듯, 모드효과로 조사방식에 따라 값에 차이가 발생하는 문제는 있지만, 응답자와의 접촉이 점차 어려워지는 상황에서, 혼합조사의 방법은 조사에서 매우 매력적이라 다양한 조사에서 많이 활용되고 있다.

혼합조사는 다양한 조사방식을 적용함으로써 응답자의 편의성을 고려하는 장점이 있지만, 혼합조사 역시 응답자를 접촉하지 못할 경우 활용하기 어려운 한계가 있다. 즉 표본으로 선정된 대상자에게 연락을 하여 조사대상자로 선정된 것을 알리고 편리한 방식으로 조사에 응하도록 하여야 하는데, 사전에 접촉할 수 있는 정보가 없다면 혼합조사 방식을 적용하기 어렵다.

따라서 혼합조사를 위해서는 조사대상자에게 연락할 수 있는 정보가 있어야 하며, 조사대상자는 모집단을 대표할 수 있어야 한다. 즉, 모집단을 반영할 수 있는 표본들이 있

11) 여기서 면접원기입 방식은 연구원이 질문하고 응답자가 응답한 결과를 면접원이 기입하는 방법과, 면접원이 설명하고 응답자가 설명을 듣고 응답하는 방식 모두를 의미함

어야하며, 표본으로 뽑힌 조사대상자들에게 연락을 할 수 있는 정보가 있어야 다양한 조사방식으로 진행할 수 있다.

4. 표본추출틀(표본틀)

사회조사의 통계는 대부분 표본을 뽑아 모집단을 추정하는 통계이다. 따라서 표본은 모집단을 대표할 수 있어야 한다.¹²⁾ 국가에서 생산하는 공식통계는 더욱 대표성과 정확성 등을 고려하게 된다. 따라서 표본을 잘 뽑을 수 있는 방안을 마련하여야 하며, 표본을 잘 뽑기 위해서는 표본설계(sampling design)도 중요하지만 표본을 추출하는 표본틀(sampling frame)이 중요하다. 표본틀은 조사의 목적에 맞는 모집단을 포괄(coverage)하는 조사대상자들로 구축되어야 한다. 그러나 개인정보 등의 문제로 이를 마음대로 구축하는 것은 어렵다. 즉, 조사를 하고 싶어도 조사대상의 목록(list)을 확보하는 것이 쉽지 않다.

통계청에서 국가통계로 승인하는 통계의 대부분은 표본틀이 명확하게 있는 통계이다. 국민대상 통계는 「인구주택총조사」의 조사구를 이용하고 있으며, 사업체 조사는 통계청의 「전국사업체조사」의 사업체 목록을 이용한다. 특별한 대상의 경우도 표본틀이 명확하여야 하는데, 문화체육관광부의 「예술인 실태조사」의 경우에는 예술인 협·단체의 회원 명부와 예술인경력증명시스템의 예술인활동증명 신청을 한 대상을 통해 예술인의 표본틀을 구축하여 활용하고 있다. 또한 「장애인 생활체육조사」의 경우는 ‘전국 등록 장애인’의 대상을 이용한다. 특정 산업의 사업체조사의 경우에도 마찬가지로, 「스포츠산업조사」의 경우에는 스포츠산업 특수분류를 이용하여, 통계청의 「전국사업체조사」에서 스포츠산업에 해당하는 사업체를 분류하여 스포츠산업 표본틀을 이용한다.

표본틀을 이용하여 조사하는 대부분의 조사통계에서는 국민대상조사보다는 사업체 대상 조사에서 혼합조사가 많이 활용된다. 규모가 있는 사업체의 경우, 사업체를 방문하여도 응답을 할 담당자를 만나는 것이 쉽지 않아, 전화를 걸어 조사에 대한 취지를 설명하고 응답하도록 독려한 후, 이메일이나 팩스를 이용하여 조사를 하는 경우가 증가하고 있다. 따라서 대면조사, 이메일, 전화, 팩스 중에서 응답자들이 선호하는 방법으로 조사

12) 표본설계를 통하여 표본조사를 통해 추정이 가능하도록 표본을 추출함

를 진행하는 혼합조사의 방법이 최근 사업체조사의 일반적인 방법이다.

조사통계에서 대표성을 충족시키기 위해서는 우선적으로 표본들이 있어야 하는데, 현재 대부분의 표본들은 통계청에서 관리하고 있다. 따라서 개인정보 문제 등으로 응답자들을 컨택할 수 있는 방법이 없어, 조사환경의 변화에 대처할 수 있는 방안이 마련되었어도 조사가 이뤄지기 어렵다. 조사환경 변화에 대응하기 위해서는 조사방식의 변화와 더불어, 조사대상자를 추출하는 표본들이 우선적으로 활용가능하도록 구축될 필요가 있다.

제3절 소결

앞서 조사환경에 영향을 주는 주요 사회환경의 변화들을 살펴보았다. 본 연구에서는 주요 변화로 4가지를 정리하였는데, 전통적 조사방법 측면에서 변경이 불가피한 변화로 ①인구구조의 변화와 ②코로나19의 확산을, 새로운 조사방법 혹은 전통적 조사방법에서 탈피하는 기반을 마련하게 된 변화로 ③디지털화(스마트기기의 일상화)와 ④데이터관련 법령의 개정 및 시행을 살펴보았다. 이를 정리하면 <표 2-9>와 같은데, 물론 이러한 변화에 따른 대응 방식을 정답으로 볼 수만은 없다. 조사방법은 각기 그 특성에 맞는 장점과 단점들이 혼재하고 있기 때문이다.

<표 2-9> 사회환경 변화에 따른 조사방법의 대응 방식

구분	주요 사회환경 변화			
	인구구조의 변화	코로나19 확산	디지털화	데이터 관련 법령의 개정 및 시행
변화 내용	1인가구의 증가, 맞벌이부부의 증가, 아파트의 증가	사회적 거리두기 시행, 대면접촉에 대한 거부감 증가	스마트폰 보급률 95%, 인터넷인프라 OECD 국가 중 1위	데이터3법 개정, 공공데이터법 개정, 데이터기반행정법 제정
조사 환경	부재가구 다수 발생, 아파트내 진입 불가	공식통계의 대부분인 대면면접조사 진행이 어려워짐	온라인조사 방식에 대한 시스템 구축은 이미 마련되어짐	행정자료 활용 및 조사의 연계 가능성 확대
발생되는 문제점	원표본 유지율 하락, 응답률 하락	조사기간 및 비용의 증가, 심각한 경우 조사 중단	-	-
대응 방안	조사방식의 변경 필요		새로운 조사방식으로의 변화 필요	
대응 방식	조사방식의 다변화 (혼합조사)	비대면조사 전환	온라인(웹, 모바일 등) 조사 확대	행정자료(빅데이터) 활용 및 연계
고려 사항	모드효과 파악 및 추정	대표성있는 표본들과 응답대상자 접촉을 위한 정보 확보	연계변수 및 활용 절차 고려, 빅데이터의 경우 대표성 확보 방안 고려	

특히 하나의 조사에 여러 조사방법을 활용하는 혼합조사는 조사의 효율성을 높이고 응답률을 높이기 위한 방안으로 각광을 받고 있는데, 이 방법에는 모드효과가 발생하여

또 다른 불안정성으로 작용된다. 따라서 혼합조사를 이용할 때는 모드효과를 파악하여야 하며 이를 고려한 추정 또한 필요하게 된다.

우리나라의 승인통계들은 대다수의 조사통계가 면접조사(약 79.0%)를 통해 진행되고 있다. 이들의 메타데이터를 살펴보면 약 24% 이상의 통계에서 여러 조사방법이 혼합되어 있다. 하지만 이들은 혼합조사로 설계된 것이 아닌, 단일조사 설계이나 응답자의 편의를 위해 다른 조사방법도 병행하고 있는 것이다. 마찬가지로 문화·관광 관련 승인통계 역시, 조사통계는 모두 면접조사이나 다른 조사방법들을 병행하여 진행하고 있다. 이러한 조사방식들은 응답률을 높이는데 어느정도 효과가 있는 것으로 알려져 있다. 하지만 여기에는 모드효과로 인한 편의가 발생하게 되는데 이는 4장의 사례분석Ⅲ 과 5장의 모드효과에서 자세히 다루고자 한다.

한편, 조사통계와 행정자료를 연계하여 활용하면, 응답자의 응답부담을 줄여 응답률 향상과 함께 통계의 유사·중복 문제도 함께 해결할 수 있다. 국가통계포털은 우리나라 모든 국가승인통계의 결과를 제공하여 포괄범위가 넓지만, 메타데이터와 통계결과표만을 제공하기 때문에 실제 데이터 활용 측면 보다는, 제공된 결과를 확인하는 차원에서 이용되어지고 있다. 그에 반해 공공데이터포털은 데이터를 제공하는 역할을 하고 있어 데이터의 활용, 분석, 연구 등 그 활용성이 높다고 할 수 있다. 통계데이터센터의 경우 한 차원 더 높은 서비스를 이용할 수 있다. 행정자료의 데이터 제공은 물론 이와 연계·분석이 가능하도록 시스템을 마련하고 공간을 제공하는 등의 서비스를 하고 있기 때문이다. 이는 3장에서 조금 더 자세히 다루고자 한다.

행정자료를 활용한다고 하더라도 인식이나 다양한 차원의 정보를 얻는 것은 쉽지 않다. 이러한 부분은 조사통계를 통해서 이뤄지게 되는데, 사회조사에서 조사통계는 대표성을 충족시켜야 한다. 대표성을 충족시키기 위해서는 모집단을 대표할 수 있는 표본들을 바탕으로 과학적인 표본설계를 통해 표본을 추출하여 조사를 진행하여야 한다.

사회환경이 변화함에 맞춰 다양한 조사방식으로 조사를 잘 진행할 수 있기 위해서는 표본들에서 뽑힌 표본(응답대상자)에게 다양한 방식으로 조사를 하여야 하는데, 이를 위해서는 접촉할 수 있는 정보가 있어야 한다. 즉, 조사대상자들의 정보를 갖고 있는 표본들이 있어야 사회환경의 변화에 맞도록 조사방식을 변화시킬 수 있을 것이다. 따라서 조사환경에 대응할 수 있는 표본들을 갖추는 방식을 마련하는 방안을 마련하는 것은 매우 중요하다.

사회환경이 증가하고 있지만, 조사예산이 현실적이지 않은 부분도 조사가 잘 진행되기 어려운 하나의 요소로 작용하고 있다. 조사원 수당이 다른 국가들보다 매우 낮으며, 응답자의 인센티브는 점점 줄어들고 있어 응답거절률이 증가하는 하나의 요인이 되고 있다. 따라서 조사예산을 물가상승률과 인건비 등을 고려하여 반영할 필요가 있다.

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

제3장

국내 통계 현황 및 사례검토

제1절 통계 현황

1. 조사통계 현황

가. 국가승인통계 현황

2021년 2월 기준 국가승인통계는 총 1,238종이며, 이 중 조사통계는 539종에 달하며, 이를 조사방법별로 살펴보면 대다수가 면접조사로 진행되고 있음을 알 수 있다(면접조사 79.0%, 인터넷조사 6.3%, 우편(FAX)조사 3.5%, 전화조사 1.9% 등). 이렇게 승인통계의 대부분이 면접조사인 이유는 다른 조사방법(전화, 인터넷조사)은 표본들이 부재한 경우가 많기 때문이다. 전화조사의 경우는 개인정보 문제 등으로 전화번호부가 없어지면서 전화조사의 표본들이 없어졌다고 할 수 있고, 인터넷조사의 경우는 인터넷을 이용하는 대상들에 대한 목록 자체가 없다. 이렇게 표본들이 없는 경우는 조사대상(조사모집단)이 목표모집단을 충분히 포괄하는지에 대한 포함률(coverage rate)을 평가할 수 없다. <표 3-1>은 국가승인통계 중에서 조사통계에 대한 조사방법별 비율을 제시한 것이다.

<표 3-1> 국가승인통계(조사통계) 조사방법별 비율

조사방법	통계 수(종)	비율(%)
면접조사	426	79.0
인터넷조사	34	6.3
우편(FAX)조사	19	3.5
전화조사	10	1.9
기타	50	9.3
합계	539	100.0

자료: 통계정책관리시스템, 국가승인통계현황 메타데이터 재가공

표본들이 없는 경우는 조사대상을 조사기관(또는 연구기관)에서 임의로 정하게 되기 때문에, 조사의 일관성을 유지하기 어려우며, 포함률(coverage rate)이 충족되는지를

판단하기 어려워 대표성 여부를 판단하기 어려운 문제가 있다.

또한 국가승인통계의 메타데이터를 분석하면 혼합조사가 약 24% 이상을 차지하는 것으로 나타났다.¹³⁾ 하지만 혼합조사를 체계적으로 활용한 것이 아닌, 단순히 응답률을 높이기 위해 적용한 것이다. 메타데이터를 검토해보면 조사방법은 특정 조사방법을 원칙으로 정한 단일조사로 되어있으나, 응답자의 선택에 따라 다양한 방법을 병행하는, 병행조사가 대다수이다. 특히 이러한 병행조사의 많은 경우가 사업체조사인데, 연락처를 알고 있지만 조사대상자를 만나기 쉽지 않아 응답자 편의를 위한 방법으로 조사하는 경우가 많기 때문이다.

국민을 대상으로 하는 통계는 가구방문 조사를 이용하는데, 조사대상자의 집 주소 이외의 정보를 파악할 수 없기 때문에, 대면조사 방식 이외의 조사방법을 활용하기 어려워 단일조사로 진행된다. 물론 집 주소 이외의 다른 정보를 알면 다른 조사방식도 가능하다. 특정 대상을 조사하는 경우에 조사대상자들의 대한 정보가 있는 표본틀이 있다면 혼합조사를 적용하고 있다. ‘예술인실태조사’를 예를 들어 설명하면, ‘예술인실태조사’의 표본틀이 있으며, 표본틀을 이용하여 표본설계와 표본을 추출한다. 이때 예술인들에 대한 연락처가 있기 때문에 연락처를 제공 받아¹⁴⁾, 대면조사를 하기 어려운 경우에 온라인조사를 병행한다. 따라서 대면조사와 온라인조사를 병행하여 조사를 수행하고 있다.

다음으로 비승인통계를 살펴보면, 비승인통계는 공식통계가 아닌 이유로, 전국사업체조사나 인구총조사의 결과를 이용할 수 없다. 따라서 대표성 있는 표본틀 없이 조사를 하는 경우가 대부분이다. 비승인통계의 경우 일회성 조사와 연속성 조사로 구분할 수 있다. 먼저 정책적 이슈 등에 의한(시의성이 필요한)일시적인 조사에서는 전화조사 또는 인터넷조사(조사업체의 패널 활용 조사 포함)를 이용하는 경우가 많다. 또한, 두 조사방법을 혼합해서 조사를 하는 경우도 많다. 지속적으로 연속성 있게 진행되는 조사의 경우는 대면조사 또는 전화조사를 이용하는 경우가 많다. 설문문항 수가 많지 않은 경우 전화조사¹⁵⁾로 조사하는 경우가 많으며, 설문문항이 많은 경우에는 대면조사를 이용한다. 연락처가 있는 경우 전화 컨택을 통해 온라인조사나 전화 조사 등 응답자가 선호하는

13) 메타데이터상의 조사방법 외에 ‘조사방법 비교’란에 기입된 내용을 확인하여 병행조사 여부를 산출한 비율로 24.1%인 것으로 나타남. 하지만, 해당 비교란이 입력되지 않은 경우가 63%이기 때문에 이를 입력한 담당자의 불성실 기입의 가능성 혹은 누락 여부에 따라 실제 병행조사 비율은 더 높을 수 있음

14) 표본으로 뽑혀 조사대상으로 선정된 대상의 연락처, 즉 조사에 한정해서만 연락처 정보를 제공받음

15) 이때의 전화조사는 대부분 RDD(random digit dialing) 방식으로 진행됨

방법을 병행하여 조사하고 있다. 그러나 연락처가 없는 경우는 방문조사¹⁶⁾로 진행되어지고 있다.

나. 문화·관광통계 현황

문화·관광통계 분야의 주요 승인통계는 15종을 살펴보면, 조사통계는 12종이며 3종은 보고통계이다. 작성주기를 보면 11종의 문화예술 분야에서 7종이 1년이며, 2년주기가 1종, 3년주기가 3종이다. 관광분야는 ‘관광사업체조사’만 1년 주기이며, ‘외래관광객조사’¹⁷⁾와 ‘주요관광지점 입장객통계’는 월주기이며, ‘국민여행조사’¹⁸⁾는 매달 조사하지만 분기별로 공표하고 있다.

본 연구에서 관심 있는 조사통계를 살펴보면, 국민대상통계가 4종이며, 사업체가 4종이고, 특정대상의 통계가 4종이 있다. 조사대상이 국민인 경우는 통계청의 ‘인구총조사’의 자료를 활용하며, 조사구를 추출하고 가구를 방문하여 가구원을 조사하고 있다. 사업체의 경우는 대부분 면접조사를 원칙으로 하고 있지만, 이메일, 팩스를 병행하여 조사하고 있다. 특정대상의 조사는 조사목록을 이용하여 전화컨택 후 조사대상자와 조사방법과 조사시기를 조정하여 조사를 진행하고 있다. 이에 대한 정리는 <표 3-2>에 제시하였다.

<표 3-2> 문화·관광 분야의 주요 승인통계 현황

분야	통계명	통계구분	작성대상	작성주기
문화·예술	국민여가활동조사	조사(면접)	국민	1년
	국민문화예술활동조사	조사(면접)	국민	1년
	예술인실태조사	조사(면접, 온라인)	예술인	3년
	근로자휴가조사	조사(면접, 이메일, 팩스)	사업체 종사자	1년
	공연예술조사	조사(면접, 이메일, 팩스)	공연기관/단체	1년
	전국도서관통계	보고	도서관	1년
	한국수어활용조사	조사(면접)	청각장애인	3년
	지역문화현황통계	보고	229개 자치단체	3년

16) 비승인통계이면서 국민을 대상으로 하는 조사는 통계청의 집계구를 이용하는 경우가 많은데, ‘인구총조사’를 이용한 조사구는 승인통계에서만 제공하고 있어 활용할 수 없기 때문임

17) 코로나19로 2020년과 2021년도는 일시적으로 1년 주기로 공표하는데, 익년 5월에 공표하도록 하였다

18) 국민여행조사는 이용자의 활용성을 확대하기 위해 2021년도부터 분기로 잠정치를 공표하며, 확정치는 익년 5월에 공표한다.

분야	통계명	통계구분	작성대상	작성주기
	국민독서실태조사	조사(면접)	국민	2년
	콘텐츠산업조사	조사(면접, 이메일, 팩스)	콘텐츠산업 사업체	1년
	광고산업조사	조사(면접, 이메일, 팩스)	광고산업 사업체	1년
관광	국민여행조사	조사(면접)	국민	매월
	관광산업조사	조사(면접)	관광진흥법 사업체	1년
	외래관광객조사	조사(면접)	방한외래객	매월
	주요관광지점 입장객통계	보고	서울 제외한 관광지	매월

문화·관광분야의 사업체조사인 ‘근로자휴가조사’의 경우는 ‘전국사업체조사’에서 공공분야를 제외한 전체사업체가 대상이 되기 때문에 모든 산업의 사업체가 조사대상이 되지만, 다른 조사의 경우 해당하는 사업체만을 선별하여 조사목록을 별도로 작성하여 조사하고 있다. 따라서 조사목록에 조사목적의 대상자를 잘 선정하여 구축하는 것이 매우 중요하며, 이에 따라 추정한 결과값에 차이가 발생할 수 있다.

특정대상 조사의 경우를 살펴보면, ‘예술인실태조사’의 경우 예술인 협/단체에게 회원 명부를 받고, ‘예술인경력증명시스템’에서 예술인 증명을 받은 대상자 목록을 합쳐서 조사대상목록을 만들고, 표본설계를 통해 조사대상을 선정하여 조사를 하고 있다. ‘한국수어활용조사’의 경우는 보건복지부의 장애인 등록 데이터베이스에 기재된 법정 장애 유형 중에서 청각장애인 1~3등급에 해당하는 20세 이상의 등록 장애인의 목록을 만든 후, 표본으로 선정된 대상의 목록을 만들어 조사를 수행한다. ‘공연예술조사’의 경우 공연행정기관과 공연단체의 목록을 만들어 조사를 진행하고 있다. ‘외래관광객조사’는 방한외래객을 대상으로 하기 때문에, 어떠한 국가의 사람이 언제 얼마나 입국하고 출국하는지에 대한 정보가 없다. 따라서 출국하는 대상들에게 한국을 방문한 목적, 경로 등에 대한 조사를 하고 있다.

문화·관광 분야의 조사를 살펴보면 국민을 대상으로 한 조사와 ‘외래관광객조사’를 제외하고는 모두 별도로 조사목록을 만든 후, 정보를 받아 조사를 진행하고 있다. 따라서 조사목록을 얼마나 조사모집단에 근접하게 만들었는지에 따라 조사결과를 추정한 값의 대표성과 정도(precision)가 결정된다. 따라서 조사목록을 관리하고 갱신(update)하는 부분이 매우 중요하다.

최근 코로나일상에서 대면조사가 어려워 국민대상과 방한외래객을 대상으로 하는 5종의 통계는 조사를 대체할 방법을 찾는 것이 어려운 반면, 목록이 있는 10종의 통계의

경우는 표본으로 선정된 대상에게 전화 또는 이메일을 보내서 조사에 대한 취지를 설명하고 온라인 또는 전화 등으로 유도할 수 있다. 따라서 조사목록이 있는 경우에는 조사 방법과 조사시기를 다양화 할 수 있다.

2. 행정자료 현황

조사통계는 통계를 작성함에 있어 많은 예산과 시간이 소요되고 있고, 갈수록 현장조사의 어려움이 커지고 있는 실정이다. 이러한 조사환경의 변화에 따라 통계청뿐만 아니라 통계작성기관들은 통계의 생산방식을 조사통계 중심에서 행정자료를 활용하는 방식으로 (통계작성 체계를) 전환하고 있다. 데이터를 다루는 많은 사람들은 행정자료를 국세청의 국세데이터, 의료보험 데이터 등의 한정적 범위에서 생각하는 경향이 있으나, 데이터 관점에서 행정자료는 대부분 전산시스템 등으로 행정처리를 통해 발생하는 모든 데이터를 의미한다. 통계청의 경우 행정자료를 공공행정자료와 민간행정자료로 구분하는데, 공공행정자료는 공공에서 생산하는 자료이고, 민간행정자료는 최근 빅데이터로 부르고 있는 데이터가 주류를 이루고 있다.

최근 행정자료는 조사자료의 단점인 시의성과 응답자들이 응답하기 꺼려하는 민감한 정보를 직접적으로 알 수 있다는 이유로 활용이 증가하고 있다. 특히 빅데이터는 기존에 파악하지 못했던 실제 (국민 또는 사업체)가 활동하면서 발생하는 데이터이기 때문에, 활용성은 점점 커지고 있다. 그러나 특정 분야로 구분하고자 할 경우 명확하게 구분하기 어려운 경우가 많아 특정 값을 구하는 것보다는 변화하는 정도 등을 파악하여 활용하고 있다.

본 연구에서의 행정자료는 조사를 통해 파악하기 어려운 데이터를 보완하는 관점에서 살펴보고자 한다. 다수의 기존 조사 설문지를 살펴보면 응답자들이 꺼려하는 설문문항으로 인해 응답의 피로도가 높아지고, 이에 따라 응답하는 도중에 표본이 탈락하는 경우가 적지 않게 일어나고 있다. 많은 경우 돈과 관련된 질문이 이러한 경우가 많은데, 사람의 경우 소득, 세금, 부동산 가격, 부채 등이 있을 것이고, 사업체의 경우는 매출액, 부채, 법인세 등이 있을 것이다.

이러한 응답하기를 꺼려하는 설문문항을 행정자료에서 활용할 수 있다면, 조사가 훨씬 수월하게 진행될 것이다. 이렇게 조사에서 행정자료를 결합하여 통계를 작성하게 되

면 응답하기에 민감한 일부 설문문항을 조사하지 않고도 더 정확한 통계값을 구할 수 있기 때문에, 응답자의 응답 피로도를 낮추고 응답률 향상과 유사통계 중복의 문제도 함께 해결할 수 있는 장점이 있다. 이에 행정자료 및 공공데이터들의 현황을 살펴 향후 활용할 통계를 살펴 볼 필요가 있다. 행정자료를 제공하는 많은 인터넷 사이트가 있지만, 여기서는 국가통계포털, 공공데이터포털, 통계데이터센터를 살펴보았다.

가. 국가통계포털

통계청의 국가통계포털(KOSIS; Korean Statistical Information Service)은 우리나라 승인통계를 한 곳에 모아 이용자가 원하는 통계를 한 번에 찾을 수 있도록 One-Stop 서비스를 제공하고 있으며, 경제·사회·환경 등 30개 분야에 걸쳐 주요 국내 통계를 제공하고 있다. (국가통계포털, <https://kosis.kr/>)

[그림 3-1] KOSIS 제공 국내통계 분야



자료: 국가통계포털(<https://kosis.kr/>)

현재 396개의 통계작성기관이 작성하는 1,239 종의 국가승인통계를 수록하고 있는데(2021. 5. 21. 기준), 통계 작성의 기관종류별 현황을 살펴보면 <표 3-3>과 같다. 하지만 해당 서비스는 통계개요, 표본설계 등과 같은 해당통계 전반의 통계설명자료와 함께 그 결과인 통계표 위주로 수록되어 있어, 실제 행정자료를 결합하여 작성하거나 분석을 위한 자료로 쓰는 것은 어려운 부분이 있다. 즉, 원하는 주제나 키워드의 통계 결과를 인용하거나 확인하는 용도로 사용하기에 적절히 구성되어 있어, 필요한 통계값의 차원(dimension)과 동일한 형태로 값이 주어지는 경우를 제외하고는 활용하기 어렵다.

〈표 3-3〉 통계작성기관 종류별 승인통계현황

(단위: 종)

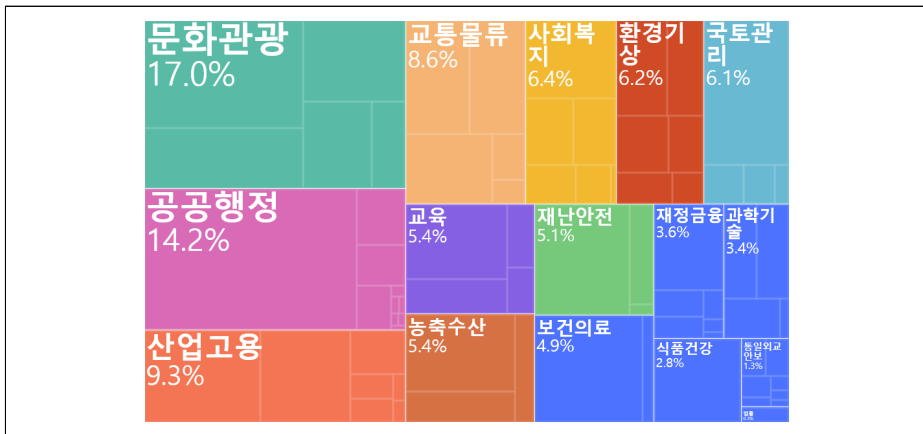
기관종류(기관 수)	작성 통계 수	작성방법별		
		조사	보고	가공
중앙행정기관(38)	423	218	152	53
지방자치단체(243)	615	210	264	141
금융기관(7)	21	8	5	8
공사/공단(33)	60	25	33	2
연구기관(29)	47	34	5	8
협회/조합(25)	33	26	5	2
기타기관(21)	40	19	14	7
합계(396)	1,239	540	478	221

자료: 나라통계, 승인통계 작성기관별 현황 재가공

나. 공공데이터포털

행정안전부와 한국지능정보사회진흥원은 공공데이터포털을 운영하여 공공기관이 생성 또는 취득하여 관리하고 있는 공공데이터¹⁹⁾를 제공하고 있다. 여기서 제공하는 데이터의 16개 분류별 국가데이터맵을 살펴보면 [그림 3-2]와 같고, 그 중 하위 분류인 ‘문화’와 ‘관광’ 파트는 [그림 3-3]과 같다.

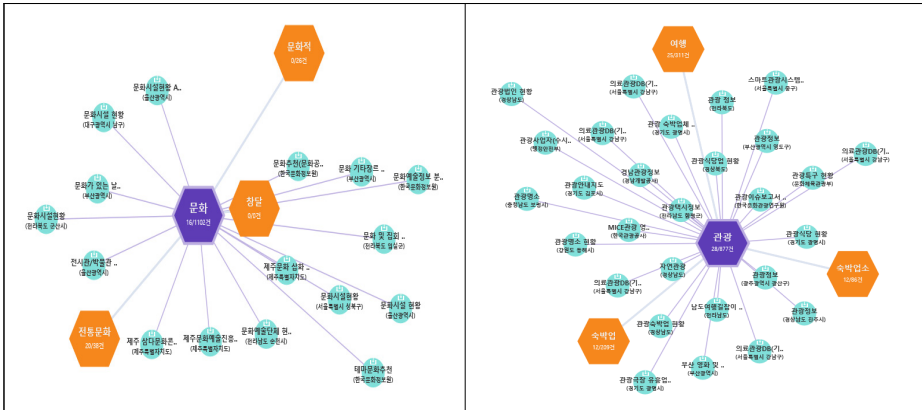
[그림 3-2] 분류별 데이터 보유 비율



자료: 공공데이터포털(www.data.go.kr)

19) 여기서 공공데이터란 데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보를 말함(공공데이터법 제2조 제2호)

[그림 3-3] 문화, 관광 개방데이터 관계도



자료: 공공데이터포털(www.data.go.kr)

현재 총 57,034건의 공공데이터를 제공하고 있는데²⁰⁾, 제공하고 있지 않는 공공데이터 역시 해당 포털에 제공 신청서를 기재하여 제출하면, 심의를 거쳐 제공여부를 결정 후에 이를 제공하고 있다. 공공데이터법 제21조에 의거 공공데이터의 효율적 제공을 위하여 통합제공시스템 즉 공공데이터포털을 구축·관리하고 활용을 촉진하고자 하는 만큼, 일방향의 결과 제공이 아닌, 활용성 높은 자료제공의 서비스를 하고 있다.

공공데이터포털의 정보는 데이터에 대해 이해하고 사용한다면²¹⁾, 통계를 대체할 수 좋은 수단이 될 수 있을 것이다. 공공데이터포털의 데이터는 다운로드를 통해 실제 분석자가 가공할 수 있는 형태이기 때문에, 그 차원(dimension)변수를 활용할 수 있다면, 해당하는 설문문항을 빼고 조사를 할 수 있을 것이다.

다. 통계데이터센터

통계청은 통계데이터센터(<https://data.kostat.go.kr/>)를 통해 통계조사 응답자의 비밀을 보호하면서 이용자가 심층적인 분석을 할 수 있도록 다양한 방법으로 통계자료를 제공하고 있다. 즉, 조사의 최종 결과값이나 마이크로데이터를 제공하는 것에 그치는 것

20) 2021. 5. 30. 기준이며, 세부적으로는 파일데이터(39,983건), 오픈API(7,336건), 표준데이터셋 122개 (9,715건)를 제공하고 있음

21) 특정 분야의 데이터만 있는 경우와 전체 분야의 데이터가 있는 등 데이터의 범위와 데이터를 분석할 수 있는 차원 변수를 살펴봐야 함

이 아니라 실제 데이터 간의 연계 분석을 할 수 있도록 구축된 DB에 접속하여 작업을 할 수가 있다. 이는 자체 보유한 자료를 반입하여 센터에서 제공하는 행정자료와 연계하는 것도 가능하며, DB내 구축된 행정자료 간의 연계 분석도 가능하게 된다. 다만 통계데이터센터에서 제공하는 데이터는 개인정보문제가 발생할 수 있기 때문에 보안을 철저히 하고, 데이터 반출 등에 있어 매우 까다로운 심의 등의 프로세스를 만들어 사용하고 있다. 이용자들이 개인정보문제 등이 발생하지 않는 범위 내에서 심층적인 분석을 한 후 이를 통계값으로 반출하여 이용하게 된다.

〈표 3-4〉는 통계청에서 구축한 자료 중에서 통계데이터센터에서 이용할 수 있는 자료의 목록인데²²⁾, 통계작성 연도에 차이가 있기 때문에 분석 이전에 이에 대한 확인이 필요하다.

〈표 3-4〉 통계데이터센터 제공 자료 요약

구분		자료명	
행정통계자료 (14종)	경제·사회	사업장기초DB(4대보험) 사업장기초DB(법인) 종사자-기업체_연계DB 육아휴직_사용현황DB	기업통계등록부DB(사업자등록번호기준) 기업통계등록부DB(대표자기준) 기업통계등록부DB_분기(대표자기준) 기업통계등록부DB_분기(사업자등록번호기준)
	인구·가구	등록센서스_인구DB 등록센서스_가구DB	등록센서스_주택DB
	농림어업	농업DB 임업DB	어업DB
민간자료 (34종)	경제·사회	1000대_상권_정보 성연령별_카드매출1 성연령별_카드매출2 성연령별_카드해외매출 신용정보(대출) 신용정보(소득) 신용정보(카드소비)	업종별_카드매출 영업_중_업소 온라인가격정보 유통고객_구매상품_정보 평균소득소비(전입인구) 평균소득소비(전출인구) 평균소득소비(통근인구)
	인구·가구	가구소득 성연령별_유동인구 성연령별_유입인구 성연령별_주거인구 성연령별_직장인구	시간단위_유동인구 요일별_유동인구 코로나_성연령별_유동인구 코로나_유입지NULL생략_유동인구

22) 국가 승인통계 또는 중요한 이슈가 있는 경우에만 승인을 받아 활용할 수 있음

구분		자료명	
	건물·지가	건축물정보 공동주택정보 표준공시지가	기초정보_인구집중유발시설 주택(단독/다세대/기타)
	차량	도로별_차량통행량	시군구별_차량등록대수
	기초정보	법정동코드_정보	행정동코드_정보
	기타	SNS_데이터	모바일상품권

자료: 통계데이터센터(<https://data.kostat.go.kr/>)

특히 기업통계등록부(Statistical Business Registers; SBR)는 조사자료와 행정자료를 상호 연계·융합하여 구축한 우리나라 모든 기업체(사업체)를 포괄하는 경제통계 부문 통합모집단으로 2017년 2월(2015년 기준) 시험 구축한 이래 매년 작성하여 서비스하고 있다. 즉, 사업을 영위하기 위해 국세청에 등록된 사업자와 통계청에서 현장 조사된 사업체를 통합하여 구축한 우리나라 전체 사업체, 기업체 목록으로 우리나라의 모든 통계 단위(기업체, 사업체)를 고유번호로 관리하여 경제통계의 통합모집단 및 표본추출용으로 관리하는 DB이기 때문에 행정통계자료 중 그 활용성이 매우 높다. 이에 기업통계등록부(SBR)의 항목으로 구성된 주요 변수들을 살펴보면 <표 3-5>²³⁾와 같다.

<표 3-5> 기업통계등록부 주요 항목

항목명	항목 설명
기준년도	자료 귀속 년도
사업자등록번호	사업자등록번호
법인등록번호	법인등록번호
사업체고유번호	사업체고유번호
조직구분	조직구분코드(1:개인,2:법인)
조직형태코드	조직형태코드 (1:개인사업체,2:회사법인,3:회사이외법인,4:비법인단체,5:국가지방자치단체)
...	
등록일자	사업자등록일자
개업일자	영업시작일
폐업일자	폐업일자

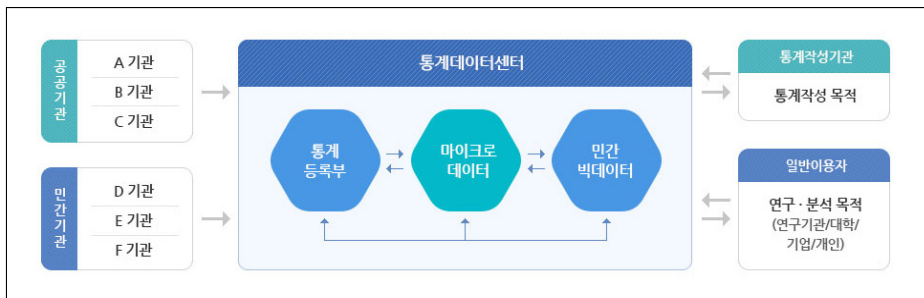
23) 기업통계등록부의 주요 항목 일부이며, 자료 활용을 위해 전체 항목 구성을 보고자 할 경우, 통계데이터센터(<https://data.kostat.go.kr/>)에서 제공하는 ‘행정통계자료 설명자료’를 참고할 것

항목명	항목 설명
폐업여부	폐업여부
...	
시도코드	시도코드
시군구코드	시군구코드
행정구역분류코드	행정구역분류코드
...	
BR사업체_매출액(배분)	기업체 단위의 매출액을 기업 내 BR사업체별로 배분한 매출액
BR사업체_총종사자(합계)	기업체 단위의 종사자를 기업 내 BR사업체별로 배분한 종사자의 합계
...	
기업내 사업체수	기업내 사업체수
BR기업체	BR기업체(1:대표)
BR기업체_고유번호	BR기업체 고유번호
BR기업체_활동여부	BR기업체의 활동여부(1:활동,2:비활동,3:폐업)
BR기업체_매출액	BR기업체 매출액
BR기업체_종사자(합계)	BR기업체의 상용과 임시및일용근로자 합계
...	

자료: 통계데이터센터(<https://data.kostat.go.kr/>)

[그림 3-4]는 통계데이터센터의 자료 구조를 도식화하여 나타낸 것이다.

[그림 3-4] 통계데이터센터 자료 구조



자료: 통계데이터센터(<https://data.kostat.go.kr/>)

행정자료 활용이 가능한 대표적인 기관인 통계청의 통계데이터센터 이용 방법은 크게 3가지로(분석센터 이용, 주문형분석서비스 이용, 명부서비스 이용) 나누어 볼 수 있는데, 이를 이용하기 위한 절차는 <표 3-6>²⁴⁾과 같다.

24) 명부서비스 이용의 경우, 조사를 위한 표본설계 및 추출을 위한 내용이므로, 행정자료 이용에 대한 내용에

〈표 3-6〉 행정자료 이용을 위한 절차 요약

분석센터	주문형분석서비스
①홈페이지를 통한 이용 신청 ②자료제공여부 검토 ③이용수수료 납부 및 좌석 배정 ④센터 방문 및 자료 분석 ⑤분석결과 반출신청 ⑥분석결과 반출여부 검토 ⑦분석결과 다운로드 ⑧연구결과 제출 및 기술노트 공유	①주문형 서비스 이용을 위한 전문가 상담 ②상담 후 홈페이지를 통한 이용신청 ③이용 승인 및 소요기간 산출 ④수수료 납부 ⑤작업자 배정 및 자료 생성 ⑥자료 이용

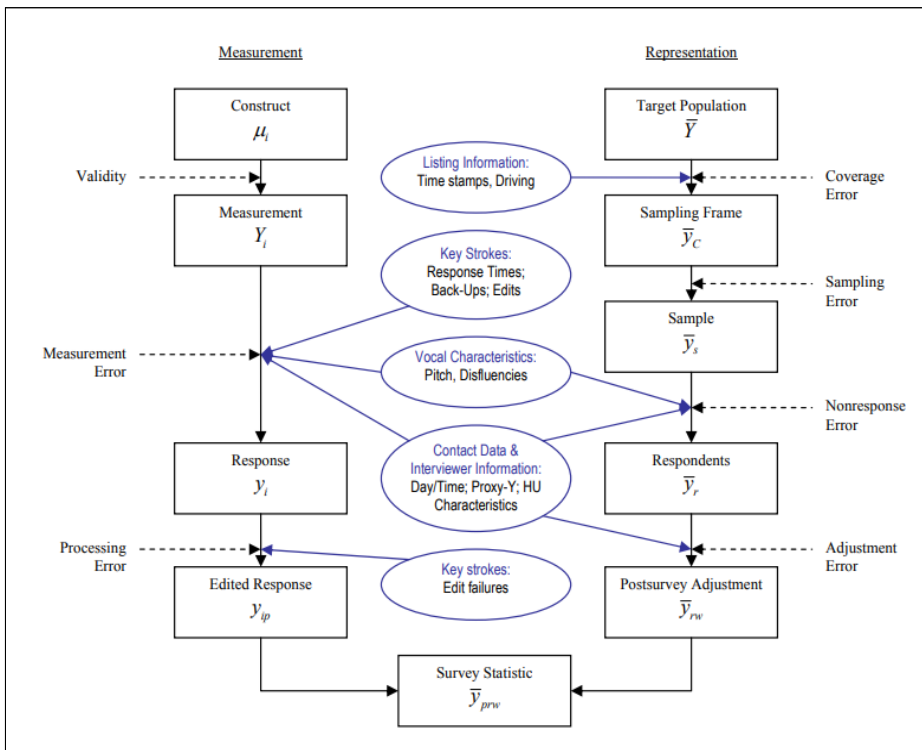
서는 다루지 않음

제2절 선행연구 및 사례검토

1. 조사과정자료(Paradata)

조사과정자료란 조사를 진행하는 과정에서 자연적으로 발생하는 파생자료로, 조사가 진행되는 과정을 객관적으로 파악하고 효율적으로 관리하기 위해 수집되어진다. 이러한 조사과정자료를 분석하여 조사표 혹은 조사방법의 개선을 유도하기도 하고, 조사 과정에서 발생하는 다양한 오차들을 축소시킬 수가 있다.

[그림 3-5] 총조사오차의 구성 및 조사과정자료와의 관계



자료: Kreuter et al.(2010)

우리나라의 경우 통계청은 2011년 5월 조사과정자료 수집 가이드라인을 제시하고, 성과지표에 조사과정자료 수집활동을 반영하기로 함에 따라, 각 지방청은 이와 관련 계획을 수립하였는데 이는 <표 3-7>과 같다.

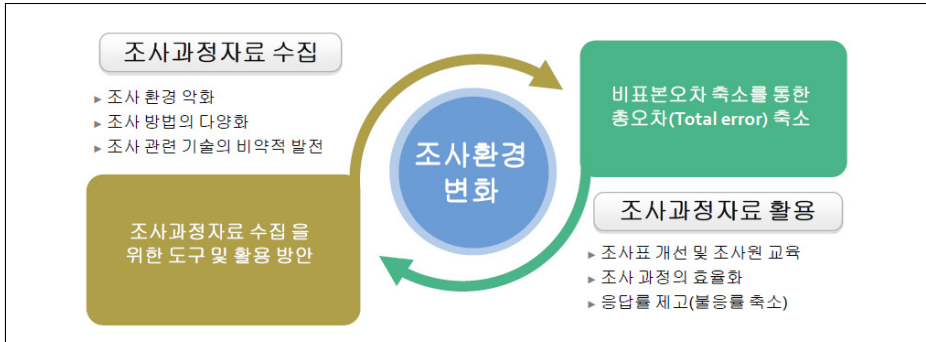
<표 3-7> 2011년 조사과정자료 관련 사업

지방청	사업명
경인청	<ul style="list-style-type: none"> - 조사환경 개선을 위한 건설경기동향조사 Paradata 구축 방안 - 조사환경 개선을 위한 경제활동인구조사 파라데이터 수집 방안 - 조사환경 개선을 위한 사회조사 파라데이터 수집 방안 ...
충청청	<ul style="list-style-type: none"> - 조사환경 개선을 위한 파라데이터 기본항목 수집 - 소비자물가조사 조사환경 개선을 위한 파라데이터 수집 계획 - 통계조사 대상처 응답성향조사 파라데이터 분석 ...
호남청	<ul style="list-style-type: none"> - 2011년 광업제조업동향조사 기초자료(paradata) 수집 및 분석 - 서비스업동향조사 응답자 친화적 통계조사방식 추진(제주사무소) - 내수면양식어업조사 조사환경 파라데이터 수집 ...
동남청	<ul style="list-style-type: none"> - 조사환경 악화대응 - 조사환경 실태 파악 및 개선을 위한 paradata 수집 - 조사환경 실태파악 paradata 자료수집 실시(농어가통계조사) ...
동북청	<ul style="list-style-type: none"> - 사업체부문 현장조사환경 계량화 및 DB 구축 - 2011년 경제활동인구조사 현장조사환경 실태파악 - 농어업통계조사 조사환경 분석을 통한 조사방법 선진화 방안 발굴 ...

자료: 임경은(2012a) 재인용

이러한 조사과정자료를 통해 최근에는 무응답 편의에 대한 연구가 많이 활용되고 있으나(홍민기 외, 2014), 조사과정자료를 활용할 목적으로 분석한 사례는 많지 않으며, 기초적 분석에 머물러있는 것이 현실이다(심규호 외, 2016). 과거에 오차를 줄이기 위한 노력은 이론적 측정이 가능한 표본오차에만 집중되어 왔고, 비표본오차를 줄이기 위한 노력은 체계화 되지 못했다. 즉, 조사과정자료의 중요성을 인지하고 수집하고 있으나 체계적으로 적용하지는 못하는 실정인 것이다(강유경, 2019). 따라서 비표본오차 축소를 통한 총오차 축소는 반드시 필요한 과정이며, 이러한 과정이 조사의 개선으로 이어져 조사의 품질을 향상시킨다고 볼 수 있다.

[그림 3-6] 조사과정자료의 수집과 활용



자료: 임경은(2012b)

이에 검토한 조사과정자료 활용 사례들을 바탕으로, 4장 현황분석에서는 수집가능한 조사과정자료를 이용하여, 현 작성통계들의 조사수행단계에서의 문제점을 파악하고 개선방안들을 도출하고자 한다. 이러한 조사과정의 개선 역시 큰 틀에서 조사방법의 개선이라고 볼 수 있기 때문이다.

〈표 3-8〉 조사과정자료를 활용한 분석 사례

구분	활용 자료	분석 목적	결과
1)	한국종합사회조사	응답거부와 접촉불가를 포함한 무응답 행위를 검토	도시거주자, 젊은 층, 그리고 지대가 높은 지역은 무응답, 접촉 불가, 조사의 어려움을 증가시키는 특성으로 나타남
2)	사업체패널조사	첫 컨택 반응이 조사에 미치는 영향 검토	첫 반응 시 강한 거절의사를 보인 경우 조사 성공률이 낮음 첫 컨택시 바쁨이나 부정적 반응인 경우 응답시간이 짧고 오류율이 높은 경향
3)	미국 국민건강면접조사	조사과정자료 분석을 통해 그 필요성을 검토	다양한 변수(조사응답률, 조사종단사유, 방문시도 횟수, 전화시도횟수, 면접시간 등)에 대해 유의미한 차이를 확인하고 개선 방안 도출
4)	2018 정부역할과 삶의 질에 대한 국민인식조사	설문조사 과정에서 면접원이 응답자의 응답시간과 응답행태에 미칠 수 있는 영향을 분석	응답자의 응답시간, 응답행태 등에 면접원의 특성(업무강도, 압박감 등)에 따른 영향이 적지 않음을 확인

자료: 1) 김지범 외(2014), 2) 홍민기 외(2014), 3) 진재현(2012), 4) 이슬기 외(2019)

2. 혼합조사와 모드효과

사회환경이 변화됨에 따라 조사환경도 같이 변화되고 있다. 전통적인 조사방식인 대면조사와 전화조사 방식에서 온라인조사 방식이 도입되었고, 개인정보의 보호가 중요하게 됨에 따라, 조사대상자를 접할 수 있는 기회가 줄어들었다. 이러한 이유로 조사대상이 응답하기 편리한 방법을 선택하게 함으로써 응답률을 높이는 방법인 혼합조사방식이 널리 이용되고 있다.

최근의 코로나19로 대면조사를 진행하는 것이 많이 어려워졌다. 따라서 가구방문조사를 제외한²⁵⁾, 많은 조사들이 온라인조사가 포함된 혼합조사로 대체되었다. 그러나 혼합조사의 경우 조사방법에 따른 모드효과(mode effects)가 있기 때문에 이를 파악하여 조사결과에 반영하여야 한다는 다양한 연구결과들이 있다.

따라서 혼합조사에서 나타나는 모드효과에 대해 이해하고 사례를 검토하는 한편, 이후 4장에서는 현황분석에서는 문화와 관련된 조사 2종(국민여가활동조사, 국민문화예술활동조사), 관광과 관련된 조사 1종(국민여행조사)의 사례를 바탕으로, 조사방법에 따른 차이를 살펴보고자 한다. 또한 비승인통계로 작성되고 있는 ‘문화체육관광 기업경기동향조사’와 ‘문화체육관광 소비자동향조사’를 진행하면서 나타난 조사방법에 따른 차이를 분석하고자 한다.²⁶⁾ 앞서 언급한 바와 같이 코로나19 발생 이후 통계청이 작성하는 통계를 포함한 많은 조사들이 면접조사의 어려움을 겪고 이를 타개하기 위해 온라인조사로 대체하거나 다양한 조사방식을 병행하여 조사하였다. 따라서, 다양한 조사방법이 혼합된 병행조사에서 각 조사방법에 따라 실제로 그 결과값에 차이가 발생하는가를 검토하는 과정이라고 볼 수 있다.

25) 가구방문조사의 경우, 시험적으로 온라인패널, 전화조사 등으로 일부 대체하려고 하였지만, 포괄성 등의 대표성 문제와 결과의 상이성이 나타났으며, 유치조사의 방법으로 온라인조사와 화상을 통한 조사 등을 시도하였지만, 조사기간이 많이 걸리고 대체율이 크게 늘어나는 문제가 발생한 것으로 나타남(조사업체의 전문가 자문회의 내용 중)

26) 승인통계의 경우는 임의적으로 조사방법을 변화하는 것은 불가능하고, 조사시기와 조사비용 등이 본 연구와 맞지 않았으나, ‘문화체육관광 기업경기동향조사’와 ‘문화체육관광 소비자동향조사’의 경우는 통계청의 승인을 받기 위해 일부 조사방법의 변화가 필요하였기 때문에, 일부에 한해 시범적으로 온라인조사를 병행하였음. 그러나 조사결과에 영향을 미치는 부분은 본 연구에서만 활용하도록 시범조사로 운영하였기 때문에, 기존 공표하는 자료의 시계열 문제는 발생하지 않도록 하였음

가. 모드효과의 이해

혼합조사는 두 개 이상의 조사방법을 이용하여 하나의 조사를 진행하는 것이다. 따라서 포함오차(coverage error) 및 무응답오차(non-response error) 축소가, 저비용의 조사방법으로 표본 규모를 확대를 통해 표본오차(sampling error)를 축소하는 등의 장점이 있다. 그러나 조사방법이 다름에 따라서 발생하는 문제도 있는데, 백지선 등(2016)은 혼합조사는 조사방법 간 응답결과 차이가 나타나는 모드효과(mode effects)로 인해 단일조사와 혼합조사 자료의 비교 가능성 저해 및 품질에 대한 우려가 있으므로 조사방법의 자료 품질에 미치는 영향 평가가 필요하다고 하였다.

모드효과는 선택효과(selection effects)와 측정효과(measurement effects)로 구분한다. 선택효과는 다양한 조사방법 중 하나를 선택함으로써 발생하는 차이를 의미하며, 측정효과는 동일한 대상자가 서로 다른 조사방법으로 동일한 조사에 참여할 때 발생하는 측정오차를 의미한다.

모드효과의 선택효과와 측정효과를 자세히 파악하기 위해 수식으로 표현하여 설명하면 다음과 같다.²⁷⁾ 변수 Y 의 범주 j 의 모비율 π_j 에 대한 선택효과는 다음과 같다.

$$S_a(\pi_j) = P(Y_a = j \mid M = a) - P(Y_a = j \mid M = b) \quad (3.1)$$

여기서 a 와 b 는 혼합조사에서 서로 다른 조사방법을 의미하며, S_a 는 a 방법을 선택한 효과를 의미하며, Y_a 는 a 방법으로 측정한 Y 의 값을 의미하고, M 은 조사방법을 의미하는 변수이다. 변수 Y 의 범주 j 의 모비율 π_j 에 대한 측정효과는 다음과 같다.

$$M_b(\pi_j) = P(Y_b = j \mid M = b) - P(Y_a = j \mid M = b) \quad (3.2)$$

식 (3.1)을 살펴보면 선택효과는 동일한 대상이 방법 a 를 선택하였을 때, 범주 j 를 선택할 확률과 방법 b 를 선택하였을 때, 범주 j 를 선택할 확률의 차이이다. 즉, 서로 다른 방법으로 응답하겠다고 하였을 때, 동일한 응답을 선택할 확률의 차이가 선택효과이다. 식 (3.2)를 살펴보면 측정효과는 동일한 대상이 방법 b 를 선택하였을 때, 서로 다른 방법으로 측정하였을 때, j 를 선택할 확률의 차이를 나타낸다.

식 (3.1)과 (3.2)에서 $P(Y_a = j \mid M = b)$ 는 실제로 관측이 불가능한 경우이기 때문

27) 백지선 등(2016)이 사용한 수식을 사용하여 각색하였음

에 추정을 하여야 된다. 추정이 가능할 때 선택효과와 측정효과를 산출할 수 있게 된다. 이를 위해서 아래의 전확률 공식을 이용한다.

$$P(Y_a = j) = P(Y_a = j \mid M = a)P(M = a) + P(Y_a = j \mid M = b)P(M = b) \quad (3.3)$$

여기서 $P(Y_a = j)$ 는 단일조사를 이용하 관측된 값이다. 그러면 식 (3.3)을 이용하여 $P(Y_a = j \mid M = b)$ 을 추정할 수 있다. 단, 이를 위해서는 2개의 가정이 필요하다. 하나는 단일조사와 혼합조사의 표본 및 응답자들이 동일한 모집단을 대표한다는 것이고, 다른 하나는 조사방법 a 의 측정효과가 두 자료 간 동일하다고 가정하는 것이다(백지선 등, 2016). 따라서 동일한 모집단을 대표한다는 가정을 만족하기 위해 추정할 때, 중요한 인구통계학적 변수(나이, 성별, 지역 등)에 대하여 사후에 모집단조정을 하여 맞춰주게 된다.²⁸⁾

일반적으로 모드효과는 집단의 구성에 따라 차이가 발생한다. 연령대가 높은 경우에는 대면조사를 선호한다면 연령대가 젊은 집단에서는 온라인 조사 등을 선호하는 경우가 많다. 어떠한 경우든 혼합조사에는 조사방법이 다름에 따라서 값의 차이가 발생하게 된다. 즉, 선택효과와 측정효과는 조사결과에 영향을 주게 된다. 따라서 선택효과와 측정효과를 추정하여 조정하지 않는다면 조사결과는 편향되어 나타나게 될 것이다. 선택효과와 측정효과를 파악하여, 혼합조사의 결과에 반영하는 것은 혼합조사의 결과를 추정하는데 매우 중요한 과정이라 할 수 있다.

나. 모드효과 평가에 대한 이해

조사여건은 시간이 지날수록 어려워질 것으로 예상되며, 사회조사가 학술연구뿐만 아니라 정책수립 및 집행과정에 기초자료를 제공한다(한신갑, 2012)는 관점에서 조사를 이용한 정보획득의 필요성이 지속된다면 혼합조사의 활용성은 더욱 확대될 것이다. 그러나 혼합조사에서 조사방법간의 모드효과의 차이가 발생됨에 따라 혼합조사에 불편추정량(unbiased estimator)을 얻기 위해서는 모드효과에 대한 정확한 평가가 이뤄져야 한다.

정확한 모드효과를 살펴보려면 선택효과와 측정효과를 파악하여야 한다. 이는 식

28) 일반적으로 레이킹 비(raking ratio) 조정을 통해 모집단을 동일하게 조정함

(3.1)과 (3.2)의 방법으로 산출하는 것이 원칙이지만, 일반적으로 동일한 조사대상에게 다른 방법으로 조사를 하는 경우는 없다. 따라서 두 조사방법의 조사대상자의 분포가 모집단을 대표하는지를 의미하는 대표성 가정을 충족할 때, 두 표본의 차이는 조사방법의 효과, 즉 측정효과라고 할 수 있다(백지선 등, 2017). 다음은 백지선 등(2017)이 제시한 대표성 가정 충족을 위한 방법들이다.

첫째, 사전에 구축된 응답자 리스트를 활용해, 각 조건에 배분하여 동질성을 확보하는 것이다(Dillman et al., 2009)

둘째, 스크리닝을 통해 최대한 동질성을 확보하는 것이다. 이는 조사방법이 여러 가지가 불가능한 집단은 혼합조사방법에서 배제해야 한다는 것이다.

셋째, 사전에 (가구)모집단 정보가 없다면 사전 리쿠르팅을 통해 표본을 구성한 후 배분하는 것이다(Kackle et al., 2010).

넷째, 보조변수 기준으로 반복비례기중²⁹⁾ 처리함으로써 두 조건을 동질적으로 만드는 것이다(Vannieuwenhuyze et al., 2010).

다섯째, 표본집단간의 동질성 확보가 어렵기 때문에 시간차를 두고 동일표본을 다시 조사하는 것이다.

모드효과의 평가는 결국 선택효과와 측정효과로 분리하는 것이라 할 수 있다. 선택효과는 응답자가 선호하는 방법으로 조사하는 것이기 때문에, 특정계층 또는 성향이 있는 대상들이 특정 조사방법을 선호하는 경향이 있게 된다. 본인이 선호하는 방법으로 조사를 하기 때문에 응답물에는 긍정적인 영향을 미치지만, 모집단 내의 특정 성향 집단들간의 선호하는 조사방법의 차이로 인한 조사결과의 차이가 있게 된다. 측정효과는 동일한 조사방법을 선택하거나 권고 받은 대상들에게 다른 조사방법으로 조사를 하였을 경우, 응답한 결과의 차이를 측정오차가 발생한 것으로 볼 수 있다.

실제 동일한 대상에게 다른 조사방법으로 두 번 조사를 하게 하는 것은 현실에 불가능하다. 따라서 모드효과를 선택효과와 측정효과로 구분하기 위한 방법을 Vannieuwenhuyze 등(2010)이 제안한 방법으로 설명하면 다음과 같다. 대면조사와 인터넷조사 두 조사방법을 이용한 혼합조사를 가정하여 보자. 혼합조사 결과 대면조사를 이용하여 조사한 대

29) 가중값을 반복적으로 수정해서 가중값이 수렴하는 값을 찾아 적용하는 방법으로, 보통 레이킹 비(raking ratio) 추정이라고도 함

상의 결과(㉔), 인터넷조사를 이용하여 조사한 대상의 결과(㉕), 인터넷조사를 하기로 하고 대면조사를 진행한 대상의 결과(㉖)가 있다고 가정하면 다음과 같이 각 효과를 측정할 수 있다,

$$\blacksquare \text{ 모드효과} = \text{㉕} - \text{㉔}$$

$$\blacksquare \text{ 선택효과} = \text{㉖} - \text{㉔}$$

$$\blacksquare \text{ 측정효과} = \text{㉕} - \text{㉖}$$

이렇게 각 효과를 측정하게 되면, 조사방법을 선택한 계층간의 차이를 확인할 수 있으며, 조사방법간의 측정효과 차이를 살펴서 실제 응답하는 차이가 있는지를 파악할 수 있다. 전체 모드효과만을 측정하였을 경우에는 정확한 진단을 하기 어렵다.

다. NordChild survey의 조사방법별 차이비교

1) 조사에 대한 개요

‘북유럽 국가 어린이들의 건강과 복지에 대한 조사’(이하 NordChild)는 1984년과 1996년에 실시되었다. 이 조사는 북유럽 5개 국가³⁰⁾의 부모들을 대상으로, 종이설문지를 우편으로 보내는 방법으로 조사되었다. 하지만 이러한 우편조사는 사회환경의 변화에 따라 조사 응답률이 낮아지는 문제점이 드러났다. 이에 웹 기반 기술을 활용함으로써 기존 종이 설문지의 대안을 마련하고자 하였는데, Lena Hohwu et al.(2013)은 데이터 수집 방법을 확대하여 4가지(종이, 종이/웹, 웹, 웹/테블릿)방법을 비교함으로써, 어떤 설문조사 방법이 효율적인지 파악하고자 하였다.

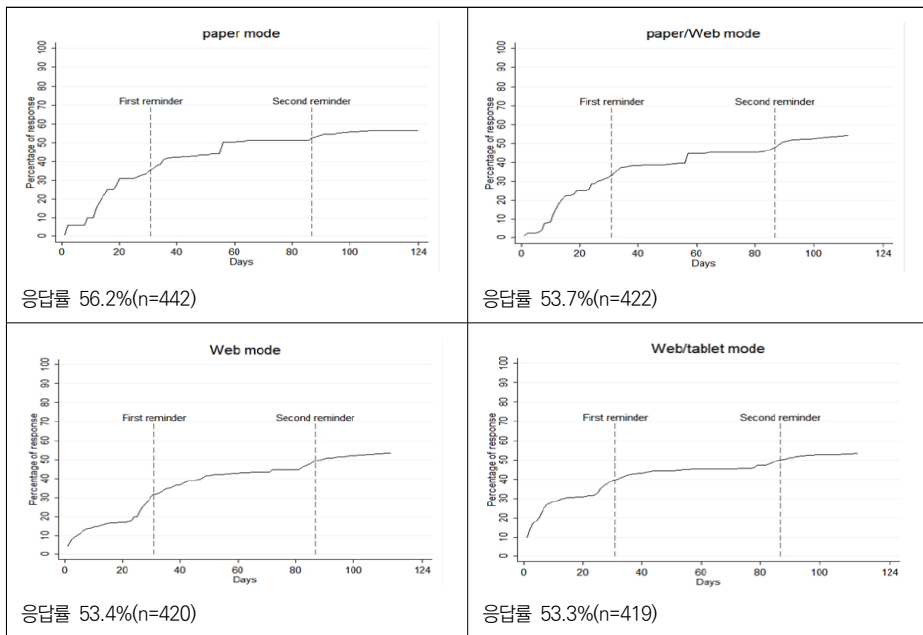
2011년에 실시된 NordChild survey는 핀란드, 아이슬란드, 노르웨이, 스웨덴의 경우 기존과 같이 종이설문지로 실시한데 반해, 덴마크는 위에서 언급한 4가지 방식으로 실시하였다. 응답자 수를 살펴보면 1984년에 총 15,339명 중 10,291명이 응답해 67.09%의 응답률을, 1996년에는 총 15,254명 중 10,667명이 응답해 69.93%의 응답률을 보였다. 이에 반해 2011년에는 총 15,948명 중 7,805명이 응답해 응답률이 48.95%로 급감한 모습을 보였고, 5개 국가가 모두 비슷한 감소세를 보였다.

30) 덴마크, 핀란드, 아이슬란드, 노르웨이, 스웨덴

2) 조사의 4가지 모드별 응답률 차이 분석

덴마크는 2~17세 자녀를 둔 부모 3,148명을 랜덤으로 조사모드에 따라 4개의 그룹으로 나눠 조사를 실시했다. 이 중 1,704명이 응답해 54.13%의 응답률을 보여, 북유럽 5개 국가의 평균 응답률(48.95%) 대비 높은 응답률을 보였다. 특이한 것은 덴마크에서 응답률이 가장 높은 모드는 종이설문지였다는 점이다(통계적으로 유의한 차이는 아니었음). 다른 3개 모드는 서로 비슷한 응답률을 보였고, 종이설문지와 비교했을 때, 종이/웹 모드의 응답률은 9%(OR³¹⁾ 0.91, 95% CI³²⁾ 0.74-1.10), 웹 모드와 웹/태블릿 모드의 응답률은 11%(OR 0.89, 95% CI 0.73-1.09)까지 감소하였다. 물론 전체 응답률에서는 종이설문지를 활용한 우편조사의 응답률이 가장 높게 나왔지만, 여기서는 다양한 모드의 설계와 함께 조사과정자료를 분석하여 3가지 함의를 도출할 수 있다.

[그림 3-7] 조사 소요 일자별 응답률의 변화 추이



자료: Lena Hohwu et al.(2013)

- 31) OR은 오즈비(odd ratio)라고 부르며, 여기서는 기준조사(종이설문지) 대비 비교대상인 조사의 응답률을 의미함. 범위는 0에서 무한대(∞)이며, 1이면 기준조사방법과 같다는 의미이며, 1보다 작으면서 0에 가까울수록 기준방법보다 더 작다는 의미이고, 1보다 클수록 기준방법보다 더욱 크다는 의미임
- 32) CI는 신뢰구간(confidence interval)을 의미하며, 신뢰구간 안에 있으면 차이가 없다는 것을 의미하며, 신뢰구간을 벗어나면 차이가 있다는 것을 의미함

[그림 3-7]은 조사 소요 일자별 응답률의 변화 추이를 제시한 것이다. 이를 살펴보면, 첫 번째로, 조사 초기인 약 10일 전후까지 웹/태블릿 모드의 응답률이 타 조사모드에 비해 월등히 높다는 점이다. 이는 응답의 즉각성이 높고 모드가 혼합되어 선택적 응답이 가능했기 때문에 상대적으로 짧은기간의 조사에 적합할 수 있음을 시사한다. 또한 첫 번째 리마인드(약 30일 전후) 이전의 총 응답률은 34.75%(n=1,094)였는데, 각 모드별로는 종이 모드 35.0%(n=276), 종이/웹 모드 32.8%(n=258), 웹 모드 31.6%(n=249), 웹/태블릿 모드 39.6%(n=311)의 응답률을 보였다. 이 시기까지 응답이 빠르게 이루어진 순서는 ‘웹/태블릿 모드 > 종이모드 > 종이/웹 모드 > 웹 모드’로 나타났다.³³⁾

두 번째로, 첫 문항에 응답한 응답자가 마지막 문항까지 응답한 확률을 살펴보면, 종이/웹 모드가 98.6%로 가장 높았으며, 종이 모드(96.7%), 웹/태블릿 모드(93.0%), 웹 모드(90.6%) 순으로 나타났다. 단일 모드인 종이 모드와 웹 모드가 각각 96.7%, 90.6%였기 때문에 이 둘의 혼합조사에서도 단순히 그 둘 사이의 확률이 나올거라 예상할 수 있지만, 그보다 높은 98.6%가 나온 것이다. 이는 응답자들이 선호하는 혹은 본인의 상황에 맞게 선택적으로 모드를 취할 수 있다는 편의성에 기인하여, 혼합조사가 응답률에 어느 정도 긍정의 효과가 있을 수 있다고 볼 수 있다.

세 번째로, 비용측면에서 종이 모드의 비용(€9.02)이 웹 모드 비용(€4.55)의 약 2배가 높았으며, 가장 비싼 모드로 나타났다. 실제 조사 현장에서는 한정된 조사비용 내에서 얼마나 효율성 있게 조사하는가가 중요하게 작용하며, 이는 표본 수 산정에도 영향을 미치기 때문에 조사여건에 맞는 모드를 사용하는 것도 조사설계에 고려할 사항이라고 볼 수 있다.

라. 병행적 혼합모드 시험조사

비용절감, 응답률 향상 등을 위해 다양한 조사모드를 활용하는 혼합조사가 확대되고 있으나, 이는 선택효과와 측정효과가 혼재된 모드효과 발생에 대한 우려가 있다. 이에 통계개발원은 혼합조사의 모드효과를 파악하기 위해 2014년 병행적 혼합모드 시험조사를 실시하였다.

모집단은 서울에 거주하는 만 13세 이상의 가구원으로 총 1,600개 표본가구를 추출

33) 웹/태블릿 모드의 응답률은 웹 모드의 응답률에 비해 유의미하게 더 높았음(p=.001)

하였다. 이를 위해 서울지역의 집계구를 표본추출틀로 활용하여 서울 4개 광역집계구층별 집계구 내 가구를 기준으로 40개 표본 집계구를 확률비례계통추출방법으로 추출한 후 각 집계구별로 40가구를 추출하는 다단계 층화집락표본추출 방법을 사용하였다. 시험조사에 적용한 조사방법은 자기기입방식인 종이조사와 웹조사이며, 조사방법을 선택할 수 있는 집단과 선택할 수 없는 집단으로 구분하였다. 선택가능 집단은 종이조사와 웹조사 중 조사대상자가 원하는 방법으로 조사(병행 선택형 혼합조사)가 진행되며, 선택불가 집단은 주어진 조사방법으로 조사(병행할당형 혼합조사)가 진행되었다. 추출된 1,600 표본가구를 각각 선택가능 집단 800가구, 종이조사 집단 400가구, 웹조사 집단 400가구로 랜덤하게 배정하였다.

종이조사와 웹조사 간 모드효과를 파악하기 위하여 위의 시험조사 자료를 이용하여 조사방법 간의 계량적 분석을 한 결과, 조사모드 간 응답분포 및 평균비교와 비례오즈모형에서는 일부 항목에서만 유의미한 차이와 모드효과 발생 가능성이 큰 것으로 나타났다. 하지만 병행 선택형 혼합조사의 모드효과를 선택효과와 측정효과를 분리 추정한 결과는 일부 항목을 제외한 대부분의 항목에서 유의미한 것으로 나타났다. 이를 통해 선택계층의 효과를 분리하였을 때, 조사항목에 따라 과대 또는 과소 응답하는 측정효과가 존재하는 것을 확인하였다.

마. 유럽사회조사(European social survey; ESS)³⁴⁾

ESS는 2002년부터 2년 주기의 횡단면 조사로 진행되는 조사이다. 유럽 30개국의 국가 간 비교를 위해 CAPI 면접조사 방법을 사용하였으나, 응답률 저하와 면접조사 비용 등의 문제를 해결하기 위해 전화조사 도입 가능성을 검토하였다. 이를 위해 2003, 2005, 2008년에 실험연구를 실시하였다.

특히 2008년의 실험연구를 살펴보면 네덜란드에서 4차 ESS와 병행하여 진행되었는데, 정치적 관심 항목에 대해 면접조사(CAPI), 전화조사(CATI), 웹조사(CAWI)로 이루어지는 자기기입방식의 조사를 통해 이들의 선택효과와 측정효과를 분리 추정하고 모드효과를 평가하였다. 실험을 위해 단일조사(2,674명)와 혼합조사(878명) 표본을 각각 추출하고, 혼합조사 표본에 대해서는 위의 3가지 조사방식 중 하나를 선택하도록 하였다.

34) 백지선 등(2016) 재인용

주요 분석변수로 사용한 문항인 ‘정치적 관심’은 4가지의 응답범주로(전혀 없음-거의 없음-꽤 있음-매우 있음) 구성되어 있으며, 실험이전에 측정효과와 선택효과가 모두 있을 것으로 예상되었다. 이는 이러한 정치적 관심의 항목이 해당 국가에서는 ‘시민의 의무’로 생각되고 있다는 점 때문에, 면접조사의 방식에서 관심도를 높게 응답하는 성향의 측정효과가 나타나고, 이들 응답자들 중 무응답자들의 상당부분이 자기기입방식의 응답으로 구성되어 관심도가 낮게 나타나는 선택효과가 있을 것으로 예상된 것이다.

선택효과와 측정효과를 분리한 모드효과 추정 결과, 선택효과는 통계적으로 유의하지 않은 것으로 나타났지만, 자기기입방식으로 응답한 응답자들은 정치적 관심이 있다는 항목에 과소 응답하였고, 관심이 없다는 항목에 과대 응답하는 측정효과가 나타났다.

바. 미국 일반 공공 설문조사들의 혼합모드 분석

일반적으로 단위무응답이 발생하게 되면, 조사가 되지 않은 대상으로 인해 표본이 모집단의 포괄성(coverage)을 충족하지 못하는 문제가 발생된다. 이러한 이유로 대부분 단위무응답이 발생할 것을 대비하여 유사한 성향의 대체표본을 사전에 선정한다. 항목무응답의 경우는 조사문항간의 관계에 따라 다르지만, 대체적으로 무응답을 그대로 두고 분석할 경우 편의(bias)가 발생한다. 따라서 무응답 대체방법(nonresponse imputation method)을 적용하여 무응답이 없도록 만든 후 분석한다.

웹조사, 우편조사와 같은 모드에서도 항목 무응답은 조사 데이터 품질에 영향을 미치는 주요 문제가 될 수 있으며, 경우에 따라 단위 무응답만큼 심각할 수 있다. 하지만 웹조사와 우편조사(가구 조사)에서 항목 무응답 결정요인에 대해서는 거의 알려져 있지 않다. Messer et al.(2012)³⁵⁾는 미국의 일반 공공 가구 조사에서 웹과 우편의 혼합조사를 추가 설계하여 항목 무응답 차이를 평가하고 이러한 무응답의 감소를 위한 잠재적인 방법을 식별하였다.

조사자료는 ‘Lewiston & Clarkston Quality of Life Survey’(이하 LCS), ‘Washington Community Survey’(이하 WCS), ‘Washington Economic Survey’(이하 WES)를 사용했으며, 이 조사는 미국의 우체국 배송 순서 파일을 이용하여 주소 기반 샘플링을 사

35) ①웹+우편 모드, ②질문 유형(ex. 사실적, 태도, 행동)+형식(ex. 명목, 순차적, 다중 항목, 개방형 질문 등), ③응답 인구 통계(ex. 성별, 연령, 교육, 인종 및 소득)의 3가지로 나누어 분석하였으나, 여기서는 모드에 따른 무응답률에 대한 내용만 다루고자 함

용했고, 각 조사에서 표본으로 추출된 응답자는 ①우편 전용 응답, ②2주 후 웹 후속 조치가 전송되는 우편 응답(ex.우편+웹), ③2주 후 우편 후속 조치가 있는 웹 응답(ex.웹+우편)으로 진행되었다.³⁶⁾

아래 <표 3-9>과 같은 각 실험에서, 우편 전용과 우편 후속 조치한 경우와 비교할 때, 혼합모드 I 조사 방식에서의 웹에서 가장 낮은 항목 무응답률 보였다.

<표 3-9> 조사설계 및 모드에 따른 응답률/무응답률 비교

구분		LCS		WCS		WES	
		우편전용	웹+우편	우편전용	웹+우편	우편전용	웹+우편
총 단위 응답률		66.3%	55.1%	50.4%	40.1%	62.2%	50.2%
총 항목 무응답률		5.0%	3.6%	4.2%	4.2%	8.1%	8.0%
혼합모드 I	사용된 모드	Mail	Web	Mail	Web	Mail	Web
	단위 응답률	64.4%	40.8%	49.2%	25.0%	62.2%	32.6%
	항목 무응답률	5.0%	2.7%	4.2%	2.7%	8.1%	6.1%
혼합모드 II	사용된 모드	Web	Mail	Web	Mail	Web	Mail
	단위 응답률	1.9%	14.3%	1.3%	15.1%	-	17.5%
	항목 무응답률	DNC**	6.2	DNC**	6.9	-	11.6

*혼합모드 I : 우편 + 웹 후속조치 / 혼합모드 II : 웹 + 우편 후속조치

**DNC: did not calculate due to small sample size

3. 행정자료와 빅데이터의 활용

가. 전통적인 설문조사와 빅데이터 서베이

사회 현상과 변화, 인간의 심리 등을 연구하는 인문사회 분야에서 과학적 연구방법론은 현대에서 가장 중요한 요소이며, 이를 위한 사회조사 방법론은 지속 발전해 왔다. 정치, 경제, 사회, 문화 등 각 분야에서 객관적이고, 실증적인 자료를 수집하는 것이 중요하고, 설문조사는 그러한 목적을 달성하기 위해서 설계되어 대부분의 인문사회 연구에 활용되고 있으며, 선거조사, 여론조사, 시장조사 등 목적에 따라 다양하게 이뤄지고 있다. 예를 들면, 관광 분야에서는 얼마나 많은 관광객이 어떠한 관광행태를 보이는지, 관광 목적과 만족도는 어떠한지, 향후 관광 계획과 선호 관광지는 어떠한지 등 다양한 내

36) 각 웹과 우편 설문지는 시각적 설계 및 레이아웃의 효과를 최소화하기 위해 매우 유사하게 설계함

용을 파악하기 위해 설문조사가 활용되고 있다.

설문조사는 과학적으로 설계하여, 객관적이고, 실증적인 조사결과를 획득하는 것을 목적으로 한다. 이를 위해 구조화된 조사표 등 객관적인 방법들이 적용되어야 하며, 가장 중요한 요소는 조사대상, 즉 표본에 대한 추출이 모집단을 대표할 수 있어야 한다. 이론적으로는 모집단의 전수조사를 실시하는 것이 가장 정확한 조사자료를 얻을 수 있으나, 모집단 전수조사는 대부분의 경우, 막대한 조사비용과, 장기간의 조사기간, 신속한 조사의 어려움 등으로 인해 현실적으로 불가능한 경우가 많다. 그렇기에 대부분의 설문조사는 모집단을 대표할 수 있는 표본을 대상으로 진행되며, 이럴 경우 표본의 설계와 추출에서 객관성을 확보하는 것이 중요하다. 설문조사 표본의 객관성을 확보하는 대표적인 방법론으로 무작위표본추출법(임의추출법)이 활용되고 있다. 기존의 선행연구 결과를 토대로 모집단의 구성 특성과 조사목적의 관련성을 알고 있는 경우, 다양한 형태의 층화추출법이 활용되기도 한다.

설문조사는 기존에 현지조사의 형태를 가진 대면조사로 시작하여, 전화조사로 발전하였으며, 인터넷 조사, 모바일 조사 등 새로운 접근 방법이 활용되고 있다. 그러나 모집단을 대표할 수 있는 객관적인 표본의 확보에 대한 이론적인 부합성은 대면조사 또는 전화조사를 통한 무작위표본추출법 적용을 여전히 선호하고 있다. 특히 공공기관의 정기적인 설문조사는 과거 조사결과와의 연속적인 연구 진행, 통계결과 추적 등을 중요시하여, 대면조사 방법론이 지속 활용되고 있다.

그러나 새로운 사회현상의 출현, 급변하는 조사 환경, 조사방법의 간편성 등 여러 가지 측면에서 인터넷 조사와 모바일 조사가 점점 더 많이 활용되고 있다. 특히 민간 기업들은 마케팅 전략 수립, 신제품 출시 방향, 고객 만족도 개선 등 특수하고 다양한 목적을 위해 설문조사를 활용하는데, 기존의 전통적인 대면조사, 전화조사 보다는 간편한 인터넷 조사, 모바일 조사를 점점 더 많이 활용하고 있다. 앞서 2장에서 살펴본 바와 같이 대면조사, 전화조사, 인터넷 조사 등 각 조사방법별로 장점과 단점이 존재하지만, 적어도 기업측면에서는 인터넷 조사, 모바일 조사가 더욱 선호되고 있는데, 이러한 현상은 코로나19가 발생한 2020년 이후 더욱 가속화 되고 있다. 전세계적인 팬데믹 상황에서, 대면조사를 실시하기에는 그 비용이 막대하고, 효율성은 낮기 때문이다.

하지만 인터넷 조사, 모바일 조사는 모집단을 대표할 수 있는 표본인 것인가에 대하여 지속적인 도전을 받아 왔다. 현대에 설문조사를 수행하는 조사기관들은 기존에 확보하고

있는 조사패널을 활용해, 다양한 층화추출법을 적용하여 인터넷 조사, 모바일 조사를 실행하고 있다. 그러나 국내 리서치사들이 보유한 조사패널은 수십만명에 불과하며(액티브 패널은 수만명에 불과), 각 리서치사들간 조사패널도 중복되는 것으로 추정되어 모집단을 충분히 대표할 수 있는가에 대한 논쟁의 소지가 존재한다. 이에 모집단을 국민으로 설정하여 설문조사를 통해 국가 통계를 생산하는 공공기관 등의 경우 여전히 대면조사, 전화조사의 방법론이 우선 활용되고 있다. 하지만 코로나19로 인한 대면조사의 어려움과, 전화조사의 한계성으로 인해 공공기관의 경우도 마찬가지로 새로운 조사방법론에 대한 시도와 연구가 필요한 상황에 직면해 있다.

한편, 기존의 설문조사 방식이 아닌, 빅데이터 분석 방법론의 활용이 많아지고 있다. 4차산업혁명 시대의 도래로 인해 인간의 사회활동은 인터넷과 모바일, 각종 센서들에 의해 데이터로 표현되고 있으며, 이러한 빅데이터를 활용한 분석은, 기존의 설문조사를 통한 자료수집 분석과는 또 다른 활용 가치가 높기 때문이다. 빅데이터 분석 방법의 가장 큰 차이점은 설문조사에 의한 자료수집 대비, 압도적인 크기의 데이터를 통해 모집단을 분석할 수 있다는 것이다. 국내 환경에서 볼 때, 각 분야에서 독과점 형태의 지배 사업자인 대기업들이 보유한 빅데이터는 국민들의 문화·관광 활동 등을 분석하는 중요한 도구로 활용될 수 있다. 또한 이러한 빅데이터가 장기간에 축적되어 있을 경우, 장기간의 설문조사에 의한 시계열 연구를 대체할 수 있는 가능성도 보유하고 있다. 대면조사, 전화조사 등 전통적인 방법에서는 조사대상 표본의 기억에 의한 답변에 의존할 수 밖에 없지만, 이에 대비하여 특정 대상의 행동에 따라 장기간 수집된 빅데이터는 더욱 정확한 실재값을 반영할 수도 있기 때문이다. 또한 설문조사는 조사설계, 조사실행, 조사결과 수집 등 절차와 시간이 필요한 것에 비해, 빅데이터 분석은 훨씬 빠른 시간에 대용량의 분석을 할 수 있으며, 이는 현대에 이르러 대용량의 빅데이터 분석 기술의 발전에 따라 효용성이 높아졌다.

빅데이터 분석은 인간의 사회생활을 분석할 수 있지만, 인간의 심리 변화, 사회현상의 원인과 과정을 세밀하게 분석하기 어려운 단점도 존재한다. 즉, 빅데이터로 수집된 분야에서의 인간 행동의 결과는 정확하게 분석할 수 있지만, 그 이면에 있는 인간의 심리를 분석하기 어려운 것이다. 예를 들어 카드회사의 빅데이터로, 누가(카드사용자 프로필 특성 등), 언제(소비 시점), 어디서(소비 장소), 무엇을(소비 가맹점 업종) 소비했는지를 분석할 때, 어떠한 설문조사보다 정확한 분석결과를 얻을 수 있다. 그러나 소비 행동 이유, 소비 만족도, 소비의 과정, 소비 계획 등 인간의 심리가 반영된 부분의 분석에는 한

계가 명확하다.

이에 빅데이터 분석방법론과 설문조사가 결합한 형태의 ‘빅데이터 서베이’라는 새로운 조사방법론이 대안으로 활용되기도 한다. 국내에서 개인정보를 수집·활용·제공하기 위해서는 개인정보보호법에 의해 반드시 개인의 동의를 얻어야 한다. 국내에서 개인에 대한 빅데이터를 보유하고 있는 대기업들은 마케팅과 서비스 제공 목적으로 고객과 커뮤니케이션 할 수 있도록 ‘개인의 동의’를 받은 수백만명 이상의 고객을 확보하고 있다. 그러한 기업들 중 일부 회사는 고객으로부터 인터넷과 모바일을 통해 설문조사를 할 수 있는 시스템을 갖추고 있으며, 이때 수백만명의 동의 받은 고객들을 조사패널로 활용할 수 있다. 빅데이터를 보유하고 있는 기업의 고객 설문조사 결과와 빅데이터를 결합하여 분석하는 방법론을 ‘빅데이터 서베이’라고 할 수 있으며, 이는 빅데이터 분석의 장점과 대면조사 등 전통적인 설문조사의 장점을 일정부분 모두 취할 수 있다. 즉, 설문조사를 통해 획득한 응답결과와 더불어 응답자의 빅데이터(예를 들면, 카드사의 소비패턴, 고객 특성 데이터)를 융합하여 연구가 가능하다.

‘빅데이터 서베이’는 설문조사 측면에서, 수백만명 이상의 고객을 대상으로 하기에 표본 추출에서 모집단의 대표성, 객관성 확보에 기존의 리서치 회사의 패널 조사 방식보다 강점을 가진다. 하지만, 전 국민을 대상으로 하는 무작위표본추출법을 주요한 방법론으로 적용하는 대면조사에 대비하여, 표본의 대표성, 객관성 측면에서 ‘빅데이터 서베이’가 미흡하다는 의견도 있다. 이러한 주장의 대표적인 견해는 수백만명 또는 수천만명의 고객 일지라도 해당 정보를 보유한 특정기업의 편중현상으로 인해 모집단의 대표성 측면에서 한계가 있다는 것이다. 하지만 특정기업의 고객집단이 모집단과 일부 다른 구성을 갖고 있더라도, 다양한 층화추출법을 적용하면, 최대한 근접하게 모집단을 반영할 수 있는 표본을 구성할 수 있다. 예를 들면, 특정기업 고객집단의 지역별 구성비가 모집단(예: 국민)의 지역별 구성비와 다를지라도, 모집단의 지역별 구성비에 맞도록 표본을 추출하여 조사결과를 획득하는 방식이다. 만약 국민의 지역별, 성별, 연령대별 구성과 동일한 구성비의 표본을 원할 경우, 특정기업집단의 수백만명의 고객 중에서 층화추출을 통해 적용할 수 있다. 또한 특정기업집단의 고객의 전반적인 소득수준이 높은 특성이 있다면, 모집단의 소득수준 구간별 구성비에 맞도록 표본을 추출하는 방식으로 표본의 객관성과 대표성을 높일 수 있다.

특히 카드사의 경우 경제활동인구 전반을 고객으로 확보하고 있기에, 경제활동인구를

대상으로 하는 설문조사의 경우 모집단을 대표할 수 있는 표본 추출에 유용하다. 예를 들면, 카드사는 14세 이상의 경우에 체크카드를 발급받을 수 있기에 13세 이하의 고객은 없지만, 15세 이상 혹은 성인을 대상으로 하는 기존의 문화·관광 관련 조사의 경우 대체 가능성이 존재한다. 즉, 과학적 표본 설계를 통해 ‘빅데이터 서베이’를 적용할 경우, 충분히 모집단을 대표할 수 있는 객관성 있는 표본 추출이 가능하다.

‘빅데이터 서베이’의 가장 큰 장점은 기존의 방법론에서 어려웠던, 특정 이슈에 연동된 표본, 특정 유형의 정교한 표본 추출이 가능하다는 것이다. 이를테면, 카드사의 고객들을 통해 ‘빅데이터 서베이’를 실행할 경우, 카드 소비패턴에 따른 고객유형을 사전에 분석하여, 조사 목적별 맞춤형 표본을 추출하여 조사할 수 있다. 예를 들면, 백화점을 주로 이용하였으나, 최근에는 온라인 쇼핑을 더욱 많이 이용하는 고객을 카드 소비패턴으로 파악해서 표본 추출하여 모바일을 통해 실시간 설문조사를 할 수 있다. 또한, 특정 관광지를 방문한 관광객을 소비패턴을 통해 표본 추출하여 조사할 수 있는데, 기존에 현장조사의 시간적 제약을 일정부분 극복할 수 있다. 예를 들면, 최근 5년간 각 년도별로 제주도를 방문한 관광객을 카드 빅데이터 분석을 통해 표본 추출하여, 현 시점에서 한꺼번에 모바일 설문조사 할 수 있다. 특히 시의성이 중요한 트렌드 분석이나 혹은 정책효과를 빠르게 판단하기 위한 자료수집을 위해서 ‘빅데이터 서베이’가 더욱 유용하게 활용될 가능성이 있는 것이다. 요컨대 ‘빅데이터 서베이’는 국내 환경에서 대면조사를 대체 또는 보완하기 위해 충분히 시도할 가치가 있다. 표본 추출에 있어서, 모집단을 고려한 충분한 과학적 설계가 전제된다면, 빅데이터 분석과의 융합 분석의 강점, 특수하거나 정교한 표본 추출의 강점, 시공간의 제약을 극복할 수 있는 모바일 조사의 장점 등을 활용 할 수 있을 것으로 보인다.

한편, ‘빅데이터 서베이’는 모바일 조사를 주로 활용하는데, 대면조사 수준의 성실한 응답결과를 얻기 위해서는 응답고객에게 충분한 보상을 지급하고, 구조화된 조사표 구성, 응답 편의성을 고려한 기술적 조치(예: 1문항 1Page, 객관식 문항) 역시 고려해야 할 사항이다.

나. 빅데이터를 이용한 조사 및 활용

최근에 들어오면서 빅데이터의 중요성은 점차 중요하게 활용되고 있다. 그러나 빅데

이터의 이용은 매우 제한적이다. 가장 큰 이유를 개인정보문제로 생각하고 있지만, 실제로는 빅데이터가 가지고 있는 구조적인 문제가 크다. 빅데이터는 용량은 크지만 관심변수가 하나 또는 몇 개 안되는 경우가 많다. 가입자의 정보 등을 이용해서 다양한 차원으로 분석함에 따라 다양한 형태로 가공되어 다양한 형태의 정보를 제공할 뿐이다. 또 하나는 너무나 많은 데이터가 있기 때문에, 이를 이용한 고급 분석은 거의 사용하지 않고 있다. 데이터 핸들링과 분석을 하기 위해 이를 지원할 수 있는 시스템과 프로그램이 있어야 하기 때문이다.

무엇보다도 빅데이터는 결과(fact)는 존재하지만, 원인과 인식 등 결과와 연계할 변수들이 없다는 한계가 있다. 이러한 데이터들이 있다면 더욱 활용하기 좋은 데이터가 되고, 이를 분석하여 가치 있는 정보를 제공하게 될 것이다.

빅데이터와 반대로 조사데이터는 조사비용(시간과 예산)이 많이 들에 따라 조사된 데이터의 수는 빅데이터에 비해 아주 적으며, 특정 시점에서의 응답자들이 답한 결과이다. 그러나 그 시점에서의 다양한 결과(fact)를 제시하고, 원인과 인식, 생각 등을 파악할 수 있다.

이러한 이유로 빅데이터와 조사데이터를 연계하여 활용하고자 하는 방안에 대한 연구가 지속적으로 진행되고 있다. 빅데이터의 동적(active)인 데이터와 조사데이터의 정적(static)인 데이터의 결합, 일부 변수와 다양한 변수의 결합은 분석가들의 관점에서는 매력적이라 할 수 있다. 그러나 개인정보의 문제가 있어 이러한 결합은 어렵다. 가명데이터를 이용하여 결합한다고 하더라도, 빅데이터와 조사데이터의 결합을 위해서는 사전에 데이터 결합을 위한 기준을 마련해야 하는 등의 문제가 있게 된다.

최근 가명정보를 이용한 데이터 결합은 많은 연구자(분석가)들의 주목을 받고 있다. 그러나 가명정보를 이용한 결합은 결합도 어렵지만, 결합한 이후의 분석을 위한 데이터를 다시 분할해야 하는 등의 문제가 있어 쉽게 접근하기가 어렵다. 따라서 각각의 데이터를 생성 후 결합하기 보다는, 빅데이터의 정보에 조사를 추가하여 정보를 확대하는 방안이 분석을 위한 설계에 더 쉽다. 최근에 카드회사에서 카드고객에게 동의를 받아, 동의한 대상으로 표본설계를 통해 표본추출하고, 표본으로 뽑힌 고객들에게 조사하여, 그 결과를 추정하는 방법을 활용하고 있다.³⁷⁾ 고객들의 정보로 표본추출을 하지만, 표본으로 추출된 대상들의 성별, 연령, 지역, 소득구간 등의 정보만을 제공하고 있어 개인정

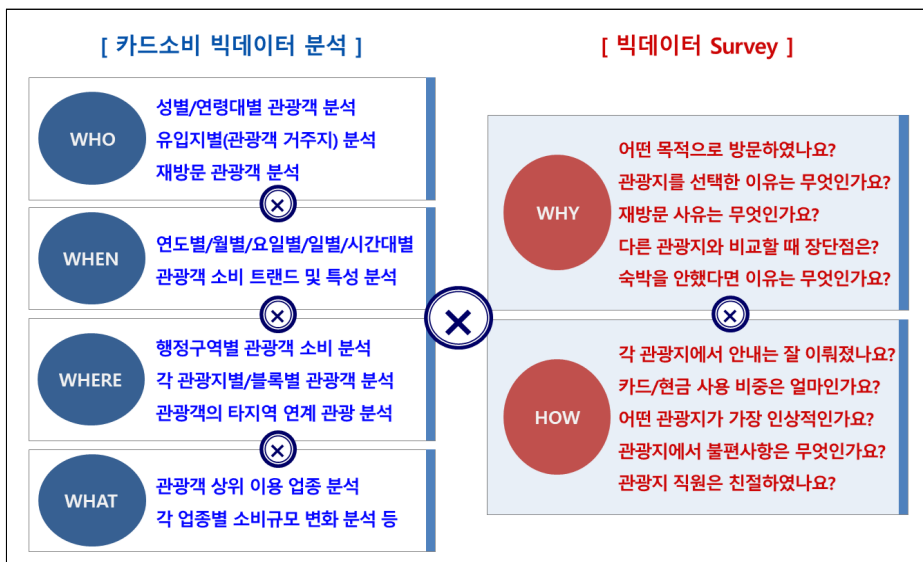
37) 신한카드에서 제공한 정보를 바탕으로 표본이론을 접목하여 설명함

보문제가 없이, 추출된 대상의 소비행태 등의 정보를 활용할 수 있다. [그림 3-5]는 카드 고객을 이용하여 표본조사한 후, 빅데이터와 조사데이터를 연계하여 분석한 것을 설명한 자료이다.

카드사의 고객들의 동의를 통해 마련된 조사대상(약 600만명)으로 이 결과를 국민전체로 추정하는 것은 가능할 것으로 판단된다. 물론, 이러한 부분은 실제 추정하여 값을 비교하고 검증하는 방법이 필요할 것이다.

표본조사한 결과이기 때문에, 빅데이터의 값을 그대로 활용하는 것이 아니고, 가중값을 산출하여 추정하여 활용하게 된다. 그러나 실제 카드사 전체의 값과 차이가 오차범위 내에서 제시되고 있어, 표본을 추출하여 조사를 한 결과는 빅데이터의 정보를 가지고 있다고 할 수 있다.

[그림 3-8] 빅데이터 서베이를 이용한 관광분석



자료: 신한카드 내부자료

이러한 정보를 공신력있게 활용하기 위해서는 실제 조사한 결과가 우리나라 전체의 정보와 연계할 수 있는 지에 대한 부분이다. 또한 조사방법을 온라인조사로만 하고 있는데, 이러한 부분의 값이 대면조사 등과 비교할 때 어느 정도 차이가 나는지 등의 연구를 통해 정합성 등의 검증을 할 필요가 있다.

다. 행정자료를 이용한 조사 및 활용

많은 국가에서 센서스 및 기타 공공의 통계 생성을 위해 조사 방법의 다양화가 이루어지고 있는데, 이는 조사비용의 확보 문제, 응답부담과 응답자의 반발에 따른 응답률의 감소 등 조사환경의 변화와 주로 관련되어 있다(조순기 등, 2008).

특히 등록부에 기반한 센서스(register-based census: 이하 등록센서스)는 센서스 조사 방법의 변경에 있어 중요한 흐름 중 하나라고 할 수 있다. 등록제도가 발달한 일부 북유럽 국가를 중심으로 등록센서스로의 전환이 이루어졌는데 이는 현장조사가 아닌 행적목적으로 구축해 놓은 등록부를 통해 개별 단위에 관한 자료를 수집하는 방식을 말하는데, 등록 기반 통계의 오랜 전통이 존재하는 덴마크, 핀란드, 스웨덴과 같은 북유럽 국가들이 EU의 회원국이 되었을 때 유럽통계청(Eurostat)에서는 공식적 통계 생산의 '새로운' 방법에 대한 논의를 시작하였다.³⁸⁾

등록센서스 실시에 가장 선구적인 국가는 덴마크로, 1976년에 불완전한 등록센서스를 실시한 이후 1981년에 세계 최초로 완전한 등록센서스를 실시하였다(Borchsenius, 2001). 덴마크는 오늘날 세계에서 가장 복잡하고, 고도로 통합되어 있으며, 효율적으로 전산화된 행정 시스템을 가지고 있는 국가 중 하나이다. 특히 개인의 거주지 등록제도가 일찍 시작되었으며, 1970~1980년대에 이미 광범위한 전국적인 행정 시스템이 계획·설립되었다. 센서스에서 행정자료 활용이 앞선 덴마크, 핀란드 등은 등록부 기반 통계 생산에 오랜 경험을 가지고 있으며 다른 국가들에 비해 통계생산 목적을 위해 행정자료를 보다 광범위하게 활용하고 있다(조순기 등, 2008). 이에 우리나라의 등록센서스 사례와 행정자료를 활용하여 작성한 통계사례를 살펴보면 다음과 같다.

1) 통계청의 인구주택총조사³⁹⁾

행정자료를 이용한 조사의 대표적 사례는, 우리나라에 사는 모든 내·외국인과 주택의 규모 및 특징을 알기 위한 국가의 기본적인 통계조사인 통계청의 인구주택총조사를 들 수 있다. 기존 전수조사로 실시하던 방식에서 2015년부터 등록센서스⁴⁰⁾ 방식으로 최초

38) 유럽통계청에 의해 조직된 1997년의 세미나는 등록 기반 통계가 EU내에서 공식 통계를 생산하는 방법으로 수용되게 되는 전환점이 되었음(Wallgren and Wallgren, 2007)

39) 인구주택총조사 홈페이지(<https://www.census.go.kr/>)를 바탕으로 정리하여 재작성함

40) 전국의 모든 가구를 직접 방문하지 않고, 주민등록부, 건축물대장 등 행정자료를 이용하여 인구·가구·주택

전환하였는데, 등록센서스로 행정자료를 활용한 인구, 가구, 주택에 관한 통계는 매년 생산되고 있지만, 지역의 복지·경제·교통 등 정책수립에 필요한 실질적인 자료를 수집하기 위해 5년마다 우리나라 전체 가구의 20% 표본을 선정하여 현장조사를 실시하고 있다. 즉, <표 3-10> 과 같이, 전수항목에 해당하는 16개 항목은 매년 등록센서스로 행정자료를 통해 수집되며, 표본항목의 경우는 5년 주기로 현장조사+행정자료를 통해 수집된다.

〈표 3-10〉 2020 인구주택총조사 조사항목

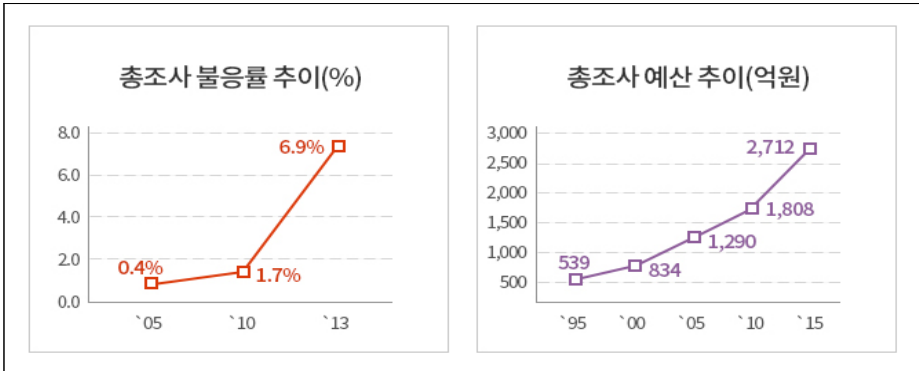
구분	전수항목(16)	표본항목(55)			
		현장조사(45)			행정자료(10)
인구 (34)	-성명	-성명	-활동 제약	-현 직업 근무연수	-1년 전 거주지
	-성별	-성별	-활동 제약 돌봄	-근로 장소	-5년 전 거주지
	-생년월일	-생년월일	-통근·통학 여부	-혼인 상태	
	-가구주와의 관계	-가구주와의 관계	-통근·통학 장소	-혼인 연월	
	-국적	-국적	-이용 교통 수단	-출산 자녀 수	
	-입국 연월	-입국 연월	-통근·통학 소요시간	-자녀 출산 시기	
	-1년 전 거주지	-교육 정도	-경제활동상태	-추가 계획 자녀 수	
	-국적취득연도	-교육 영역	-종사상지위	-결혼 전 취업 여부	
		-출생지	-산업	-경력 단절	
		-아동 보육	-직업	-사회 활동	
가구 (16)	-가구 구분	-가구 구분	-소방시설 보유 여부	-거주 기간	-주거시설 형태
	-주거시설 형태	-1인 가구 사유	-사용 방 수	-주거 전용·영업 겸용 여부	-타지 주택 소유 여부
		-혼자 산 기간	-난방시설		
		-반려(애완) 동물	-주차장소	-점유 형태	
		-마시는 물	-건물 및 거주층	-임차료	
주택 (6)	-거처의 종류				-거처의 종류
	-총 방수				-총 방수
	-주거시설 수				-주거시설 수
	-주거용 연면적				-주거용 연면적
	-대지면적				-대지면적
	-건축 연도				-건축 연도

자료: 2020 인구주택총조사 홈페이지(<https://www.census.go.kr>)

이러한 등록센서스가 도입된 이유는 바로 대내외적인 조사환경의 변화와 조사비용의 증대에 있다. 1인 가구 및 맞벌이 부부의 증가, 사생활 보호의식 강화 등으로 응답자가 부재중이거나 답변을 거부하는 사례가 증가하였고, 전국 단위의 대규모 현장조사로 인해 비용도 크게 증가한 반면, 대규모 공공데이터의 전산화로 통계작성 및 활용은 가능해졌기 때문이다. [그림 3-9]는 인구주택총조사의 불응률과 예산 추이를 제시한 것이다.

에 대한 통계를 생산하는 새로운 방식의 인구주택총조사를 말함

[그림 3-9] 총조사의 불응률 및 예산 추이



자료: 2020 인구주택총조사 홈페이지(<https://www.census.go.kr>)

이렇듯 통계청은 사회경제적 여건 변화에 따라 2015년부터 부처 간 행정자료의 공유·협력을 기반으로 등록센서스 방식을 도입한 후, 이에 따라 인구주택총조사를 진행하고 있는 것이다.

2) 통계청의 기업통계등록부 자료를 이용한 조사

국가 승인통계 중 사업체를 대상으로 하는 통계는 대부분 「전국사업체조사」 결과를 표본틀로 하여 표본설계를 하고, 추출된 사업체의 전화번호와 주소를 제공받아 컨택한 후, 조사설계 시 정한 방식으로 조사를 수행한다. 표본설계를 할 때, 업종과 사업체규모를 층화(stratification)변수로 이용하고 지역변수를 내재적 층화변수로 활용하여 전국적으로 골고루 뽑히도록 한다. 사업체 규모는 매출액 등의 금액을 이용하면 더욱 정교한 설계가 가능하지만, 대부분 종사자규모를 변수로 이용하여 조사하고 있다. 매출액 정보를 이용하기 어려운 경우가 많고, 매출액과 같은 정보는 표준화하기도 어려워 종사자규모 변수를 이용하고 있다.⁴¹⁾

통계청의 통계데이터센터에서는 필요한 경우, 통계청에서 구축한 기업통계등록부 자료를 이용하여 표본설계를 할 수 있도록 지원하고 있다. 기업통계등록부는 통계 작성 지원과 조사모집단 제공을 목적으로 작성되었으며, 재화와 서비스를 생산하는 모든 기업(법인기업, 개인기업)을 포괄하고 있다. 기업통계등록부에서 제공하는 정보는 기업체, 사

41) 국가승인통계의 사업체 조사를 할 경우, MDIS에서 매출액 정보 활용이 가능함

업체 단위별로 기본정보와 산업분류, 종사자, 매출액이 있으며, 기본정보에는 사업체 주소, 대표자 성별, 개/폐업일 등의 정보가 포함되어 있다.

기업통계등록부 자료는 기업체와 사업체의 다양한 정보를 제공하고 있어 개인정보문제가 있기 때문에, 절차를 거쳐서 활용하도록 하고 있다. 표본설계를 위한 명부서비스 이용절차는 다음과 같다.⁴²⁾

- ① 공문선행
- ② 표본설계 여부에 따른 이용신청
- ③ 표본설계가 필요한 경우, 분석센터 이용신청
- ④ 센터 방문 후 자료분석 및 표본 설계
- ⑤ 표본데이터 1차 반출 신청
- ⑥ 반출파일에 대한 확인 및 명부 작업
- ⑦ 분석센터 방문하여 표본데이터 확인
- ⑧ 최종 표본데이터로 2차 반출 신청
- ⑨ 반출 승인 완료
- ⑩ 파기확인서 업로드

통계청의 기업통계등록부를 이용하게 되면, 해당 자료에 포함된 정보를 조사 설문문항에 연계하여 통계를 산출할 수 있다. 그러나 기업통계등록부 자료를 만들기 위해 다양한 기관의 정보를 결합하여 생성하므로 시의성이 늦기 때문에 최근의 결과를 이용하는 데 어려움은 있다. 하지만 개별 통계가 아닌 전체의 현황을 살펴보고자 한다면 설문문항에 매출액 등의 정보를 이용할 수 있다.

「전국사업체조사」자료는 국세청에 신고된 사업체와 비교하면, 규모가 작은 사업체가 많이 누락되어 있다. 따라서 「중소기업 실태조사⁴³⁾」와 같은 경우 규모가 작은 사업체가 많기 때문에 기업통계등록부 자료를 이용하여 표본을 추출하고, 조사를 하였다. 따라서 기업통계등록부 자료를 같이 연계하여 활용함에 따라 민감한 설문문항을 줄일 수 있어 조사의 효율성을 높였다. 단, 개인정보 문제로 통계를 산출할 경우 개별 단위의 값은 반

42) <https://data.kostat.go.kr/>

43) 1회성으로 2017년도에 작성된 적이 있음

출이 안 되고, 통계값만 반출 승인을 받고 제공된다. 따라서 원데이터(raw data)를 이용하고자 하는 경우에는 향후 활용에 한계가 있다. 이러한 이유로 많은 정보를 활용할 수 있는 장점이 있어도 이를 이용하는 것을 주저하게 되는 경우들이 발생된다.

유사한 사례로, 중소기업벤처부의 '장애인기업실태조사'⁴⁴⁾의 경우가 있다. 이 조사의 가장 큰 애로사항은 우리나라 기업 중 장애인이 대표인 기업을 확인할 수 조사모집단이 존재하지 않아 정확한 모집단 규모를 파악하는데 어려움이 있다는 점과 조사명부가 존재하지 않아 현장조사를 하는데 있어서도 어려움이 있다는 점이었다.

조금 더 자세히 살펴보자면, 2017년 조사까지는 통계청의 전국사업체조사를 기본 추출틀로 하여 보건복지부의 장애인등록명부의 성명과 출생년도(나이)를 활용해 전국사업체조사의 대표자 성명과 출생년도 자료를 연계한다. 또한 장애인기업 후보모집단을 작성하여 일부 사업체를 대상으로 장애인기업 여부를 전화조사를 통해 판별조사를 하고, 장애인기업으로 확인되는 비율을 활용하여 장애인기업을 추정하는 방법으로 적용하였다. 이 때 장애인기업으로 확인되는 비율은 7.9%에 불과하여 정확한 장애인기업을 추정하는데 한계가 있고, 장애인기업에 대한 모집단을 작성할 수 없어 장애인기업에 대한 모집단 정보를 파악하는데 어려움이 있었던 것이다. 하지만 2019년 조사에서는 기업통계등록부 자료의 대표자성명, 성별, 출생년월일 항목을 활용하여 보건복지부의 장애인등록명부를 연계하여 후보모집단으로 작성하고, 장애인기업이 아닌 기업을 판별조사를 통해 후보모집단에서 제외하여 조사모집단을 구축하였다. 일부 판별이 잘못된 기업이 조사모집단에 포함될 가능성은 있지만 리스트로 구축된 조사모집단을 통해 장애인기업의 산업, 기업규모, 지역 등의 모집단 분석이 가능하고 표본을 보다 효율적으로 추출할 수 있게 된 것이다. 이는 향후 기업통계등록부의 개인대체식별번호와 보건복지부의 개인대체식별번호로 연계할 수 있다면 보다 완벽한 조사모집단을 구축할 수 있을 것으로 예상해 볼 수 있다.

기업통계등록부 자료이외에도 4대보험자료 등이 있는데, 이러한 자료를 이용하여 국민대상 조사를 할 수 있다면, 소득 등에 대한 정보를 이용할 수 있을 것이고, 조사통계의 활용을 높이는 계기가 될 것이다. 기업통계등록부와 4대 보험자료와 같은 다양한 데이터를 기반으로 조사를 할 수 있는 방안은 조사 이후의 데이터를 연계하여 활용한 것보다

44) 장애인기업실태조사는 장애인기업의 일반사항, 재무상황, 창업사항 등 전반적 실태 파악을 통해 장애인기업 정책수립의 기초자료 제공하는 것이 목적인 조사로서 대표자가 장애인인 기업을 대상으로 5,000개 표본 기업을 추출하여 2년마다 실시하는 조사임

더 좋은 정보를 제공하는 방법이 될 수 있다. 단, 개인정보문제를 해결할 수 있는 자료활용을 위한 보안 관련 정책 마련이 선행되어야 한다.

3) 통계청의 기업통계등록부 자료를 이용한 가공통계

행정통계자료를 조사통계에 적용한 것 외에도 가공통계를 작성하는데 활용한 사례도 다수 있는데, 그 중 국가승인통계에 적용된 사례로 ‘지자체행정통계’를 들 수 있다. 지자체별로 행정자료 및 기존 통계자료를 연계·결합하는 방법으로 가공통계를 작성하고 있고, 대표적으로 여성통계와 노인등록통계 등의 통계가 지자체별로 작성되고 있다.

여성통계는 김포시, 안양시, 의왕시, 아산시, 서산시, 당진시 등 지자체별로 15종의 국가승인통계가 작성되고 있다. 예를 들어 김포시 여성통계⁴⁵⁾는 통계청의 등록센서스 인구DB, 가구DB, 주택소유DB, 경제활동DB, 사업체DB, 기업체DB, 신혼부부통계 등 자료와 국민연금공단, 공무원연금공단, 사학연금공단, 고용노동부, 국민건강보험공단 등의 행정자료를 활용하고 있다.

또한 노인등록통계는 지자체별로 19종의 국가승인통계가 작성되고 있다. 시·도는 대구광역시, 울산광역시, 전라남도 등에서 작성을 하고 있고, 시·군·구에서는 김포시, 고양시, 오산시, 의왕시 등에서 노인통계를 작성하고 있다. 예를 들어 김포시 노인등록통계⁴⁶⁾는 행정통계자료인 등록센서스 인구DB, 가구DB와 기존 통계자료인 사망원인통계DB, 인구동향조사, 건강보험자료, 고용보험자료 등을 활용하여 작성하고 있다.

4) 한국문화관광연구원의 문화체육관광산업통계

문화·관광 분야에서 행정자료를 활용하여 작성한 대표적인 사례는 한국문화관광연구원에서 작성·생산하고 공표하고 있는 문화체육관광산업통계(승인번호 제434001호)를 들 수 있다. 이 통계는 통계청의 기업통계등록부DB를 활용하여 문화체육관광산업 영위 사업체의 경영활동 현황(사업체 수, 종사자 수, 매출액 등)을 산출하는 가공통계인데, 이를 산출하는 과정을 요약해보면 다음과 같다.⁴⁷⁾

45) 여성(전 연령)의 경제사회적 삶의 수준 등 여성의 삶을 다차원적으로 분석하여, 삶의 질 향상을 위한 여성 정책 수립의 기초자료 제공하기 위해 작성되는 가공통계

46) 65세 이상 노인의 경제사회적 삶의 수준 등 노인의 삶을 다차원적으로 분석하여, 노인의 삶의 질 향상을 위한 노인복지정책 수립에 필요한 기초 자료 제공하기 위해 작성되는 가공통계

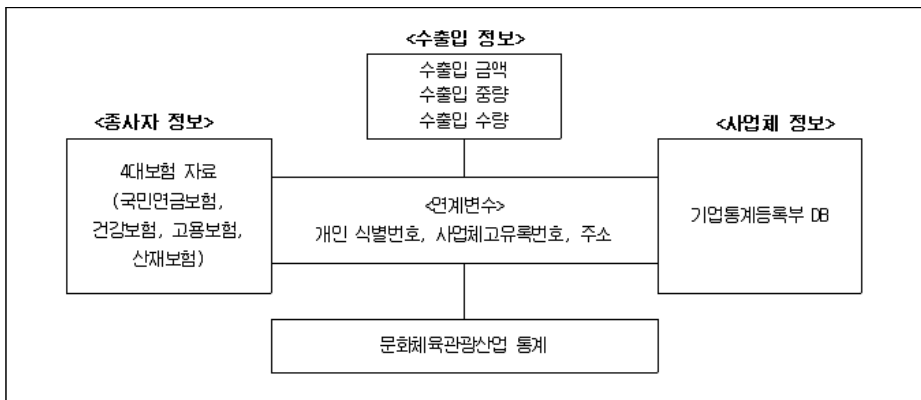
〈표 3-11〉 문화체육관광산업통계 산출 과정

구분	단계	내용
2017년 추정	①	문화체육관광 관련 산업분류에 맞는 사업체를 통계청의 전국사업체조사 자료에서 추출하여 2017년 기준의 문화체육관광산업 표본들을 구축
	②	문화체육관광산업 표본들의 사업체고유번호와 기업통계등록부자료의 사업체고유번호를 연계하여 문화체육관광산업에 해당하는 사업체의 경영활동 DB를 생성
	③	문화체육관광산업 표본들과 기업통계등록부자료의 일부가 매칭되지 않는 경우가 있어 이에 대한 부분은 가중값으로 처리하여 추정
	④	가중값을 처리할 때 기준은 중분류(19개 구간)와 종사자 수 구간(7개 구간) 기준으로 다음과 같이 산정 가중값 = (해당구간의 매칭된 사업체 수) / (해당 구간의 전국사업체 자료 수)
	⑤	DB와 연계하여 산출된 값에 2018년도 가중값 적용하여 최종 결과값을 산출
2018년 추정	⑥	전국사업체조사 자료의 연도별 사업체 수의 변화를 표로 정리
	⑦	문화체육관광 관련 표본들을 연도별로 중분류와 종사자구간으로 정리하여, 전국사업체조사 자료의 추이와 문화체육관광 관련 표본들 구간별 추이의 관계식을 찾음
	⑧	관계식에 따라 2018년도 문화체육관광 관련 모집단 규모를 중분류와 종사자구간별로 추정
	⑨	2017년 문화체육관광산업 표본들의 사업체고유번호와 기업통계등록부자료의 사업체고유번호를 연계하여 문화체육관광산업에 해당하는 사업체의 경영활동 DB를 생성
	⑩	문화체육관광산업 표본들과 기업통계등록부자료가 매칭되지 않는 경우가 있어 이에 대한 부분은 가중값 처리하여 추정
	⑪	DB와 연계하여 산출된 값에 2018년도 가중값 적용하여 최종 결과값을 산출

자료: 한국문화관광연구원(2020)

또한 해당 통계를 산출하기 위한 자료연계 구조를 살펴보면 [그림 3-10]과 같다.

[그림 3-10] 자료연계 상세 구조도



47) 문화체육관광산업통계는 경영활동 현황 외에도 4대보험자료를 활용한 일자리 현황, 수출입통관자료를 활용한 수출입 실적 현황이 포함되어 있으나, 여기서는 SBR을 활용한 경영활동 현황에 대한 내용만 정리

이러한 문화체육관광산업통계는 2016년 통계청에서 통계작성지정기관 확대를 통한 국가통계 개발 사업의 일환으로 작성되었다. 특히 작성 배경은 ①문화·체육·관광 관련 산업 통계에 대한 수요 증가 ②현장조사 환경 악화 ③행정자료를 활용한 통계생산의 국제적 추세 등을 들 수 있다. 당시 정부는 일자리 창출 및 경기활성화를 위한 창조 경제의 핵심으로 그 중요성이 증가했음에도 불구하고, 다양한 산업이 융복합 되어있는 문화체육관광 산업 정보를 종합적으로 파악하는데 한계가 있었다. 기존 승인통계와 개별지표가 있었지만 산업별 사업체만을 대상으로 하여 문화체육관광산업 전체에 대한 종합적 통계정보 파악이 어려웠기 때문이다. 우리나라 전체 산업별 특성 파악을 목적으로 한 고용노동부의 사업체노동력조사나 통계청의 경제활동인구조사 등 역시 산업구조가 세분화되어 있는 문화체육관광 관련 산업에 대한 세부 정보 파악이 불가능 했다. 이와 더불어 통계조사의 증가, 조사항목의 확대 등으로 기업의 응답부담이 가중되고, 응답기피, 조사단가 상승에 따른 조사비용 증가 역시 조사환경을 악화시키는 주요 원인이 되었다. 또한 국제적 추세를 보더라도 UN, EU 등 국제기구 역시 응답부담 감소를 위해 행정자료 활용을 적극 권장하는 등 이러한 행정자료를 이용한 통계생산은 세계적인 트렌드였다.

현재까지 4회차 발간한 문화체육관광산업통계는 직접 조사를 수행하지 않고 행정자료를 활용해 통계를 생산함으로써, 예산에 대한 절감 효과는 물론, 문화·체육·관광 산업의 특성을 종합적으로 파악한 통계를 생산하였고, 이를 기초자료로 관련 산업의 동향과 성장세를 가늠하는 지표로 활용하고 있다.

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

제4장

현황 분석

제1절 분석 개요

조사통계는 필연적으로 불확실성, 즉 오차(error)를 내재한다. 이는 일반적으로 표본 오차(sampling error)와 비표본오차(non-sampling error)로 분류되는데(Mazzi et al., 2021), 이러한 표본오차와 비표본오차를 모두 설명하는 것을 총조사오차(Total Survey Error; TSE)라 한다. 즉 조사 데이터의 설계, 수집, 처리 및 분석의 과정에서 발생할 수 있는 모든 오차의 합인 총조사오차를 최소화하여 데이터의 정확성을 향상시킬 수 있는 것이다(Biemer et al., 2014). 그러나 총조사오차의 중요성에도 불구하고 과거 이에 대한 연구가 미흡했는데, 이는 복잡하고 비용이 많이 들며 분석하기 어려울 뿐더러 정교한 통계 모델을 필요로 하기 때문이다.⁴⁸⁾ 따라서 Biemer(2010)는 설문조사에서 직면하게 될 문제점에 대해 조사과정자료(Paradata)의 혁신적인 사용을 강조하고 있다.

한편, 2020년 2월부터 시작된 코로나19의 확산은 마스크 착용과 함께 ‘사회적 거리 두기’를 시행하는 등 일상생활에서 많은 변화가 있어왔다. 이는 대면조사가 많은 비중을 차지하고 있는 국가승인통계에도 영향을 미치게 되었고, 많은 조사가 일시적으로 중단되는 현상이 있었다. 또한 코로나19가 조기 종식되지 않고 장기화의 기미를 보이자, 통계청뿐만 아니라 많은 연구, 조사기관 등은 대체 조사방법을 찾아 진행하려는 노력을 하였다. 통계청의 경우는 「인구총조사」를 기반으로 10% 조사구를 통계작성 기반으로 관리하고 있는데, 관리하는 대상들의 정보를 이용하여 컨택을 통해 전화조사 또는 온라인 조사로 대체하여 조사하였다. 이렇게 직접 조사대상을 관리할 수 있는 방안이 있다면, 면접 조사 방식이 어려운 경우 안내를 통해 다른 조사방식으로 진행할 수 있다. 이는 표본틀이 동일하기 때문에, 조사방식은 달라도 모집단의 포괄성(coverage)이 동일하다. 따라서 통계작성 목적이 동일한 조사에서 다른 조사방식들을 혼합한 조사가 가능하다고 할 수 있다.

48) Smith(1990)와 Platek et al.(2001)은 불확실성의 측정 중 표본오차와 비표본오차를 통합하는 데 지난 50년간 진전이 없었다는 점에 주목하고 있음

이에 4장에서는 ①조사과정자료 분석, ②조사방식에 따른 차이 분석, ③혼합모드 분석 각각에 대해 실제 조사되고 있는 문화·관광 통계들을 바탕으로 현 작성 통계들에 적용한 현황분석을 제시하고자 한다.

〈표 4-1〉 현황 분석의 틀

구분	목적
현황분석 Ⅰ) 조사과정자료 분석	조사과정 중에 수집되는 자료를 통한 조사수행 단계에서의 문제점 파악 및 개선방안 도출
현황분석 Ⅱ) 조사방식에 따른 차이 분석	조사방식에 따라 결과값의 차이가 발생하는지 여부 검토
현황분석 Ⅲ) 혼합모드 분석	혼합조사에서 모드효과가 어떻게 발생하는지 파악

제2절 조사과정자료 분석

1. 국민여행조사

국민여행조사의 경우, 2020년도 조사는 코로나19로 일정기간 동안 조사를 멈췄다가 진행할 수밖에 없는 상황이었다. 국민여행조사는 과거에 경험한 여행을 토대로 그 행태에 대한 내용을 조사하기 때문에, 조사가 늦어지면 화상오차가 크게 발생하기 쉽고, 특히 여행횟수와 여행비용 등의 응답을 제대로 하는데 어려움이 있다. 이러한 이유로 2018년도부터 월별조사로 조사주기와 조사내용을 변경해서 조사를 진행하고 있다. 따라서 코로나19가 장기화됨에 따라 대체조사방안에 대한 필요성이 제기되었다.

국민여행조사의 표본을 추출하기 위해 제공받는 정보는 MDIS의 조사구 데이터이다. 이도 사전 승인 후 권한을 부여받고 원격시스템으로 접속하여 표본설계 관련한 정보를 분석하고, 표본을 추출하여, 추출된 데이터를 반출하여 제공받고 있다. 따라서 면접조사를 하지 않을 경우, 표본으로 추출된 가구와 접촉할 수 있는 전화 등의 정보가 없어 활용이 어렵다.

이런 이유로 조사업체에서 보유하고 있는 온라인 패널(panel)을 이용하여 조사를 일부 진행하는 방안을 고려하였다. 그러나 온라인 패널이 우리나라 국민을 대표할 수 있도록 분포되었는지, 즉 포괄성이 충족되는지를 판단할 필요가 있기 때문에, 온라인 패널의 활용성 여부를 판단할 필요가 있다.

온라인 패널을 이용한 국민여행조사 시범조사의 개요는 다음과 같다.

- 패널 수: 약 80만명 / 적극적인 조사참여자: 약 12만명(15%)
- 조사방법: 표본설계에 따른 조사대상에게만 온라인 컨택을 하여 조사 수행
- 표본설계 방법: 80만 패널을 우리나라 국민으로 추정할 수 있도록 성별, 연령별, 지역별로 제공근 비례배분하여 추출

- 기간: 1개월(4월1일~4월30일)
- 응답사레비: 1만원 현금지급

표본은 1주일간 4회를 컨택하고, 응답이 없을 경우 다른 대체표본을 발송하는 방식으로 진행하였다. 단, 원표본은 2주간 유지하였으며, 이후 대체표본은 1주일간 유지하였다. 그러나 응답률이 저조하여 전화 컨택원을 교육한 후 독려하였음에도 응답률은 6.83%에 불과하였다. 따라서 국민여행조사의 규모를 판단할 때 적절하지 않다고 판단하였다.

전체 패널 80만명 중 최근에 접속한 대상자인 적극적인 패널(active panel)의 응답률은 25.52%였지만, 최근 접속하지 않은 패널의 응답률은 3.03%에 불과하였다. 이러한 결과를 볼 때 패널수는 12만명으로 고려하는 것이 적절하다고 판단하였다. 따라서 5만명을 조사하는 국민여행조사를 감당하기에는 패널 수가 부족하다고 판단하였다. <표 4-2>는 국민여행 온라인패널 시범조사 진행현황을 제시한 것이다. 이를 보면 6개월 이내 접속한 적이 없는 비접속자들이 95%에 해당하여 이들은 패널로서의 역할을 한다고 보기 어려웠다.

〈표 4-2〉 국민여행 온라인패널 시범조사 조사진행 현황

	구분		계	조사완료	중도탈락	비접속	조사 완료 비율
전체	최근 6개월	접속자	1,352	345	188	819	25.52%
		비접속자	6,644	201	133	6,310	3.03%
	전체		7,996	546	321	7,129	6.83%
원표본	최근 6개월	접속자	340	94	56	190	27.65%
		비접속자	1,660	80	46	1,534	4.82%
대체표본1	최근 6개월	접속자	335	85	51	199	25.37%
		비접속자	1,665	47	38	1,580	2.82%
대체표본2	최근 6개월	접속자	321	87	39	195	27.10%
		비접속자	1,677	41	27	1,609	2.44%
대체표본3	최근 6개월	접속자	356	79	42	235	22.19%
		비접속자	1,642	33	22	1,587	2.01%

모든 온라인패널에게 조사의 취지를 설명하고, 응답대상으로 뽑혔을 경우 우리나라 관광산업과 정책에 도움이 될 것이라는 홍보를 하였다면 응답률과 조사결과의 편의(bias)는 많이 줄어들 수 있을 것이라 생각된다. 결론적으로 온라인 패널은 국민 전체를

포괄한다고 판단하기 어려운 부분이 있다. 따라서 이러한 부분에 대한 보완이 필요하며, 온라인 패널을 활용할 경우 사전에 홍보방안과 더불어 패널을 관리할 수 있는 방안 마련이 필요하다고 할 수 있다.

2. 국민여가활동조사 및 국민문화예술활동조사⁴⁹⁾

가. 조사원 선발

국민여가활동조사(국민문화예술활동조사)는 전국의 만15세 이상을 대상으로 한다. 따라서 조사업체는 전국 단위의 조사를 원활히 수행 할 수 있도록 지방마다 실사조직을 갖추고 슈퍼바이저가 상주하며 조사원을 관리한다. 조사원에게 업무량이 과다하게 배정되거나 혹은 자의에 의해 과다한 조사를 소화하려고 할 경우, 조사원에 의한 부정적 개입이 발생하여 비표본오차가 발생할 여지가 커지게 된다. 따라서 1일 업무량 기준으로 조사원 1인당 성공(조사완료)기준을 설정하고 이를 고려하여 적절하게 할당하여 관리를 하였다.⁵⁰⁾ 마찬가지로 현장조사를 총괄하여 관리하는 슈퍼바이저 역시, 관리자 1인당 조사원 수가 과다하게 배정되지 않도록 조사원이 많은 서울지역은 3명을 배정하여 관리하도록 진행하였다.

〈표 4-3〉 현장조사 관리자 업무량

지역 실사 사무소	투입 조사원 수 (A)	현장조사 관리자 수 (B)	관리자 1인당 조사원 수 (A/B)
전체	160명	9명	18명
수도권	72명	3명	24명
부산	23명	1명	23명
대구	14명	1명	14명
대전	22명	1명	22명
광주	18명	1명	18명
강원	6명	1명	6명
제주	5명	1명	5명

49) 두 조사는 표본설계가 동일하며, 동일 조사기관에서 수행하여 함께 기술함

50) 조사원들은 프리랜서의 개념으로 일하고 있기 때문에, 하나의 조사를 전문적으로 하도록 관리를 한다고 해도 다른 업체의 다른 조사를 병행하는 경우가 상당히 있음. 또한 실제 현장에서 조사를 완료하기 위해 많은 컨택과 재방문이 반복이 되기 때문에 이러한 다양한 요소들을 고려하여 1일 업무량을 산정함

〈표 4-4〉 조사원 업무량 배정 결과

지역	조사원	총 할당 부수	1인당 평균 부수	1일 업무량(working day 50일 기준)
전체	160명	10,000부	63부	1.3부
수도권	72명	4,840부	67부	1.3부
부산	23명	1,560부	68부	1.4부
대구	14명	1,040부	74부	1.5부
대전	22명	1,110부	50부	1.0부
광주	18명	1,020부	56부	1.1부
강원	6명	310부	51부	1.0부
제주	5명	120부	24부	0.5부

〈표 4-4〉는 국민여가활동조사(국민문화예술활동조사)의 조사원의 업무량을 제시한 표이다. 너무 많은 조사를 수행함에 따른 조사품질 저하의 우려와는 달리, 1일당 1.3부의 조사가 완료됨을 보여 1일 조사량이 너무 작은 것으로 나타났다. 특히 제주의 경우 0.5부로 1일에 1부도 완료하지 못한 것으로 나타났다.

코로나19로 인해 조사의 진행이 어려워, 조사홍보와 유치조사⁵¹⁾ 방법 등을 같이 활용하여 조사하였음을 감안하여도 매우 느리게 조사가 진행됨을 알 수 있다. 이는 국민여가활동조사(국민문화예술활동조사)의 조사에 집중하지 않았다고 생각할 수도 있는 부분이다. 따라서 향후 조사원을 선발해서 조사원교육을 시킬 경우, 조사원에게 조사의 집중도를 높일 수 있도록 하고, 조사관리자(survey supervisor)에게는 조사관리에 대한 책임을 주지시킬 필요가 있다.

또 다른 한편으로, 조사성공률이 얼마나 낮은가를 짐작하게 해준다. 때문에 앞서 언급한 행정자료의 활용을 통해 응답자들의 응답 피로도를 낮추고, 조사방식의 다변화를 통해 응답의 편의성을 높이며, 적절한 보상(현재는 약 3,000~4,000원 상당의 답례품을 제공)을 통한 응답유도가 필요함을 시사한다.

나. 조사구 대체

국민여가활동조사(국민문화예술활동조사)는 1차 추출단위로 조사구를 추출하고, 2차

51) 유치조사는 조사대상에게 설문지 등을 남겨두고 재방문하여 회수하거나, 언제 방문하겠다는 의사를 남기고 다시 방문하여 조사하는 등의 방법을 의미함

추출단위로 가구를 추출한다. 추출된 가구에서는 생일자법을 활용해 조사대상을 선정해 1명의 가구원을 조사하게 된다. 따라서 표본 대체는 3가지 종류가 있을 수 있다. 조사구 대체의 경우 해당 조사구가 재개발 지역이어서 조사구 전체가 공사현장 이라던가 혹은 아파트 단지의 경우 경비실에서 조사원의 출입 자체를 거부하여 부득이 조사가 불가능한 경우 등 다양한 원인이 있을 수 있다. 하지만 이러한 특수한 경우를 제외한 경우에, 조사구 대체가 이루어졌다는 것은 해당 조사구내에서 10가구의 조사를 완료하지 못해 추가적인 대체 조사구에서 조사를 이어가야 하는 경우이다. 결국 조사구 대체는 실제 최종 조사단위인 가구원에 대한 대체 비율에 비해 낮게 나타날 수밖에 없으므로 이를 견지하여 살펴볼 필요가 있다.

〈표 4-5〉 조사구 대체 현황(국민여가활동조사)

지역	전체				
	원 조사구 (개)	대체 조사구 전체(개)	대체 조사구 -아파트(개)	대체 조사구 -주택(개)	대체율 (%)
전국	1,000	82	35	47	8.2
서울특별시	192	10	5	5	5.2
부산광역시	68	6	4	2	8.8
대구광역시	49	0	0	0	0.0
인천광역시	54	2	1	1	3.7
광주광역시	29	8	3	5	27.6
대전광역시	30	6	0	6	20.0
울산광역시	22	0	0	0	0.0
세종특별자치시	6	0	0	0	0.0
경기도	238	19	9	10	8.0
강원도	31	5	1	4	16.1
충청북도	32	3	1	2	9.4
충청남도	43	11	5	6	25.6
전라북도	37	1	1	0	2.7
전라남도	36	2	2	0	5.6
경상북도	55	4	2	2	7.3
경상남도	66	5	1	4	7.6
제주특별자치도	12	0	0	0	0.0

〈표 4-5〉는 국민여가활동조사의 조사구 대체 현황을 제시한 것인데, 대체율이 예년에 비해 높게 나타났다.⁵²⁾ 특히 광주광역시(27.6%), 충청남도(25.6%), 대전(20.0%)로 20% 이상 대체된 것으로 나타나, 조사가 원활하게 진행되지 않았다는 것을 알 수 있다.

〈표 4-6〉 조사구 대체 현황(국민문화예술활동조사)

지역	전체				대체율 (%)
	원 조사구 (개)	대체 조사구 전체(개)	대체 조사구 -아파트(개)	대체 조사구 -주택(개)	
전국	1,000	102	48	54	10.2
서울특별시	192	19	11	8	9.9
부산광역시	68	13	5	8	19.1
대구광역시	49	1	1	0	2.0
인천광역시	54	3	1	2	5.6
광주광역시	29	3	0	3	10.3
대전광역시	30	7	2	5	23.3
울산광역시	22	5	3	2	22.7
세종특별자치시	6	0	0	0	0.0
경기도	238	14	7	7	5.9
강원도	31	9	4	5	29.0
충청북도	32	1	1	0	3.1
충청남도	43	11	5	6	25.6
전라북도	37	1	1	0	2.7
전라남도	36	1	0	1	2.8
경상북도	55	10	5	5	18.2
경상남도	66	4	2	2	6.1
제주특별자치도	12	0	0	0	0.0

이렇듯 조사구가 대체가 많이 이뤄졌다는 것은, 가구대체가 매우 크게 이뤄졌다는 것을 의미한다. 조사기간 중에 코로나19로 인해 거리두기가 강화된 것으로 인해 조사거절 등이 특히 많이 발생되었다. 대체가 많이 발생한 것은 결국 비표본오차가 많이 발생되었을 가능성이 있다.

한편 조사구 대체가 필요한 경우 슈퍼바이저에게 보고하고, 이를 기록하도록 매뉴얼

52) 일부 조사구의 경우 가구수가 적은경우가 있으며, 지역적으로 성별, 연령대의 비율에 차이가 있어 할당된 조사대상을 찾지 못해 조사구가 일부 대체가 되기는 하나, 조사구내에서 조사가 완료되는 경우가 대부분임

화 되어있다. 이 기록을 살펴보면 조사구 대체 사유의 대부분은 ‘조사 거절’(70% 내외)이다. 다음으로 ‘노년층 거주 비중이 높음’(13% 내외)인데, 이는 표본배분 시 국민에 대한 성별*연령대별 비율을 고려하여 쿼터를 할당한 것이 하나의 요인으로 작용한 것으로 판단된다. 특별히 노년층 거주 비중이 높아 대체비율이 높아졌다는 것은 젊은층을 조사하기가 어려웠다는 점을 반증한다. 노년층의 조사 쿼터는 찻으나 젊은층의 조사 완료율이 미흡하여, 젊은 조사대상자를 찾기 위해 조사구를 대체한 것으로 보인다. 조사구를 대체한 다양한 요인이 있었겠지만, 코로나19로 인해 조사거절이 크게 발생된 부분이 주된 사유라고 할 수 있다. <표 4-7>은 조사구 대체 사유를 제시한 것이다.

〈표 4-7〉 조사구 대체 사유

조사 불가능 사유	국민여가활동조사		국민문화예술행동조사	
	개수(개)	비율(%)	개수(개)	비율(%)
조사 거절	62	75.6	68	66.7
부재가구가 많아서 조사가 어려움	2	2.4	3	2.9
노년층 거주 비중이 높음	10	12.2	14	13.7
상가밀집 지역	4	4.9	13	12.7
재개발 지역	4	4.9	4	3.9
총 합계	82	100.0	102	100.0

다. 현장조사 진행 자료

전국단위로 볼 때, 전체 접촉 가구 수는 ‘국민여가활동조사’가 61,046명, ‘국민문화예술행동조사’가 62,319명이었으며, 조사를 완료한 성공 수는 각각 10,088명, 10,099명으로, 접촉 대비 조사완료율은 약 16%대 인 것으로 나타났다. 조사를 위해 조사대상 가구를 방문한 접촉현황을 보면, 두 조사가 모두 ‘부재’, ‘거절’, ‘성공’, ‘기타’, ‘가구없음’ 순으로 나타났다. <표 4-8>과 <표 4-9>는 각 조사별로 가구 접촉현황을 제시한 것이다.

조사원이 표본으로 선정된 가구에 방문했을 때, 조사를 진행하지 못한 이유로 부재인 경우가 가장 많이 나타났다. 이렇듯 가구 방문 시 부재로 인한 조사 실패가 많은 이유는 지역적인 특성(지리적 위치, 도시규모 등)이나 주택유형에 따른 주 연령층(젊은 층이 많이 사는 원룸의 경우 등)을 만나지 못한다는 것을 의미한다. 따라서 이러한 원인을 파악하여 요일이나 시간대별로 방문전략을 세워, 조사대상과의 접촉률을 높일 필요가 있다.

〈표 4-8〉 가구 접촉 현황(국민여가활동조사)

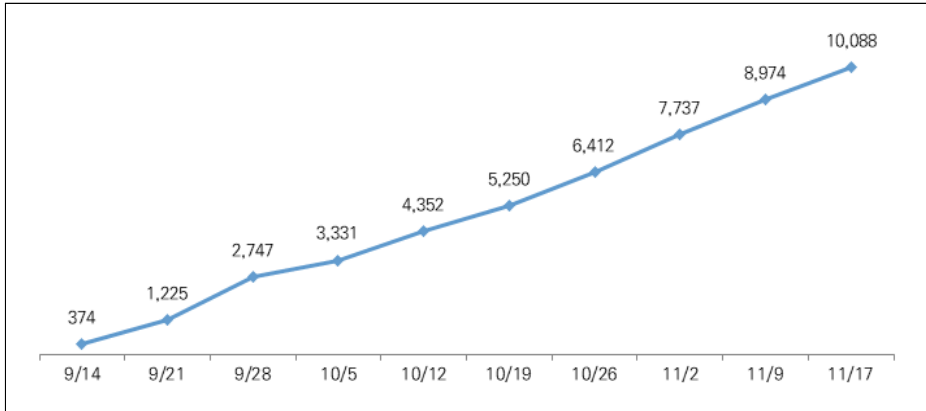
구분	전체접촉 가구 수	성공	거절	부재	가구 없음	기타
합계	61,046	10,088	18,310	29,515	655	2,478
서울	12,107	1,931	3,986	6,164	2	24
부산	3,836	681	735	1,756	0	0
대구	2,983	490	854	1,680	96	714
인천	3,036	556	1,120	1,353	2	5
대전	1,881	303	660	851	67	0
광주	1,845	294	449	952	71	79
울산	1,152	221	295	636	0	0
세종	385	61	125	187	12	0
경기	14,598	2,408	4,909	7,214	42	25
강원	1,660	320	439	879	22	0
충북	1,903	321	685	822	75	0
충남	2,324	430	792	1,011	92	0
전북	2,412	375	645	1,021	50	321
전남	2,173	360	416	1,031	80	286
경북	3,857	557	756	1,510	10	1,024
경남	3,370	660	822	1,884	4	0
제주	1,336	120	622	564	30	0

〈표 4-9〉 가구 접촉 현황(국민문화예술활동조사)

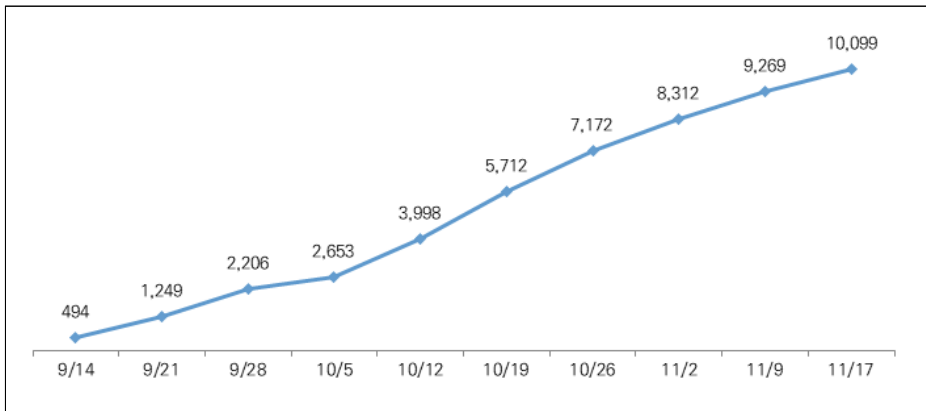
구분	전체접촉 가구 수	성공	거절	부재	가구 없음	기타
합계	62,319	10,099	18,023	31,844	438	1,915
서울	10,665	1,928	4,102	4,586	44	5
부산	5,268	686	1,014	3,568	0	0
대구	4,345	491	1,017	2,347	36	454
인천	2,633	549	1,018	1,066	0	0
대전	1,520	303	339	675	157	46
광주	1,797	293	371	954	7	172
울산	1,747	222	336	1,189	0	0
세종	417	62	71	284	0	0
경기	13,405	2,412	5,186	5,768	33	6
강원	1,943	314	721	888	11	9
충북	1,905	329	296	1,204	36	40
충남	2,123	436	435	969	55	228
전북	2,128	376	439	1,090	17	206
전남	1,997	361	335	983	31	287
경북	4,495	549	1,067	2,409	8	462
경남	5,209	667	985	3,554	3	0
제주	722	121	291	310	0	0

[그림 4-1]과 [그림 4-2]는 ‘국민여가활동조사’와 ‘국민문화예술활동조사’의 주간단위 별 조사 진행률을 제시한 것이다. 조사진행률은 일정하게 진행된 것으로 나타났다. 이는 예년대비 조사진행이 느리게 진행된 것이며⁵³⁾, 조사여건이 어려웠지만, 꾸준히 조사는 진행된 것을 알 수 있다.

[그림 4-1] 주간 단위 실사 진행률(국민여가활동조사)



[그림 4-2] 주간 단위 실사 진행률(국민문화예술활동조사)



〈표 4-10〉과 〈표 4-11〉은 두 조사의 최종 조사 완료된 표본수를 제시한 것이다. 지역별로 목표표본수가 완료된 것을 알 수 있다. 일부지역은 쿼터별 할당을 채우기 위해 목표대비 추가조사된 것으로 나타났다.

53) 보통 추석주간을 포함해서 45일 가량 조사를 진행하였음

〈표 4-10〉 최종 조사 완료 표본 수 현황(국민여가활동조사)

구분	목표 표본 수	완료 표본 수	목표 대비 완료 비율(%)
전국	10,000	10,088	100.9
서울특별시	1,920	1,931	100.6
부산광역시	680	681	100.1
대구광역시	490	490	100.0
인천광역시	540	556	102.9
광주광역시	290	294	101.4
대전광역시	300	303	101.0
울산광역시	220	221	100.5
세종특별자치시	60	61	101.6
경기도	2,380	2,408	101.2
강원도	310	320	103.1
충청북도	320	321	100.3
충청남도	430	430	100.0
전라북도	370	375	101.3
전라남도	360	360	100.0
경상북도	550	557	101.3
경상남도	660	660	100.0
제주특별자치도	120	120	100.0

〈표 4-11〉 최종 조사 완료 표본 수 현황(국민문화예술활동조사)

구분	목표 표본 수	완료 표본 수	목표 대비 완료 비율(%)
전국	10,000	10,099	101.0
서울특별시	1,920	1,928	100.4
부산광역시	680	686	100.9
대구광역시	490	491	100.2
인천광역시	540	549	101.7
광주광역시	290	293	101.0
대전광역시	300	303	101.0
울산광역시	220	222	100.9
세종특별자치시	60	62	103.3
경기도	2,380	2,412	101.3
강원도	310	314	101.3
충청북도	320	329	102.8
충청남도	430	436	101.4
전라북도	370	376	101.6
전라남도	360	361	100.3
경상북도	550	549	99.8
경상남도	660	667	101.1
제주특별자치도	120	121	100.8

제3절 조사방식에 따른 차이 분석

1. 조사 방법(mode)별 차이

가. 문화체육관광 기업경기동향조사

1) 조사에 대한 개요

문화체육관광 기업경기동향조사는 문화체육관광 분야에서 영위하는 사업체들의 경영 활동 실적에 대한 전망을 파악하여 지표를 작성하고 있다. 문화체육관광산업 분류는 문화예술, 문화산업, 스포츠, 관광 4개로 분류하며, 중분류 업종은 19개로 구분하고 있다. 이는 <표 4-12>에 제시하였다.

<표 4-12> 조사범위

대분류(산업)	중분류(업종)
문화예술	① 문화유산 및 문화시설, ② 문학 및 출판, ③ 공연예술, ④ 시각예술, ⑤ 공예예술
문화산업	① 출판, ② 음악, ③ 영화 및 방송, ④ 광고, ⑤ 게임, ⑥ 시각그래픽 및 캐릭터
스포츠	① 스포츠시설, ② 스포츠용품, ③ 스포츠서비스
관광	① 관광 숙박 및 음식점업, ② 여행사업 및 운수업, ③ 문화 오락 및 레저산업 ④ 관광 쇼핑업, ⑤ 국제회의 및 전시업

조사내용은 크게 일반현황, 조사항목 사업체 정보 등으로 구분된다. 일반현황은 산업 분야, 종사자규모, 소재지, 조직형태, 사업체 구분 등으로 구성되며, 경기 판단하는 조사 항목은 매출규모, 채산성, 투자규모, 자금사정, 고용사정, 상품단가, 전체 업항으로 구분되며, 이를 각각 현황 및 전망으로 구분하여 질문한다. 여기에다가 호전 및 악화에 대한 이유를 묻고 있다. 이는 <표 4-13>에 제시된 것과 같다.

〈표 4-13〉 조사항목

구분	조사내용
사업체 특성	▶산업분야 ▶종사자 규모 ▶소재지 ▶조직형태 ▶사업체 구분
경기 판단	▶매출규모 현황 및 전망 ▶투자규모 현황 및 전망 ▶고용사정 현황 및 전망 ▶채산성 현황 및 전망 ▶자금사정 현황 및 전망 ▶상품단가 현황 및 전망 ▶전체 업황(대상업체의 경기) 현황 및 전망 (호전 및 악화 이유)

조사방법은 전화조사를 원칙으로 하고 있으며, 요청이 있을 경우 이메일 또는 팩스로 진행하고 있다. 조사대상은 문화체육관광 분야에 영위하는 5인 이상의 사업체를 대상으로 하고 있다. 표본은 2,200개 사업체이며, 분기별로 조사를 진행하고 작성하고 있다.

2) 전화조사와 온라인조사의 모드별 차이 분석

‘문화체육관광 기업경기동향조사’는 주로 전화조사로 진행되어 왔으며, 최근에 사업체들의 응답률이 저조해지고 있어 대체표본을 이용한 조사비율이 다소 증가하였다. 따라서 응답률을 높일 수 있는 방안을 마련하기 위해 기존의 전화조사와 새로운 방법인 온라인조사를 각각 수행하여 결과를 비교하였다.

기존의 조사방법에서 일부의 표본을 온라인 조사로 진행하였다. 전화조사와 온라인조사의 비율은 3:1로 하여 추출하였다. 따라서 산업별로 표본의 차이가 있으며, 조사방법 별로도 차이가 발생한다.⁵⁴⁾

〈표 4-14〉는 ‘문화체육관광 기업경기동향조사’의 전체업황에 대한 2020년 1분기 현황을 제시한 값이다. 전체표본에서 전화조사 표본의 비중(약 75%)이 크기 때문에, 전체결과와 전화조사 결과의 차이는 적으며, 온라인조사의 결과는 다소 차이가 난다. 그러나 전화조사와 온라인조사의 통계적 차이를 검정하기 위해 독립표본 t-검증을 실시한 결과, 차이는 있지만 통계적으로 유의하다고 판단할 수 있는 차이는 아니었다.⁵⁵⁾ 표본조사에

54) 이번 절에서는 공표 단위인 산업별 차이를 제시하고자, 산업별로 값을 제시하였으나 5장에서는 전체적인 관점에서 값을 비교하도록 함. 5장은 조사방법별 차이를 명확하게 제시하기 위해 많은 표본을 활용하는 것이 타당함

55) 업종별로 조사모집단의 차이가 있으며, 전화조사와 온라인조사의 표본수가 약 3:1로 분배하였기 때문에, 관광쇼핑업, 국제회의 및 전시업, 음악산업 등은 표본수가 적게 작성됨. 그러나 독립표본 t-검정은 적은 표본수에도 검정력이 높은 로버스트(robust)한 방법이기 때문에 검정 결과를 신뢰할 수 있음

서의 각 조사방법의 차이가 있는지를 파악하기 위한 것이기 때문에, 독립표본 t-검정은 표본값을 이용하였다.

〈표 4-14〉 2021년 1분기 기업경기 실적(BSI)의 조사방법별 비교

업종	전체		전화조사		온라인조사		t-값 (유의확률)
	평균	표준편차	평균	표준편차	평균	표준편차	
문화유산 및 문화시설	101.33	18.73	102.07	19.91	99.12	14.32	0.131 (0.896)
문학 및 출판	74.37	41.15	77.59	37.36	64.66	49.65	1.597 (0.113)
공연	82.69	42.00	84.86	42.28	74.81	39.64	0.873 (0.385)
시각예술	76.04	40.97	73.96	41.42	80.85	39.16	-0.769 (0.443)
공예	79.94	42.30	77.98	42.20	85.61	41.83	-0.903 (0.368)
출판산업	87.03	36.02	88.47	36.76	81.59	30.72	1.141 (0.256)
음악산업	85.51	52.60	128.70	39.96	75.00	53.41	0.568 (0.583)
영화산업 방송산업	88.02	31.07	88.30	28.29	86.04	40.95	0.202 (0.841)
광고산업	73.35	40.50	75.65	40.69	67.77	39.71	0.753 (0.453)
게임산업	87.03	39.58	89.24	41.13	79.74	38.22	0.743 (0.461)
시각그래픽아 트 및 캐릭터	82.92	43.12	80.27	38.41	90.77	55.51	-0.543 (0.591)
스포츠 시설업	83.49	41.76	82.78	44.39	86.09	31.57	0.529 (0.598)
스포츠 용품업	90.14	41.84	93.03	40.94	81.66	42.64	0.974 (0.332)
스포츠 서비스업	82.31	45.25	84.06	44.15	77.50	48.45	0.263 (0.793)
관광숙박업	75.09	42.89	73.55	42.86	79.05	43.21	0.184 (0.854)
관광식당업	74.97	49.54	77.21	47.55	70.67	55.65	1.248 (0.215)
여행사업 및 관광운수업	62.46	45.18	61.57	45.48	65.61	43.44	-0.201 (0.841)
문화오락 및 레저산업	97.86	33.57	96.58	31.88	102.40	39.25	-0.552 (0.582)
관광쇼핑업	91.57	45.28	85.40	45.99	100.32	41.15	-1.486 (0.154)
국제회의 및 전시업	80.17	44.40	80.56	47.93	74.00	34.63	0.358 (0.723)

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-15〉는 2020년 2분기 전망을 제시한 값이다. 전망도 현황도 같이 전체값과 전
화조사의 지수값이 비슷하며, 온라인조사와는 조금씩 차이가 있다. 두 방법의 차이가 있
는지를 비교해 봤을 때, 공예업과 스포츠유통업은 유의수준 0.05에서 차이가 있는 것으
로 나타났으며, 스포츠서비스업과 관광쇼핑업은 유의수준 0.1에서 유의한 차이가 있는
것으로 나타났다. 전망에 대한 두 조사방법을 비교하는 독립표본 t-검정도, 현황과 마찬
가지로 표본을 이용하여 두 집단을 비교한 것이다.

〈표 4-15〉 2021년 2분기 기업경기 전망(BSI)의 조사방법별 비교

업종	전체		전화조사		온라인조사		t-값 (유의확률)
	평균	표준편차	평균	표준편차	평균	표준편차	
문화유산 및 문화시설	104.87	26.57	106.69	24.24	99.12	32.21	0.877 (0.382)
문학 및 출판	85.53	34.89	87.21	35.16	81.46	33.97	0.305 (0.761)
공연	99.66	39.52	99.02	37.99	102.31	44.79	-0.053 (0.958)
시각예술	94.35	31.46	92.84	28.42	98.28	38.44	-1.134 (0.259)
공예	95.27	34.72	90.75	34.82	106.60	30.61	-2.132 (0.035)*
출판산업	93.53	31.35	94.83	31.36	88.43	29.58	0.982 (0.328)
음악산업	112.03	36.85	96.76	28.40	125.00	25.11	-1.451 (0.177)
영화산업 방송산업	99.92	29.52	100.34	29.62	96.68	31.10	0.885 (0.378)
광고산업	94.11	35.09	96.94	31.31	85.34	43.29	0.833 (0.410)
게임산업	101.64	34.98	103.29	39.44	96.86	24.00	0.733 (0.467)
시각그래픽아트 및 캐릭터	98.18	33.91	96.30	29.54	102.31	43.89	0.161 (0.873)
스포츠 시설업	100.10	29.58	100.65	30.82	98.28	25.15	0.542 (0.589)
스포츠 용품업	102.51	38.15	108.37	32.27	85.01	47.62	2.073 (0.044)*
스포츠 서비스업	86.69	38.48	91.97	38.40	70.23	34.55	1.863 (0.065)
관광숙박업	100.98	39.31	103.07	41.96	93.56	29.94	1.089 (0.278)
관광식당업	94.57	42.21	96.36	40.04	95.05	43.16	1.031 (0.305)
여행사업 및 관광운수업	79.84	43.64	79.77	42.79	80.27	46.51	0.069 (0.945)
문화오락 및 레저산업	113.61	29.98	112.73	28.32	116.81	35.29	-0.279 (0.782)
관광쇼핑업	112.14	40.75	103.90	44.76	133.42	23.68	-1.980 (0.062)
국제회의 및 전시업	110.24	28.47	114.01	22.53	100.00	33.78	1.462 (0.154)

* p<0.05, ** p<0.01, *** p<0.001

나. 국민여행조사

1) 조사에 대한 개요

국민여행조사는 우리나라 국민의 여행실태를 종합적으로 파악하여 국가 관광에 관한 정책수립과 연구 및 분석 등을 위한 기초 자료를 제공하고, 궁극적으로 국민의 복지 및 삶의 질을 제고하는 목적으로 작성되고 있다.

조사내용은 크게 여행 총량, 여행 행태, 여행 평가 등으로 구분되며, 이에 대한 세부 지표 항목은 <표 4-16>에 제시된 것과 같다.

<표 4-16> 조사항목

조사 영역	세부 항목
여행 총량	▶여행 경험률 ▶여행 횟수 ▶여행 일수 ▶여행 지출액
여행 행태	▶여행 시기 ▶여행 방문지 ▶여행 방문지 선택 이유 ▶여행 정보 획득 경로 ▶참고한 인터넷 사이트 ▶여행사 상품 구매 여부 및 시기 ▶사전예약 여부 및 유형 ▶여행 동반자 수 및 유형 ▶여행지에서의 활동 ▶주요 이동 수단 ▶주요 숙박시설 ▶여행관련 항목별 지출 비중
여행 평가	▶전반적 만족도 ▶재방문 의향 ▶타인 추천 의향
여행 제약 요인	▶여행을 다녀오지 않은 이유

조사방법은 가구방문을 통한 1:1면접조사이며 구조화된 설문지를 태블릿PC에 탑재하여 활용하고 있다. 특히, 여행에 대한 회상오차를 줄이고 공표주기를 단축하여 시의성 있는 자료제공을 하기 위해서 2018년부터는 기존 연간조사에서 월 조사로 변경하여 진행하고 있다.

조사대상은 국내에 거주하는 만 15세 이상 국민이며, 통계청의 '인구주택총조사'의 조사구 자료를 이용하여⁵⁶⁾ 매월 4,200가구를(해당 가구의 가구원 1명을 조사) 조사하고 있다. 따라서 연간 50,400명이 조사되고 있어, 신뢰성 있는 통계를 산출하고 있다고 할 수 있다.

56) 매년 최근의 자료를 활용하는데, 2021년 조사에서는 2019년 '인구총조사' 자료를 활용하여 조사를 진행함

2) 대면조사와 온라인패널 조사의 모드별 차이 분석

코로나19로 인해 국민여행조사의 가구방문조사를 대체하는 조사방안을 마련하고자 조사업체의 온라인조사를 진행하였다. 이때 조사업체의 온라인패널은 우리나라 국민들 전체에서 대표성 있게 조사된 마스터샘플(master sample)이라는 가정하에서 진행하였다.⁵⁷⁾

〈표 4-17〉과 〈표 4-18〉은 온라인 패널 조사와 면접조사의 결과를 비교한 것이다. 두 결과는 코로나19로 인해 조사시기의 차이는 발생하지만, 조사기준은 2020년 1월로 동일 기간에 국민여행 경험 등에 대한 질문한 조사결과이다.⁵⁸⁾ 여행 경험률에서부터 해외여행 경험률 모두 온라인 패널을 이용한 조사가 높게 나타났다. 특히 숙박여행 경험률과 해외여행 경험률이 높게 나타났는데, 이는 온라인 패널 응답자들이 적극적 패널(active panel) 위주로 응답이 이뤄진 것도 하나의 원인이라고 할 수 있다.⁵⁹⁾

〈표 4-17〉 국민여행 온라인 패널 시범조사 조사결과

		사례수	국내여행 경험률	국내숙박 여행경험률	국내당일 여행경험률	해외여행 경험률
전체		(546)	67.77	46.34	42.67	20.15
성별	남자	(208)	72.12	50.48	43.75	22.60
	여자	(338)	65.09	43.79	42.01	18.64
연령	15~19세	(77)	62.34	38.96	40.26	18.18
	20대	(158)	65.19	44.94	45.57	23.42
	30대	(139)	68.35	51.80	41.73	22.30
	40대	(76)	71.05	50.00	40.79	19.74
	50대	(57)	64.91	42.11	35.09	8.77
	60대	(34)	85.29	47.06	52.94	23.53
	70세이상	(5)	80.00	40.00	60.00	0.00
표본 구분	원표본	(174)	66.67	46.55	39.66	17.82
	대체표본	(372)	68.28	46.24	44.09	21.24

57) 조사업체에서 운영하고 있는 온라인패널은 자발적으로 참여하기 때문에, 선택편의(selection bias)가 있고, 연구자가 설계하여 모집한 것이 아니기 때문에 포괄성을 충족하기 어려움

58) 두 조사는 표본수에서 차이가 있으며, 지역별, 성별, 연령별의 표본의 차이가 있어 절대적인 비교는 어렵지만, 큰 맥락에서만 차이를 살펴보고자 제시함

59) 해외여행 경험률을 보면 얼리어답터(early adapter)들의 성향이 있는 것으로 보임

〈표 4-18〉 국민여행 대면조사 조사결과

		사례수	국내여행 경험률	국내숙박 여행경험률	국내당일 여행경험률	해외여행 경험률
전체		(4,025)	60.77	26.97	38.02	4.41
성별	남자	(2,004)	61.06	26.66	38.59	4.59
	여자	(2,021)	60.47	27.28	37.45	4.23
연령	15~19세	(257)	61.15	24.51	38.51	3.13
	20대	(599)	60.17	29.96	35.58	4.96
	30대	(617)	66.06	32.46	39.12	6.59
	40대	(745)	63.24	31.75	36.78	5.11
	50대	(774)	62.97	27.70	40.35	4.46
	60대	(545)	56.05	20.37	37.95	3.50
	70세이상	(488)	51.82	15.36	37.73	1.13

2. 조사 도구(tool)별 차이

가. 국민여가활동조사

1) 조사에 대한 개요

국민여가활동조사는 다양하게 변화되는 국내 여가환경변화에 따라, 국민의 여가수요에 미치는 활동실태를 분석하여 생활양식의 변화 및 삶의 질 수준을 파악하기 위해 매년 조사를 실시하고 있다.

조사내용은 크게 여가활동 참여실태, 평일 및 휴일 여가활동, 동호회 활동 및 사회성 여가활동, 여가공간, 여가인식 및 만족도 등으로 구분된다. 이에 대한 세부 지표 항목은 〈표 4-19〉에 제시된 것과 같다.

〈표 4-19〉 조사항목

조사 영역	세부 항목
여가활동 참여실태	▶참여한 여가활동 ▶여가활동 목적 ▶가장 만족스런 여가활동 ▶여가 경력 ▶여가비용 및 희망 여가비용
평일 및 휴일 여가활동	▶평일/휴일 여가시간 ▶평일/휴일 유형별 여가활동 ▶여가시간 ▶휴가활용 ▶여가비용 및 희망 여가비용
동호회 활동 및 사회성 여가활동	▶동호회 활동 ▶사회성 여가활동 ▶스마트기기 이용시간 ▶스마트기기를 활용한 여가활동
여가공간	▶이용한 여가공간 ▶이용 희망 여가공간 ▶여가공간 만족도
여가인식 및 만족도	▶정책 중요도 ▶일과 여가생활 간 균형 ▶향후 여가시간 활용 ▶여가생활 만족도 ▶여가인식 ▶행복수준

조사방법은 가구방문을 통한 1:1면접조사이며 구조화된 종이 설문지를 활용하고 있다. 다만, 2018년에는 조사원에 의한 비표본오차를 줄이고 데이터편집의 소요시간과 오입력으로 인한 데이터의 오염을 방지하기 위해 TAPI를 활용한 조사과정의 전자화를 추진한 바 있다. 기존 종이설문지를 활용한 조사를 기본으로 하되, 일부 TAPI를 활용한 시범적 조사를 진행하였고, 2019년에는 전면 TAPI로 조사를 진행하였다. 그러나 2020년에는 코로나19가 확산되어, 다시 종이설문지로 회귀하였다.⁶⁰⁾ 이에 PAPI와 TAPI의 조사도구별 결과차이는 동일한 기준에서의 분석이 가능한 최신자료인 2018년 자료로 분석하였다.

조사대상은 국내에 거주하는 만 15세 이상 인구이며, 통계청의 ‘2016년 인구총조사’ 조사구 자료를 이용하여 1,000개의 조사구를 추출하고, 해당 조사구에서 10인을 조사하였다(10가구를 방문하여 조사대상으로 할당된 조건에 맞는 개인을 조사함). 이러한 방식으로 매년 전국의 총 10,000여 명을 추출하여 조사를 진행하며, 조사결과를 가지고 통계를 작성하고 있다. 마찬가지로 TAPI 운영을 위한 시범조사는 위의 본조사의 유효표본과는 별도로 동일문항의 추가조사를 진행하였다.

2) PAPI와 TAPI의 조사도구별 결과차이 분석

조사도구에 따른 차이만을 보기 위해 인구특성 비율에 큰 차이가 나지 않도록 쿼터를 관리하여 조사하였다. 그러나 TAPI의 경우 태블릿PC를 이용한다는 점에서, 고연령층이 설문내용을 스크린을 통해 보는데 어려움이 있어 PAPI에 비해 적은 비율이 조사되었다. 응답자 특성별 비율은 <표 4-20>과 같다.

해당 조사의 주요지표 항목에 대해 조사도구별(PAPI 와 TAPI) 차이를 살펴보면, 여가 시간과 여가비용, 휴가 일수, 지속적인 여가활동 비율은 각 결과값에 유의미한 차이를 보이지 않았고, 여가활동 개수, 동호회 활동 경험, 자원 봉사 활동 경험, 스마트기기 활용 시간, 여가생활 만족도 항목의 경우는 유의미한 차이를 보였다. 이들 중 유의미한 차이를 보인 항목에 대한 결과의 비교는 아래 <표 4-21>에서 <표 4-27>에 제시된 것과 같다.

60) 종이설문지의 경우 면접원과 응답자가 설문지를 각각 보면서 진행이 가능하나, TAPI로 진행 시 하나의 도구(태블릿)를 같이 보고 진행되기 때문에 밀접 접촉이 불가피함. 따라서 이에 대한 위험성을 최소화하고자 다시 종이설문지로 조사를 진행함

〈표 4-20〉 응답자 특성 비교

구분		PAPI		TAPI	
		사례 수(명)	비율(%)	사례 수(명)	비율(%)
전체		1,157	100.0	150	100.0
성별	남성	572	49.4	76	50.7
	여성	585	50.6	74	49.3
연령별	20대	200	17.3	20	13.3
	30대	222	19.2	30	20.0
	40대	243	21.0	23	15.3
	50대	227	19.6	33	22.0
	60대	141	12.2	41	27.3
	70세 이상	124	10.7	3	2.0

〈표 4-21〉 지난 1년간 참여한 여가활동 개수 비교

구분		PAPI		TAPI		t-값 (유의확률)
		사례 수(명)	여가활동 개수	사례 수(명)	여가활동 개수	
전체		1,157	18.8	150	13.7	8.155 (0.000)***
성별	남성	572	18.5	76	13.3	5.916 (0.000)***
	여성	585	19.2	74	14.1	5.640 (0.000)***
연령별	20대	200	21.4	20	14.3	4.496 (0.000)***
	30대	222	20.7	30	15.3	3.913 (0.000)***
	40대	243	19.1	23	17.2	1.106 (0.279)
	50대	227	18.2	33	12.7	5.165 (0.000)***
	60대	141	16.9	41	11.3	4.848 (0.000)***
	70세 이상	124	13.9	3	9.7	4.160 (0.018)*

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-22〉 동호회 활동 참여 경험 여부 비교

구분		PAPI		TAPI		χ^2 (유의확률)
		사례 수(명)	참여비율(%)	사례 수(명)	참여비율(%)	
전체		1,157	15.1	150	6.7	7.819 (0.005)**
성별	남성	572	19.2	76	9.2	4.553 (0.033)*
	여성	585	11.1	74	4.1	3.535 (0.060)
연령별	20대	200	16.5	20	5.0	1.840 (0.175)
	30대	222	15.3	30	10.0	0.596 (0.440)
	40대	243	14.4	23	4.3	1.815 (0.178)
	50대	227	19.8	33	6.1	3.685 (0.055)
	60대	141	9.2	41	7.3	0.143 (0.705)
	70세 이상	124	12.1	3	-	0.412 (0.521)

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-23〉 자원봉사 활동 참여 경험 여부 비교

구분		PAPI		TAPI		χ^2 (유의확률)
		사례 수(명)	참여비율(%)	사례 수(명)	참여비율(%)	
전체		1,157	8.6	150	1.3	9.716 (0.002)**
성별	남성	572	4.9	76	-	3.888 (0.049)*
	여성	585	12.1	74	2.7	5.935 (0.015)*
연령별	20대	200	9.0	20	-	1.960 (0.161)
	30대	222	5.9	30	-	1.852 (0.174)
	40대	243	6.6	23	4.3	0.176 (0.675)
	50대	227	11.5	33	3.0	2.197 (0.138)
	60대	141	13.5	41	-	6.169 (0.013)*
	70세 이상	124	5.6	3	-	0.179 (0.672)

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-24〉 스마트기기 활용 시간_평일 비교

구분		PAPI		TAPI		t(유의확률)
		사례 수(명)	평균 시간	사례 수(명)	평균 시간	
전체		1,157	1.3	150	1.6	-3.612 (0.000)***
성별	남성	572	1.4	76	1.8	-3.029 (0.003)**
	여성	585	1.3	74	1.5	-1.996 (0.049)*
연령별	20대	200	2.1	20	2.0	0.360 (0.723)
	30대	222	1.6	30	2.1	-2.283 (0.028)*
	40대	243	1.4	23	1.9	-2.668 (0.013)*
	50대	227	1.1	33	1.3	-1.787 (0.081)
	60대	141	0.8	41	1.3	-3.058 (0.003)**
	70세 이상	124	0.5	3	1.0	-6.804 (0.000)***

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-25〉 스마트기기 활용 시간_휴일 비교

구분		PAPI		TAPI		t(유의확률)
		사례 수(명)	평균 시간	사례 수(명)	평균 시간	
전체		1,157	1.9	150	2.2	-2.171 (0.030)*
성별	남성	572	1.9	76	2.7	-3.657 (0.000)***
	여성	585	1.9	74	1.7	1.101 (0.273)
연령별	20대	200	2.9	20	2.8	0.308 (0.758)
	30대	222	2.3	30	2.8	-1.605 (0.118)
	40대	243	2.0	23	3.1	-3.731 (0.000)***
	50대	227	1.5	33	1.6	-0.327 (0.745)
	60대	141	1.2	41	1.5	-0.948 (0.345)
	70세 이상	124	0.8	3	1.3	-0.854 (0.478)

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-26〉 일과 여가생활 간 균형 비교

구분		PAPI		TAPI		t(유의확률)
		사례 수(명)	평균 (7점척도)	사례 수(명)	평균 (7점척도)	
전체		1,157	3.83	150	3.57	2.485 (0.013)*
성별	남성	572	3.83	76	3.55	1.800 (0.072)
	여성	585	3.83	74	3.58	1.708 (0.088)
연령별	20대	200	3.66	20	3.75	-0.318 (0.750)
	30대	222	3.63	30	3.97	-1.526 (0.128)
	40대	243	3.67	23	3.65	0.062 (0.951)
	50대	227	3.77	33	3.00	3.128 (0.003)**
	60대	141	4.11	41	3.56	2.364 (0.021)*
	70세 이상	124	4.56	3	4.00	0.947 (0.438)

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-27〉 여가생활에 대한 전반적인 만족도 비교

구분		PAPI		TAPI		t(유의확률)
		사례 수(명)	평균 (7점척도)	사례 수(명)	평균 (7점척도)	
전체		1,157	4.33	150	4.66	-3.629 (0.000)***
성별	남성	572	4.32	76	4.71	-3.172 (0.002)**
	여성	585	4.34	74	4.61	-1.991 (0.050)*
연령별	20대	200	4.41	20	4.75	-1.152 (0.262)
	30대	222	4.23	30	4.67	-1.906 (0.065)
	40대	243	4.24	23	4.78	-2.604 (0.015)*
	50대	227	4.30	33	4.48	-1.142 (0.259)
	60대	141	4.43	41	4.68	-1.378 (0.172)
	70세 이상	124	4.51	3	4.67	-0.179 (0.874)

* p<0.05, ** p<0.01, *** p<0.001

제4절 혼합조사의 모드효과 분석

1. 문화체육관광 소비자동향조사

가. 선택형 혼합조사 모드효과 분석

① 선택 조사방법별 응답성향 및 응답결과 차이분석

전화조사와 모바일조사의 선택성향을 파악하기 위해 선택형 혼합조사 자료의 인구통계학적 변수를 분석하여, 조사방법 선택에 영향이 있었는지 살펴보았다. 그 결과는 <표 4-28>에 제시하였다.

<표 4-28> 소비자동향조사의 선택한 조사방법에 따른 인구통계학적 변수 비교

인구사회학적특성		조사방법		카이제곱 ⁶¹⁾ (유의확률)
		전화조사	모바일조사	
연령대	19-29	6.0	9.8	15.587*** (0.003)
	30-39	13.5	25.5	
	40-49	24.3	26.5	
	50-59	27.8	23.5	
	60세 이상	28.5	14.7	
가구주여부	가구주	79.0	72.5	(0.183)
	가구주아님	21.0	27.5	
결혼여부	미혼	21.8	29.4	(0.115)
	기혼	78.3	70.6	
학력	초등학교 졸업 이하	4.3	0	12.387* (0.015)
	중학교 졸업	4.3	1.0	
	고등학교 졸업	24.3	17.6	
	대학 재학/졸업	55.3	61.8	
	대학원 재학/졸업	12.0	19.6	
자영업여부	예	26.5	15.9	(0.056)
	아니오	73.5	84.1	

인구사회학적특성		조사방법		카이제곱 ⁶¹⁾ (유의확률)
		전화조사	모바일조사	
소득	월 99만원 이하	15.3	6.9	16.722* (0.010)
	월 100만원-199만원 이하	10.3	2.9	
	월 200만원-299만원 이하	14.5	13.7	
	월 300만원-399만원 이하	16.8	15.7	
	월 400만원-499만원 이하	11.0	20.6	
	월 500만원-599만원 이하	11.5	15.7	
	600만원 이상	20.8	24.5	

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

연령대와 학력, 소득은 유의수준 0.05에서 유의하게 차이가 있는 것으로 나타났다. 모바일조사를 선택한 대상의 경우 연령대가 낮을수록 선택한 비율이 높으며, 연령대가 높을 경우, 특히 60대 이상의 경우 상대적으로 낮게 나타났다. 학력의 경우, 대졸이상에 서는 모바일조사, 고졸이하에서는 전화조사의 비율이 높게 나타나, 학력이 높을수록 모바일조사를 선택한 비율이 높은 것으로 나타났다. 소득의 경우 400만원 이상에서는 모바일조사를, 400만원 미만에서는 전화조사를 선택한 비율이 높아, 소득이 높을수록 모바일조사를 선택한 비율이 높게 나타났다.

〈표 4-29〉는 조사항목 16개의 소비자 지출 동향에 대한 항목을 현황과 전망을 구분해 정리한 것이다. 이를 보면 현황과 전망에서 다른 분포를 보이는데, 현황에서는 ‘스포츠용품 및 기구’와 ‘미디어기기 구입비’, ‘국내여행비’, ‘도서구입비’ 4개 항목을 제외한 나머지 12개 항목에서 조사방법별로 유의한 차이가 있는 것으로 나타났다. 전망에서는 ‘오락시설 및 지출비용’과 ‘온라인 문화생활비’, ‘스포츠 경기 관람료’, ‘아웃도어용품 구입비’, ‘해외여행비’, ‘음악미술학원비’, ‘여행관광비’ 7개 항목에서 조사방법별로 유의한 차이가 있고, 9개 항목에서는 차이가 없는 것으로 나타났다.

‘오락시설 및 지출비용’과 ‘스포츠 경기 관람료’, ‘아웃도어용품 구입비’, ‘해외여행비’, ‘음악미술학원비’, ‘여행관광비’는 현황에서는 유의수준 0.001에서 유의하고 전망에서는 유의수준 0.05수준에서 유의한 것으로 나타나, 조사방법을 선택한 대상들의 성향이 현황에서 매우 유의한 차이가 있다고 판단한 경우, 전망에서는 상대적으로 덜 차이

61) 2×2는(가구주여부, 결혼여부, 자영업여부) Fisher’s exact test를 수행하였기 때문에 검정통계량 값이 없고 유의확률 값만 제시함. 이는 2×2 표의 경우 카이제곱 검정보다 Fisher’s exact test의 검정력(power)이 더 좋을 때문에 사용하였음

가 있다고 느끼는 것으로 보인다. 또한 ‘온라인 문화생활비’는 현황에서는 차이가 없는 것으로 나타났는데, 전망에서는 모바일 조사대상자가 향후 더 좋아질 것 이라고 응답한 것으로 나타났다.

〈표 4-29〉 소비자동향조사 선택형 조사방법에 따른 조사항목의 차이분석

구분	항목	전화조사	모바일조사	t-값(유의확률)
		평균	평균	
현황	오프라인 문화생활비	74.50	58.33	3.043(0.003)**
	오락시설 지출비용	81.88	55.88	5.046(0.000)***
	온라인 문화생활비	107.75	108.82	-0.219(0.827)
	스포츠시설 이용료	86.25	75.00	2.135(0.035)*
	스포츠용품 및 기구	96.25	89.22	1.338(0.183)
	스포츠 경기 관람료	88.50	67.65	4.380(0.000)***
	아웃도어용품 구입비	98.13	78.92	3.717(0.000)***
	미디어기기 구입비	108.25	117.16	-1.803(0.074)
	국내여행비	95.13	84.31	1.840(0.068)
	해외여행비	91.00	62.25	5.356(0.000)***
	도서구입비	110.13	108.33	0.421(0.674)
	오락용품 구입비	103.25	91.18	2.525(0.013)**
	음악미술학원비	98.38	82.35	3.703(0.000)***
	스포츠 학원비	96.75	85.29	2.772(0.006)**
	여행관광비	92.21	68.70	4.900(0.000)***
	종합지출	95.69	83.58	3.839(0.000)***
전망	오프라인 문화생활	99.63	94.61	1.043(0.299)
	오락시설 지출비용	98.88	90.20	1.995(0.048)*
	온라인 문화생활비	105.25	113.24	-2.231(0.027)*
	스포츠시설 이용료	101.63	96.08	1.228(0.222)
	스포츠용품 및 기구	103.50	100.00	0.915(0.361)
	스포츠 경기 관람료	100.38	90.20	2.558(0.012)*
	아웃도어용품 구입비	107.63	96.57	2.310(0.023)*
	미디어기기 구입비	102.88	110.29	-1.837(0.068)
	국내여행비	111.00	108.82	0.459(0.647)
	해외여행비	95.63	83.33	2.488(0.014)*
	도서구입비	107.50	111.76	-1.024(0.308)
	오락용품 구입비	102.50	99.02	0.768(0.444)
	음악미술학원비	99.25	91.18	2.044(0.043)*
	스포츠 학원비	99.88	99.02	0.229(0.819)
	여행관광비	100.12	90.79	2.158(0.033)*
	종합지출	101.64	98.06	1.228(0.222)

* p<0.05, ** p<0.01, *** p<0.001

② 선택형 모드효과 분석

모드효과 분석을 위해서는 선택효과와 측정효과를 분리하여야 한다. 그런데 3장에서 살펴본 식 (3.1)과 (3.2)에서 관측불가능하다고 한 부분인 $Y_a = j \mid M = b$ 의 값이 있어야 한다. 본 실험에서는 선택권 없이 조사한 혼합 I 에서 전화조사를 한 집단의 관측결과를 이용하였다. 이론상으로는 모바일조사를 하기로 하고 전화조사를 한 대상을 이용하여야 하지만, 현실적으로 이런 방식의 조사 진행이 어렵기 때문에, 여기서는 선택하지 않고 할당한 혼합 I 에서의 전화조사 대상자들을 이와 동일하다고 가정하고 분석하였다. 이러한 가정을 하면 식 (3.3)의 산출이 가능하다. 모드효과를 분해 한 값들은 표<4-30>에 제시하였다.

〈표 4-30〉 소비자동향조사 항목의 모드효과분해

구분	항목	혼합조사			전화조사 ③	모드효과 분해		
		전화 ①	모바일②	전체		모드 효과 ②-①	선택 효과 ③-①	측정 효과 ②-③
현황	오프라인 문화생활비	74.50	58.33	71.22	74.35	-16.17	-0.15	-16.02
	오락시설 지출비용	81.88	55.88	76.59	81.90	-25.99	0.03	-26.02
	온라인 문화생활비	107.75	108.82	107.97	109.30	1.07	1.55	-0.48
	스포츠시설 이용료	86.25	75.00	83.96	90.95	-11.25	4.70	-15.95
	스포츠용품 및 기구	96.25	89.22	94.82	97.85	-7.03	1.60	-8.63
	스포츠 경기 관람료	88.50	67.65	84.26	88.60	-20.85	0.10	-20.95
	아웃도어용품 구입비	98.13	78.92	94.22	97.15	-19.20	-0.97	-18.23
	미디어기기 구입비	108.25	117.16	110.06	108.45	8.91	0.20	8.71
	국내여행비	95.13	84.31	92.93	93.75	-10.81	-1.38	-9.44
	해외여행비	91.00	62.25	85.16	90.40	-28.75	-0.60	-28.15
	도서구입비	110.13	108.33	109.76	106.80	-1.79	-3.33	1.53
	오락용품 구입비	103.25	91.18	100.80	100.90	-12.07	-2.35	-9.72
	음악미술학원비	98.38	82.35	95.12	99.55	-16.02	1.18	-17.20
	스포츠 학원비	96.75	85.29	94.42	96.20	-11.46	-0.55	-10.91
	여행관광비	92.21	68.70	87.43	91.38	-23.50	-0.83	-22.68
	종합지출	95.69	83.58	93.23	95.84	-12.11	0.15	-12.26
전망	오프라인 문화생활	99.63	94.61	98.61	102.15	-5.02	2.53	-7.54
	오락시설 지출비용	98.88	90.20	97.11	100.30	-8.68	1.43	-10.10
	온라인 문화생활비	105.25	113.24	106.87	104.75	7.99	-0.50	8.49
	스포츠시설 이용료	101.63	96.08	100.50	104.20	-5.55	2.58	-8.12

구분	항목	혼합조사			전화조사 ③	모드효과 분해		
		전화 ①	모바일②	전체		모드 효과 ②-①	선택 효과 ③-①	측정 효과 ②-③
	스포츠용품 및 기구	103.50	100.00	102.79	104.10	-3.50	0.60	-4.10
	스포츠 경기 관람료	100.38	90.20	98.31	100.00	-10.18	-0.38	-9.80
	아웃도어용품 구입비	107.63	96.57	105.38	105.85	-11.06	-1.78	-9.28
	미디어기기 구입비	102.88	110.29	104.38	100.80	7.42	-2.08	9.49
	국내여행비	111.00	108.82	110.56	110.75	-2.18	-0.25	-1.93
	해외여행비	95.63	83.33	93.13	94.90	-12.29	-0.72	-11.57
	도서구입비	107.50	111.76	108.37	106.05	4.26	-1.45	5.71
	오락용품 구입비	102.50	99.02	101.79	101.40	-3.48	-1.10	-2.38
	음악미술학원비	99.25	91.18	97.61	100.40	-8.07	1.15	-9.22
	스포츠 학원비	99.88	99.02	99.70	99.25	-0.86	-0.63	-0.23
	여행관광비	100.12	90.79	98.22	99.53	-9.33	-0.59	-8.75
	종합지출	101.64	98.06	100.91	101.50	-3.57	-0.14	-3.44

* p<0.05, ** p<0.01, *** p<0.001

모드효과는 모바일조사와 전화조사를 한 값의 차이이기 때문에 ②-①로 구할 수 있으며, 이는 <표 4-29>에서 분석하여 많은 항목에서 차이가 있는 것을 확인하였다. <표 4-30>은 모드효과를 선택효과와 측정효과로 분해한 결과인데, 같은 방법으로 조사한 값들의 차이인 선택효과는 ③-①로 구해지며, 측정효과는 ②-③으로 구해진다. ‘종합지출 현황’을 예로 하여 선택형 혼합조사의 모드효과를 살펴보면 모드효과는 -12.11이며, 선택효과와 측정효과를 분리하여 순수한 모드효과는 -12.26이 존재하는 것으로 분석된다. 선택효과가 발생한 주요한 이유는 모바일조사를 선택한 대상자들이 상대적으로 젊은 층의 비중이 높은 것아 발생하는 것으로 생각할 수 있다.

나. 할당형 혼합조사 모드효과 분석

할당된 소비자동향조사의 전화조사와 모바일조사 각각의 응답비율은 <표 4-31>과 같다. 선택형 혼합조사와는 다른 결과가 나타난 것을 알 수 있는데, 연령대와 소득의 응답 분포는 차이가 없는데, 학력과 자영업여부의 비율에는 차이가 있었다.

〈표 4-31〉 소비자동향조사의 할당한 조사방법별 인구통계학적 변수 비교

인구사회학적특성		조사방법		카이제곱 (유의확률)
		전화조사	모바일조사	
연령대	19-29	9.1	8.0	1.683 (0.794)
	30-39	16.6	16.4	
	40-49	24.1	24.6	
	50-59	26.5	29.0	
	60세 이상	23.7	22.0	
가구주여부	가구주	78.7	79.0	(0.947)
	가구주아님	21.3	21.0	
결혼여부	미혼	24.9	27.6	(0.261)
	기혼	75.1	72.4	
학력	초등학교 졸업 이하	4.7	0.4	27.047*** (0.000)
	중학교 졸업	3.4	2.0	
	고등학교 졸업	21.8	24.2	
	대학 재학/졸업	57.4	59.8	
	대학원 재학/졸업	11.9	13.6	
자영업여부	예	31.7	24.9	(0.020)*
	아니오	68.3	75.1	
소득	월 99만원 이하	12.2	8.4	9.425 (0.151)
	월 100만원-199만원 이하	11.7	13.0	
	월 200만원-299만원 이하	14.8	15.8	
	월 300만원-399만원 이하	15.6	18.6	
	월 400만원-499만원 이하	11.7	13.8	
	월 500만원-599만원 이하	11.9	10.4	
	600만원 이상	22.1	20.0	

* p<0.05, ** p<0.01, *** p<0.001

할당형 혼합조사의 경우, RDD 방식으로 응답자를 선정해서, 1,000명은 전화로 조사하고 500명은 문자를 보내서 조사하기 때문에 응답자는 랜덤(random)하게 선정된다. 그러나 선정된 대상이 응답하기 싫은 경우, 다른 대상을 선정해서 조사하기 때문에 모든 인구통계학 변수에서 동일한 비율로 조사되기는 어렵다.

〈표 4-32〉는 16개의 소비자동향조사의 조사항목에 대하여 분석한 것이다. ‘스포츠경기 관람료 현황’에 대한 문항을 제외하고는 모두 전화조사와 모바일조사 두 조사방법에 따라 차이가 있는 것으로 나타났는데, 많은 항목에서 유의수준 0.001에서 유의한 차이를 보이고, 일부는 유의수준 0.01에서 유의한 차이를 보여 매우 다른 결과를 보였다.

〈표 4-32〉 소비자동향조사 할당형 조사방법에 따른 조사항목 차이분석

구분	항목	전화조사	모바일조사	t-값(유의확률)
		평균	평균	
현황	오프라인 문화생활	74.35	60.70	5.166(0.000)***
	오락시설 지출비용	81.90	61.80	3.541(0.000)***
	온라인 문화생활비	109.30	100.10	8.223(0.000)***
	스포츠시설 이용료	90.95	73.00	6.512(0.000)***
	스포츠용품 및 기구	97.85	82.90	3.830(0.000)***
	스포츠 경기 관람료	88.60	69.50	1.587(0.113)
	아웃도어용품 구입비	97.15	84.30	7.487(0.000)***
	미디어기기 구입비	108.45	100.50	5.904(0.000)***
	국내여행비	93.75	76.10	6.310(0.000)***
	해외여행비	90.40	63.30	5.018(0.000)***
	도서구입비	106.80	95.60	8.683(0.000)***
	오락용품 구입비	100.90	86.00	7.379(0.000)***
	음악미술학원비	99.55	80.30	5.466(0.000)***
	스포츠 학원비	96.20	80.30	5.008(0.000)***
	여행관광비	91.38	67.04	3.287(0.001)**
	종합지출	95.84	78.96	3.037(0.002)**
전망	오프라인 문화생활	102.15	94.10	6.830(0.000)***
	오락시설 지출비용	100.30	86.00	4.765(0.000)***
	온라인 문화생활비	104.75	101.60	11.229(0.000)***
	스포츠시설 이용료	104.20	91.30	7.801(0.000)***
	스포츠용품 및 기구	104.10	93.60	4.868(0.000)***
	스포츠 경기 관람료	100.00	84.10	2.796(0.005)**
	아웃도어용품 구입비	105.85	94.90	6.818(0.000)***
	미디어기기 구입비	100.80	94.60	6.148(0.000)***
	국내여행비	110.75	98.90	9.172(0.000)***
	해외여행비	94.90	76.70	7.156(0.000)***
	도서구입비	106.05	100.30	7.841(0.000)***
	오락용품 구입비	101.40	89.00	5.631(0.000)***
	음악미술학원비	100.40	86.40	11.203(0.000)***
	스포츠 학원비	99.25	88.20	7.705(0.000)***
	여행관광비	99.53	83.19	11.014(0.000)***
	종합지출	101.50	89.99	7.458(0.000)***

* p<0.05, ** p<0.01, *** p<0.001

여기서 유의한 차이는 다른 변수들의 영향을 포함하고 있을 수 있다. 예를 들면 전화 조사를 한 대상들과 온라인조사를 한 대상의 연령의 차이가 있다면, 조사방법의 차이에 는 연령에 의한 차이도 내포하고 있게 된다. 따라서 다른 변수들의 차이를 배제하고 순수한 조사방법의 차이를 파악할 필요가 있다.

〈표 4-33〉은 인구통계학적 변수를 통제하고, 순수한 조사방법의 효과를 파악하기 위해 회귀분석을 이용하여 분석한 결과를 제시한 것이다. 독립변수로는 조사방법, 연령대, 가구주여부, 혼인상태, 학력, 자영업여부, 소득, 이렇게 7개의 변수를 사용하였다. 이 중 에서 현재 관심이 있는 조사방법의 효과를 나타내는 회귀계수만을 제시하였으며, 다른 인구통계학적 변수들은 유의한 변수들만을 유의한 통제변수 칸에 제시하였다. 예를 들어 ‘오프라인 문화생활비’의 회귀계수가 12.832로 유의수준 0.001 수준에서 유의하게 나타났는데, 이는 다른 인구통계학적 변수들을 통제하였을 때, 전화조사가 온라인조사보다 12.832점을 더 높게 나타났다는 것을 의미한다.

〈표 4-33〉 소비자자동향조사에서 인구사회학적 변수를 통제한 조사방법 효과분석

항목	현황		전망	
	베타계수 (유의확률)	유의한 통제변수	베타계수 (유의확률)	유의한 통제변수
오프라인 문화생활비	12.832(0.000)***	연령, 자영업여부	8.009(0.000)***	학력
오락시설 지출비용	19.156(0.000)***	연령, 자영업여부	13.657(0.000)***	연령, 학력, 소득, 자영업여부
온라인 문화생활비	8.060(0.000)***	연령, 학력, 소득, 자영업여부	2.973(0.086)	학력, 소득
스포츠시설 이용료	17.414(0.000)***	연령, 소득 자영업여부	12.568(0.000)***	연령, 학력, 소득
스포츠용품 및 가구	14.892(0.000)***	연령, 학력, 소득	10.465(0.000)***	연령, 학력, 소득
스포츠 경기 관람료	17.857(0.000)***	연령,	15.720(0.000)***	연령, 학력 가구주, 소득
아웃도어용품 구입비	12.217(0.000)***	소득	10.793(0.000)***	학력, 소득
미디어기기 구입비	6.977(0.002)**	연령, 소득	5.655(0.003)**	학력, 소득
국내여행비	17.999(0.000)***	소득 자영업여부	11.882(0.000)***	학력, 소득, 자영업여부
해외여행비	26.595(0.000)***	소득	18.009(0.000)***	소득
도서구입비	10.560(0.000)***	학력, 소득	5.855(0.001)**	연령, 학력, 소득

항목	현황		전망	
	베타계수 (유의확률)	유의한 통제변수	베타계수 (유의확률)	유의한 통제변수
오락용품 구입비	13.967(0.000)***	연령, 혼인상태, 학력, 소득	11.476(0.000)***	연령, 학력, 소득
음악미술학원비	18.837(0.000)***	연령, 소득	13.642(0.000)***	연령, 학력, 소득
스포츠 학원비	15.991(0.000)***		10.948(0.000)***	학력, 소득
여행관광비	24.082(0.000)***	소득, 자영업여부	16.218(0.000)***	연령, 소득
종합지출	16.376(0.000)***	학력, 소득, 자영업여부	11.280(0.000)***	학력, 소득, 자영업여부

* p<0.05, ** p<0.01, *** p<0.001

2. 문화체육관광 기업경기동향조사

가. 선택형 혼합조사 모드효과 분석

① 선택 조사방법별 응답성향 및 응답결과 차이분석

전화조사로 조사되어왔던 기업경기동향조사를 웹조사와 병행행하여 조사하도록 하였다. 응답자에게 전화조사와 웹조사 중에서 선택하여 조사를 하도록 하였다. 조사대상이 되는 사업체들의 특성에 따라 차이가 있는지 살펴보았으며, 그 결과는 <표 4-34>에 제시하였다.

<표 4-34>을 살펴보면 사업체특성에 따라 조사방법을 선택한 차이는 없는 것으로 나타났다. <표 4-28>의 소비자동향조사의 결과와 같이 살펴보면, 응답성향은 사업체의 특성별로 발생하지는 않으며, 응답자의 특성에 따라 차이가 발생한다는 것을 유추해 볼 수 있다.⁶²⁾

62) 기업경기동향조사는 사업체를 대상으로 하는 조사이기 때문에 응답자의 성별 이외에는 응답자의 정보를 묻지 않기로 되어 있어, 응답자별 특성을 분석하지 못하는 한계가 있음

〈표 4-34〉 기업경기동향조사의 선택한 조사방법별 사업체특성별 변수 비교

사업체 특성		조사방법		카이제곱 ⁶³⁾ (유의확률)
		전화조사	웹조사	
종사자 규모 ⁶⁴⁾	5~9명	45.2	41.5	1.090 (0.580)
	10~19명	31.4	36.8	
	20~49명	23.4	21.7	
산업구분	문화예술산업	27.0	34.9	2.880 (0.410)
	문화산업	21.0	17.9	
	스포츠산업	24.2	19.8	
	관광산업	27.8	27.4	
조직형태	개인사업체	16.1	17.0	1.849 (0.763)
	회사 법인	68.1	68.9	
	회사 이외 법인	2.1	3.8	
	국가/지방자치단체	10.6	7.5	
	비법인 단체	3.1	2.8	
사업체 구분	단독사업체	59.6	53.2	1.070 (0.586)
	본사(점)	22.6	27.3	
	지사(점), 영업소, 공장	17.8	19.5	
응답자 성별	남성	38.7	34.9	(0.509)
	여성	61.3	65.1	

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-35〉는 조사항목 7개의 사업체의 동향에 대한 항목을 현황과 전망을 구분해 정리한 것이다. 이를 보면 현황에서는 투자규모와 외부고용사정, 상품단가에서 전화조사의 동향조사 결과가 높게 나타났으며, 전망에서는 투자규모와 외부고용사정에서 전화조사의 동향조사 결과가 높게 나타난 것을 알 수 있다.⁶⁵⁾

63) 여기서도 2×2(성별)표는 Fisher's exact test를 수행하였기 때문에 검정통계량 값이 없고 유의확률 값만 제시함.

64) 50인 이상의 기업은 모집단의 개수가 적으며, 기업경기동향조사의 혼합Ⅰ 조사가 본 조사이기 때문에, 혼합Ⅰ 조사를 먼저 진행한 후, 혼합Ⅱ 조사를 진행하였음. 따라서 선택형 조사인 혼합Ⅱ 조사에서는 50인 이상이 사업체 조사가 된 표본이 적기 때문에 50인 미만의 사업체만 대상으로 하여 분석하였음

65) 〈표 5-9〉의 분석에서는 가중값을 사용하지 않았음. 표본의 값에서 두 조사방법의 차이가 있는지를 살펴보고자 하였음

〈표 4-35〉 기업경기동향조사 선택형 조사방법에 따른 조사항목의 차이분석

구분	항목	전화조사	웹조사	t-값(유의확률)
		평균	평균	
현황	매출규모	73.77	77.36	-0.678(0.498)
	투자규모	88.31	77.83	2.435(0.015)*
	외부 고용사정	89.87	80.19	2.634(0.009)**
	채산성	81.69	77.83	0.832(0.406)
	자금사정	75.97	75.94	0.007(0.995)
	상품단가	100.13	90.09	2.780(0.006)**
	전반적 업황	82.08	74.53	1.648(0.100)
전망	매출규모	94.55	100.00	-1.258(0.209)
	투자규모	95.45	85.38	2.861(0.004)**
	외부 고용사정	97.79	87.74	3.093(0.002)**
	채산성	95.71	88.68	1.693(0.091)
	자금사정	92.08	86.79	1.297(0.195)
	상품단가	100.26	96.23	1.296(0.195)
	전반적 업황	92.08	86.79	0.010(0.992)

* p<0.05, ** p<0.01, *** p<0.001

② 선택형 모드효과 분석

기업경기동향조사에서 모드효과 분석을 위해서는 선택효과와 측정효과를 분리하여야 한다. 선택효과와 측정효과를 명확하게 분리하기 위해서는, 소비자동향조사에서와 같이 식(3.1)과 식(3.2)에서 관측불가능하다고 한 부분인 $Y_a = j \mid M = b$ 의 값이 있어야 한다. 본 실험에서는 $Y_a = j \mid M = b$ 의 값을 선택권 없이 조사한 혼합 I에서 전화조사를 한 집단의 관측결과가 이와 유사한 상황이기 때문에 이 값을 이용하였다. 모드효과를 분해 한 값들은 표(4-36)에 제시하였다.

〈표 4-36〉을 보면 기업경기동향조사에서는 대체적으로 측정효과가 선택효과보다 크다는 것을 알 수 있다. 측정효과는 매출규모의 경우 매우 적게 나타났으나, 현황에서는 투자규모와 외부 고용사정, 전반적 업황에서 크게 나타났고, 전망에서는 투자규모와 전반적 업황의 경우는 절대값이 모드효과보다 더 크게 차이가 나타나는 것을 알 수 있다.

〈표 4-36〉 기업경기동향조사 항목의 모드효과분해

구분	항목	혼합조사			전화조사 ③	모드효과 분해		
		전화 ①	모바일②	전체		모드 효과 ②-①	선택 효과 ③-①	측정 효과 ②-③
현황	매출규모	73.77	77.36	74.54	77.36	3.59	3.59	0.00
	투자규모	88.31	77.83	86.05	92.37	-10.48	4.06	-14.54
	외부 고용사정	89.87	80.19	87.78	91.72	-9.68	1.85	-11.53
	채산성	81.69	77.83	80.86	84.24	-3.86	2.56	-6.41
	자금사정	75.97	75.94	75.97	80.95	-0.03	4.98	-5.01
	상품단가	100.13	90.09	97.96	99.85	-10.04	-0.28	-9.76
	전반적 업황	82.08	74.53	80.45	84.66	-7.55	2.58	-10.13
전망	매출규모	94.55	100.00	95.72	99.47	5.45	4.92	0.53
	투자규모	95.45	85.38	93.28	98.93	-10.08	3.48	-13.55
	외부 고용사정	97.79	87.74	95.62	98.61	-10.06	0.81	-10.87
	채산성	95.71	88.68	94.20	99.35	-7.04	3.63	-10.67
	자금사정	92.08	86.79	90.94	94.15	-5.29	2.08	-7.36
	상품단가	100.26	96.23	99.39	101.28	-4.03	1.02	-5.05
	전반적 업황	92.08	86.79	90.94	99.11	-5.29	7.03	-12.32

나. 할당형 혼합조사 모드효과 분석

이번에는 전화조사와 웹조사를 1,700개와 500개를 할당하여 조사한, 기업경기동향 조사의 할당형 혼합조사의 모드효과를 살펴보도록 한다. 먼저 응답한 사업체들의 특성에 따른 차이가 있는지 살펴보았으며, 그 결과는 〈표 4-37〉에 제시하였다. 〈표 4-37〉을 보면 조직형태에서는 조사방법에 따라 유의한 차이가 있는 것으로 나타나지만, 다른 특성에서는 유의한 차이가 나타나지 않는 것을 알 수 있다.

〈표 4-37〉 기업경기동향조사의 할당한 조사방법별 사업체특성별 변수 비교

		조사방법		카이제곱 (유의확률)
		전화조사	웹조사	
종사자 규모	5~9명	32.2	34.9	2.425 (0.658)
	10~19명	26.8	26.0	
	20~49명	23.6	21.5	
	50~99명	10.7	9.9	
	100명 이상	6.8	7.8	
산업구분	문화예술산업	28.8	30.2	1.110 (0.775)
	문화산업	27.8	25.8	
	스포츠산업	17.8	18.8	
	관광산업	25.6	25.2	
조직형태	개인사업체	12.3	13.0	34.355 (0.000)
	회사 법인	70.6	67.6	
	회사 이외 법인	1.4	5.8	
	국가/지방자치단체	12.9	11.2	
	비법인 단체	2.8	2.3	
사업체 구분	단독사업체	59.8	54.9	2.845 (0.241)
	본사(점)	24.4	27.4	
	지사(점), 영업소, 공장	15.8	17.7	
응답자 성별	남성	40.8	41.1	0.918
	여성	59.2	58.9	

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-38〉은 기업경기동향조사 14개 조사항목에 대하여 전화조사와 모바일조사에 차이가 있는지를 분석한 결과를 제시한 것이다. 투자규모와 외부고용사정은 현황과 전망에서 모두 유의수준 0.001 수준에서도 유의하게 나타나, 모드별 차이가 큰 것으로 나타났다. 그리고 현황에서는 채산성, 자금사정, 상품단가는 유의수준 0.05 수준에서 유의한 것으로 나타났으며, 매출규모와 전반적 업황은 유의하지 않은 것으로 나타났다. 전망에서는 투자규모, 외부 고용사정, 채산성, 자금사정이 유의수준 0.001수준에서 유의하게 나타났으며, 매출규모는 유의수준 0.01에서 유의한 것으로 나타났다. 그리고 상품단가와 전반적 업황은 모드별 차이가 없는 것으로 나타났다.

〈표 4-38〉 기업경기동향조사 선택형 조사방법에 따른 조사항목의 차이분석

구분	항목	전화조사	웹조사	t-값(유의확률)
		평균	평균	
현황	매출규모	77.361	76.07	0.561(0.575)
	투자규모	92.37	81.01	6.004(0.000)***
	외부 고용사정	91.72	81.78	5.972(0.000)***
	채산성	84.24	79.85	2.212(0.027)*
	자금사정	80.95	76.07	2.445(0.015)*
	상품단가	99.85	97.001	2.120(0.034)*
	전반적 업황	84.66	82.07	1.276(0.202)
전망	매출규모	99.47	94.48	2.607(0.009)**
	투자규모	98.93	89.05	6.458(0.000)***
	외부 고용사정	98.61	89.24	6.861(0.000)***
	채산성	99.35	90.31	5.183(0.000)***
	자금사정	94.15	87.69	3.628(0.000)***
	상품단가	101.28	100.48	0.733(0.463)
	전반적 업황	99.11	96.32	1.596(0.111)

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-39〉 기업경기동향조사에서 사업체 특성 변수를 통제한 조사방법 효과분석

항목	현황		전망	
	베타계수 (유의확률)	유의한 통제변수	베타계수 (유의확률)	유의한 통제변수
매출규모	-1.226(0.585)	종사자규모, 산업구 분, 조직형태, 사업체 구분	-5.272(0.006)**	종사자규모, 조직형 태, 사업체구분
투자규모	11.328(0.000)***	종사자규모, 산업구 분, 성별	-10.259(0.000)***	산업구분, 조직형태,
외부 고용사정	-10.165(0.000)***	산업구분, 조직형태, 사업체구분, 성별	-9.779(0.000)***	산업구분, 조직형태, 사업체구분, 성별
채산성	-4.521(0.021)*	종사자규모, 산업구 분, 조직형태, 사업체 구분, 성별	-9.464(0.000)***	산업구분, 조직형태, 사업체구분, 성별
자금사정	-4.845(0.012)***	종사자규모, 산업구 분, 조직형태, 사업체 구분, 성별	-6.755(0.000)***	산업구분, 조직형태, 사업체구분, 성별
상품단가	-3.083(0.022)*	산업구분, 성별	-1.021(0.346)	산업구분, 사업체구분
전반적 업황	-2.320(0.242)	산업구분, 조직형태, 사업체구분, 성별	-2.936(0.090)	종사자규모, 조직형 태, 사업체구분, 성별

* p<0.05, ** p<0.01, *** p<0.001

〈표 4-39〉는 기업경기동향조사에서 사업체 특성변수를 통제한 상태에서, 순수한 조사 방법의 효과를 파악하기 위해 최귀분석을 이용하여 분석한 결과를 제시한 것이다. 독립변수로는 종사자 규모, 산업구분, 조직형태, 사업체구분, 응답자 성별 5개의 변수를 사용하였다. 이 중에서 조사방법의 효과를 나타내는 베타계수를 제시하고, 유의한지를 나타내는 유의확률을 제시하였다. 또한 유의하게 나타나는 통제변수를 같이 제시하였다. 베타계수의 값은 웹조사의 기업경기동향지수가 전화조사의 기업경기동향조사와의 차이를 의미한다. 즉, 현황의 외부 고용사정 값은 웹조사방식이 전화조사방식보다 -10.165점 차이가 나는 효과가 있다는 것을 의미한다. 이러한 차이를 알기 위해서는 유의한 영향을 미치는 산업구분, 조직형태, 사업체구분, 성별 변수를 통제해야 하는 것을 의미한다.

제5절 소결 및 시사점

1. 소결

4장에서는 조사방법에 따른 차이분석을 위한 다양한 사례를 살펴보았다. 먼저 조사과 정자료 분석을 살펴보면, 기존의 (가구방문) 대면조사 방식을 온라인 조사방식으로 변경하고자 할 때, 전국민을 대상으로 하는 표집틀을 마련하는 것은 어렵다. 따라서 이에 대한 대안으로 조사업체의 온라인 패널을 활용하는 방안을 고려하는 경우가 많다. 그러나 온라인패널 현황을 분석한 결과 최근 6개월 내 한번이라도 접속을 한 패널이 약 15% 밖에 되지 않았고, 나머지 대부분인 85%의 패널은 패널로 참여되고 있는지에 대한 현황 파악이 되고 있지 않다. 따라서 온라인패널의 표본틀에 대한 정보가 불분명하다고 판단 되기 때문에, 이를 이용한 조사는 모집단의 포괄성에 대한 진단이 어려워 모수추정을 위한 통계에는 활용이 어렵다.

가구방문조사의 경우, 코로나19로 인해 예년에 비해 조사구대체가 많이 발생하였다. 따라서 가구의 대체는 더 크게 발생되어, 낮은 응답률로 인한 비표본오차로 통계품질에 대한 신뢰가 떨어진다.

조사방법별 차이를 살펴보면, 각 모드별로 차이가 달라지고 있는 것을 확인할 수 있다. 물론 각 조사마다 항목별로 그 차이에 유의미한 경우와 그렇지 않은 경우가 혼재되어 있었지만, 많은 항목에서 차이가 나고 있음을 확인하였고, 조사방법 뿐만 아니라 조사를 위해 사용하는 도구(종이설문지, 태블릿 등)에 따라서도 응답결과가 달라짐을 확인하였다.

그러나 3장의 혼합조사 사례를 살펴보면 단일조사에 비해 항목 무응답률이 낮게 나타났음을 알 수 있었다. 따라서 응답률이 점차 낮아지고 있는 조사환경에서 응답률을 높이기 위한 방안으로 혼합조사를 이용한 것에는 효과가 있다는 것을 알 수 있었다.

2. 시사점

조사방법을 다르게 적용하더라도 동일한 조사의 경우, 모집단은 같아야 한다. 즉, 동일한 조사를 다른 방법으로 하더라도, 표본은 모집단을 포괄하여야 한다. 따라서 표본들은 모집단을 충분히 반영하여야 한다. 이러한 이유로 국민을 대상으로 하는 조사의 경우, ‘인구총조사’를 활용하는 방법으로 조사한 방법을 승인통계로 인정하고 있다.

본 연구에서 제시하고 있는 조사업체의 온라인패널의 경우, 자발적인 참여로 구성된 패널이기 때문에, 온라인패널 구성이 선택편의가 있을 뿐만 아니라, 가입 후 오랫동안 참여를 안 한 대상들을 온라인패널이라 하기 어렵다. 온라인패널을 이용하여 국민대상의 조사로 활용하기 위해서는 온라인패널을 구성하고 있는 대상들의 분포를 분석하여, 우리나라 국민을 포괄하기 위한 방안에 대한 연구가 필요하다. 즉, 국민대상 조사의 마스터 샘플(master sample)로 활용하기 위해서는 온라인패널로 활용가능한 정확한 대상을 파악하고, 파악된 대상의 특성 분석을 하여 모수추정할 때 특성을 고려한 (보정 방안을 고려하여) 추정방법을 마련하여야 한다.

가구방문조사의 경우, (코로나와 같은 사회환경의 변화가 있어) 조사가 어려운 경우 표본대체가 많이 발생된다. 표본대체가 많이 발생된다는 것은 조사의 정도(precision)가 떨어지게 된다. 따라서 다양한 방법으로 조사에 대한 홍보방안을 마련하고, 과거 조사과정자료(paradata)를 분석하여 응답 대상자의 특성에 맞는 방문 전략이 필요함을 알 수 있다. 원표본 유지율을 높이기 위한 방안으로, 노후화된 조사구의 사전 답사 등을 통한 정보의 업데이트가 필요하며, 지역 및 지리적 특성을 고려한 조사방안을 수립하여야 한다. 또한 최근 1인가구가 증가함으로 조사의 어려움이 커지고 있다. 이러한 다양한 특성의 가구에 대한 조사 성공을 위한 전략을 사전에 수립하여야 한다.

조사방법별로 모드효과의 차이가 있어, 결과에 차이가 발생한다. 또한 조사도구(종이 설문지, 태블릿 등)에 따라서도 응답결과가 달라진다. 하지만 어떤 응답이 정답인지는 알 수 없다. 따라서 각 조사방법별로 모드효과를 분석하여 어느 정도의 영향이 있는가를 살펴 볼 필요가 있다. 특히 조사의 응답률을 높이기 위한 방안으로 여러 조사방법을 병행하는 혼합조사에서는, 이러한 조사방법의 차이에 따른 효과뿐만 아니라 응답자가 조사방법을 선택할 때 발생하는 선택효과에 대한 영향도 파악할 필요가 있다.

혼합조사에서 조사방법별로 조사결과의 차이가 발생하지만, 단일조사보다는 무응답

률이 낮게 나타났다. 이는 응답자들이 원하는 방법으로 조사를 할 있도록 선택할 수 있기 때문이다. 따라서 혼합조사를 적극 활용하여 응답률을 높일 수 있도록 하되, 각 방법별 차이를 보정하여 추정할 수 있는 방안 마련이 필요하다.

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

제5장

혼합조사 추정 방안

제1절 혼합조사 추정방법

두 가지 이상의 조사방법을 이용하여 하나의 조사를 하는 혼합조사 방법은 단일조사 처럼 하나의 조사결과로 추정할 경우, 서로 다른 조사방법을 이용하였기 때문에 선택효과와 측정효과가 발생하게 된다. 따라서 혼합조사의 추정은 각각의 방법을 고려하여 추정할 필요가 있다.

혼합조사 추정방법으로 많이 이용하는 것은 최적결합을 이용하는 방법과 성향점수를 이용한 방법이 많이 이용되고 있는데, 본 연구에서도 두 가지 방법으로 추정한 결과를 제시하고자 한다.⁶⁶⁾

1. 최적결합방법

동일한 조사라 하더라도 조사방법이 다를 경우 추정결과는 다르게 된다. 이러한 차이는 실제로 측정방법에 따라 응답자가 조사내용을 다소 다르게 인식하는 경우도 있을 것이고, 실제로는 같은 값이지만, 조사방법을 선택한 대상이 다른 성향을 가지기 때문에 다르게 나타날 수도 있다. 따라서 동일한 조사라 하더라도 다른 조사방법의 특성을 고려하여 각 방법의 결과를 함께 이용한다면 더욱 좋은 추정값을 얻을 수 있을 것이다.

최적결합방법(optimal weighting method)은 서로 다른 조사방법의 결과를 통합하는 방법이며, 서로 다른 조사방법으로 인해 도출된 결과의 분산을 이용하여 최소평균제곱오차(minimum mean square error)가 되도록 결과를 결합하게 된다. 이때 각 조사방법에 따른 결과의 분산은 불편성(unbiasedness)을 만족하여야 한다.

혼합조사에서 조사방법을 두 개 이용한다고 가정하면 결합 추정량과 분산은 다음과

66) 본 연구에서 제시하는 혼합조사 방법의 표현은 임경은 등(2013)이 사용한 방법을 토대로 정리함

같게 된다. 각 조사방법은 서로 독립(independent)적으로 조사를 진행한다. 따라서 두 조사결과는 서로 독립이기 때문에 공분산(covariance)은 0이 된다.

$$f = w_1 f_1 + (1 - w) f_2 \quad (5.1)$$

$$\begin{aligned} \text{var}(f) &= w^2 \text{var}(f_1) + (1 - w)^2 \text{var}(f_2) \\ &= w^2 \sigma_1^2 + (1 - w)^2 \sigma_2^2 \end{aligned} \quad (5.2)$$

여기서 f_1 과 f_2 는 서로 다른 조사방법의 추정량이다. 예를 들면 f_1 은 대면조사방법의 추정량이고, f_2 는 인터넷조사 방법의 추정량을 고려할 수 있다. 또한 w 는 전체추정량인 f 의 오차 분산을 최소로 하는 최적 가중값(optimal weight)이며 다음과 같다.

$$w = \frac{\sigma_2^{-2}}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (5.3)$$

이를 일반화하면 다음과 같게 표현된다. M 개의 조사방법을 이용한 혼합조사이며, 각 조사방법은 서로 독립적으로 진행된다고 가정한다.

$$f = w_1 f_1 + w_2 f_2 + \dots + w_M f_M \quad (5.4)$$

$$w_m = \frac{\sigma_m^{-2}}{\sigma_1^{-2} + \sigma_2^{-2} + \dots + \sigma_M^{-2}} \quad (5.5)$$

여기서 $w_1 + w_2 + \dots + w_M = 1$ 이다.

최적결합방법의 식을 잘 살펴보면 분산이 작은 조사방법의 조사결과의 가중치를 높이고, 분산이 큰 조사방법의 조사결과의 비중을 적게 작동하도록 하는 방법이라는 것을 알 수 있다. 즉, 분산이 작은 조사방법은 안정적인 방법이고 분산이 큰 조사방법은 상대적으로 불안정한 방법이라는 개념에서 도출한 방법이다.

2. 성향점수를 이용한 추정방법

성향점수보정(propensity score adjustment) 방법은 관측연구(observational

study)에서 서로 다른 두 집단을 비교할 때, 공변량(covariate)들이 사전에 결정되어 연구자가 조절(control)할 수 없다. 따라서 공변량의 분포가 달라 실제 실험계획(experimental design)의 결과에 공변량이 영향을 주는 선택편의(selection bias)가 나타나는데, 이를 보정하기 위한 방법이며 Rosenbaum과 Rubin(1983, 1984), D'Agostino(1998)에 의해 제안되었다.

혼합조사 역시 조사의 효율을 높이기 위해, 조사방법을 응답자들이 선택하게 하기 때문에, 응답자들의 성별, 연령, 거주지역 등의 공변량들의 분포를 연구자가 조절할 수 없는 경우이다. (이러한 공변량을 연구자가 조절한다면, 응답자들이 원하지 않는 방법으로 조사방법이 결정되기 때문에, 혼합조사의 가장 큰 장점인 응답률이 높아지지 않을 것이다.) 따라서 성향점수보정방법을 적용할 수 있다. 성향점수는 통제집단(control group)과 실험집단(experimental group)으로 구분하여 통제집단을 기준으로 실험집단의 결과를 비교하여 실험효과를 측정하며, 실험효과를 정확하게 측정하기 위해 두 집단의 공변량의 분포를 안정(균형)시킨다.

따라서 혼합조사의 경우에도 서로 다른 조사방법의 기준이 되는 참조조사(reference survey)가 존재한다는 가정에서 시작되는데, 참조조사는 관측연구에서의 대조집단의 역할을 하게 된다. 따라서 참조조사는 비교하는 조사와 비슷한 시점에 실행되어야 하며, 가구면접조사와 같이 신뢰성이 인정된 전통적인 조사방법에 의한 것으로 높은 응답률을 갖는 양질의 조사이어야 한다(김영원 등, 2010).

성향점수를 이용한 추정 방법을 설명하기 위해 두 조사방법을 대면조사(P)와 온라인 조사(I) 두 방법을 예로 설명하면 다음과 같다. 통합표본은 $s = s^P \cup s^I$ 이고, 표본의 크기는 $n = n^P + n^I$ 가 된다. 여기서 s^P 는 대면조사 표본이고, s^I 는 온라인조사 표본이 된다. 그리고 n^P 는 대면조사의 표본크기이고, n^I 는 온라인조사의 표본크기가 된다. 성향점수의 식은 다음의 방법으로 추정된다.

$$\ln \left[\frac{e(\mathbf{x})}{1 - e(\mathbf{x})} \right] = \alpha + \beta^T f(\mathbf{x}) \quad (5.6)$$

여기서 \mathbf{x} 는 공변량 벡터이고 $f(\mathbf{x})$ 는 공변량 벡터의 함수이다. 성향점수를 산출할 때 사용하는 함수는 로짓모형(logit model)을 주로 이용한다.⁶⁷⁾

67) 두 개의 방법에 대한 성향점수를 산출할 경우에는 로지스틱 회귀모형을 주로 이용하고, 세 개 이상의 방법

성향점수는 각 조사방법으로 산출된 전체자료를 산출하는데, i 번째 조사대상자의 인터넷조사 참여 성향점수는 $e(\mathbf{x}_i) = P(i \in s^I \mid \mathbf{x}_i)$ 로 계산된다. 이는 i 번째 조사대상자의 공변량(\mathbf{x}_i)이 주어졌을 때, i 번째 조사대상자가 인터넷조사에 참여할 확률이 된다. 성향점수보정 방법은 매칭(matching), 층화(stratification), 공분산조정 등의 방법이 있는데 여기서는 몇 개의 층⁶⁸⁾으로 구분한 후, 각 층에서 대면조사와 인터넷조사의 가중치의 비중이 같도록 조정하여, 대면조사와 온라인조사의 분포가 일치하도록 한다.

성향점수를 산출하면 추정된 성향점수의 크기순으로 정렬한 후, C 개의 층으로 구분하는데, 각 층의 크기는 보통 같도록 한다. c 번째 층의 표본은 $s_c = s_c^P \cup s_c^I$ 로 표현되며, 표본크기는 $n_c = n_c^P + n_c^I$ 개가 된다. 다음으로 각 층의 조정인자(adjustment factor)를 계산한다.

$$f_c = \frac{\sum_{k \in s_c^P} d_k^P / \sum_{k \in s^P} d_k^P}{\sum_{j \in s_c^I} d_j^I / \sum_{j \in s^I} d_j^I} \quad (5.7)$$

여기서 d_j^I 는 $j(j = 1, 2, \dots, n^I)$ 번째 응답자의 온라인조사 기본 가중값이고, d_k^P 는 $k(k = 1, 2, \dots, n^P)$ 번째 응답자의 대면조사 기본 가중값이 된다. 여기서는 대면조사가 참조조사의 역할을 한다. 기본 가중값이 선택확률의 역수라면 식 (5.7)은 다음과 같이 표현될 수 있다.

$$f_c = \frac{\sum_{k \in s_c^P} d_k^P / \sum_{k \in s^P} d_k^P}{\sum_{j \in s_c^I} d_j^I / \sum_{j \in s^I} d_j^I} = \frac{\hat{N}_c^P / \hat{N}^P}{\hat{N}_c^I / \hat{N}^I} \quad (5.8)$$

c 번째 층 내에서, 온라인조사의 표본의 j 번째 응답자에 대한 성향점수보정인자는 다음과 같이 표현된다.

$$d_j^{I,PSA} = f_c d_j^I = \frac{\hat{N}_c^P / \hat{N}^P}{\hat{N}_c^I / \hat{N}^I} d_j^I \quad (5.9)$$

에 대한 성향점수를 산출할 경우에는 다항 로지스틱 회귀모형을 이용함
68) Cochran(1968)은 5개의 층으로 구분함

단순확률추출과 같이 모든 응답자들이 갖는 기본가중치가 같거나 이를 알 수 없는 경우에는 다음과 같이 조정인자를 사용한다.

$$f_c = \frac{n_c^P / n^P}{n_c^I / n^I} \quad (5.10)$$

식 (5.9)와 같이 성향점수보정을 이용하면 성향점수로 이용한 층의 온라인조사와 대면조사의 표본의 분포가 같아진다. 이때 온라인조사에서 c 번째 층에 해당하는 모집단 크기에 대한 추정값은 다음과 같다.

$$\hat{N}_c^{I.PSA} = \sum_{j \in s_c^I} d_j^{I.PSA} = \hat{N}^I \frac{\hat{N}_c^P}{\hat{N}^P} \quad (5.11)$$

온라인조사의 표본(s^I)로부터 얻어지는 관심 변수(y)에 대한 평균은 다음과 같이 구해진다.

$$\hat{y} = \frac{\sum_c \sum_{j \in s_c^I} d_j^{I.PSA} y_j}{\sum_c \sum_{j \in s_c^I} d_j^{I.PSA}} \quad (5.12)$$

제2절 혼합조사 추정을 위한 시험조사

1. 시험조사 설계

혼합조사방법은 응답률을 높이고 조사에 드는 시간과 예산을 줄일 수 있는 효율적인 방법이지만, 개인별 환경과 상황에 따라 선호하는 방법에 따른 선택효과와 조사방법의 근본적인 차이로 인한 측정효과로 인한 차이가 발생할 수 있다. 따라서 본 연구에서는 ‘문화체육관광 동향조사’를 기반으로 전화조사와 온라인조사를 혼합한 조사를 설계하였다.⁶⁹⁾

‘소비자동향조사’는 우리나라 만 19세 이상의 국민을 대상으로 조사하고 있다. 기존의 ‘소비자동향조사’는 RDD 방식의 전화조사 방식으로 진행하여왔으나, 시험조사를 위해 할당형 혼합조사와 선택형 혼합조사를 각각 설계하였는데, 할당형 혼합조사는 전화조사 1,000표본, 모바일조사⁷⁰⁾ 500표본을 조사하였다. 선택형 혼합조사의 경우 502표본을 조사 하였는데, RDD로 전화를 걸어 전화조사와 모바일조사를 선택하게 하였다. 전화조사는 400표본 모바일조사는 102표본이 응답하였다.

‘기업경기동향조사’는 문화체육관광 사업체 표본틀⁷¹⁾을 이용하여 조사하는데, 표본으로 추출된 사업체들에게 전화조사 방법으로 진행하여왔다.⁷²⁾ 기업경기동향조사도 연구를 위해 할당형 혼합조사와 선택형 혼합조사를 진행하였다. 할당형 혼합조사는 전화조사가 1,700개소, 온라인(이메일)조사 500개소로 구성하여 조사하였고, 선택형 혼합조사는 513개소를 조사하였는데, 전화조사는 406개소, 온라인조사가 107개소 응답하였다. 조사에 대한 내용은 <표 5-1>에 정리하여 제시하였다.

69) 해당 시험조사는 예산의 효율적 활용을 위해 새로운 조사로 설계하여 진행하지 않고, 기존 조사에 시험조사를 추가 설계하여 진행함. 즉, 선택형 혼합조사(혼합Ⅱ)를 각각 500표본씩 추가 조사함

70) RDD로 된 전화번호에 문자로 조사의 취지를 설명하고, 설문에 접속할 수 있는 URL(uniform resource locator) 보내 조사를 진행함

71) 표본틀은 통계청의 ‘전국사업체조사’의 명부에서 문화·체육·관광 영역에 해당하는 사업체들을 선정하여 문화체육관광 사업체 표본틀을 구축함

72) 사업체들이 원할 경우, 이메일과 팩스를 이용하였지만, 극히 일부이며 대부분 전화조사로 진행되어 음

〈표 5-1〉 조사설계 비교

구분		소비자동향조사	기업경기동향조사
모집단		우리나라 전 국민	해당분기에 문화체육관광 관련업을 영위한 사업체
조사 대상		전국 17개 시도 만 19세 이상 가구주 및 가구주의 배우자	문화체육관광 표준산업구분에 해당되는 문화예술, 문화산업, 스포츠, 관광사업체 중 상시 종사자 수 5인 이상 사업체
기존	표본 크기	1,500명	2,200개소
	조사 방법	RDD를 통한 번호 추출 후 CATI를 활용한 전화조사	제공된 리스트를 통해 전화조사
혼합 I	표본 크기	전화조사: 1,000명 온라인(모바일)조사: 500명	전화조사: 1,700개소 온라인(이메일)조사: 500개소
	조사 방법	2가지 조사방법별 따로 설계하여, 각각 전화/온라인조사	2가지 조사방법별 따로 설계하여, 각각 전화/온라인조사
혼합 II	표본 크기	500명(유효표본 502명)	500개소(유효표본 513개소)
	조사 방법	RDD를 통한 번호 추출하여 조사대상을 선택한 후, 응답자가 전화조사 or 온라인(모바일)조사를 선택하여 응답진행	추출된 표본리스트를 통해 전화 컨택 후, 응답자가 전화조사 or 온라인(이메일)조사를 선택하여 응답진행

2. 혼합조사 추정

가. 소비자동향조사 혼합조사 추정

1) 최적결합방법에 의한 추정

앞서 소비자동향조사의 모드효과 분석을 통해 조사방법별로 값의 차이가 있다는 것을 확인하였다. 따라서 조사방법별로 동일한 가중값을 적용하면 잘못된 값을 도출하게 된다. 따라서 정도(precision)를 높일 수 있도록 조사방법별로 값을 적용하여야 하는데, 이러한 때에 최적결합방법을 이용한다. 최적결합방법은 분산이 작은 방법의 비중을 더 많이 반영하는 방법이다. 〈표 5-2〉는 선택형 혼합조사에서 최적결합방법과 기존의 방법으로 추정한 값을 같이 제시한 것이고, 〈표 5-3〉은 할당형 혼합조사에서 최적방법과 기존의 방법을 추정한 값을 제시한 것이다.

〈표 5-2〉 소비자동향조사 선택형 혼합조사의 최적결합 추정값 비교

구분	항목	전화	온라인	결합 추정값 ①	기존 추정값 ②	차이 ①-②
현황	오프라인 문화생활	72.24	55.95	65.24	69.89	-4.65
	오락시설 지출비용	79.49	48.93	66.54	75.09	-8.55
	온라인 문화생활비	111.23	110.16	110.79	111.08	-0.29
	스포츠시설 이용료	93.14	73.05	83.72	90.25	-6.53
	스포츠용품 및 기구	102.03	86.49	94.72	99.79	-5.07
	스포츠 경기 관람료	90.59	64.72	83.72	86.87	-3.15
	아웃도어용품 구입비	92.35	69.09	82.71	89.00	-6.29
	미디어기기 구입비	108.76	115.87	111.34	109.78	1.56
	국내여행비	94.54	74.21	87.41	91.62	-4.20
	해외여행비	90.60	53.18	82.14	85.21	-3.07
	도서구입비	109.79	108.73	109.39	109.64	-0.24
	오락용품 구입비	108.80	84.46	100.94	105.29	-4.36
	음악미술학원비	97.33	73.57	93.62	93.91	-0.29
	스포츠 학원비	95.70	78.57	92.55	93.24	-0.69
	여행관광비	91.75	59.33	84.00	87.08	-3.08
	종합지출	96.49	78.31	92.51	93.87	-1.36
전망	오프라인 문화생활	96.64	97.57	96.97	96.78	0.19
	오락시설 지출비용	96.63	87.43	93.58	95.31	-1.73
	온라인 문화생활비	104.52	112.52	107.17	105.68	1.49
	스포츠시설 이용료	99.36	87.64	95.87	97.67	-1.80
	스포츠용품 및 기구	96.54	92.39	94.60	95.94	-1.34
	스포츠 경기 관람료	98.76	86.71	95.32	97.03	-1.70
	아웃도어용품 구입비	105.06	88.21	100.43	102.63	-2.21
	미디어기기 구입비	100.65	105.77	102.42	101.39	1.03
	국내여행비	108.45	98.39	104.39	107.00	-2.62
	해외여행비	94.06	81.66	90.99	92.27	-1.28
	도서구입비	109.33	106.53	108.49	108.93	-0.44
	오락용품 구입비	102.29	93.33	99.93	101.00	-1.06
	음악미술학원비	98.34	87.27	96.18	96.75	-0.56
	스포츠 학원비	99.45	94.35	98.13	98.72	-0.58
	여행관광비	98.27	86.55	95.43	96.58	-1.15
	종합지출	99.53	93.76	98.19	98.70	-0.51

〈표 5-3〉 소비자동향조사 할당형 혼합조사의 최적결합 추정값 비교

구분	항목	전화	온라인	결합 추정값 ①	기존 추정값 ②	차이 ①-②
현황	오프라인 문화생활	74.31	61.14	68.48	69.91	-1.43
	오락시설 지출비용	81.74	62.30	73.80	75.24	-1.44
	온라인 문화생활비	109.02	100.78	105.97	106.27	-0.30
	스포츠시설 이용료	90.96	73.33	83.96	85.07	-1.10
	스포츠용품 및 기구	98.14	83.49	92.52	93.25	-0.73
	스포츠 경기 관람료	88.52	70.56	82.32	82.52	-0.20
	아웃도어용품 구입비	97.66	85.51	93.01	93.60	-0.59
	미디어기기 구입비	107.99	100.75	105.15	105.57	-0.42
	국내여행비	94.25	76.39	86.85	88.28	-1.43
	해외여행비	90.15	63.48	82.98	81.24	1.74
	도서구입비	106.57	96.10	102.74	103.07	-0.33
	오락용품 구입비	100.78	86.33	96.64	95.95	0.69
	음악미술학원비	99.83	80.49	95.75	93.37	2.38
	스포츠 학원비	96.39	80.21	92.55	90.99	1.56
	여행관광비	91.35	67.25	84.44	83.30	1.14
	종합지출	95.81	79.30	91.70	90.30	1.40
전망	오프라인 문화생활	102.27	94.56	99.65	99.69	-0.04
	오락시설 지출비용	100.10	86.68	96.05	95.62	0.43
	온라인 문화생활비	104.54	101.62	103.64	103.57	0.07
	스포츠시설 이용료	104.49	92.11	100.39	100.35	0.04
	스포츠용품 및 기구	104.17	94.09	100.76	100.81	-0.05
	스포츠 경기 관람료	99.88	84.72	95.24	94.82	0.42
	아웃도어용품 구입비	105.81	95.38	102.59	102.32	0.26
	미디어기기 구입비	100.84	95.50	98.79	99.06	-0.27
	국내여행비	110.88	99.77	106.81	107.17	-0.36
	해외여행비	94.71	76.49	90.49	88.62	1.87
	도서구입비	106.26	100.37	104.37	104.29	0.08
	오락용품 구입비	101.28	89.45	98.21	97.33	0.88
	음악미술학원비	100.58	86.46	97.84	95.86	1.97
	스포츠 학원비	99.51	88.45	97.54	95.81	1.72
	여행관광비	99.44	83.29	95.55	94.04	1.51
	종합지출	101.53	90.35	99.38	97.79	1.59

〈표 5-2〉와 〈표 5-3〉에서 일부항목에서는 기존추정값과 결합추정값의 차이가 크지 않지만, 현황에서 오락시설 지출비용 같은 항목은 다소 차이가 난다. 차이가 크게 나타나는 항목의 경우는 두 집단의 분산이 크게 차이가 나타난다는 것을 의미하며, 이러한 경우 안정된 조사의 값을 따르는 것이 타당할 것이다.

2) 성향점수를 이용한 추정

성향점수를 이용한 보정기법은 원래 관측연구(observational study)에서 비교집단 간의 집단별 선택에 따른 영향으로 집단 간의 개체 특성의 차이를 균등화(balancing)함으로써 본연의 연구의 평가를 할 수 있도록 하기 위한 방법이다. 혼합조사에서도 모드에 따라 응답자의 특성 차이가 발생하기 때문에, 조사방법간의 응답자 특성의 차이를 보정하는 방법을 적용함으로써 효율이 뛰어난 추정을 할 수 있게 된다.

지금까지 진행해 왔던 소비자동향조사는 전화조사 방식이었기 때문에 기준조사(reference survey)를 전화조사방법, 비교조사를 모바일조사방법으로 설정하였다. 본 연구에서는 조사방법을 2가지로 정하였기 때문에, 두 조사방법 선택에서 오는 차이를 비교하여 균형을 맞추기 위한 성향점수(propensity score)를 로지스틱 회귀분석(logistic regression)의 예측값을 활용하였다.

할당형 혼합조사와 선택형 혼합조사에서 선택형 혼합조사가 연구자가 응답자를 조절할 수 없기 때문에, 성향점수를 이용한 추정은 선택형 혼합조사에 적용하는 것이 타당하지만, 응답자의 거절 등의 이유로 대체과정에서 할당형 혼합조사에서도 선택형 혼합조사처럼 일부는 연구자가 응답자를 결정하지 못하는 문제가 발생한다. 따라서 할당형 혼합조사와 선택형 혼합조사 모두 성향점수를 이용한 추정방법을 적용하도록 한다.

① 선택형 조사방법에서 성향점수 이용한 추정

성향점수를 활용하기 위한 공변량(covariate)은 연령대, 가구주여부, 결혼여부, 학력, 가구원수, 소득, 자영업여부 7개의 변수를 활용하였으며, 각 표본수의 구성비에 따라 학력, 소득은 범주를 재조정하였다. 〈표 5-4〉에 성향점수모형에 들어갈 공변량과 범주, 각 조사방법별 빈도(frequency)를 제시하였다. 로지스틱 회귀분석의 독립변수는 연속형 데이터만 들어야 하기 때문에 명목변수(nominal data)는 더미변수(dummy variable)⁷³⁾

73) 더미변수는 0과 1로만 구성되기 때문에, 4개의 범주가 있는 명목변수의 경우에는 3개의 더미변수가 만들

로 변환하여 활용하였다.

〈표 5-4〉 소비자동향조사 선택형 혼합조사에서 공변량과 범주

인구사회학적 특성		전화 조사	모바일 조사	인구사회학적 특성		전화 조사	모바일 조사
응답자 구분	가구주	316	74	가구원 수	1인 가구	117	26
	가구주의 배우자	84	28		2-3인 가구	197	54
연령대	30대 미만	78	36		4인 이상 가구	86	22
	40대	97	27	소득	월 200만원 미만	102	10
	50대	111	24		월 200만원-299만원 이하	58	14
	60대 이상	114	15		월 300만원-399만원 이하	67	16
결혼 여부	미혼	87	30		월 400만원-499만원 이하	44	21
	기혼	313	72		월 500만원-599만원 이하	46	16
학력	고등학교 졸업 이하	131	19	자영업 여부	600만원 이상	83	25
	대학 재학/졸업	221	63		예	75	13
	대학원 재학/졸업	48	20		아니오	208	69

〈표 5-5〉 소비자동향조사 선택형 로지스틱 회귀모형 추정

성향변수		계수	Odds	유의확률
응답자 구분	가구주	-1.134	0.322	0.002**
	가구주의 배우자			
연령대	30대 미만	-0.032	0.969	0.935
	40대	0.117	1.124	0.761
	50대	-0.157	0.855	0.696
	60대 이상			
결혼여부	미혼	1.245	3.474	0.003**
	기혼			
학력	고등학교 졸업 이하	-0.329	0.720	0.466
	대학 재학/졸업	-0.029	0.972	0.937
	대학원 재학/졸업			
가구원 수	1인 가구	0.108	1.114	0.851
	2-3인 가구	0.459	1.583	0.286
	4인 이상 가구			
소득	월 200만원 미만	-0.946	0.388	0.115
	월 200만원-299만원 이하	0.404	1.497	0.301
	월 300만원-399만원 이하	-0.509	0.601	0.212
	월 400만원-499만원 이하	0.971	2.640	0.028**
	월 500만원-599만원 이하	0.181	1.199	0.726
	600만원 이상			
자영업 여부	예	-0.175	0.839	0.687
	아니오			

* p<0.05, ** p<0.01, *** p<0.001

여저 구성된다. 더미변수로 만들어지지 않는 항목은 모형에서 다른변수의 결과를 해석할 때 비교기준이 된다.

〈표 5-5〉는 로지스틱 회귀모형 추정결과를 제시한 것이다. 유의한 변수는 응답자 구분, 결혼여부, 소득 3개 변수이며, 이 모형으로 구한 예측확률값을 성향점수로 활용한다. 성향점수를 이용하여 보정하는 방법들 중 집단을 구분하는 방법을 사용하였는데, Cochran(1968)이 제시한 5분위수(quintile points)를 이용하여 다섯 개의 구간으로 나누어 각각 20%의 데이터가 들어가도록 하였다.

〈표 5-6〉은 소비자동향조사의 선택형 혼합조사에서 성향점수보정한 값을 제시하고 기존의 추정값과의 차이를 비교한 것을 제시한 것이다. 온라인 조사를 전화조사의 성향과 비슷하게 조정하여 추정한 값을 이용한 성향점수보정값이 기존추정값과 크게 차이가 없는 것으로 나타났다. 즉 성향의 영향이 적다는 것을 의미한다.

〈표 5-6〉 소비자동향조사 선택형 혼합조사의 성향점수 보정 추정값 비교

구분	항목	전화	온라인	성향점수 보정 ①	기존 추정값 ②	차이 ①-②
현황	오프라인 문화생활	72.24	55.95	71.32	69.89	1.43
	오락시설 지출비용	79.49	48.93	75.81	75.09	0.72
	온라인 문화생활비	111.23	110.16	108.78	111.08	-2.30
	스포츠시설 이용료	93.14	73.05	91.13	90.25	0.88
	스포츠용품 및 기구	102.03	86.49	98.75	99.79	-1.04
	스포츠 경기 관람료	90.59	64.72	86.95	86.87	0.08
	아웃도어용품 구입비	92.35	69.09	88.94	89.00	-0.06
	미디어기기 구입비	108.76	115.87	106.60	109.78	-3.18
	국내여행비	94.54	74.21	92.87	91.62	1.25
	해외여행비	90.60	53.18	84.79	85.21	-0.42
	도서구입비	109.79	108.73	107.55	109.64	-2.09
	오락용품 구입비	108.80	84.46	104.13	105.29	-1.17
	음악미술학원비	97.33	73.57	93.94	93.91	0.03
	스포츠 학원비	95.70	78.57	92.50	93.24	-0.73
	여행관광비	91.75	59.33	87.15	87.08	0.07
	종합지출	96.49	78.31	93.24	93.87	-0.63
전망	오프라인 문화생활	96.64	97.57	96.51	96.78	-0.27
	오락시설 지출비용	96.63	87.43	95.02	95.31	-0.29
	온라인 문화생활비	104.52	112.52	103.95	105.68	-1.73
	스포츠시설 이용료	99.36	87.64	97.42	97.67	-0.25
	스포츠용품 및 기구	96.54	92.39	95.62	95.94	-0.33

구분	항목	전화	온라인	성향점수 보정 ①	기존 추정값 ②	차이 ①-②
	스포츠 경기 관람료	98.76	86.71	96.89	97.03	-0.14
	아웃도어용품 구입비	105.06	88.21	102.00	102.63	-0.63
	미디어기기 구입비	100.65	105.77	100.67	101.39	-0.71
	국내여행비	108.45	98.39	106.41	107.00	-0.59
	해외여행비	94.06	81.66	91.80	92.27	-0.47
	도서구입비	109.33	106.53	107.92	108.93	-1.01
	오락용품 구입비	102.29	93.33	101.29	101.00	0.29
	음악미술학원비	98.34	87.27	95.99	96.75	-0.76
	스포츠 학원비	99.45	94.35	98.30	98.72	-0.42
	여행관광비	98.27	86.55	96.07	96.58	-0.51
	종합지출	99.53	93.76	98.14	98.70	-0.55

② 할당형 조사방법에서 성향점수 이용한 추정

할당형 조사에서도 공변량(covariate)은 연령대, 가구주여부, 결혼여부, 학력, 가구원 수, 소득, 자영업여부 7개의 변수를 활용하였으며, <표 5-7>에 성향점수모형에 들어갈 공변량과 범주, 각 조사방법별 빈도(frequency)를 제시하였다. 성향점수를 로지스틱 회귀분석을 활용하기 때문에 명목변수(nominal data)는 더미변수(dummy variable)로 변환하여 활용하였다.

<표 5-7> 소비자동향조사 할당형 혼합조사에서 공변량과 범주

인구사회학적 특성		전화 조사	모바일 조사	인구사회학적 특성		전화 조사	모바 일조사
응답자 구분	가구주	787	395	가구원 수	1인 가구	274	144
	가구주의 배우자	213	105		2-3인 가구	496	249
연령대	30대 미만	257	122		4인 이상 가구	230	107
	40대	241	123	소득	월 200만원 미만	239	107
	50대	265	145		월 200만원-299만원 이하	148	79
	60대 이상	237	110		월 300만원-399만원 이하	156	93
결혼 여부	미혼	249	138		월 400만원-499만원 이하	117	69
	기혼	751	362		월 500만원-599만원 이하	119	52
학력	고등학교 졸업 이하	307	133		600만원 이상	221	100
	대학 재학/졸업	574	299	자영업 여부	예	240	97
	대학원 재학/졸업	119	68		아니오	518	292

〈표 5-8〉은 로지스틱 회귀모형 추정결과를 제시한 것이다. 유의한 변수는 소득과 자영업 여부 2개 변수이다. 할당형이기 때문에 성향의 구분이 명확한 변수가 덜 나타난 것으로 보인다. 예측확률값을 성향점수로 활용하여 5분위수(quintile points)를 이용하여 다섯 개의 구간으로 나누어 각각 20%의 데이터가 들어가도록 하였다.

〈표 5-8〉 소비자동향조사 할당형 로지스틱 회귀모형 추정

성향변수		계수	Odds	유의확률
응답자 구분	가구주	-0.030	0.971	0.838
	가구의 배우자			
연령대	30대 미만	-0.016	0.984	0.931
	40대	0.104	1.110	0.530
	50대	0.247	1.280	0.105
	60대 이상			
결혼여부	미혼	0.094	1.099	0.586
	기혼			
학력	고등학교 졸업 이하	0.202	1.223	0.144
	대학 재학/졸업	0.259	1.296	0.193
	대학원 재학/졸업			
가구원 수	1인 가구	0.163	1.177	0.446
	2-3인 가구	0.105	1.111	0.482
	4인 이상 가구			
소득	월 200만원 미만	0.051	1.053	0.801
	월 200만원-299만원 이하	0.170	1.186	0.400
	월 300만원-399만원 이하	0.362	1.436	0.052
	월 400만원-499만원 이하	0.166	1.180	0.322
	월 500만원-599만원 이하	-0.279	0.757	0.044*
	600만원 이상			
자영업 여부	예	-1.114	0.328	0.000***
	아니오			

* p<0.05, ** p<0.01, *** p<0.001

〈표 5-9〉는 소비자동향조사의 할당형혼합조사에서 성향점수 보정한 값을 제시하고 기존의 추정값과의 차이를 비교한 것을 제시한 것이다. 선택형에 비해 기존값의 차이가 별로 없는 것을 알 수 있다. 이는 할당형이기 때문에, 선택평향이 적기 때문에 개인특성에 따른 성향점수의 차이가 크지 않기 때문에 성향점수의 보정이 크지 않다는 것으로 판단할 수 있다.

〈표 5-9〉 소비자동향조사 할당형 혼합조사의 성향점수 보정 추정값 비교

구분	항목	전화	온라인	성향점수 보정 ①	기존 추정값 ②	차이 ①-②
현황	오프라인 문화생활	74.31	61.14	69.72	69.91	-0.19
	오락시설 지출비용	81.74	62.30	75.06	75.24	-0.18
	온라인 문화생활비	109.02	100.78	105.80	106.27	-0.46
	스포츠시설 이용료	90.96	73.33	84.76	85.07	-0.31
	스포츠용품 및 기구	98.14	83.49	92.86	93.25	-0.39
	스포츠 경기 관람료	88.52	70.56	82.49	82.52	-0.03
	아웃도어용품 구입비	97.66	85.51	93.19	93.60	-0.41
	미디어기기 구입비	107.99	100.75	105.03	105.57	-0.55
	국내여행비	94.25	76.39	88.06	88.28	-0.22
	해외여행비	90.15	63.48	81.20	81.24	-0.04
	도서구입비	106.57	96.10	102.69	103.07	-0.38
	오락용품 구입비	100.78	86.33	95.57	95.95	-0.38
	음악미술학원비	99.83	80.49	93.39	93.37	0.02
	스포츠 학원비	96.39	80.21	90.93	90.99	-0.05
	여행관광비	91.35	67.25	83.20	83.30	-0.10
	종합지출	95.81	79.30	90.05	90.30	-0.25
전망	오프라인 문화생활	102.27	94.56	99.16	99.69	-0.54
	오락시설 지출비용	100.10	86.68	95.41	95.62	-0.21
	온라인 문화생활비	104.54	101.62	103.25	103.57	-0.31
	스포츠시설 이용료	104.49	92.11	99.91	100.35	-0.45
	스포츠용품 및 기구	104.17	94.09	100.48	100.81	-0.32
	스포츠 경기 관람료	99.88	84.72	94.69	94.82	-0.12
	아웃도어용품 구입비	105.81	95.38	101.93	102.32	-0.39
	미디어기기 구입비	100.84	95.50	98.90	99.06	-0.16
	국내여행비	110.88	99.77	106.59	107.17	-0.58
	해외여행비	94.71	76.49	88.51	88.62	-0.11
	도서구입비	106.26	100.37	104.04	104.29	-0.25
	오락용품 구입비	101.28	89.45	97.00	97.33	-0.33
	음악미술학원비	100.58	86.46	96.03	95.86	0.17
	스포츠 학원비	99.51	88.45	95.77	95.81	-0.04
	여행관광비	99.44	83.29	93.79	94.04	-0.25
	종합지출	101.53	90.35	97.55	97.79	-0.24

나. 기업경기동향조사 혼합조사 추정

1) 최적결합방법에 의한 추정

기업경기동향조사의 모드효과 분석을 통해 조사방법에 따른 값의 차이가 있는 것을 확인하였다. 따라서 조사방법별로 동일한 가중값을 적용하면 정확한 추정값이 될 수 없다. 따라서 최적결합방법을 적용함으로써 각각의 분산을 고려하여 분산이 작은 방법의 비중을 더 많이 반영함으로써 정도(precision) 높은 값을 얻게 된다.

〈표 5-10〉은 선택형 혼합조사에서 최적결합방법과 기존의 방법으로 추정한 값을 같이 제시한 것이고, 〈표 5-11〉은 할당형 혼합조사에서 최적결합방법과 기존의 방법을 추정한 값을 제시한 것이다. 값을 살펴보면 기존추정값과 크게 차이가 발생하지는 않는 것을 알 수 있다. 특히 할당형의 경우 선택형에 비해 기존의 값과 차이가 발생하지 않고 있다. 이는 선택형일 때 두 조사의 결과에 차이가 크게 발생한다는 것을 의미한다.

〈표 5-10〉 소비자동향조사 선택형 혼합조사의 최적결합 추정값 비교

구분	항목	전화	온라인	결합 추정값 ①	기존 추정값 ②	차이 ①-②
현황	매출규모	70.39	71.85	71.11	70.72	0.39
	투자규모	87.80	75.98	82.77	85.12	-2.34
	외부 고용사정	90.76	79.77	86.19	88.26	-2.07
	채산성	78.13	76.69	77.38	77.80	-0.42
	자금사정	74.39	74.21	74.30	74.34	-0.04
	상품단가	102.72	88.42	97.45	99.47	-2.02
	전반적 업황	79.83	70.39	75.16	77.69	-2.53
전망	매출규모	93.56	97.14	95.21	94.37	0.84
	투자규모	95.00	82.10	90.65	92.07	-1.42
	외부 고용사정	99.60	86.58	95.63	96.65	-1.02
	채산성	94.86	86.08	90.98	92.86	-1.89
	자금사정	93.28	84.57	89.29	91.30	-2.01
	상품단가	101.39	95.00	99.69	99.94	-0.25
	전반적 업황	95.10	95.43	95.27	95.17	0.10

〈표 5-11〉 소비자동향조사 할당형 혼합조사의 최적결합 추정값 비교

구분	항목	전화	온라인	성향점수보정①	기존 추정값②	차이①-②
현황	매출규모	72.56	74.39	73.46	73.01	0.45
	투자규모	90.75	80.34	86.62	88.19	-1.57
	외부 고용사정	90.45	79.23	85.99	87.69	-1.71
	채산성	81.51	77.49	79.45	80.52	-1.07
	자금사정	78.09	74.06	76.14	77.10	-0.96
	상품단가	100.52	100.28	100.43	100.46	-0.03
	전반적 업황	80.74	80.19	80.47	80.60	-0.13
전망	매출규모	95.93	93.64	94.89	95.37	-0.48
	투자규모	98.14	87.40	94.35	95.50	-1.16
	외부 고용사정	97.78	87.68	94.49	95.30	-0.80
	채산성	97.85	87.86	93.42	95.39	-1.97
	자금사정	92.53	86.78	89.97	91.12	-1.15
	상품단가	102.60	103.42	102.82	102.81	0.01
	전반적 업황	96.30	94.96	95.69	95.97	-0.28

2) 성향점수를 이용한 추정

기업경기동향조사의 성향점수를 이용한 보정방법도 소비자동향조사의 성향점수를 이용한 보정방법과 같은 방법으로 추정한다. 기업경기동향조사도 기존에 진행해왔던 조사 방식이 전화조사 방식이기 때문에 기준조사(reference survey)를 전화조사방법, 비교조사를 웹조사 방법으로 설정한다. 조사방법이 2가지이므로 로지스틱 회귀분석을 적용하며, 예측값을 성향점수로 활용하도록 한다.

① 선택형 조사방법에서 성향점수 이용한 추정

기업경기동향조사의 성향점수를 얻기 위한 공변량은 종사자구간, 산업구분, 조직형태, 사업체구분, 응답자성별 5개를 이용한다. 모드효과를 선택하는 효과는 응답자의 성향이 더 중요하지만, 본 연구에서 사업체조사이기 때문에 사업체의 특성을 이용하여 분석하였다. 〈표 5-12〉에 성향점수모형에 들어갈 공변량과 범주, 각 조사방법별 빈도(frequency)를 제시하였다.

〈표 5-12〉 기업경기동향조사 선택형 혼합조사에서 공변량과 범주

사업체 특성		전화 조사	모바일 조사	사업체 특성		전화 조사	모바일 조사
종사자 구간	5~9명	174	44	조직 형태	개인사업체	62	18
	10~19명	121	39		회사 법인	262	73
	20~49명	90	23		회사 이외 법인	8	4
산업 구분	문화예술산업	104	37		국가/지방자치단체	41	8
	문화산업	81	19		비법인 단체	12	3
	스포츠산업	93	21	사업체 구분	단독사업체	161	41
	관광산업	107	29		본사(점)	61	21
성별	남성	149	37		지사(점), 영업소, 공장	48	15
	여성	236	69				

〈표 5-13〉는 로지스틱 회귀모형 추정결과를 제시한 것이다. 산업구분, 조직형태, 사업체구분, 응답자성별 4개 변수가 유의한 영향을 미치는 것으로 나타났다. 로지스틱 모형으로 나타난 예측값을 크기순으로 나열하여 5분위수(quintile points)를 이용하여 다섯 개의 구간으로 나누어 각각 20%의 데이터가 들어가도록 하였다.

〈표 5-13〉 기업경기동향조사 선택형 로지스틱 회귀모형 추정

성향변수		계수	Odds	유의확률
종사자 구간	5~9명	0.003	1.003	0.931
	10~19명	0.041	1.042	0.299
	20~49명			
산업 구분	문화예술산업	0.149	1.161	0.000***
	문화산업	-0.375	0.688	0.000***
	스포츠산업	-0.146	0.864	0.000***
	관광산업			
조직형태	개인사업체	0.182	1.199	0.060
	회사 법인	0.075	1.078	0.448
	회사 이외 법인	0.972	2.644	0.000***
	국가/지방자치단체	0.076	1.079	0.472
	비법인 단체			
사업체 구분	단독사업체	0.153	1.166	0.000***
	본사(점)	0.549	1.731	0.000***
	지사(점), 영업소, 공장			
응답자 성별	남성	-0.087	0.917	0.001**
	여성			

* p<0.05, ** p<0.01, *** p<0.001

〈표 5-14〉는 기업경기동향조사의 선택형 혼합조사에서 성향점수 보정한 값을 제시하고 기존의 값과 비교한 것인데, 기존추정값과 비슷하게 나타났다.

〈표 5-14〉 기업경기동향조사 선택형 혼합조사의 성향점수 보정 추정값 비교

구분	항목	전화	온라인	성향점수 보정 ①	기존 추정값 ②	차이 ①-②
현황	매출규모	70.39	71.85	71.09	70.72	0.37
	투자규모	87.80	75.98	85.60	85.12	0.49
	외부 고용사정	90.76	79.77	88.24	88.26	-0.02
	채산성	78.13	76.69	77.57	77.80	-0.23
	자금사정	74.39	74.21	74.63	74.34	0.28
	상품단가	102.72	88.42	99.63	99.47	0.16
	전반적 업황	79.83	70.39	78.02	77.69	0.33
전망	매출규모	93.56	97.14	94.43	94.37	0.06
	투자규모	95.00	82.10	92.12	92.07	0.05
	외부 고용사정	99.60	86.58	96.18	96.65	-0.46
	채산성	94.86	86.08	92.88	92.86	0.01
	자금사정	93.28	84.57	91.35	91.30	0.05
	상품단가	101.39	95.00	100.20	99.94	0.27
	전반적 업황	95.10	95.43	95.45	95.17	0.28

② 할당형 조사방법에서 성향점수 이용한 추정

할당형 조사에서도 공변량은 동일하게 종사자구간, 산업구분, 조직형태, 사업체구분, 응답자성별 5개를 이용한다. 〈5-15〉에 성향점수모형에 들어갈 공변량과 범주, 각 조사방법별 빈도(frequency)를 제시하였다.

〈표 5-15〉 기업경기동향조사 할당형 혼합조사에서 공변량과 범주

사업체 특성		전화 조사	모바일 조사	사업체 특성		전화 조사	모바일 조사
종사자 구간	5~9명	542	180	조직 형태	개인사업체	207	67
	10~19명	451	134		회사 법인	1,190	349
	20~49명	397	111		회사 이외 법인	23	30
	50~99명	180	51		국가/지방자치단체	218	58
	100인 이상	115	40		비법인 단체	47	12

사업체 특성		전화 조사	모바일 조사	사업체 특성		전화 조사	모바일 조사
산업 구분	문화예술산업	485	156	사업체 구분	단독사업체	725	208
	문화산업	468	133		본사(점)	296	104
	스포츠산업	300	97		지사(점), 영업소, 공장	192	67
	관광산업	432	130	성별	남성	687	212
					여성	998	304

〈표 5-16〉은 로지스틱 회귀모형 추정결과를 제시한 것이다. 산업구분, 조직형태, 사업체구분 3개 변수가 유의한 영향을 미치는 것으로 나타났다. 로지스틱 모형으로 나타난 예측값을 크기순으로 나열하여 5분위수(quintile points)를 이용하여 다섯 개의 구간으로 나누어 각각 20%의 데이터가 들어가도록 하였다.

〈표 5-16〉 기업경기동향조사 할당형 로지스틱 회귀모형 추정

성향변수		계수	Odds	유의확률
종사자 구간	5~9명	0.036	1.037	0.638
	10~19명	-0.038	0.963	0.630
	20~49명	-0.128	0.880	0.113
	20~99명	-0.140	0.869	0.146
	100인 이상			
산업 구분	문화예술산업	0.242	1.273	0.000***
	문화산업	0.094	1.098	0.011*
	스포츠산업	0.111	1.117	0.002**
	관광산업			
조직형태	개인사업체	0.150	1.162	0.084
	회사 법인	0.458	1.581	0.000***
	회사 이외 법인	1.879	6.546	0.000***
	국가/지방자치단체	0.003	1.003	0.975
	비법인 단체			
사업체 구분	단독사업체	-0.379	0.684	0.000***
	본사(점)	-0.168	0.846	0.000***
	지사(점), 영업소, 공장			
응답자 성별	남성	-0.015	0.985	0.510
	여성			

〈표 5-17〉은 기업경기동향조사의 할당형 혼합조사에서 성향점수 보정한 값을 제시하고 기존의 값과 비교한 것이다.

〈표 5-17〉 기업경기동향조사 할당형 혼합조사의 성향점수 보정 추정값 비교

구분	항목	전화	온라인	성향점수 보정 ①	기존 추정값 ②	차이 ①-②
현황	매출규모	72.56	74.39	72.91	73.01	-0.10
	투자규모	90.75	80.34	88.02	88.19	-0.17
	외부 고용사정	90.45	79.23	87.59	87.69	-0.10
	채산성	81.51	77.49	80.41	80.52	-0.11
	자금사정	78.09	74.06	76.98	77.10	-0.12
	상품단가	100.52	100.28	100.20	100.46	-0.26
	전반적 업황	80.74	80.19	80.53	80.60	-0.08
전망	매출규모	95.93	93.64	95.28	95.37	-0.09
	투자규모	98.14	87.40	95.44	95.50	-0.06
	외부 고용사정	97.78	87.68	95.21	95.30	-0.09
	채산성	97.85	87.86	95.35	95.39	-0.05
	자금사정	92.53	86.78	90.99	91.12	-0.13
	상품단가	102.60	103.42	102.61	102.81	-0.19
	전반적 업황	96.30	94.96	95.90	95.97	-0.07

다. 추정방법의 비교

전화조사와 모바일조사(웹조사) 방법으로 조사함으로써 생기는 모드효과의 차이를 고려하여 보정하여 추정하는 방법으로 결합추정방법과 성향점수를 이용한 방법을 살펴보았다. 여기서는 두 방법의 추정값의 차이와 분산값의 차이를 살펴보면서 추정방법을 살펴보고자 한다.

1) 소비자동향조사의 결과비교

〈표 5-18〉에서는 소비자동향조사의 선택형 혼합조사의 추정방법의 차이를 비교하였는데, 추정값의 차이는 결합추정방법이 더 크게 나타났으며, 성향점수의 차이가 작게 나타났다. 분산추정에서는 결합추정방법이 분산을 매우 낮게 만들어주기 때문에 조사의 안

정성을 높여준다는 것을 알 수 있다.

〈표 5-18〉 소비자동향조사 선택형 혼합조사의 추정방법 비교

구분	항목	추정값차이		분산추정			분산차이	
		기존 - 결합	기존 - 성향	기존 ①	결합 ②	성향 ③	①-②	①-③
현황	오프라인 문화생활	4.65	-1.43	1894.81	1016.86	1852.38	877.95	42.43
	오락시설 지출비용	8.55	-0.72	1669.08	853.35	1615.93	815.73	53.14
	온라인 문화생활비	0.29	2.30	1832.37	1014.38	1866.11	818.00	-33.74
	스포츠시설 이용료	6.53	-0.88	2396.71	1225.42	2370.66	1171.29	26.05
	스포츠용품 및 가구	5.07	1.04	1941.98	997.09	2024.80	944.89	-82.82
	스포츠 경기 관람료	3.15	-0.08	1063.75	576.83	1013.97	486.92	49.78
	아웃도어용품 구입비	6.29	0.06	1848.72	987.62	1814.85	861.10	33.87
	미디어기기 구입비	-1.56	3.18	1214.19	695.79	1401.01	518.40	-186.82
	국내여행비	4.21	-1.25	2110.72	1195.08	2032.58	915.64	78.13
	해외여행비	3.07	0.42	1272.84	634.01	1268.42	638.82	4.41
	도서구입비	0.25	2.09	1002.33	573.60	1090.21	428.74	-87.87
	오락용품 구입비	4.35	1.16	1464.35	816.25	1505.26	648.10	-40.92
	음악미술학원비	0.29	-0.03	752.43	354.67	757.00	397.76	-4.58
	스포츠 학원비	0.69	0.74	692.00	360.09	748.51	331.90	-56.51
	여행관광비	3.08	-0.07	1139.25	587.02	1078.13	552.24	61.12
	종합지출	1.36	0.63	409.72	211.39	436.88	198.33	-27.16
전망	오프라인 문화생활	-0.19	0.27	1179.45	685.11	1172.10	494.34	7.35
	오락시설 지출비용	1.73	0.29	933.46	539.95	900.40	393.51	33.06
	온라인 문화생활비	-1.49	1.73	754.01	436.62	773.63	317.39	-19.62
	스포츠시설 이용료	1.80	0.25	1125.68	653.63	1100.43	472.05	25.25
	스포츠용품 및 가구	1.34	0.32	1486.78	777.66	1483.57	709.12	3.22
	스포츠 경기 관람료	1.71	0.14	745.94	429.27	738.50	316.67	7.44
	아웃도어용품 구입비	2.20	0.63	957.71	543.50	971.90	414.22	-14.19
	미디어기기 구입비	-1.03	0.72	792.01	458.88	779.66	333.13	12.35
	국내여행비	2.61	0.59	1339.42	742.10	1331.08	597.32	8.34
	해외여행비	1.28	0.47	922.36	527.70	922.22	394.66	.14
	도서구입비	0.44	1.01	810.49	477.11	823.16	333.38	-12.67
	오락용품 구입비	1.07	-0.29	834.64	484.53	775.05	350.10	59.58
	음악미술학원비	0.57	0.76	504.22	272.79	516.45	231.43	-12.23
	스포츠 학원비	0.59	0.42	545.85	318.31	558.01	227.55	-12.16
	여행관광비	1.15	0.51	752.14	428.29	737.55	323.85	14.59
	종합지출	0.51	0.56	312.19	178.28	318.33	133.90	-6.15

〈표 5-19〉에서는 소비자동향조사의 할당형 혼합조사의 추정방법의 차이를 비교하였는데, 추정값에서는 두 방법간의 차이가 느껴지지 않았다. 분산추정에서는 결합추정방법이 분산을 매우 낮게 만들어주기 때문에 조사의 안정성을 높여준다는 것을 알 수 있다.

〈표 5-19〉 소비자동향조사 할당형 혼합조사의 추정방법 비교

구분	항목	추정값차이		분산추정			분산차이	
		기존 - 결합	기존 - 성향	기존 ①	결합 ②	성향 ③	①-②	①-③
현황	오프라인 문화생활	5.83	4.58	2195.78	1107.29	2188.07	1088.49	7.71
	오락시설 지출비용	7.93	6.68	1832.40	900.15	1842.03	932.25	-9.63
	온라인 문화생활비	3.05	3.21	1626.47	822.88	1645.44	803.58	-18.97
	스포츠시설 이용료	6.99	6.20	1731.59	854.98	1758.98	876.61	-27.39
	스포츠용품 및 가구	5.62	5.28	1663.01	828.49	1678.84	834.52	-15.83
	스포츠 경기 관람료	6.20	6.03	1361.05	650.19	1355.55	710.86	5.50
	아웃도어용품 구입비	4.65	4.46	1592.81	799.97	1602.66	792.85	-9.85
	미디어기기 구입비	2.84	2.97	1674.54	854.58	1695.68	819.96	-21.14
	국내여행비	7.40	6.19	2029.10	1008.37	2031.57	1020.73	-2.47
	해외여행비	7.17	8.95	1570.37	656.34	1563.17	914.02	7.20
	도서구입비	3.83	3.88	1492.43	748.37	1502.89	744.06	-10.46
	오락용품 구입비	4.14	5.20	1233.77	566.04	1259.68	667.73	-25.92
	음악미술학원비	4.08	6.44	1076.58	409.78	1081.28	666.80	-4.70
	스포츠 학원비	3.85	5.46	1022.60	423.50	1033.86	599.10	-11.26
	여행관광비	6.91	8.14	1298.61	557.63	1296.99	740.98	1.62
	종합지출	4.12	5.76	607.20	245.65	627.91	361.55	-20.71
전망	오프라인 문화생활	2.62	3.11	1399.00	696.05	1415.31	702.94	-16.31
	오락시설 지출비용	4.05	4.69	1263.91	594.36	1278.63	669.55	-14.72
	온라인 문화생활비	0.90	1.29	1005.07	490.89	1035.75	514.18	-30.68
	스포츠시설 이용료	4.10	4.58	1285.73	624.97	1309.58	660.76	-23.85
	스포츠용품 및 가구	3.42	3.69	1200.68	591.63	1215.91	609.05	-15.23
	스포츠 경기 관람료	4.65	5.19	1236.83	578.81	1246.28	658.02	-9.45
	아웃도어용품 구입비	3.22	3.87	1236.17	593.19	1259.77	642.98	-23.60
	미디어기기 구입비	2.05	1.95	1194.64	609.56	1205.88	585.08	-11.23
	국내여행비	4.07	4.29	1750.65	878.67	1774.10	871.99	-23.44
	해외여행비	4.22	6.20	1345.14	551.12	1341.65	794.02	3.49
	도서구입비	1.89	2.22	1094.11	537.56	1105.86	556.55	-11.76
	오락용품 구입비	3.07	4.28	1005.85	446.24	1027.78	559.61	-21.93
	음악미술학원비	2.74	4.55	885.88	330.91	881.80	554.97	4.08
	스포츠 학원비	1.97	3.74	850.31	306.98	856.17	543.33	-5.86
	여행관광비	3.88	5.64	1123.61	470.76	1131.98	652.85	-8.36
	종합지출	2.14	3.98	543.72	201.25	562.62	342.46	-18.90

소비자동향조사의 선택형과 할당형에서 결합추정방법이 분산을 낮게 만들어 안정적인 추정값을 제시하는 것으로 나타났다. 추정값에서는 선택형과 할당형의 결과에서 차이가 있는 것으로 나타났다. 모드효과보다 측정효과의 값이 크게 나타나기 때문에 분산을 줄여 측정결과의 효율성을 높여주는 결합추정방법이 현재조사결과에서는 더 타당한 것으로 보인다.

2) 기업경기동향조사의 결과비교

〈표 5-20〉에서는 기업경기동향조사의 선택형혼합조사의 추정방법별 차이를 비교하였다. 추정값에서는 성향점수를 이용한 보정방법이 기존값과 가깝게 나타났으며, 분산추정에서는 결합추정방법이 분산을 크게 안정시킨 것으로 나타났다.

〈표 5-20〉 기업경기동향조사 선택형 혼합조사의 추정방법 비교

구분	항목	추정값차이		분산추정			분산차이	
		기존 - 결합	기존 - 성향	기존 ①	결합 ②	성향 ③	①-②	①-③
현황	매출규모	-0.72	-0.70	2333.86	1176.31	2328.14	1157.56	5.72
	투자규모	5.03	2.20	1611.80	844.90	1594.40	766.90	17.39
	외부 고용사정	4.56	2.51	1185.09	622.77	1198.60	562.31	-13.51
	채산성	0.74	0.56	1879.30	922.92	1874.89	956.38	4.40
	자금사정	0.09	-0.24	1799.20	921.32	1762.82	877.88	36.38
	상품단가	5.26	3.09	1113.32	585.58	1129.49	527.74	-16.17
	전반적 업황	4.67	1.81	1775.62	884.75	1746.01	890.86	29.60
전망	매출규모	-1.65	-0.87	1564.59	810.47	1561.82	754.11	2.76
	투자규모	4.35	2.88	1065.22	563.22	1038.99	501.99	26.23
	외부 고용사정	3.97	3.42	839.81	436.32	861.62	403.50	-21.80
	채산성	3.88	1.98	1515.49	791.00	1509.40	724.49	6.09
	자금사정	3.99	1.93	1341.79	691.24	1329.37	650.55	12.42
	상품단가	1.69	1.18	889.87	462.77	846.56	427.10	43.31
	전반적 업황	-0.17	-0.35	1406.81	695.00	1395.89	711.80	10.92

〈표 5-21〉은 기업경기동향조사의 할당형 혼합조사의 추정방법별 비교를 한 것이다. 추정값에서는 성향점수를 이용한 방법이 기존값과의 차이가 적은 것으로 나타났으며, 분산에서는 결합추정방법이 더 안정적으로 나타났다.

즉 기업경기동향조사에서도 소비자동향조사와 거의 유사한 결과가 나타났다. 따라서 성향점수를 이용한 방법의 분산이 낮아지지 않았기 때문에, 성향점수를 이용한 추정방법의 결과를 활용하기는 어려울 것으로 보인다.

〈표 5-21〉 기업경기동향조사 할당형 혼합조사의 추정방법 비교

구분	항목	추정값차이		분산추정			분산차이	
		기존 - 결합	기존 - 성향	기존 ①	결합 ②	성향 ③	①-②	①-③
현황	매출규모	-0.45	0.10	2243.86	556.22	2242.44	1687.64	1.42
	투자규모	1.57	0.17	1567.65	328.37	1565.99	1239.28	1.66
	외부 고용사정	1.71	0.10	1178.93	245.78	1183.76	933.15	-4.82
	채산성	1.07	0.11	1686.15	425.43	1690.08	1260.72	-3.94
	자금사정	0.96	0.12	1735.03	425.63	1740.03	1309.40	-5.00
	상품단가	0.03	0.26	927.49	193.89	928.25	733.61	-0.76
	전반적 업황	0.13	0.08	1783.29	438.57	1787.12	1344.72	-3.83
전망	매출규모	0.48	0.09	1554.87	368.27	1557.73	1186.60	-2.87
	투자규모	1.16	0.06	1024.99	190.43	1027.93	834.56	-2.94
	외부 고용사정	0.80	0.09	832.69	141.40	832.66	691.30	0.03
	채산성	1.97	0.05	1278.12	292.42	1280.26	985.70	-2.14
	자금사정	1.15	0.13	1341.66	311.35	1344.07	1030.31	-2.41
	상품단가	-0.01	0.19	575.66	77.45	575.69	498.21	-0.03
	전반적 업황	0.28	0.07	1277.78	301.53	1277.98	976.25	-0.21

제3절 소결 및 시사점

1. 소결

3장에서는 혼합조사의 선행연구와 사례들을 검토하였는데, 특히 최근의 조사환경이 점차 어려워지고 있어, 전통적인 전화조사 또는 가구방문조사 등의 단일방법이 아닌 여러 방법을 혼합하여 조사를 진행하는 혼합조사에서 각 조사방법의 효과의 차이를 파악하는 방법을 설명하였다. 모드효과는 선택효과와 측정효과로 구분되기 때문에 이 두 효과를 파악하여야 한다. 각 효과를 제대로 파악하기 위해서는 동일한 조사를 동일한 조사 대상에게 다른 조사방식으로 두 번 조사할 수 있다면 정확한 효과를 파악할 수 있지만, 이는 현실적으로 불가능하다. 따라서 기준조사(reference survey) 또는 예고한 것과 다른 방법으로 조사한 결과가 있다면 각 효과를 측정할 수 있다.

기존에 했던 방식이 대면조사(㉠)이고, 새로운 방식의 인터넷조사한 결과(㉡), 기준조사결과 또는 인터넷조사로 진행하기로 하고 대면조사를 진행한 결과(㉢)가 있다면 각 효과를 다음과 같은 식으로 측정할 수 있다

$$\blacksquare \text{ 모드효과} = \text{㉡} - \text{㉠}$$

$$\blacksquare \text{ 선택효과} = \text{㉢} - \text{㉠}$$

$$\blacksquare \text{ 측정효과} = \text{㉡} - \text{㉢}$$

따라서 모드효과를 분석하기 위해서는 위와 같은 방식으로 조사연구를 진행하면 될 것이다.

한편, 4장에서는 실제 혼합조사에서 발생하는 모드효과를 분석하고, 5장에서는 두가지 추정방안을 이용해 실제 추정값을 제시하였다. 혼합조사는 응답자가 원하는 방식으로 조사를 진행하기 때문에 적은 비용(예산, 시간 등)으로 높은 응답률을 가져오는 방식으로 조사를 진행할 수 있을 것이다. 그러나 선택효과와 측정효과 등의 모드효과로 오는 차이가 있기 때문에 두 값을 동일한 가중값을 적용해서 측정하는 것은 적절하지 못하다. 혼합

조사 추정방법으로 가장 많이 사용되는 것은 최적결합방법과 성향점수를 이용한 방법이 있다.

최적결합방법은 분산이 적은 조사방법이 안정된 방법이라는 가정하에 분산이 큰 조사방법은 적은 비율을 적용하고, 분산이 작은 조사방법은 많은 비율을 적용하여 결합하여 추정하는 방법이다. 분산의 관점에서는 매우 효율이 높으며, 결과값도 안정적으로 제공하며, 무엇보다도 쉽게 산출할 수 있는 장점이 있다.

성향점수를 이용한 추정방법은 기준조사가 있을 경우, 다른 조사방법을 기준조사방법으로 조정하여 산출하는 방법이다. 이를 위해서는 적절한 공변량이 있어야 하는 문제가 있다. 공변량이 적절이 있다면 이를 이용하여 값을 잘 보정할 수 있지만, 공변량이 적절하지 않다면 보정에는 한계가 있을 수 있는 방법이다.

본 연구에서 사용한 소비자동향조사와 기업경기동향조사는 각각 선택형과 할당형으로 전화조사와 모바일조사(웹조사)를 진행한 결과를 이용하여 모드효과와 추정방법을 적용하였다. 모드효과에서는 선택효과보다는 측정효과가 크게 나타났으며, 전화조사가 모바일조사(웹조사)보다 크게 나타났다. 이는 면접원효과가 포함된 것으로 판단된다. 면접원의 설명이 있는지에 따라, 상대방이 있기 때문에 편하게 응답하지 못하는 등의 이유가 포함된 것으로 판단된다.

추정방법에서는 모두가 최적결합방법이 적절한 것으로 나타났다. 최적결합방법 결과는 분산을 크게 낮춤으로써 효율성 높은 결과를 제시하는 것으로 나타났다. 성향점수를 이용한 추정방법은 조사의 특성을 유지하기 위해서 기존의 조사방법으로 진행되었기 때문에 공변량이 충분하지 않았던 문제가 있다. 따라서 최적의 응답자 특성을 반영하여 보정하는 것이 어려움이 있는 것으로 보인다.

2. 시사점

혼합조사로 이루어진 조사방법은 각 조사방법으로 인해 차이가 발생하는 것을 알 수 있다. 따라서 조사한 값을 그대로 반영한다면 조사방법들 중에서 어떤 방법을 더 많이 활용했는지에 따라서 차이가 발생할 수밖에 없다. 이러한 문제를 해결하기 위해서는 모드효과를 파악하고 적절한 추정방법을 적용해야 할 것이다.

모드효과를 파악하기 위해서는, 선택효과와 측정효과 등을 추정할 수 있도록 설계를 하는 것이 중요하다. 이때 기준조사(reference survey)가 있는지? 있다면 어떻게 기준 조사를 활용하는지에 대한 고려도 필요하다.

추정방법은 보통 최적결합추정방법을 사용하는 것이 좋을 것이다. 최적결합추정방법은 특별한 제약조건없이 사용될 수 있으며, 그 결과값은 항상 분산을 줄여주기 때문에 효율성을 높여준다. 성향점수를 이용한 추정방법을 하기 위해서는 응답자의 성향을 파악할 수 있기 위한 충분한 공변량이 있어야 하는데, 설문내용이 긴 조사에서 응답사례가 있는 조사에서 가능할 것이다. 그러나 성향점수를 이용한 추정방법을 잘 적용할 수 있다면, 새로운 조사방법을 기준조사와 동일한 성향을 가진 응답자의 결과처럼 조정할 수 있기 때문에 동일한 조사의 결과와 유사한 효과를 가져다 줄 것이다.

혼합조사는 쉽게 적은 비용으로 조사를 할 수 있는 방법이지만, 각 방법별로 효과의 차이가 있기 때문에, 모드효과를 잘 파악할 수 있어야 정확한 추정이 가능하다. 따라서 조사를 진행하기 전에 어떻게 조사를 진행할지에 대한 충분한 조사설계 방안을 마련할 필요가 있다.

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

제6장

결론 및 제언

제1절 조사설계 개선 방안

전통적으로 자료를 수집하는 주요 방법이었던 설문조사(survey)는 연구자가 표본설계(sampling design)를 하여 조사하기 때문에, 객관적인 통계를 작성한다는 장점이 있지만 자료를 수집하고 통계를 산출하는데 시간이 많이 걸리고 예산비용도 많이 사용되는 단점이 있어왔다. 또한 많은 조사통계는 시의성이 늦다는 지적이 있어왔는데, 정보기술(IT: information technology)의 발전으로 개인정보 활용의 문제가 대두되며 설문조사의 제약이 생겨, 공공의 행정자료와 민간의 빅데이터로 많은 통계가 대체되고 있다.

이러한 이유로 조사통계가 감소할거라 예상되어 왔지만, 공공을 기반으로 한 사회조사는 민간에서의 조사와는 다르게 기존의 조사체계가 유지되고 있다. 하지만 전통적인 조사방식이 IT와 융합하면서 온라인 조사가 발전하고 있으며, 빅데이터가 등장하면서 이와 조사를 연계하여 설계함으로써 정보의 활용성을 높이는 방향의 시도들이 이루어지고 있다. 이렇듯 조사방식이 다양해지고 발전하면서 시의성이 높아지고 있지만, 사회조사의 기본이 되는 대표성과 객관성을 담보할 수 있는지에 대한 우려도 높아지고 있다. 따라서 조사방법별 결과의 차이 분석과 모드효과를 파악하는 등 이에 대해 보완하는 연구가 진행되어왔다.

2020년에 나타난 코로나19의 영향은 전통적인 조사방법의 주를 이루는 대면조사를 하기 어려운 상황으로 만들었다. 이러한 이유로 IT기반의 웹조사 방식으로의 조사가 증가하였지만, 다양한 이유로 웹조사의 대표성은 의심을 받고 있는 실정이다.

본 연구에서는 대표성 있는 조사방법을 진행한다는 가정 하에서, 사회조사방법의 변화를 제시하였다. 대표성 있는 조사통계를 생산하기 위해서는 기본적으로 표본틀(sample frame)이 있어야 한다. 표본틀이 있어야 연구자가 모집단을 대표성 있게 추정할 수 있도록 표본설계를 할 수 있다.

조사방법은 추출된 표본, 즉 응답자들을 접할 수 있는 정보가 있는지에 따라서 결정된

다. 추출된 대상을 사전에 접할 수 있는 전화번호나 이메일 등의 정보가 있다면 응답자들이 원하는 방법으로 조사를 할 수 있다. 그러나 대부분의 국민대상 조사는 조사구를 추출하여 요도⁷⁴⁾를 통계청에서 받고, 가구방문을 통해서 조사가 진행된다. 그렇기 때문에 전통적인 조사방식인 대면조사 방식으로 조사가 어려울 경우, 전화번호나 이메일 등의 정보가 있다면 이를 이용해 사전접촉을 하여 혼합조사방법을 적용할 수 있을 것이다. 하지만 가구방문조사의 경우, 사전에 컨택 할 수 있는 정보가 없기 때문에 이를 구축할 수 있는 방법이 필요하다.⁷⁵⁾ 전자는 혼합조사 설계 방안으로 후자는 조사패널방안 구축을 통해 진행할 수 있을 것이다.

1. 혼합조사 설계 방안

혼합조사의 경우에는 사전에 조사대상자 명부가 있는 경우, 즉 선정된 조사대상자에게 연락한 후, 다양한 조사방식 중에서 선택하도록 하여 조사를 진행하는 방법이다. 따라서 표본설계 후, 표본으로 선정된 대상에게 연락할 수 있는 개인정보를 포함한 표본들이 있는 경우에 활용할 수 있는 방법이다.

따라서 법에 명시된 조사이거나 개인의 동의를 받은 대상들을 모아 조사패널(survey panel)을 구축하였을 때 활용할 수 있다. 법에 명시된 조사로는 ‘예술인실태조사’가 있는데, 「예술인 복지법」제4조의3 제2항에 ‘문화체육관광부장관은 제1항에 따른 실태조사 실시를 위하여 필요한 경우 관계 중앙행정기관·지방자치단체 및 공공기관의 장, 예술인과 계약을 체결한 개인, 법인 또는 단체 등에게 관련 자료의 제출을 요구할 수 있다.’ 라고 명시되어 있다. 이 경우 ‘자료의 제출을 요구받은 자는 특별한 사유가 없으면 이에 따라야 한다’고 명시하고 있어, 「개인정보보호법」제18조 제2항에 따라, 실태조사를 위한 예술인들의 정보를 제공받을 수 있도록 하였다. 제공받은 예술인들의 정보를 가지고, 예술분야, 지역, 성별, 연령 등을 고려하여 표본을 추출한다. 표본으로 선정된 예술인들에게 연락을 취해서, 원래의 조사방법인 대면조사 방법을 권유하고, 어려울 경

74) 조사구내의 가구 위치를 파악할 수 있도록 작성한 지도, 즉 조사의 편의를 위해 작성된 지도임

75) 정보가 없다면 어떠한 방식으로든 대면접촉이 일어날 수밖에 없다. 따라서 비대면 조사를 위해서는 최소한 조사대상으로 선정되었음을 알릴 수 있는 정보가 있어야 함

우 웹조사방식으로 진행하고 있다.

또 다른 방식인 조사패널을 이용한 방식은, 조사대상자를 모집하여 조사를 위한 개인 정보 활용에 대한 동의를 구한 후, 이들을 대상으로 표본을 선정하여 조사를 진행하는 방법이 있다. 현재 조사업체들이 하고 있는 온라인 패널조사가 이러한 방식으로 대상자를 모집하여 조사하고 있다.

조사대상자의 정보를 이용하여 조사를 수행할 경우, 표본으로 뽑힌 조사대상자가 응답자가 원하는 방식으로 조사를 진행할 수 있기 때문에, 응답률이 높은 시의성있는 통계를 생산할 수 있다. 이러한 혼합조사방식은 응답자를 접하기 어려운 사회환경에서 다양한 매체를 이용하여 조사할 수 있어, 최근조사에서는 혼합조사방식을 많이 도입하여 활용하려고 노력하고 있다.

그러나 혼합조사를 사용할 경우, 어떤 조사방식을 선택했는지에 따른 선택효과, 측정 과정에서 오는 측정효과 등 조사방식별로 모드효과가 다르게 나타난다는 것을 많은 사례연구에서 제시하고 있다. 이는 본 연구의 4장에서도 확인하였다. 따라서 혼합조사를 사용하여 조사할 경우 모드효과를 파악할 수 있는 방법으로 측정할 필요가 있다.

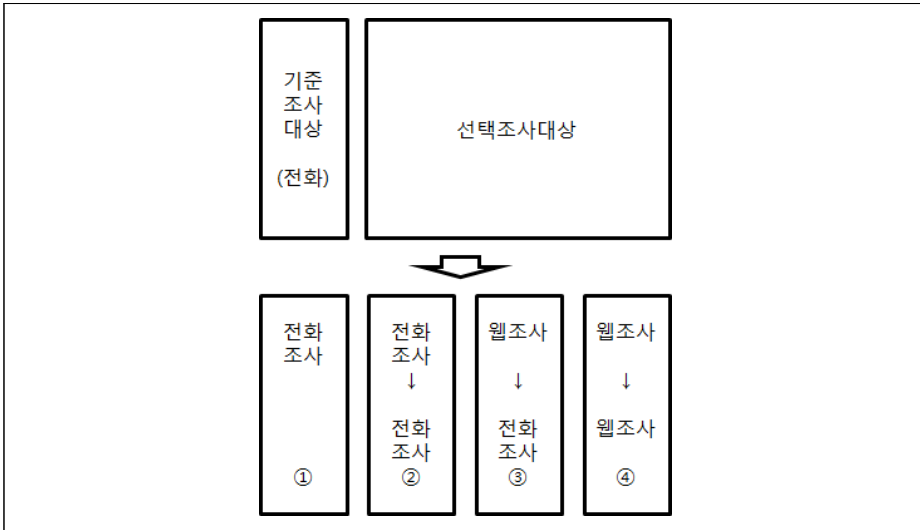
모드효과를 파악하고, 다양한 기준으로 보정하기 위해서는 다음과 같이 조사를 수행하는 것을 제안하도록 한다. 우선적으로 기준이 되는 조사방식을 선정한다. 원래 조사하던 방식이 있다면 이를 기준조사(reference survey)로 선정한다. 만약에 원래 조사하던 방식이 없다면 전통적인 조사방식인 대면조사나 전화조사를 기준조사로 선정하도록 한다.⁷⁶⁾ [그림 6-1]은 혼합조사 방법의 조사대상 선정을 표현한 것인데, 설명을 위하여 전화조사와 웹조사 2가지 방법을 고려하였다.⁷⁷⁾

표본을 선정한 후, 기준조사대상과 선택조사대상을 선정한다. 기준조사대상은 기존의 조사방식으로 조사를 하며, 선택조사대상은 전화조사와 웹조사 중에서 본인이 선택할 수 있도록 한다. 그러나 대안조사인 웹조사를 선택한 대상중에 일부는 양해를 구해 전화조사를 실시하도록 한다. 이렇게 선정하여 조사를 하게 되면, 조사결과를 이용하여 모드효과를 분석하고, 최적결합추정법과 성향점수를 이용한 추정법을 적용하여 추정을 할 수 있다.

76) 이는 조사원이 조사의 정확함 지침에 따라 설명하고 응답을 받는 타계식 방법으로 응답과정에서 발생하는 비표본오차가 적게 나타난다고 알려져 있기 때문이다

77) 이는 4장의 실험과 연결하여 설명하기 위해 전화조사와 웹조사로 정하였음

[그림 6-1] 혼합조사 대상 선정 방법



모드효과 분석은 다음과 같이 할 수 있다.

- 모드효과: ④의 추정값 - (①+②)의 추정값
- 선택효과: ③의 추정값 - (①+②)의 추정값
- 측정효과: ④의 추정값 - ③의 추정값

혼합조사의 추정방법 중에서 최적결합추정은 다음과 같이 구할 수 있다.

- 최적결합추정값 = (④의 분산)/((①+②+③의 분산)+④의 분산)×(①+②+③의 추정값) + (①+②+③의 분산)/((①+②+③의 분산)+④의 분산)×(④의 추정값)

성향점수를 이용한 추정방법은 대안조사인 웹조사 결과를 기준조사인 전화조사방식과 동일한 결과를 제시하기 위한 방법이다. 따라서 성향점수를 잘 적용할 수 있도록 적절한 공변량을 측정하여야 한다. 공변량을 측정한 후, ①의 공변량과 ④의 공변량을 이용하여 성향점수를 만들어 웹조사방법의 결과를 보정하도록 한다.⁷⁸⁾ 만약 표본이 적거나, 성향점수를 이용한 보정방법을 적용하지 않는다면 [그림 6-1]의 ①은 고려할 필요가 없다.

78) ①+②+③을 합하여 기준조사대상의 역할을 하도록 하는 방법도 있다. 그러나 기존의 방법과 동일한 방법을 적용하고자 할 경우에는 ①의 대상을 선정할 필요가 있다.

2. 문화·관광 분야의 조사 표본틀 구축 방안

가. 문화·관광 분야의 표본틀 활용 현황

현재 문화·관광 분야의 공식통계로 승인받은 조사통계의 조사대상은 국민, 사업체와 특정대상 이렇게 3가지 영역으로 구분할 수 있다. 우선, 특정대상 조사를 살펴보면, 14개 예술분야에 종사하는 예술인들을 대상으로 조사하는 ‘예술인실태조사’, 법정 장애 유형 중에서 청각장애 1-3등급에 해당하는 20세 이상 등록 장애인을 조사하는 ‘한국수어 활용조사’, 한국을 방문 후 출국하는 외국인을 대상으로 하는 ‘외래관광객조사’, 그리고 ‘관광진흥법’에 의거 등록·허가·신고·지정된 7개 업종에 해당하는 사업체를 대상으로 하는 ‘관광산업통계’가 있다. 이중에서 ‘예술인실태조사’는 「예술인복지법」제4조의 3에 명시되어 있으며, ‘한국수어활용조사’는 「한국수화언어법」제9조에 명시되어 있고 ‘관광산업통계’는 「관광진흥법」제47조의 2에 명시되어 있다. 법에 명시되어 있는 조사들은, 조사를 위한 자료를 공공기관에 요청하여 제공받을 수 있도록 되어 있다.⁷⁹⁾ 따라서 조사 대상에 대한 정보를 제공받아 표본틀(sample frame)을 구축하고, 조사를 위한 목적으로 조사대상에게 연락한 후, 설문조사에 대한 응답을 받기 위한 방법과 일정을 조율하고 조사를 진행한다.⁸⁰⁾

국민과 사업체의 승인통계 경우, 표본추출을 위한 표본틀은 통계청에서 구축한 MDIS(Microdata Integrated Service, 마이크로데이터 통합서비스)에 신청하여, 일정 비용을 지불하고 이용하고 있다. 이중에서 사업체의 경우는 표본설계를 하여 추출된 사업체들의 정보에 전화번호, 주소 등을 정보를 제공받아 조사대상이 되는 사업체를 조사할 수 있도록 하고 있다.

국민을 대상으로 하고 있는 조사는, 개인을 대상으로 하는 조사나 가구를 대상으로 하는 조사와 상관없이 조사구를 추출하여 선정된 가구를 방문하여 조사하도록 되어 있다.⁸¹⁾ 따라서 표본으로 추출된 조사구의 요도를 통계청으로부터 제공받아, 사전에 규칙

79) 이때 제공된 조사대상의 정보는 법에 명시된 조사에 한해서만 사용하여야 하며, 그 외의 목적으로 사용할 경우 「개인정보보호법」에 위배됨

80) 대부분의 조사는 기본조사방안이 있기 때문에, 특별한 사정이 있을 경우를 제외하고는 기준조사방법으로 조사를 진행하고 있음

81) 조사구 표본추출은 통계청의 MDIS를 이용하며, 2021년 9월 현재 활용하고 있는 표본틀은 ‘2019년 인구총조사’ 자료임

을 정해 선정된 가구를 요도에 체크해서 조사를 진행하고 있다. 국민을 대상으로 하는 조사의 경우는 표본추출단위가 조사구이고, 조사단위가 가구에 사는 개인이기 때문에 조사단위에 대한 정보를 제공받지 않고 있다. 이러한 이유로 응답대상과 접할 수 있는 방법은 가구를 방문하여 대면조사를 하는 방법 아니면 없다.

공식통계에서 표본들은 두 가지 방법으로 활용된다는 것을 알 수 있다. 첫 번째는 법령에 조사를 명시하고 조사를 위한 자료를 제공받을 수 있도록 하는 것이다. 두 번째는 통계청의 MDIS서비스를 제공받는 방법이다. 그러나 두 방법 모두 활용에 한계가 있다. 먼저 법령에 있는 조사의 경우, 해당연도의 조사에만 활용하고 폐기하여야 한다는 것이다. 따라서 해당조사연도에 조사대상자를 요청하고, 자료를 취합하여 표본들을 구축하고 조사를 한 후에는 폐기하여야 한다.

국민을 대상으로 하는 조사는 조사구를 이용하기 때문에, 반드시 가구를 방문하여 해당 가구원⁸²⁾을 조사해야 한다. 따라서 코로나19와 같은 문제로 대면조사가 어려운 경우에 활용할 수 없다는 문제가 있다. 즉, 조사대상자와 만나 조사방법과 조사일정 등을 조율할 수가 없다. 따라서 대면조사가 어려운 상황에서 조사구를 이용한 조사는 진행되기 어렵다.

문화와 관광분야는 사회가 발전할수록 중요하게 생각되는 분야이며, 문화와 관광의 정책이나 향유 등의 정보를 알고자 하는 요구가 증가하고 있다. 따라서 국민들에게 정책에 대한 인식이나 만족도, 국민들이 향유하는 내용이 무엇이고 얼마나 활용하는지 등을 시의성 있게 국민들의 의견을 파악할 필요가 있다. 이러한 경우 승인받은 조사가 아니기 때문에, 통계청의 MDIS를 활용하지 못하고, 활용할 수 있다 하더라도 MDIS 활용절차와 가구방문 조사를 통한 통계가 나오는 기간이 많이 걸리며, 비용도 많이 들게 된다. 따라서 대부분의 시의성이 필요한 비승인 조사들은 RDD(random digit dialing)방식의 전화조사나 조사업체의 온라인패널을 이용하여 진행되고 있다.

그러나 이러한 조사들의 경우, 응답률이 낮고 대체율이 높은 이유로 선택편의(selection bias)⁸³⁾가 발생한다. 특히 정량적인 조사결과와의 차이는 매우 크게 발생하는 것으로 알려져 있다. 따라서 이들 통계는 인식조사와 같은 정성적인 질문의 조사, 정량의 경우 변화의 관점에서 통계를 활용하는 것이 적절하다.

82) 가구원 중에서 대상을 선정하는 가장 많이 활용하는 방법으로 생일자법이 있는데, 가구원들 중에서 생일이 가장 빠른 대상으로 하는 방법과 조사시기와 가장 가까운 가구원을 선정하는 방법이 있음

83) 연구자(또는 조사자)가 응답자들을 통제하지 못함으로써 생기는 편의. 즉, 응답자와 응답거절자간의 성향의 차이로 인해 생기는 결과값의 차이

나. 문화·관광 분야의 표본틀 구축 및 활용 방안

문화·관광분야의 표본틀을 지속적으로 활용하기 위해서는 표본틀을 별도로 관리할 필요가 있다. 우선 법상에 있는 통계의 경우, 조사를 위한 자료를 제공받을 수 있도록 법에 명시되어 있지만, 매년 조사가 끝나면 폐기해야 하기 때문에, 매년 새롭게 조사대상의 정보를 받아 활용해야 한다. 따라서 표본틀을 관리할 수 시스템을 마련하고, 조사를 위해 관리할 수 있는 제도가 마련될 필요가 있다.

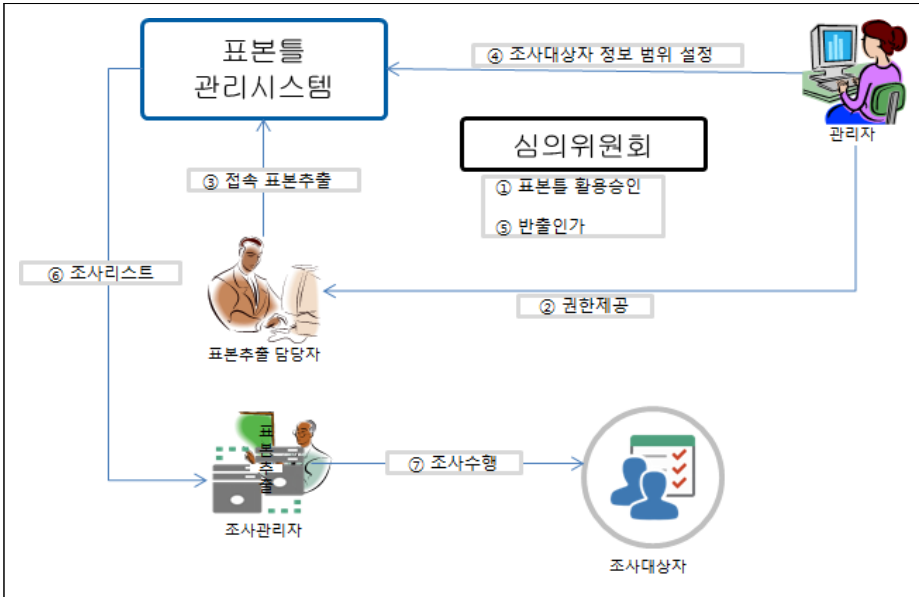
현재 사업체 표본틀의 경우, 한국문화관광연구원이 문화·체육·관광 관련 산업의 분류체계를 관리하고 있으며, 이를 기반으로 표본틀을 구축하여 활용하고 있다.⁸⁴⁾ 따라서 사업체 관련된 조사의 경우, 조사 설계 이후 빠르게 표본을 추출하여 조사를 착수할 수 있다.

향후 문화·관광 분야의 원활한 조사를 위하여 표본틀 활용할 수 있는 시스템을 구축할 필요가 있다. [그림 6-2]는 표본틀 관리시스템을 구축하였을 경우, 이를 활용하는 과정을 도식화한 것이다. 번호순서대로 진행을 하게 되는데, 이들을 각각 설명하면 다음과 같다.

- ① 표본틀에는 개인정보가 있기 때문에 표본틀 관리시스템 활용을 위한 심의위원회를 구성하도록 하며, 심의위원회는 표본틀 활용을 위한 요청을 받으면 이를 승인하는 절차를 거쳐 승인을 함
- ② 표본틀 활용승인이 나면, 관리자는 표본추출 담당자에게 시스템을 활용할 수 있는 권한을 줌
- ③ 표본추출 관리자는 표본틀 관리시스템에 접속하여 표본추출을 하며, 필요한 추가 정보를 관리자에게 요청함
- ④ 관리자는 표본추출 담당자가 요청한 정보의 범위를 설정하고 이를 제공할 수 있도록 하며, 이를 반출해도 되는지 심의위원회에 심의를 요청
- ⑤ 심의위원회에서는 반출정보에 대한 부분을 심의한 후 반출여부를 결정
- ⑥ 반출승인이 났을 경우 조사리스트를 조사관리자에게 제공
- ⑦ 조사관리자는 조사방법에 맞게 조사원 또는 온라인시스템을 이용하여 조사대상자들에게 조사를 진행하여 응답을 받음

84) 현재 이를 기반으로 '스포츠타산업조사'와 '광고산업조사', '문화체육관광 일자리조사' 그리고 '문화체육관광 기업경기동향조사' 등에서 활용하고 있음

[그림 6-2] 표본틀 활용과정



다. 문화·관광 분야의 국민대상 조사패널 구축

국민대상의 조사는 통계청의 MDIS 서비스를 이용하는 방법과 RDD방식의 전화조사, 조사업체의 온라인패널을 이용한 조사가 있다. 이 중에서 RDD방식의 전화조사는 국민들 모두에게 전화가 걸릴 수 있다. 이런 이유로, 모집단의 포괄성을 인정한다고 하더라도⁸⁵⁾ 연구자(조사설계자)가 조사대상자를 전혀 통제하지 못하는 상태로 조사되어지기 때문에, 선택편의(selection bias)가 크게 발생할 수밖에 없다. 또한 응답 거절자에 대한 정보가 없기 때문에 대체된 응답자들과의 유사성도 파악하기 어려움이 있다.

조사회사의 온라인패널은 패널을 모집할 때, 자발적 참여자들로 구성되었기 때문에, 연령별로 구성원 비중의 차이가 크며⁸⁶⁾, 거주 지역도 차이가 있는데 주로 도시에 거주하는 사람들의 비율이 높게 구성되어 있다. 무엇보다도 자발적 참여자에 의하여 구축된 패널이기 때문에 구축할 때부터 기본적으로 선택편의가 있게 된다.

통계청 MDIS의 ‘인구총조사’ 자료를 이용하는 것은 가구에 대한 다양한 정보를 분석

85) 이렇게 표현하는 이유는 RDD방식의 전화조사가 할당된 대상에 대한 조사가 끝나면 완료되는 방식이기 때문에 정확하게는 포괄한다고 보기 어렵기 때문이다

86) 조사회사에서 운영하고 있는 온라인 패널은 10대와 60대 이상 패널의 수가 상대적으로 적음

하여 활용하기 때문에, 대표성은 매우 좋으며 연구에 따라 다양한 설계가 가능하다는 장점이 있다. 그러나 승인통계에 한해서만 사용할 수 있으며, 조사구를 추출하여 가구를 직접 방문해야 하기 때문에, 비대면으로 조사할 수 없다는 한계가 있다.

이러한 문제를 해결하기 위한 방법으로, 대표성 있는 마스터샘플을 구축하여 조사패널로 활용하는 방안을 제시하고 있다. 마스터샘플의 장점은 국민을 대표할 수 있도록 구축하고, 동의를 받고 개인의 정보를 활용하기 때문에 다양한 연구에 활용할 수 있다는 것이다. 또한 마스터샘플의 변동이 생기더라도 이를 모집단의 특성에 맞도록 보정을 할 수 있기 때문에 매우 정교한 추정을 할 수 있다.⁸⁷⁾ 조사패널로 마스터샘플을 구축하는 방법은 두 가지로 생각할 수 있다.

첫 번째는 통계청과 협의 하에 전국의 일정비율의 조사구를 추출하여, 조사구에 있는 대상들에게 조사패널에 취지를 설명하고, 참여동의를 한 대상으로 구축하는 방법이다.⁸⁸⁾ 두 번째는 현재 진행하고 있는 가구방문조사⁸⁹⁾에서 응답을 받을 경우, 조사패널 참여에 대한 정보를 얻어 패널을 구축하는 방법이다.

두 방법 모두 랜덤하게 표본설계를 통해 추출된 대상들에게 동의를 구하기 때문에 대표성 있게 마스터샘플을 마련할 수 있게 된다. 첫 번째 방법은 큰 모집단을 구축할 수 있기 때문에 더욱 신뢰할 수 있는 조사를 위해 활용할 수 있지만, 구축하는 초기비용과 패널을 유지하기 위한 비용이 크게 들게 된다. 두 번째 방법은 큰 표본의 조사에 사용하기는 어렵지만, 1~2천명 단위의 전국단위 국민대상조사에서 활용하기에는 어려움이 없을 것이다. 구축비용과 유지보수 비용이 들어가지만, 첫 번째 방법보다는 적게 든다.

현재 큰 규모의 조사는 승인통계이기 때문에 통계청의 MDIS를 이용하여 조사한다면, 두 번째 방법으로 마스터샘플을 구축하는 것이 활용하고 유지하기에 적절할 것이다. 문화·관광분야의 조사패널인 마스터샘플을 구축한다면 시의성 있으면서 대표성 있는 조사를 수행할 수 있을 것이다.

87) 이는 표본조사론의 이중추출법(double sampling, two-phase sampling)의 이론을 기반으로 제시된 것임

88) 보통 10% 이내의 조사구를 추출하여, 전화나 이메일의 정보를 제공받아 동의를 받도록 함

89) 현재 진행하고 있는 문화·관광 분야의 가구방문조사 규모는 ‘국민여행조사’ 50,000표본, ‘국민여가활동조사’ 10,000표본, ‘국민문화예술활동조사’ 10,000표본, 국민독서실태조사 6,000명으로 전체 약 76,000 표본임

제2절 조사방식 개선 방안

민감한 질문이 있는 설문조사의 경우, 중요한 질문에서 결측값이 빈번하게 발생한다. 결측값을 허용하지 않는다면 응답거절이 증가하게 된다. 민감한 질문의 경우를 살펴보면, 개인은 사생활에 해당하는 질문이 이에 해당되며, 사업체의 경우는 매출 관련된 세부적인 질문이 해당된다.

민감한 질문이 있을 경우 조사도 어렵지만, 조사한 후에도 응답에 대한 신뢰에 대한 의문이 따르게 된다. 민감한 질문이 포함된 조사의 경우에는 조사기간도 길어지고 설문 조사 1부당 단가도 높게 책정된다. 이러한 민감한 질문을 하지 않고 활용할 수 있는 정보가 있다면, 조사도 쉬워지며 조사결과에 대한 신뢰도 높아질 것이며, 활용도 역시 높아질 것이다.

행정자료는 우리가 알고 싶은 정보도 있지만, 개인정보도 가지고 있다. 개인정보는 함부로 이용하게 되면 개인정보보호법에 위배되는 문제가 발생하지만, 이를 잘 활용할 수만 있다면 효율적으로 조사를 할 수 있는 방안을 마련할 수 있다. 현재 국민을 대상으로 하는 승인통계 조사는 가구방문조사 방식으로 진행되고 있다. 행정자료를 이용할 수 있다면 표본틀을 구축하고, 추출된 개인에게 조사대상으로 선정되었다고 이메일이나, 문자, 우편 등으로 알려주고, 온라인조사 등을 활용하여 조사를 할 수 있을 것이다.

이번 절에서는 민감한 정보가 있는 경우, 행정자료(빅데이터 포함)를 조사에 활용하는 방법과 행정자료를 표본틀로 활용하여 조사에 활용하는 방법을 적용한 조사방식의 개선 방안을 제시하고자 한다. 행정자료를 이용하기 위해서는 법이나 제도 등의 정비도 따라주어야 하지만, 본 연구에서는 조사방식의 개선에 대한 내용만을 다루고자 한다.

1. 행정자료를 활용한 조사 방안

통계청에서는 행정자료를 활용할 수 있는 통계는 적극적으로 활용하는 것을 권장하고 있다. 「통계법」에서는 통계작성 할 때 행정자료 활용가능한지를 판단하고, 해당기관에 요구할 수 있다고 되어 있다. 따라서 행정자료를 활용할 수 있다면 적극적으로 활용할 수 있는 방안을 마련할 필요가 있다.

행정자료를 활용하는 방법은 크게 두 가지의 경우로 구분하여 생각할 수 있다. 첫 번째는 매크로데이터(macro data)를 이용하는 방법이다. 매크로데이터는 수집된 자료를 집계 및 요약한 자료를 말한다(박원환 외, 2004). 매크로데이터로 활용할 질문은 제외하고 설문조사를 수행한 후 그 결과에 매크로데이터의 통계를 결합하는 방법이다. 이는 통계값 자체를 활용하기 때문에 큰 어려움 없이 결합하여 활용할 수 있는 장점이 있다.

두 번째는 마이크로데이터(micro data)를 활용하는 방법이다. 마이크로데이터는 수집 자료에 포함되어 있는 올바르지 못한 내용을 수정한 유효자료(valid data)를 말한다(박원환 외, 2004). 즉 마이크로데이터는 자료정제(data cleaning) 과정을 거쳐 통계값으로 산출할 수 있는 상태의 데이터를 의미한다. 마이크로데이터를 활용하는 방법은 두 가지가 있다. 하나는 조사된 데이터에 매칭(matching)등의 방법으로 결합하는 것이고, 다른 하나는 마이크로데이터의 대상을 이용하여 추출하고 표본조사를 하는 방법이다.

조사된 데이터와 행정자료를 매칭하여 활용하는 방법은, 행정자료에서 가져올 항목에 대한 질문을 제외하고 설문지를 작성한 후, 이를 응답대상자에게 조사를 통해 데이터를 수집한다. 수집된 데이터는 정제 작업 후 행정자료와 결합한다. 결합된 자료는 하나의 데이터가 되기 때문에 서로 연계한 분석이 가능하다. 즉, 조사된 자료와 행정자료가 1대1로 매칭되기 때문에 다양한 분석이 가능하다.⁹⁰⁾ 이때 조사자료와 행정자료를 연계할 때, 정확하게 동일한 대상을 연계할 수 있는 변수⁹¹⁾가 있다면 정확매칭(exact matching)방법을 사용하고, 반면 정확하게 연계할 변수가 없다면 통계적매칭(statistical matching) 방법을 활용한다.

마이크로데이터를 표본으로 활용하는 방법⁹²⁾은 다양한 방법과 관점에서 활용이 가능

90) 정확매칭인 경우는 1대1 매칭을 사용하고, 통계적매칭인 경우 1vs1, 다vs1, 1vs다 모두 가능함

91) 주민등록번호, 이동통신 전화번호 또는 여러 변수(주소, 이름 등)들의 조합 등으로 개인정보가 될 수 있는 정보가 해당되며, 이를 활용하기 위해서는 가명정보 처리 등의 방안이 마련되어야 함

92) 이때 활용하는 마이크로데이터는 해당하는 대상들의 모집단의 특정 자료라고 가정함

하기 때문에 매력적인 방법이다. 모든 데이터를 활용할 경우, 데이터의 용량에 따른 분석시스템의 과부하, 개인정보 활용의 어려움, 추가정보 필요 등의 문제가 있을 수 있다. 따라서 일부의 마이크로데이터에서 표본을 추출하여, 추출된 표본의 자료만을 이용하여 분석하면 대용량 데이터를 활용하지 않아도 되며, 일부의 데이터를 관리하기 때문에 개인정보보호 방침을 준수하기도 용이하면서 다양하게 데이터를 분석할 수 있다.

그러나 추가적인 정보가 필요하다면, 행정자료인 마이크로데이터를 표본틀로 활용하는 것과 같다. 표본설계를 통해 표본을 추출하고 추출된 대상을 조사하게 된다. 따라서 추출된 조사대상자의 행정자료를 그대로 활용할 수 있기 때문에 조사자료와 행정자료를 연계한 정확매칭과 동일한 수준으로 정보활용이 가능하다.

하지만 마이크로데이터를 이용할 경우 개인정보 문제가 직결되는 부분은 매우 민감한 사항이다. 따라서 행정자료를 활용 가능한 수준을 정해 등급을 정하고, 정해진 등급에 맞는 수준으로 활용하도록 할 필요가 있다. 본 연구에서 조사에 활용하기 위한 마이크로데이터는 개인정보를 포함하고 있기 때문에, 보안센터에서 활용하는 것이 타당할 것이다. 보안센터의 시스템에 가명처리한 표본틀을 제공받고, 표본설계를 하여 추출된 대상의 정보를 제공받는다. 제공받은 표본의 값을 그대로 분석할 경우에는 보안센터에서 분석을 하고 분석한 결과만 반출하여 사용하면 된다. 만약 표본조사를 할 경우라면, 가명처리되어 추출된 표본을 우편, 이메일, 모바일 문자 등으로 시스템에서 자동으로 발송되도록 하여, 조사가 진행되도록 한다.⁹³⁾ 즉, 마이크로데이터의 표본을 추출하는 과정과 조사가 진행되는 과정을 분리하여 관리하는 것이다.

그래도 추가적인 개인정보 문제를 고려해야 한다면, 통계적 익명화 기법⁹⁴⁾을 적용하는 방법을 고려할 수 있다. 표본으로 활용된 데이터들이 결합하여 개인정보를 파악할 수 있을 경우를 대비하여 조사된 숫자 또는 행정자료의 값들을 통계적 익명화 처리방법⁹⁵⁾으로 변화하여 활용하는 방법으로 고려할 수 있다. 이렇게 되면 표본을 추출하여

93) 이때 휴대전화의 가상번호를 발급받아 활용하는 등의 방법으로 조사대상자가 원하는 조사방식으로 조사에 임할 수 있도록 할 수 있음

94) 통계적 익명화 기법은 다양한 방법이 있는데, 그 중 하나를 설명하면 임의의 잡음(random noise)을 추가함으로써, 각각의 값을 변환하여 알 수 없게 하지만, 평균, 분산을 동일하게 하여 분석에는 영향을 미치지 않도록 하는 방법이 있음

95) 본 연구에서는 통계적 익명화 처리방법까지 다루지 않기 때문에, 관심있는 경우, 박원환 외(2004)를 참조하거나, 영국의 정보보호위원회(ICO, The Information Commissioner's Office)의 2012년 「개인정보 익명화 실무지침(Anonymisation: managing data protection risk code of practice)」를 참조할 것을 권장함

추정하는 방법도 하나의 익명화 방법이기 때문에, 2단계로 개인정보 취급에 대한 보호정책이라 할 수 있다.

행정자료를 이용하여 분석할 경우, 사전에 이용하기 위한 다양한 분석이 필요하다. 실제로 데이터를 이용할 수 있을 정도의 시의성이 있는지⁹⁶⁾, 정보를 활용하기 위한 변수의 차원(dimension)이 맞는지 등 다양한 연구가 동반되어야 한다. 무엇보다도 개인정보 문제 부분과 데이터를 관리하는 기관과의 협조가 중요하다.

특히 표본들을 이용하는 방법은 조사방법을 개선할 수 있는 중요한 방법 중 하나이다. 예를 들어 국제청의 신고자료를 활용할 수 있다면, 현재 사업체 조사의 표본들로 활용하고 있는 통계청의 ‘전국사업체조사’ 자료에 없는 1인 사업체⁹⁷⁾의 정보가 있기 때문에, 이를 활용할 수 있다면 프리랜서가 많은 예술분야와 콘텐츠분야의 세밀한 조사에 활용할 수 있기 때문이다.

2. 빅데이터와 연계한 조사 방안

일반적인 빅데이터는 일상생활 또는 행정처리, 업무 등 다양한 활동이 자동화 또는 시스템을 통하여 진행되면서 자연스럽게 쌓이는 대용량의 데이터를 말하기도 하지만, 이를 이용하여 다양한 형태의 분석 또는 데이터의 가공을 통하여 생긴 가치(value)있는 정보를 의미하기도 한다. 실제 빅데이터는 개개인의 데이터를 분석함으로써 개개인 또는 특성별로 활용할 수 있는 유기적인(organic) 활용을 하여야 하지만, 개인정보 문제 등으로 매크로 형태의 분석으로만 활용하고 있는 것이 현재의 실정이다.

여기서는 빅데이터를 표본의 형태로 활용하는 방안을 구체적으로 제안하고자 한다. 실제 빅데이터 가입자를 이용한 조사방식은 현재 다양하게 이용되고 있다. 3장에서 살펴본 것처럼, 카드사에서는 고객의 동의를 구한 후, 이들을 대상으로 마케팅 관련 조사를 진행하고 있으며, 카드데이터와 연계하여 활용하는 방안도 진행하고 있다.

이동통신자료는 현재 선거여론조사에서 활용하고 있다. ‘중앙선거여론조사심의위원

96) 통계청에서 제공하고 있는 기업등록부(BR: business register) 자료는 2021년 9월 현재 2019년 자료를 제공하고 있어, 시의성 있는 통계를 생산하는데 활용이 어려움

97) 사업장이 없이 사업자 신고를 한 대상들 포함

회'는 이동통신사업자에게 가상번호⁹⁸⁾를 제공받아 지역, 성별, 연령대를 구분하여 조사를 진행한다. 이 경우는 선거여론조사로만 활용하고 있지만, 가상번호를 제공받아 조사하는 방안은 가명 처리된 정보를 이용하는 방안이기 때문에, 다양하게 정보를 활용할 수 있는 방안으로 확장할 수 있을 것이다.

대부분의 빅데이터는 시간이 지나면서 지속적으로 쌓이는 데이터이기 때문에, 개체들의 행태 등의 변화를 파악할 수 있다. 또한 이러한 변화를 분석하여 다양한 정보를 파악하고 있다. 그러나 빅데이터는 시간의 변화에 따라 발생된 현상(fact)을 제시하고 있을 뿐, 원인과 그에 대한 생각을 파악할 수 없다. 따라서 빅데이터를 가지고 있는 회사(기관)에서 가입자들의 동의를 받은 대상자들을 분석하여 표본설계를 통해 추출하고, 인식이나 원인 등을 파악할 수 있는 설문조사를 하여 데이터를 결합 분석하면 인과관계를 파악할 수 있는 데이터로 활용이 가능할 것이다.

한 단계 더 나아가, 빅데이터는 시간의 흐름에 따라 자연스럽게 쌓이는 동적인 데이터(dynamic data)라는 특성을 이용할 수 있도록, 조사대상자의 동의를 구해 지속적으로 조사하는 패널조사 방식으로 조사할 수 있다면 더욱 활용도 높은 정보로 활용이 가능할 것이다. 하지만 빅데이터의 가입자 모두를 조사대상으로 이용하는 것은 불가능하다. 조사대상으로 삼기 위해서는 조사에 대한 동의도 받아야 하기 때문에 어려움이 있다. 또한 빅데이터 가입자들이 우리나라 국민을 대표한다고 하기 어렵기 때문에, 이를 활용하기 위해서는 모집단의 포괄성을 분석해야 한다. 이러한 이유로 마스터샘플을 만들어 이용하는 방법을 고려해야 한다.

예를 들어 카드회사의 가입자를 마스터샘플로 만드는 방법을 생각해 보면, 카드회사 가입자들의 지역, 성별, 연령, 아파트 거주여부, 소득(또는 지출) 등의 분포 등을 분석하여 참여할 대상을 선발하고 참여에 대한 동의를 받아 마스터샘플을 구축한다. 여기서 구축된 대상들의 포괄성을 분석하여 추정된 결과를 보정(또는 보완)하는 방안을 마련하여 대표성을 확보한다. 이렇게 마스터샘플로 확보된 대상들만으로 표본설계를 하고 추출된 대상들에게 표본조사를 시행하여 분석한다.

이러한 마스터샘플은 다양한 방법으로 활용이 가능하다. 예를 들면 문화공간 활용을 높이기 위해, 이용자들의 만족도를 파악하는 조사를 한다고 가정하자. 이러한 경우, 대부분

98) 가상번호는 개인정보 문제로 유효기간(10일 이내)을 설정한 후, 7일 이내 제공함

의 조사에서는 조사대상자로 뽑힌 많은 대상자들이 해당 문화공간에 접근하지 않았기 때문에, 조사대상자들 중의 일부만의 결과만을 활용하게 된다. 그러나 카드데이터⁹⁹⁾의 경우는 문화공간에서 지출한 대상을 파악하여¹⁰⁰⁾ 이들을 대상으로 조사를 할 수 있다.¹⁰¹⁾

문화공간에서 지출한 비용을 알 수 있기 때문에, 만족도와 공간에 대한 느낌, 개선사항 등을 질문한다면 다양한 정보를 파악할 수 있다. 이러한 경우 마스터샘플 중에서 문화공간을 방문한 비율을 알 수 있기 때문에, 활용현황도 나오고 해당하는 대상만을 추출할 수 있기 때문에 조사비용도 낮출 수 있게 된다.

여행한 사람의 정보를 파악하고 싶다면, 거주지를 벗어난 대상¹⁰²⁾만을 조사하여 여행 여부를 판단하고, 여행목적, 여행이유, 여행활동 등을 파악할 수 있을 것이다. 여행을 가지 않은 대상을 조사하지 않았지만, 여행을 간 비율도 파악할 수 있으며, 여행 간 사람만을 조사하기 때문에 조사예산을 더 효율적으로 이용할 수 있다.

빅데이터의 분석에서도 개인정보 문제가 발생한다. 따라서 행정자료 활용과 동일하게 마스터샘플의 정보를 관리하고, 데이터를 분석할 수 있는 시스템을 갖춘 보안센터가 있어야 할 것이다. 또한 표본으로 뽑힌 대상자들에게 조사에 대한 홍보와 응답요청을 하는 이메일 또는 문자 등을 발송하는 시스템이 필요하다. 이러한 경우 마스터샘플에 참여하는 대상들에게 어떠한 보상을 하는지가 중요하다. 참여하는 자체에 대한 인센티브를 제공하고, 조사대상으로 선정되어 응답하였을 때 보상을 주어야 할 것이다. 이 부분은 패널조사의 패널관리와 유사하다고 할 수 있다.

물론 빅데이터를 이용하는 것도 중요하지만, 넘쳐나는 정보로 인하여 개개인의 데이터를 모두 활용하는 것이 점차 힘들어지고 있다. 따라서 빅데이터의 장점을 이용하면서 단점인 원인이나 인식 등의 정적인 정보는 조사를 통해서 보완할 수 있는 방안을 마련하는 것이 필요하다. 따라서 빅데이터와 조사데이터를 연계하는 방안의 하나인 빅데이터 가입자들을 대상으로 조사하는 방안은 데이터 활용에서 매우 중요한 과제가 될 것이다.

99) 이동통신데이터의 경우는 위치기반 데이터를 통해 해당 문화공간 부근에 있었던 사람들을 대상으로 조사할 수 있음

100) 전자지불결제대행(PG :payment gateway)사 또는 VAN(value added network: 부가가치통신망)사들은 세부적으로 지불한 내역 파악이 가능하기 때문에, 이 정보를 같이 활용한다면 가능할 것임

101) 정확하게 알 수 없다면 조금 넓게 선정하여 질문을 통해 추출함

102) 카드데이터의 경우는 거주지 외에서 지출한 대상, 이동통신데이터의 경우는 거주지 이외의 기지국에서 탐지된 대상이 해당됨

제3절 제언

본 연구에서 살펴본 것처럼 사회환경이 다양한 이유로 변화하면서, 정보를 수집하는 주요한 방법인 설문조사(survey)환경은 나빠지고 있다는 것을 알 수 있다. 특히 설문조사의 중요수단이었던 전화조사와 대면조사는 점차 어려워지고 있다. 전화조사는 모바일 전화의 등장과 개인정보 유출의 문제로 전화번호부 제작이 중단되면서 RDD방식으로 채워지고 있는데, RDD 방식의 전화조사는 표본틀이 없기 때문에, 응답자의 자발적 참여에 의지하여 대표성 있는 통계조사에서는 사용하지 않고 있다. 대면조사의 경우는 1인 가구의 증가, 개인사생활의 중요성 확대, 코로나19 등의 문제로 사람들 간의 접촉이 어려워진 상황에서 제대로 조사가 진행되기 어렵다.

공공에서 진행되고 있는 사회조사의 경우는 통계의 대표성이 충족되는지가 매우 중요하다. 통계의 대표성을 확보하기 위해서는 모집단을 포괄하는 표본틀이 있어야 하며, 표본틀을 기반으로 표본설계를 한 후 랜덤하게 표본을 추출하여 조사를 하여야 한다. 따라서 표본틀을 구축하여 활용하거나, 표본틀을 대체하여 조사에 활용할 수 있는 마스터샘플이 마련되어야 한다.

조사를 위한 표본틀로 활용하는 정보를 이용하기 위해서는 개인정보를 활용하여야 한다. 이때 개인정보문제가 발생하므로 익명화(anonymity) 방법을 적절히 활용하여야 한다. 또한 정보 활용과 조사를 원활하게 하기 위한 예산의 확보도 중요하다.

이번 절에서는 사회환경이 변화에 따라 문화·관광 분야에서 조사방법 개선을 위한 필요한 사항들을 하나씩 제시하고자 한다.

① 조사 포괄성(coverage) 확보

공공분야에서 사회조사는 대표성이 있어야 한다. 이를 위한 방안으로는 크게 두 가지 방안을 생각할 수 있는데, 첫 번째가 표본틀 구축이고, 다른 하나가 마스터샘플을 마련하는 것¹⁰³⁾이다.

■ 문화·관광 관련 법을 이용한 표본틀 구축

표본틀은 표본조사를 원활하게 진행하기 위해서 가장 중요한 요소 중의 하나이다. 사업체는 ‘전국사업체조사’ 자료를 이용하고, 국민은 ‘인구총조사’ 자료를 이용하고 있다. 그러나 문화·관광분야의 조사는 예술인, 프리랜서 등의 특정대상에 대한 정보가 필요하기 때문에 이러한 대상을 조사할 수 있는 표본틀을 구축할 필요가 있다.

통계작성을 목적으로 개인정보를 활용할 수 있도록 문화·관광 관련한 법률에 명시한 후, 자료를 가지고 있는 기관에 요청함으로써 표본틀을 구축하는 방안이 있다. 예를 들어 ‘인구주택총조사’의 문항에 예술인 여부의 항목을 삽입한다면, 직업 특성으로 프리랜서 문항을 추가하여, 이들의 정보를 통계청과 협조 하에 MDIS에서 활용하는 방안이다.

■ 마스터샘플(master sample) 구축

표본틀은 조사할 수 있는 대상들 중에서 조사에 활용할 수 있는 모든 대상을 포함하여 구축한다. 따라서 법 또는 제도를 통하지 않고 표본틀을 구축하는 것은 어렵고, 구축한다고 하더라도 개인정보 유출 등의 이유로 정보 활용에 제약이 있을 수밖에 없다. 따라서 이중추출법(double sampling)의 개념을 도입하여 마스터샘플을 구축하는 방법을 생각할 수 있다.

마스터샘플은 표본틀보다는 상대적으로 매우 적은 대상을 모집하여, 표본 추출 시 활용한다. 마스터샘플로 선정할 대상도 설계에 의해 추출한 후, 동의를 받아 구축하기 때문에, 조사의 대표성을 충족하게 된다. 이것은 국민대상의 조사에서만 고려하는데, 사업체조사는 ‘전국사업체조사’자료 활용과 정보제공이 상대적으로 유연하기 때문이다.

마스터샘플을 구축하는 방법은 여러 가지로 생각할 수 있지만, 통계청의 협조를 받아 ‘인구총조사’자료의 일부를 대상으로 모집하는 방법을 생각할 수 있다. 다음으로는 민간의 빅데이터의 가입자를 이용하는 방법이 있다. 민간의 빅데이터는 시간의 변화에 따라 데이터의 변화를 파악할 수 있기 때문에, 문화·관광활동을 한 대상들만을 추출하여 조사할 수 있다는 장점이 있다. 또한 민간데이터이기 때문에 비용 등의 문제만 해결된다면 더욱 유연하게 활용할 수 있다.

103) 이는 조사패널(survey panel)을 구축한다는 것과 같은 의미임

② 개인정보 활용방안 마련

조사의 신뢰를 위해서는 개인정보를 활용하여야 한다. 그러나 개인정보문제로 특정조사에서는 활용하는데 한계가 있다. 본 연구에서 제안하고 있는 표본틀 구축이나 마스터 샘플은 표본을 추출한 후, 비대면으로도 조사할 수 있도록 조사대상자를 컨택하여 원하는 방법으로 조사를 진행할 수 있어야 한다. 따라서 조사를 위해 필요한 개인정보는 충분히 활용할 수 있어야 한다.

무엇보다도 마스터샘플을 구축하여 활용하기 위해서는 개인정보를 관리하고 분석할 수 있는 보안센터가 필요하다. 특히, 모집단의 포괄성을 충족할 수 있는지를 분석할 수 있어야하기 때문에, 개인정보를 다양하게 분석하기 위한 보안센터가 필요하다. 따라서 데이터의 익명화 처리방안 등이 마련되어야 한다.

③ 혼합조사 방안과 데이터 익명화 방안의 지속적 연구 필요

■ 혼합조사의 추정방안 연구 필요

현재에도 많은 조사가 혼합조사를 이용하여 진행되고 있지만, 향후에는 혼합조사가 대세를 이루게 될 것이다. 그러나 혼합조사는 조사방법마다 차이가 나는 문제가 있어, 이에 대한 방안 마련이 필요하다. 문화·관광분야에 종사하는 대상들은 상대적으로 이동량이 많아, 이들을 만나서 조사하는 것에 어려움이 따른다. 따라서 응답하기 편한 방법으로 조사를 할 수 있어야 저조한 응답률을 높일 수 있다.

혼합조사를 원활하게 진행한 후 이를 잘 추정할 수 있도록 설계하여야 하며, 조사된 자료를 이용하여 좋은 추정을 하는 방안에 대한 지속적인 연구가 진행되어야 할 것이다.

■ 데이터 익명화 처리에 대한 연구

더 많은 정보를 이용하기 위해서는 더욱 세밀한 정보를 활용하고 분석하여야 한다. 가명처리를 하여 활용한다고 하더라도, 데이터의 결합 또는 분석을 하게 되면 개인정보가 발견되는 경우가 생긴다. 또는 개인정보라고 생각하지 않은 정보라도 이들이 연계되며, 누군지를 파악할 수 있는 경우도 발생할 수 있다. 따라서 개인의 인구통계학적 자료 뿐만 아니라 분석하는 자료들도 익명화 처리방안 마련이 필요하다.

현재 통계청 등에서 연구가 진행되었지만, 아직은 미흡한 단계이다. 특히 개개인의 데이터 값은 달라지지만, 전체를 분석할 때, 결과가 동일하게 나오게 하는 방법은 데이터를

세밀하게 분석하기를 원할 경우 매우 중요하다. 따라서 데이터 익명화에 대한 연구는 조사 통계만이 아니라 빅데이터 등의 다양한 분야에서 데이터를 활용하는데 매우 중요하다.

④ 조사예산 확보

사회가 발전하고 있지만, 조사예산의 증가는 매우 낮거나 거의 변하지 않고 있다. 이를테면 조사수행을 위한 표본당 단가가 십수년전과 크게 달라지지 않은 것이다. 좋은 데이터를 확보하기 위해서는 충분한 사례가 반영되어야 할 것이다. 조사업체의 연구원, 조사원, 응답자 모두에게 적절하게 예산이 책정되어야 조사가 원활하게 진행되며, 양질의 데이터가 나오게 된다. 따라서 조사마다 응답하는데 걸리는 시간을 측정하여, 이에 합당한 비용을 지불할 수 있는 방안 마련이 필요하다.

또한 마스터샘플이 구축된다면, 이를 유지 관리하는 비용도 발생되며, 이들을 관리하고 분석하는 보안센터를 이용하는 비용도 발생된다. 향후 조사방안 개선이 이뤄진다면 이러한 사항들을 고려하여 예산을 확보할 수 있는 방안 마련이 필요하다.

참고문헌

〈논문〉

- 류진화, 김종훈(2003), 전자우편을 이용한 설문조사시 응답률, 응답속도, 자료의 질에 영향을 미치는 요인들에 대한 연구, *경영학연구*, 32(2), 475-498.
- 박원환, 황조연(2004), 통계자료의 비밀보호를 위한 익명화 방법들, *통계연구*, 9(2), 146-172.
- 백지선, 민경아(2016), 병행적 혼합조사의 모드효과 분석, *응용통계연구*, 29(5), 787-806.
- 백지선, 심규호, 홍영택, 장덕현(2017), 혼합조사의 모드효과 추정: 선택효과와 측정효과와의 분리 방법, *통계연구*, 22(3), 1-24.
- 이슬기, 금현섭(2019), 면접원이 설문조사 질에 미치는 영향: 응답자의 응답시간과 응답행태를 중심으로, *조사연구*, 20(3), 33-63.
- 이현우(2009), 패널조사의 유용성과 문제점: 17대 대선자료를 대상으로, *조사연구*, 10(2), 23-43.
- 채서일, 손영석(1993), 우편을 이용한 설문조사의 응답율에 영향을 미치는 요인에 관한 문헌 연구, *마케팅연구*, 8(1), 50-69.
- 한신갑(2012), 혼합식 조사와 웹패널의 (열은)빛과 (짙은)그늘, *조사연구*, 13(3), 1-31.
- 허명희, 강용수, 손은진(2004), 사회조사에서 조사방법에 따른 가중 칸 설정에 관한 연구, *조사연구*, 5(1), 1-26.
- Biemer, P.(2010), Total Survey Error design, implementation, and evaluation, *Public Opinion Quarterly*, 74(5), 817-848.
- Biemer, P., Trewin, D., Bergdahl, H., & Japec, L.(2014), A System for Managing the Quality of Official Statistics, *Journal of Official Statistics*, 30(3), 381-415.
- Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies, *Biometrics*, 24, 295-313.

- Dillman, D. A.(2017), The promise and challenge of pushing respondents to the Web in mixed-mode surveys, *Survey Methodology*, 43(1), 3-30.
- Lena Hohwü et al.(2013), Web-Based Versus Traditional Paper Questionnaires: A Mixed-Mode Survey With a Nordic Perspective, *Journal of Medical Internet Research*, 15(8), e173.
- Mazzi, G. B., Mitchell, J., & Carausu, F.(2021), Measuring and Communicating the Uncertainty in Official Economic Statistics, *Journal of Official Statistics*, 37(2), 289-316.
- Platek, Richard, and Carl-Erik Sarndal.(2001), Can a Statistician Deliver?, *Journal of Official Statistics*, 17(1), 1-20.
- Smith, T. M. Fred.(1990), Comment on Rao and Bellhouse: Foundations of Survey-based Estimation and Analysis, *Survey Methodology*, 16, 26-29.
- Tse-Hua Shih, Xitao Fan(2007), Response Rates and Mode Preferences in Web-Mail Mixed-Mode Surveys, *International Journal of Internet Science*, 2(1), 59-82.
- Vannieuwenhuyze, Loosveldt, Molenberghs(2010), A Method for Evaluating Mode Effects in Mixed-mode Surveys, *Public Opinion Quarterly*, 74(5), 1027-1045.

〈연구보고서, 단행본〉

- 김수택 외(2010), 「조사방법의 이해」, 서울: 교우사.
- 김현주, 최경은, 김형중(2020), 「인구구조 변화에 따른 국내관광 정책방향」, 서울: 한국문화관광연구원.
- 송정연, 박근희(2020), 「1인 가구의 문화·여가 및 여행 특성 심층 분석」, 서울: 한국문화관광연구원.
- 심규호, 나유리(2016), 순차적 혼합모드 시험조사의 조사과정자료 분석, 통계개발원 2016년 하반기 연구보고서 제 I 권, 145-199.
- 임경은(2012a), 조사과정자료 수집 및 활용을 위한 가이드라인 수립 방안, 통계개발원 2012년 하반기 연구보고서 제Ⅲ권, 77-143.
- 임경은, 박라나(2013), 혼합조사 추정방법 실무적용방안 검토, 통계개발원 2013년 하반기 연구보고서 제Ⅲ권, 1-69.
- 조순기, 최은영(2008), 행정자료 기반의 등록센서스 해외 사례 연구, 통계개발원

2008년 상반기 연구보고서, 1-82.

- 홍민기, 김재광, 한치록, 김기민(2014), 「패널자료 품질개선 연구(III)」, 서울: 한국노동연구원.
- Messer, Edwards & Dillman(2012), *Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Based Mixed-Mode Surveys of the General Public*, Social & Economic Sciences Research Center, SESRC Technical Report 12-001.

〈발표자료, 간행물〉

- 강유경(2019), 조사과정자료, 조사자료만큼 중요하다!, KOSTAT 통계플러스, Vol.5, 77-82. 2-3.
- 박선희(2017), 모바일조사 현황과 모바일조사표 설계방법, KIPA 조사포럼, Vol.21, 50-55.
- 박원용(2020), 비대면조사 바로알기, KOSTAT 통계플러스, Vol.12, 77-82.
- 박인호(2013), 조사모드의 효과적 활용 및 추정방법의 향후 과제, 국가통계 방법론 심포지엄 발표자료.
- 심규호, 백선현(2015), 조사과정자료 분석을 통한 조사 개선방안 연구, 국가통계 방법론 심포지엄 발표자료.
- 심규호, 나유리(2016), 조사과정 자료분석을 통한 조사 개선방안 연구: 순차적 혼합모드 시험조사를 중심으로, 통계청, 국가통계 방법론 심포지엄 발표자료.
- 임경은(2012b), 국가통계 품질 향상을 위한 조사과정자료의 역할, 국가통계방법론심포지엄 발표자료.
- 전종우(2008), 패널조사의 중요성과 발전방향, 보건복지포럼, Vol.145, 2-4.
- 진재현(2012), 미국 국민건강면접조사 조사과정자료 분석, 보건복지포럼, Vol.194, 97-106.
- Kreuter, Couper & Lyberg(2010), *The use of paradata to monitor and manage survey data collection*, 2010 Joint Statistical Meetings, Section on Survey Research Methods, 282-296.
- 통계개발원(2016), 「조사품질 향상을 위한 혼합조사 활용 방안 연구」.
- 통계개발원(2012), 「혼합조사에서의 추정 방법 개발」.
- 통계청(2020), 「2020년도 국가통계 품질관리 매뉴얼」.

- 한국문화관광연구원(2020), 「2018년 기준 문화체육관광산업통계」.
- 행정안전부(2019), 「공공데이터 관리지침」.

ABSTRACT

A Study on Improving Survey Methods in the Culture and Tourism Sector According to Changes in the Social Environment

Kim jihak, Park kwenhwa

Changes in the social environment are leading to the increase of the rejection rate of survey responses for various causes and reasons and are setting us a social task of improving or changing the survey method. The increase in single-person households and apartment households makes it difficult to meet respondents in person, making face-to-face surveys more challenging. In particular, the spread of COVID-19 accelerated the difficulty further, which raised the necessity of a turning point in traditional surveying techniques. Accordingly, recent surveys were conducted in various ways depending on the situation, such as resuming the survey after a temporary suspension, distributing surveys to minimize face-to-face process, or conducting online (mobile) surveys using e-mail or QR codes. However, for official statistics, interview surveys are still conducted mainly due to the representativeness of the sample and the absence of a sampling frame to be used for contactless surveys. Therefore, it is required to seek measures to resolve these issues which can derive statistical values that are representative and efficient in conducting surveys in accordance with the changes in the social environment.

To this end, various research methods such as literature review, case study, pilot survey test, and expert opinion gathering are used to ensure the validity of the improvement of the survey method in this study. In addition, the followings are considered: ① the use of a mixed-mode survey to increase the response rate by combining various survey methods, ② the use of administrative data and big data,

and ③ the construction of a sampling frame for the cultural tourism research.

First, to improve the survey design, a mixed methods research design and constructing a sampling frame for the tourism research are suggested. For the mixed-mode survey, this study proposes a method of estimating and correcting the mode effect generated by each survey method through a test survey. In constructing a sampling frame, the process of constructing and harnessing the sampling frame for the culture and tourism sector, and a measure to construct a representative master sample as their survey panel are presented. To improve the survey method, this study suggests using administrative data and big data. Administrative data are proposed to be used by combining data through data matching at the microdata level or using microdata as a sample. For Big data, the study presents to establish a master sample based on the data from the private sector and connect it with surveys. In order to improve the survey method, however, it is needed to ensure the comprehensiveness of the survey and prepare a measure to utilize data privacy. Furthermore, continuous studies on the mixed-mode survey and data anonymization methods should be conducted, and also securing a survey budget including ensuring sufficient cases for high-quality data is also critical.

Keyword

Changes in social environment, Mixed-mode survey, Administrative data, Big data, Master sample

집필내역

연구책임

김지학 한국문화관광연구원 차석전문위원: 제1장, 제2장, 제3장, 제4장, 제6장 연구총괄
박근화 한국문화관광연구원 선임전문위원: 제1장, 제4장, 제5장, 제6장

사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구

발행인 김대관

발행처 한국문화관광연구원
서울시 강서구 금남화로 154
전화 02-2669-9800 팩스 02-2669-9880
<http://www.kcti.re.kr>

인쇄일 2021년 10월 15일

발행일 2021년 10월 15일

인쇄인 (사)한국장애인이워크협회 일자리사업장

I S B N 978-89-6035-865-2 93300

DOI <https://doi.org/10.16937/kcti.rep.2021.04>

이 연구보고서를 인용하실 때는 다음과 같은 사항을 기재해 주십시오.

김지학·박근화(2021), 사회환경 변화에 따른 문화관광분야 조사방법 개선 방안 연구,
한국문화관광연구원



한국문화관광연구원

서울특별시 강서구 금남화로 154

전화 02-2669-9800

팩스 02-2669-9880

www.kcti.re.kr



아래의 DOI 또는 QR코드를 통해
이 보고서를 무료로 다운로드할 수 있습니다.
<https://doi.org/10.16937/kcti.rep.2021.e4>

