

15강 표본조사 사례
– 2012년 주거실태조사 –

정보통계학과 이기재교수

학/습/목/차

1. 항목 무응답 대체 사례
2. 확률응답기법
3. 표본조사 사례분석의 주요 내용
4. 표본조사 사례

무응답 대체법의 수치적 비교(1)

- 교양강좌 수강생 50명(남 30명, 여 20명)의 강의평가 자료

X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	4	5	1	1	1	1	1	2	2	4	-	2	4	1
1	2	4	1	3	5	1	3	2	2	3	4	2	-	1
1	3	3	1	-	4	1	-	3	2	1	2	2	3	2
1	-	2	1	4	-	1	3	-	2	1	2	2	2	5
1	1	4	1	3	3	1	2	4	2	-	-	2	1	3
1	2	3	1	4	1	1	4	4	2	4	3	2	2	-
1	4	-	1	2	5	1	4	5	2	2	4	2	2	2
1	-	2	1	-	3	1	-	2	2	3	3	2	4	4
1	3	2	1	4	-	1	1	1	2	4	5	2	1	1
1	3	1	1	1	4	1	2	5	2	1	-	2	1	5

▶ X : 성별, Y : 학년(8명 결측), Z : 강의평가(8명 결측)

무응답 대체법의 수치적 비교(2)

- 교양강좌 수강생 50명(남 30명, 여 20명)의 강의평가 자료
 - ▶ 학년의 항목 무응답 : 남학생 6명, 여학생 2명 발생,
응답자 학년 평균 : 남학생 2.7, 여학생 2.4
➔ 학년에 무응답한 경우 남학생은 2.7, 여학생은 2.4로 대체
 - ▶ 강의평가 항목 무응답 : 남학생 4명, 여학생 4명 발생
응답자 강의평가 평균 : 남학생 3.1, 여학생 2.9
➔ 강의평가에 무응답한 경우 남학생은 3.1, 여학생은 2.9로 대체
- 최근방 대체
 - ▶ 데이터를 유사하도록 정렬한 후 결측을 바로 앞이나 뒤의 응답자의 응답으로 대체
 - ▶ 학년과 강의평가 결측은 바로 앞의 학생이 대답한 값으로 대체

무응답 대체법의 수치적 비교(2)

■ 최근방 대체 결과

X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	4	5	1	1	1	1	1	2	2	4	5	2	4	1
1	2	4	1	3	5	1	3	2	2	3	4	2	4	1
1	3	3	1	3	4	1	3	3	2	1	2	2	3	2
1	3	2	1	4	4	1	3	3	2	1	2	2	2	5
1	1	4	1	3	3	1	2	4	2	1	2	2	1	3
1	2	3	1	4	1	1	4	4	2	4	3	2	2	3
1	4	3	1	2	5	1	4	5	2	2	4	2	2	2
1	4	2	1	2	3	1	4	2	2	3	3	2	4	4
1	3	2	1	4	3	1	1	1	2	4	5	2	1	1
1	3	1	1	1	4	1	2	5	2	1	5	2	1	5

무응답 대체법의 수치적 비교(3)

■ 회귀 대체

- ▶ 데이터를 남학생과 여학생으로 구분하여 각각 선형회귀식 적합
- ▶ 학년 무응답 대한 선형회귀모형 : 종속변수 Y, 독립변수 Z
남학생의 회귀 추정식 : $Y = 2.1 + 0.1 * Z$
4번째 학생은 학년 결측이고, 강의평가는 2이므로 대체값은 2.3
- ▶ 강의평가 무응답은 종속변수 Z, 독립변수 Y인 회귀모형 이용
남학생의 회귀 추정식 : $Z = 1.6 + 0.5 * Y$
7번째 학생은 4학년인데, 강의평가 결측 대체값은 3.6
- ▶ 여학생의 학년 무응답과 평가 무응답에 대해서도 동일 방식을 적용하여 대체함

무응답 대체법의 수치적 비교(3)

■ 회귀 대체 결과

X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	4	5	1	1	1	1	1	2	2	4	3.3	2	4	1
1	2	4	1	3	5	1	3	2	2	3	4	2	2.1	1
1	3	3	1	2.5	4	1	2.4	3	2	1	2	2	3	2
1	2.3	2	1	4	3.6	1	3	3.1	2	1	2	2	2	5
1	1	4	1	3	3	1	2	4	2	2.5	3.1	2	1	3
1	2	3	1	4	1	1	4	4	2	4	3	2	2	3.0
1	4	3.6	1	2	5	1	4	5	2	2	4	2	2	2
1	2.3	2	1	2.4	3	1	2.3	2	2	3	3	2	4	4
1	3	2	1	4	3.6	1	1	1	2	4	5	2	1	1
1	3	1	1	1	4	1	2	5	2	1	2.8	2	1	5

무응답 대체법의 수치적 비교(4)

■ 핫덱 대체

- ▶ 무응답 대체 층으로 성별 구분 이용,
- ▶ 각 무응답 대체 층에서 결측의 대체값을 랜덤 선정 후 대체

■ 핫덱 대체 결과

X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	4	5	1	1	1	1	1	2	2	4	3	2	4	1
1	2	4	1	3	5	1	3	2	2	3	4	2	2	1
1	3	3	1	3	4	1	3	3	2	1	2	2	3	2
1	3	2	1	4	3	1	3	3	2	1	2	2	2	5
1	1	4	1	3	3	1	2	4	2	3	3	2	1	3
1	2	3	1	4	1	1	4	4	2	4	3	2	2	3
1	4	4	1	2	5	1	4	5	2	2	4	2	2	2
1	2	2	1	4	3	1	4	2	2	3	3	2	4	4
1	3	2	1	4	2	1	1	1	2	4	5	2	1	1
1	3	1	1	1	4	1	2	5	2	1	4	2	1	5

무응답 대체법의 수치적 비교(5)

- 무응답 대체결과 비교

대체방법	학년		강의평가	
	평균추정	분산추정치	평균추정	분산추정치
평균대체	2.56	1.155	3.02	1.613
최근방 대체	2.62	1.342	3.08	1.749
회귀 대체	2.53	1.185	3.06	1.667
핫덱 대체	2.62	1.260	3.04	1.672

학/습/목/차

1. 항목 무응답 대체 사례

2. 확률응답기법

3. 표본조사 사례분석의 주요 내용

4. 표본조사 사례

확률화 응답기법(1)

1. 응답오차 유발하는 질문

- ▶ 응답을 꺼려하는 사적 민감한 사안 질문
 - ➔ 직접 물어 보면 거짓 응답 또는 응답거부의 가능성
- ▶ 종합소득세와 관련하여 연간 소득액을 정직하게 신고하는가?

2. 확률화 응답기법

- ▶ 사적 민감한 특정 속성(마약복용, 임신중절, 백화점 절도)의 비율 추정
- ▶ 직접 질문 시 응답 거부 또는 거짓 응답으로 비표본오차 발생
- ▶ 응답자만 알 수 있도록 두 종류의 질문 중 하나 랜덤 선정 후 응답
- ▶ 조사원은 응답자가 선택한 질문 내용은 모르고, 응답자의 '예', '아니오'라는 답변만 수집한 후 민감 특성 비율을 추정
- ▶ 응답자의 사적인 비밀을 보장하여 정직하게 응답하도록 유도

확률화 응답기법(2)

3. 확률화 응답기법 적용 및 추정

- ▶ 두 가지 유형의 질문 준비한다.

설문 1 : "나는 민감 속성그룹에 속한다"

설문 2 : "나는 민감 속성그룹에 속하지 않는다"

- ▶ 질문 선택 확률장치 : 주머니 속 흰 돌과 검은 돌 구성비 $P : (1 - P)$
- ▶ 잘 섞은 주머니에서 랜덤하게 선택한 바둑알 색깔을 보고 질문 선택
→ 꺼낸 바둑돌이 검은색이면 설문 1에 응답, 흰색이면 설문 2에 응답
- ▶ 조사원 응답자 꺼낸 바둑알 색깔 모르고, '예', '아니오' 대답만 수집
- ▶ n 명 중 m 명이 '예'라고 응답한 경우 민감 속성 비율 추정

$$\hat{\pi} = \frac{\hat{\phi} - (1 - P)}{2P - 1} = \frac{m/n - (1 - P)}{2P - 1}$$

- ▶ $n=50$ 명 조사한 결과 26명 "예"라고 응답한 경우
 $P=0.6$ 이면 민감 속성 비율 추정치는 $0.6 (= (0.52 - 0.4)/(1.2 - 1.0))$

학/습/목/차

1. 항목 무응답 대체 사례

2. 확률응답기법

3. 표본조사 사례분석의 주요 내용

4. 표본조사 사례

표본조사 사례 분석의 주요 내용

- 표본조사 사례

- ▶ 2012년 정보화 통계조사 : 사업체 대상 조사
- ▶ 2012년 주거실태조사 : 가구 대상 조사

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 명확하게 기술되었는가?
- ▶ 표본설계로 달성 가능한가?

표본조사 사례 분석의 주요 내용

■ 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 조사대상의 포괄성
- ▶ 조사목적에 부합하는가?

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 모수 추정에 대한 목표 정도의 적절성

- ▶ 통계적으로 표현했는가?

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

▶ 자료수집방법은 적합한가?

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 실제값과 측정값간 차이가 있는가?

- ▶ 보완대책은 적절한가?

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 추출틀은 완벽한가?

- ▶ 추출단위는 적합한가?

표본조사 사례 분석의 주요 내용

■ 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 표본오차 계산은 가능한가? 이를 최소화하는 방안인가?
- ▶ 추정량과 분산의 계산식은 타당한가?
- ▶ 예산과 요구 정도를 충족한가?
- ▶ 과거의 유사 표본조사가 실행되었는가?

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

▶ 조사업무를 훈련된 감독관이 관장 가능한가?

표본조사 사례 분석의 주요 내용

- 표본조사 사례 분석의 주요 내용

조사목적

모집단 정의

목표정도

자료수집방법

측정오차

추출틀

표본추출방법

조사업무 통제

자료 분석

- ▶ 적절하게 분석 가능한가?
- ▶ 명확하게 요약할 수 있는가?

학/습/목/차

1. 항목 무응답 대체 사례

2. 확률응답기법

3. 표본조사 사례분석의 주요 내용

4. 표본조사 사례

1. 조사개요(1)

- 주택정책

- ▶ 공급 양적 문제 → 주거환경,복지 등 질적 문제
- ▶ 현실성 정책수립 주거실태 정확한 파악

- 정책효과 극대화

- ▶ 기초자료의 안정적 생산 및 관리 필요

- 주거 관련 통계수요 증대

- ▶ 관련 통계자료 생산 및 축적 미흡

- 통계청 인구주택총조사 및 사회통계조사

- ▶ 조사주기, 조사항목 등의 측면에서 제한이 있음

1. 조사개요(2)

- 2012년 주거실태조사 (주택법 근거로 한국토지주택공사 시행)
 - ▶ 지역별 · 계층별 주거실태 파악 : 정부정책수립 기초자료 제공
 - ▶ 조사내용 : 주택유형 및 점유형태 등 8개 분야

1. 주택유형 및 점유형태	주택유형, 점유형태, 주택규모 등
2. 거주 주택의 특성	건축년도, 내부시설과 사용형태 등
3. 주택가격과 주거비 부담	주택가격(현재,구입시기,구입가격), 임차 상황, 대출상황 등
4. 이사 경험과 주거만족도	현 주택 거주시기, 이사 횟수, 주거 만족도 등
5. 주택에 대한 일반적인 생각	주택에 대한 생각, 3인 가족 적정규모 등
6. 가구주 된 후의 주택구입 경험과 소유현황	주택소유 여부, 주택 매각 의사와 이유 등
7. 2년 이내 이사계획과 희망 주택	현 주택 거주계획, 이사하려는 이유, 이사 희망 지역, 희망 주택 유형, 희망 주택 규모 등
8. 가구 일반 현황	동거 가구원 수, 출생연도, 혼인상태, 맞벌이 여부, 교육정도, 직업, 소득금액 등

2. 모집단 – (1) 가구관련 용어 정의

가구(Household)	취사, 취침 등 생계를 같이하는 생활단위
거처(Living quarters)	독립된 거주 단위(주택, 주택 이외 거처)
주택(Housing unit)	한 가구 살림 위한 부엌과 한 개 이상 방 갖춘 영구건물
다가구주택	여러 가구 독립생활 가능 독립소유 매매 불가(단독가구)
다세대주택	여러 가구 독립생활 가능 독립 소유 매매 가능(연립주택)
연립주택	4층 이하 다세대 주택(2~4층 빌라, 맨션 포함)
영업용(비 거주용) 건물 내 주택	영업용 목적 건물 사람 거주(주택요건)
주택 이외의 거처	주택의 요건 못 갖춘 거처(오피스텔 등)



2. 모집단

◀ 조사모집단 가구 수 분포(일부) ▶

지역	수도권 권역	읍면동	최종표본추출률						가구수 총계
			단독	아파트 86㎡미만	아파트 86㎡이상	연립	다세대	기타	
전국			6,884,621	7,086,502	1,420,624	501,957	1,243,916	596,211	17,733,831
서울	1	동	205,355	245,649	144,452	22,358	85,060	30,854	733,728
	2	동	439,739	336,305	84,734	46,107	130,695	78,855	1,116,435
	3	동	242,233	144,828	46,489	31,495	117,783	35,792	618,620
	4	동	457,224	413,208	75,337	39,744	108,835	36,161	1,130,509
부산		동	445,894	498,951	113,927	33,467	90,220	43,222	1,225,681
		읍면	9,706	18,509	3,071	1,024	2,345	1,344	35,999
대구		동	344,693	337,919	77,717	4,464	43,573	12,403	820,769
		읍면	16,407	36,343	3,162	422	2,100	987	59,421
인천	5	동	122,911	199,538	47,358	8,213	88,203	19,087	485,310
	6	동	76,470	221,790	33,590	13,400	93,909	12,119	451,278
		읍면	13,595	1,763	425	983	2,093	600	19,459

3. 모집단 총화

- ▶ 전국 단위와 16개 시도별 독립적인 통계 생산 목적
- ▶ 7개 특·광역시와 9개 도지역 : 1차 총화
- ▶ 9개 도 지역 내 동부와 읍·면부 : 2차 총화
- ▶ 수도권 생활권 기준 총화 : 서울, 인천, 경기도의 도시기본계획, 통근통행 자료 등으로 활용하여 10개 지역층으로 총화 (생활권역별 통계작성 가능)
 - ➔ 결과적으로 46개 지역 층 생성
- ▶ 주택유형별 구분(일반, 아파트조사구, 신축아파트): 3차 총화
- ▶ 전체 층 수: 138개

4. 표본크기 산정

- ▶ 16개 시도 단위 통계 작성 가능
 - ➔ 시도별 가구 수에 따라 목표오차 차등 적용
- ▶ 전체 표본가구 33,000가구 (3,300개 조사구 추출)
- ▶ 추정오차 관리 변수 : 사용면적, 가구 소득, 가구주 소득, 아파트비율
- ▶ 2010년 주거실태조사 결과를 분석하여 표본크기 및 표본배분의 기초자료로 활용함
- ▶ 시도별로 목표 상대표준오차를 차등 적용
 - ➔ 각 시도별 표본크기 산정

5. 표본배분

- ▶ 각 시도별로 결정된 표본 조사구 수는 네이만배분법을 적용하여 층별로 배분한 후 일부 조정함
- ▶ 조사구 당 조사 가구 수는 10가구, 전국 3,000개 조사구 필요
- ▶ 해당 지역 층 내에서 주택유형(일반, 아파트)에 따라 표본배분

지역	수도권 권역	읍면동	2010년 총조사 조사구수	10.11~11.4 신축아파트	시도별 표본배분	네이만배분	네이만배분 (조정후)
서울	1	동	11,234	223	621	164	164
	2	동	17,018	193		158	158
	3	동	9,813	231		158	158
	4	동	17,391	203		140	140
부산		동	19,068	264	230	176	176
		읍면	520	36		54	54
대구		동	12,981	146	154	100	100
		읍면	876	48		54	54

6. 표본추출

- ▶ 각 층에서 조사구 내 가구 수의 확률비례추출법에 따라 조사구 추출
- ▶ 계통추출에 앞서 층별로 조사구를 다음 기준에 따라 정렬
 - 서울 및 광역시 : 공동주택비율, 자가주택비율, 가구주 연령
 - 도 지역 : 공동주택비율, 비농림어가비율, 가구주 연령
- ▶ 1차 추출단위: 조사구, 2차추출단위: 가구
- ▶ 조사구에서 계통추출 또는 단순임의추출법을 적용하여 10가구 추출
- ▶ 표본조사구 대체 : 조사구 유고, 조사불능이나 특성이 변경된 경우
- ▶ 표본가구 대체 : 3회 이상 방문 후 조사불가인 경우는 인근 가구로 대체

7. 표본관리

- 모집단 변동 발생 : 아파트와 주택의 신축, 재개발 또는 재건축
 - ▶ 조사구 요도에 기초하여 조사구 내 가구변동을 추정에 반영
 - ▶ 최신 모집단 이용 및 사후층화 보정
- 조사원 교육강화 : 표본가구 임의 변경 금지
- 표본가구 관리지침 및 인센티브 : 응답률 제고 노력
- 실사관리 및 준비를 위해 충분한 시간 필요
 - ▶ 시간이 촉박하면: 재방문 및 응답자 설득 과정이 제한적임

8. 표본대체 및 무응답 처리

- 재개발 및 재건축 조사구 손실 조사 불능 : 예비조사구 교체
 - ▶ 조사구 특성 및 주변 환경 참조 : 유사 조사구
- 조사구내의 가구 : 불응, 불성실한 응답, 가구원 응답거부
 - ▶ 인근 유사가구로 교체 조사
 - ▶ 무응답률, 응답거부율, 항목별 무응답률 분석 후 대처
- 조사단위 결측 : 표본교체, 가중치 보정, 대체방법 등
 - ▶ 결측 조사단위에 무응답 보정 : 단위무응답 처리
 - ▶ 응답가구와 응답거부 가구 특성 차이 : 유사특성 가구 교체

9. 가중치 작성

- 모집단 구조와 표본 구조 일치 : 추출률, 응답률, 보조정보
- 설계 가중치
 - ▶ 층화2단확률비례계통추출 - 층별 조사구내 가구별 추출률 차이
 - ▶ 각 표본가구의 표본추출률의 역수
- 무응답 보정 가중치 : 조사구 내 가구 유사성 가정
 - ▶ 응답률의 역수로 무응답 보정
- 사후층화조정: 2010년 인구주택총조사 결과 이용
 - ▶ 사후층화 : 지역층(46), 지역(4) × 주택유형(5) × 점유형태(5)
- 최종가중치
 - ▶ 설계가중치 × 무응답조정 가중치 × 사후층화 보정

10. 모수추정

- 모평균 추정 : 표본가중평균

- ▶
$$\bar{y} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{w_{\dots}}$$

- ▶ w_{hij} : 조사 가구 가중치, y_{hij} : 조사 가구 관찰값

- ▶ $w_{\dots} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$: 조사 가구 가중값 합

10. 모수추정

- 모평균 추정량의 분산 추정량 및 상대표준오차

$$\blacktriangleright \text{var}(\bar{y}) = \sum_{h=1}^L \frac{n_h(1-f_n)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi.} - \bar{e}_{h...})^2$$

$$\text{단, } e_{hi.} = \frac{\sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \bar{y})}{w_{...}}, \quad \frac{\sum_{i=1}^{n_h} e_{hi.}}{n_h}$$

$$\blacktriangleright rse(\bar{y}) = \frac{\sqrt{\text{var}(\bar{y})}}{\bar{y}} \times 100$$



Korea National Open University
이 강의는
강의용 휴대폰(U-KNOU 서비스 휴대폰)으로도
다시 볼 수 있습니다.

다시 볼 수 있습니다.