

# CHAPTER 3

---

## 조사 방법 및 표본 추출 방법

설문 조사 방법과 2 장의 설문지 작성 방법은 밀접한 관계가 있다. 우편 조사의 경우 문항 설명이 다소 길고 전문적인 내용이 포함되어 있어도 무방하나 전화 조사의 경우 문항 개수, 지문 길이, 보기 수 등은 짧아야 한다.

적절한 표본을 추출 방법(SRS, Systematic, Clustering, Stratified)과 설문 조사 방법은 밀접한 관계가 있다. 예를 들어, 전화 조사(poll)는 층화-계통 추출법을 사용하게 되고 방문 조사를 실시한다면 층화 추출 방법을 사용하게 된다. 또 다른 예를 살펴보자. OO 백화점 고객 만족도 조사를 위하여 일대일 면접 조사를 실시한다면 계통 추출 방법(예상 일일 고객 수를 표본 수로 나누어 찾는 고객 순번과 일치하는 사람들을 선택, 예상 고객=1,000 명, 표본 수=100 인 경우 3 번째 고객, 13 번째 고객, 23 번째 고객...)을 사용하고 고객 리스트를 표본 프레임으로 이용한다면 연령, 직업을 이용하여 층화 추출법을 이용해 표본을 뽑은 후 우편 조사를 실시하면 된다.

### 3.1. 설문 조사 방법

설문 자료 수집 방법으로는 면접 조사, 전화 조사, 우편 조사, 면접 조사, 응답자 설문 기입법, 온라인 조사 등이 있다. 어떤 방법을 사용할 것인지는 연구 목적, 조사 내용, 조사 대상, 소요 시간 및 비용, 응답률, 응답 민감도 등을 고려하여 결정하여야 한다.

### 3.1.1. 누가 설문지를 기입하느냐에 따른 분류

#### (1) 응답자 자기 기입 방식 (self administrated survey)

대부분의 설문 조사는 응답자가 자기 의견을 직접 기입하는 응답자 자기 기입 방식을 채택하고 있다. 응답자가 혼자 스스로 설문지에 응답하는 방법으로 조사원이 옆에 있을 수는 있으나 응답자의 조사 과정을 관리 감독할 뿐이지 설문 응답에는 관여하지 않는다. 그러므로 응답자의 응답 능력이 높은 경우 사용하게 되며 응답 대상자들이 문맹, 문항에 대한 이해 능력이 떨어지거나 응답 기입 능력이 낮은 노인이나 어린이들에게는 사용할 수 없다.

① 일대일 감독 (one-to-one supervision): (예) Computer Assist Personal Interview, Computer Assist Self-administrated Interview 등 컴퓨터 이용 설문 조사 (한 번에 한 명의 응답자), IQ 검사

② 그룹 감독 (group administration): (예) 학기말에 실시하는 강의 평가, CAPI, CASI 등 컴퓨터 이용 설문 조사 (한 번에 여러 응답자 동시에)

이 외에도 ARS 이용한 여론 조사, 인터넷 조사, 우편조사도 자기 기입식 설문제에 해당된다.

#### (2) 조사원 기입 (interviewer survey)

조사원이 응답자를 도와 설문지에 응답을 직접 적어 주는 방법으로 조사원의 외모, 말씨, 태도 등이 응답 결과에 큰 영향을 미친다.

##### ① 전화 조사 (telephone survey)

조사원이 전화를 통해 응답자의 의견을 들은 후 설문 조사지에 적어 넣는다.

##### ② 조사원 기입 면접조사 (personal interview)

조사원이 응답자를 직접 방문하여 일대일로 만나 응답자의 답변을 설문지에 옮겨 적는 방법이다.

### 3.1.2. 조사 방법에 따른 분류

#### (1) 면접 조사

조사자가 직접 응답자를 방문하여 조사하는 것으로 조사원의 영향이 크다. 설문지는 조사자가 작성할 수도 있고 응답자가 직접 응답할 수도 있다. 조사원은 ① 응답할 수 있는 분

위기를 조성하고(사용하는 언어, 복장, 외모) ②응답자가 의견을 솔직히 밝힐 수 있는 환경을 조성하고 ③설문지 내용을 완벽하게 이해하고 있어야 하며 ④응답 환경을 기록하는 등 성실성이 있어야 한다.

면접 조사의 장점은 다음과 같다.

- ①응답률이 높다.
- ②응답자가 이해하지 못하는 문항에 대한 묻는 경우 즉각적인 설명이 가능하여 신뢰성 높은 응답을 얻어낸다.
- ③조사자가 응답자의 신뢰도, 응답 환경 등을 관찰할 수 있다.

면접 조사의 단점은 다음과 같다.

- ①조사원 영향이 크다.
- ②시간 및 비용 면에서 비효율적이다.

면접 조사 가운데 응답자를 한 곳에 모아 놓고 한 번에 조사하는 집단 조사가 있다. 비용 및 시간을 절약하는 장점은 있으나 집단 상황이 응답을 왜곡시킬 수 있고 (평에 있는 사람들과 의견이나 생각 교환)설문 조사 과정을 통제하는데 어려움이 있다.

## (2)전화 조사

여론 조사와 같이 응답자 표본 프레임이 매우 큰 경우 주로 사용한다. 주로 전화와 조사원에 의해 실시되고 있으나 요즘 컴퓨터가 모든 과정을 관리하는 CATI (Computer Assisted Telephone Interview)가 이용되기도 한다.

RDD(Random Digital Dialing) 시스템에 의해 전화가 걸리면 교육 받은 조사원들이 설문을 하게 된다.

전화 조사의 장점은 다음과 같다.

- ①신속성 및 조사 비용 절약
- ②넓은 지역 조사 가능
- ③조사자 관리가 쉽다.

전화 조사의 단점은 다음과 같다.

- ①질문 길이와 내용의 제한(응답자 신분 노출로 인하여)
- ②표본의 대표성 유지가 어렵다. (인구학적 변수 비율을 맞추기 어렵다)
- ③우편 조사보다 다소 높지만 응답률이 낮다.
- ④그림이나 시각적 도표가 들어간 설문 문항은 물을 수 없다.

### (3)우편 조사

패널이 구성되어 있거나 학회와 같이 주소 리스트가 있는 집단이나 대규모 아파트 단지 입주자 대상 설문 조사에 이용된다.

질문지를 반송 봉투와 함께 넣어 우편으로 보내거나 아파트와 같은 주거 밀집 지역에서 직접 방문하여 우편함에 넣는다.

우편 조사의 장점은 다음과 같다.

- ①최소의 비용과 노력으로 광범위 조사 가능
- ②문항이 많거나 전문적인 문항이 포함된 설문 조사 가능
- ③응답자가 충분한 시간을 가지고 응답할 수 있고 자발적 응답이라 응답의 신뢰성이 높다.

단점은 다음과 같다.

- ①회수율(응답률)이 낮고 (25%)
- ②응답자가 문항에 대한 의문이 생길 경우 질문할 곳이 없다는 것이다.

우편 조사의 가장 큰 문제는 응답률을 높이는 것이다. 우편 조사의 응답률은 조사 기관의 신뢰도, 설문지 양식 및 우송 방법, 그리고 응답자가 집단 성향(예: 전문가 집단)에 의존한다.

Dillman & Christenson(1974) 연구에 의하면 다음 절차를 따르면 우편 조사 응답률을 높일 수 있다.

- ①일주일 전에 설문 조사 발송을 알리는 엽서 보냄
- ②반송 봉투와 함께 설문지 발송
- ③일주일 후 설문 조사 환기 우편 발송

④3 주일 후 응답하지 않은 응답자에게 설문 재발송

⑤7 주 후 응답하지 않은 사람들 대상으로 재조사.

(4) 온라인 조사

전자 메일(e-mail)이나 인터넷(Web-based)을 이용한 설문 조사를 의미한다. 단기간에 저렴한 비용으로 조사할 수 있다는 장점은 있으나 조사 모집단 규정(표본 프레임, 컴퓨터 사용자, 웹 사용자만 모집단) 및 표본의 대표성(관심 집단만 설문에 응한다) 문제, 한 사람이 여러 번 응답(IP 체크에도 한계가 있다)로 인하여 신뢰할만한 조사는 아니다.

온라인 조사는 포털 사이트 사용자 DB 를 이용하거나 패널을 구성하여 운영하면 온라인 조사의 신뢰성을 높일 수 있을 것이다. 또한 설문 조사 할 때마다 점수를 부여하여 누적 점수에 의해 선물을 제공하게 되면 응답률도 높일 수 있다.

요즘 방송국에서도 자주 이용하는 ARS (Advanced Record System) 조사도 온라인 조사의 일종인데 문항이 한 개, 흥미 위주 조사(가요 순위, 토론 중 여론 조사)에 주로 이용된다.

방법	내용	비용	시간	회수	설문 내용	응답자 속성	Comment
일대일 면접조사	조사원이 응답자를 직접 방문하여 조사	X	△	O	주관식 객관식	호의적	조사 과정 발생하는 모든 상황에 대처 가능하나 조 사원의 영향이 크다.
전화조사	전화 조사 (주로 RDD 방법)	△	O	△	문항수가 적고 객관식	호의적	넓은 지역 응답 가능하나 신분 노출에 대한 거부감
우편조사	패널이나 주소 리스 트가 가능한 집단 대상	△	X	X	주관식 전문적 내용	전문가	회수율을 높일 수 있는 방안이 있어야 한다.
온라인	인터넷이나 전자 메 일 시스템 이용	O	O	△	주관식 객관식	관심	패널이나 고객 DB 사용 관심 집단만 응답해 왜곡

### 3.2. 표본 추출 방법

대부분 조사 연구는 모집단 전체에 대한 전수 조사(census)보다는 시간적, 경제 면에서 유리한 표본 조사(sampling survey)를 실시한다. 모집단(표본 프레임) 구성원 중 실제 설문에 응답하는 사람들을 표본이라 하고, 표본을 추출하는 과정을 표본 추출(sampling)이라 한다.

전수 조사가 표본 조사보다 모집단의 참 값(모수)을 항상 더 정확하게 알 수 있는 것은 아니다. 전수 조사의 경우 표본 오차는 발생하지 않으나 시간이 오래 걸려 응답 성향의 변화가 생길 수 있고 많은 조사원이 필요하므로 조사원이 영향이 커지므로 비확률 오차가 커지게 된다. 표본 조사에서 가장 중요한 것은 모집단을 잘 대표할 수 있는 표본을 추출하는 적절한 방법을 선택하는 것이다. 다음은 잘못된 표본 조사의 예이다.

**네티즌 55% ‘미국영화 안볼 것’**<2002.03.13(수)21:12>

동계올림픽 쇼트트랙 편파판정 시비 이후 미국영화를 보지 않겠다는 네티즌들이 절반을 넘는 것으로 조사됐다. 티켓예매 사이트 티켓링크가 지난 달 28일부터 9일까지 이용자 1994명에게 미국영화 관람의사를 물어본 결과 전체 응답자의 55%가 미국영화 관람에 부정적인 태도를 나타냈다. 이 가운데 724명(36.3%)은 “이번 기회에 확실한 의지를 보여줘야 한다”면서 미국영화 관람 거부 운동에 적극 동참할 뜻을 비쳤고 391명(19.6%)은 “미국이 정신 차릴 때까지 당분간 미국영화를 보지 않겠다”고 대답했다. 반면에 663명(33.3%)은 “영화는 영화일 뿐 평소처럼 보겠다”는 입장을 보였다.

이런 온라인 조사에서는 관심 집단만 주로 응답하게 되어 표본이 모집단의 속성을 제대로 반영할 수 없으므로(국민 전체의 의견이라 보기 어렵다) 조사가 왜곡될 가능성이 매우 높다. 즉 설문에 참여한 응답자(표본)는 전체 사람들의 축소판이라고 보기 어렵다. 이런 표본을 편기(bias) 있는 표본이라 한다. 다음은 그 기사의 끝 부분이다.

한편 서울 극장가에서는 쇼트트랙 1500M 결승경기 직후인 2월 23일~24일 미국영화 ‘뷰티풀 마인드’가 박스 오피스 1위에 올랐으며 그 뒤로도 할리우드 스타가 대거 출연한 ‘오션스 일레븐’이 정상을 질주하고 있다.

### 3.2.1. 통계 용어 (glossary)

- 모집단(population)

조사자가 정보를 얻기 원하는 대상 전체를 의미한다. 대통령 후보 지지율에 대한 여론 조사의 경우 모집단은 대한민국 유권자 전체이다. ○○대학교 학생 학교 만족도 조사의 경우 모집단은 2004년 현재 등록 학생이 모집단이다.

- 표본 프레임(sampling frame)

모집단 전체 응답자와 그것을 구별할 수 있는 ID 리스트를 표본 프레임이라 한다. 학생 명부, 주민등록 명부가 그 예이다. 표본 프레임이 갖추어야 할 요소는 다음과 같다.

- ① 포괄성

표본 프레임이 조사 가능 대상을 대부분 포함하고 있어야 한다.

- ② 추출 확률

모집단 구성원이 표본으로 추출될 확률이 동일해야 한다.

- ③ 효율성

조사자가 조사되길 원하는 사람들만으로 표본 프레임을 구성한다.

- 모수(parameter)

모집단의 특성을 나타내는 값을 모수라 한다. 여론 조사에서 유권자 전체 후보자 지지율이 모수이다.

- 표본(sample)

모집단의 일부로 설문을 위해 선택된 응답자를 말한다. 여론 조사에서 전체 유권자 중 전화 조사에 선택된 1,500명 응답자가 표본이다. ○○대학교 학생 학교 만족도 조사를 위하여 200명을 뽑았다면 이들이 표본이다.

- 통계량(statistic)

모수를 추정하기 위하여 표본으로부터 계산된 값을 통계량이라 한다. ○○정책에 대한 지지율을 알아보기 위하여 전화 여론 조사를 실시하였다고 하자. 이런 경우에는 ○○정책의 지지율이 모수가 되고 1600명 응답자 중 600명이 정책을 지지하였다고 가정하면 0.4가 통계량이 된다. 통계량은 모수를 추정하기 위하여 사용된다.

## (1) 표본 프레임 (sampling frame)

표본 프레임은 설문 조사 대상이 되는 전체 응답자와 그 응답자를 구별할 수 있는 ID 리스트를 표본 프레임이라 한다. 목표 모집단이라고 불리는 표본 프레임은 “연구자가 잠정적으로 그의 연구에 적용하기 위하여 가정한 이론적 집단”이다. (Sjoberg&Nett, 1968). 표본 프레임은 연구자의 조사 계획에 따라 정의되며 ID 와 응답자와는 일대일 대응 관계가 있어야 한다.

○○대학교 학생들을 대상으로 만족도 조사를 하는 경우 2004년 1학기 등록한 학생들의 이름과 그들의 학번이 (혹은 주민 번호) 있는 리스트나 학생 이름과 일련 번호가 매겨진 리스트가 표본 프레임이 된다. 성폭력 피해자 실태 조사와 같이 모집단 규정이 불가능하여 표본 프레임이 존재하지 않거나 표본 프레임이 존재하더라도 실제 조사에서 표본 응답자들을 선택할 때 표본 프레임으로 이용하는 것은 여러 이유로 어려움이 있는 경우가 빈번히 발생한다.

## ① 여론 조사의 경우 어떤 표본 프레임을 사용할 것인가?

- ▶ 주민 등록 대장 사용: 공공 기관의 조사가 아니면 사생활 보호 차원에서 얻기 어렵고 얻을 수 있더라도 표본으로 추출된 사람에 한정하면 조사 비용 시간이 많이 소요된다.
- ▶ 전화나 전기 사용자: 전화 번호나 (전화 보급률이 거의 100%이고 조사가 간편하다) 전기 사용 명부를 이용할 수 있으나 하나의 번호(ID)에 다수의 사람들이 있으므로 조사 기관 마다 자체 선택 기준이 있다.
- ▶ 인터넷 조사: 표본 프레임이 특정 집단에 한정되는 문제가 발생한다.

## ② 한남대학교 학생들의 학교 만족도 조사

- ▶ 학생 이름과 ID 가 있는 리스트를 표본 프레임으로 한다. 적절한 표본 추출 방법에 의해 통계학과 9733001 학생이 표본 추출 되었다고 하자. 그 학생을 찾아가 설문을 받는 것은 시간적, 경제적으로 효율적이지 못하다.
- ▶ 단과 대학 별로 1~2 개 학과를 선택하고 각 학과 전공 수업시간을 이용하거나 학생들의 출입이 많은 정문에서 설문 조사를 하면 된다. 도서관이나 휴게실 같은 공간은 대표성 있는 표본을 추출하지 못할 가능성이 높으므로 피하는 것이 좋다.

**(2)전수 조사와 (census) 표본 조사**

모집단 전체를 조사하는 (설문) 조사를 census(전수 조사, 총 조사)라 한다. 전수 조사는 비용이 많이 들고 시간이 오래 걸리므로 정부가 국가적인 차원에서 주로 실시한다. 인구 센서스, 광공업 총 조사가 그 예이다. 우리나라에서는 5년마다 한 번씩 실시한다.

정부 이외의 다른 기관에서는 모집단 전체를 조사하는 것은 비용이나 시간 면에서 효율적인 방법이 아니므로 모집단의 일부만을 조사하는 표본 조사를 실시한다. 표본 조사를 실시하기 위해서는 모집단 전체에 대한 정보가 있어야 하는데 이를 표본 프레임이라 한다. 이 표본 프레임으로부터 원하는 수의 표본 응답자를 선택하는 방법을 표본 추출 방법이라 한다.

**3.2.2. 표본 조사 오차**

모수와 표본으로부터 계산된 추정치가 차이(편차)가 발생하는 이유는 ①잘못된 표본 추출 방법 ②표본 오차 ③비표본 오차 때문이다. ②와 ③을 표본 조사 오차라 한다.

**(1)random sampling error (확률 표본 오차): 실제 계산 가능**

모집단 전체를 조사하지 않고 일부만 조사하기 때문에 표본으로부터 얻은 추정치와 모집단의 참 값 사이에 발생하는 차이를 확률 표본 오차라 한다. 이를 허용 오차, 혹은 오차 한계라고도 한다. 모집단의 일부인 표본으로부터 얻은 결과와 모집단의 정보와 차이를 넓은 의미의 표본 오차라 한다. 이 오차는 모집단의 응답자가 표본 프레임 작성 시 (혹은 모집단 고려 시) 제외되어 표본에 포함될 가능성이 없어짐으로써 발생하는 오류도 이에 해당되지만 이는 계산 불가능하다. 그러므로 표본 오차라 함은 계산 가능한 확률 표본 오차를 의미한다.

표본 오차는 표본의 크기  $n$ 을 크게 하면 줄어든다. 다음은 SRS 방법 사용한 경우 확률

표본 오차는  $1.96\sqrt{0.5^2/n}$  (신뢰 수준 95%,  $n$  = 표본 응답자) 이다.

표본 오차는 표본의 크기를 늘리면 줄일 수 있으나 비용과 시간이 증가하고 조사 업무의 과중으로 인하여 비표본 오차가 발생할 가능성이 커진다. 비표본 오차를 줄이려면 조사자가 조사 전 과정을 스크린 하면서 오류가 발생하지 않고 정확한 조사가 되도록 하는 노력이 필요하다.

## (2) non-sampling error (비 표본 오차)

비표본 오차가 발생하는 이유는 다양하다.

- ①절차 오차(processing error): 조사원들의 불성실이나 외형적인 영향, 응답자의 무성의한 응답, 적절하지 못한 조사 방법(예: 일대일 면접, 우편 조사, 전화 조사) 선택이나 조사 과정의 고의적 비고의적 실수, 자료 컴퓨터 입력 시 발생하는 오류 등 조사 과정에서 발생하는 오류이다. 절차 오차는 조사자가 충분히 조심한다면 줄일 수 있다.
- ②무응답(non-response)이 문제이다. 응답하는 자체를 싫어하는지 답하기 꺼려 하지 않았는지 알 방법이 없다. 그러므로 유사한 인구학적 특성을 가진 응답자들의 응답 결과를 이용하여 보정한다. 여론 조사 기관은 이에 대한 know-how 를 가지고 있다.
- ③응답 오차: 타인에게 드러나기 꺼려 하는 문항에 (예: IQ, 소득) 대해서는 높여 응답할 가능성이 있는데 이를 응답 오차라 한다. 응답 오차를 줄이기 위해서는 조사 과정에서 응답자가 비밀을 보장 받는다고 느낄 수 있도록 조사원이 멀리 떨어져 있거나 설문지를 본인이 직접 박스 함에 넣게 하는 방법이 있을 이용한다.

## 3.2.3. 표본의 크기 결정

국민 전체 대상 전화 조사 경우와 같이 모집단의 크기가 매우 큰 경우 설문 조사의 표본 크기 결정 공식은 다음과 같다.

$$\text{표본 오차} = z_{\alpha/2} \sqrt{0.5^2 / n} \text{ (SRS 표본 추출 시 표본 오차) 로부터 } n = \frac{z_{\alpha/2}^2 \times 0.5^2}{(\text{표본 오차})^2}$$

신뢰 수준이 90%이면  $z_{\alpha/2}=1.645$ , 95%이면  $z_{\alpha/2}=1.96$ , 99%이면  $z_{\alpha/2}=2.54$  이다. 신뢰 수준 95% 하에서 표본 오차가 2.5% 정도가 되려면 (대부분 여론 조사 기관이 이를 사용) 표본의 크기는 1,500 명~1,600 명 수준이다.

모집단의 크기가 아주 크지 않은 경우 표본의 크기는 얼마로 하는 것이 좋은가? 모집단 전

체의 크기를  $N$  이라 하면 적절한 표본의 크기는 이 공식  $n = \frac{z_{\alpha/2}^2 \times 0.5^2}{(\text{표본 오차})^2 + z_{\alpha/2}^2 (0.5)^2 / N}$  에

의해 결정된다. 일반적으로 모집단 크기의 10%나 200~300 을 표본의 크기로 사용한다. 만약 여론 조사가 아닌 경우에는 0.5 대신 모집단의 표준 편차( $\sigma$ : 알고 있다고 가정하면)나 분포가 균일 분포인 경우에는  $0.29R$  ( $R$ 은 범위), 좌우 대칭이면  $0.24R$ , 치우친 분포인 경우

에는  $0.21R$  을 사용하면 된다. 즉 일반 선택 문항 선택 분포가 균일하면  $n = \frac{z_{\alpha/2}^2 \times (0.29R)^2}{(\text{표본 오차})^2}$ , 만족도 문항과 같이 좌우 대칭인 문항은  $n = \frac{z_{\alpha/2}^2 \times (0.24R)^2}{(\text{표본 오차})^2}$  을 사용하면 된다.

다음은 모집단의 크기, 신뢰 수준, 표본 오차에 따른 표본 크기를 정리한 표이다.

모집단의 크기 (N)	표본의 크기					
	95%신뢰수준			99%신뢰수준		
	± 3%	± 5%	± 10%	± 3%	± 5%	± 10%
500	250	218	81	250	250	125
1,000	500	278	88	500	399	143
1,500	624	306	91	750	460	150
2,000	696	323	92	959	498	154
3,000	788	341	94	1,142	544	158
5,000	880	357	95	1,347	586	161
10,000	965	370	96	1,556	622	164
20,000	1,014	377	96	1,687	642	165
50,000	1,045	382	96	1,777	655	166
100,000	1,058	383	96	1,809	659	166

**3.2.4. 표본 추출 방법 (확률적 추출 방법)**

모집단 전체를 조사하는 것은 일반적으로 시간적, 경제적인 면에서 비효율적일 뿐 아니라 불가능하므로 모집단의 일부, 즉 표본에 (sample) 대한 자료를 수집하여 모집단 전체의 정보를 얻게 된다. 그럼 어떤 방법으로 모집단과 유사한 표본을 얻을 수 있을까?

○○대학교 학생 학교 만족도 조사를 위하여 200 명의 응답자를 뽑으려 한다고 하자. 학적부를 표본 프레임으로 하고 각 학생들이 표본으로 선택될 확률이 동일한 바방법을 사용하면 좋은 표본을 뽑게 될 것이다. 이를 확률적 표본 추출 방법이라 (probabilistic sampling method) 하며 SRS, Systematic, Clustering, Stratified 방법 등이 있다.

학적부를 이용할 수 없어 표본 응답자를 얻기 위하여 도서관이나 오정못 근처에서 조사를 한다면 한남대학교 학생을 대표하는 표본을 얻을 수 없을 것이고 많은 학생들이 표본으로 추출될 가능성이 없게(그러므로 각 학생의 추출 확률이 동일하지 않다) 된다. 이런 추출 방법을 비확률 추출법(non-probabilistic)이라 하고 편리 추출 방법(convenience sampling method), Snowball 추출 방법, 할당(Quota) 추출 방법, 목적(purposive) 추출 방법 등이 있다.

#### (1)SRS (Simple Random Sampling: 단순 임의 추출)

표본 프레임에 임의의 수를 부여하고 난수표를 이용하여 표본을 뽑는 방법이다. 난수표는 0-9 까지의 숫자를 늘어 놓은 것으로 1)표의 어떤 위치든 0-9 가 있을 확률은 같고 2)한 위치의 숫자가 다른 위치의 숫자에 영향을 미치지 않도록 만든 표로 형태는 다음과 같다.

Random digits						
19223	95034	05756	28713	96409	12531	42544
73676	47150	99400	01927	27754	42648	82425
45467	71709	77558	00095	32863	29485	82226
52711	38889	93074	60227	40011	85848	48767
95592	94007	69971	91481	60779	53791	17297
68417	55013	15529	72765	85089	57067	50211
82739	57890	20807	47511	81676	55300	94383
60940	72024	17868	24943	61790	90656	87964
36009	19365	15412	39638	85453	46816	83485

표본 프레임의 총 응답자 수가 2 자리이면 표의 아무 곳이나 시작하여 두 자리씩 끊어 차례로 읽으면 된다. 총 가능 응답자가 900 명이면 세 자리씩 읽으면 된다. 만약 표의 끝에 도달하면 표의 1 행 1 열부터 시작하거나 다음 페이지의 표로 넘어가면 된다. SRS 방법은

모집단의 각 개체(응답자가)가 선택될 가능성이 모두 같다는 가정 하에서 표본을 추출하는 방법이므로 가장 이상적인 방법이다.

○○대학교 학생(12,000 명) 고객 만족도 조사를 위하여 200 명을 SRS 방법으로 뽑는다고 하자. 표본 프레임은 학생들의 이름과 각 학생에게는 1-12000 ID 가 부여된 리스트가 있다고 하자. 난수표에서 임의의 곳을 지정한 후 ID 가 12000 까지 있으니 5 자리 숫자씩 읽으면 된다. 예를 들어 5 행 3 열부터 시작한다면 59294(없으니 뛰어 넘기), 00769 (769 학생 선택), 97191(뛰어 넘기), 48160(뛰어 넘기)...

[참고]엑셀을 이용한 난수 생성



난수 생성		
변수의 개수(V):	<input type="text" value="1"/>	확인
난수의 개수(B):	<input type="text" value="200"/>	취소
분포(D):	<input type="text" value="일양 분포"/>	도움말(H)
모수		
시작(E)	<input type="text" value="1"/>	종료(A) <input type="text" value="12000"/>
난수 시드(R):	<input type="text" value="7033001"/>	
출력 옵션		
<input type="radio"/> 출력 범위(O):	<input type="text"/>	
<input checked="" type="radio"/> 새로운 워크시트(P):	<input type="text"/>	

200: 표본의 크기

12000: 모집단 크기

Seed: 임의의 수

A
708.8484
5090.697
10671.09
9015.905
917.2114
9818.597
7075.089

→ ID 708, 5090, 10671 ... 학생을 선택하면 된다.

#### ▪장점

모집단을 대표하는 표본을 뽑을 수 있다. 가장 ideal 한 방법이다.

#### ▪단점

①실제 표본 프레임을 구하는 것이 불가능한 경우가 많다.

②표본 응답자에게 반드시 해야 하므로 시간과 비용이 많이 든다.

가장 이상적인 방법이지만 표본 프레임 리스트가 필요하므로 현실적이지 못해 설문 조사에서는 거의 이용되지 않고 실험 계획법이나 모집단 크기가 적어 추출 단위에 일련 번호 부여가 쉬운 조사에 이용된다.

#### (2)Systematic sampling (계통 추출법)

SRS 방법과 유사하지만 SRS와는 달리 난수표를 한 번만 이용한다. 1)모집단 전체에 대한 표본 프레임이 필요하고 2)각 응답자가 표본으로 뽑힐 가능성이 모두 동일하다는 점에서 SRS 방법과 동일하지만 난수를 한 번만 이용한다는 점에서 다르다.

계통 추출방법은 표본 추출 방법은 다음과 같다.

①모집단 전체 응답자에 일련 번호를 부여한다. 물론 꼭 연속 번호일 필요는 없다.

②모집단의 수를 표본의 수로 나눈다. 그 값을  $K$ 라 하자. ( $K = N/n$ )

③1 부터  $K$ 까지의 응답자 중 난수를 이용하여 한 명을 뽑고 그 다음부터는 그 난수에  $K$ 씩 더해가면서 해당되는 일련 번호를 뽑으면 된다.

○○대학교 학생(12,000 명) 만족도 조사를 위하여 표본을 200 명을 계통 추출 방법으로 뽑는다고 하자.  $12000/200=60$ 이므로 1-60 사이의 임의의 숫자를 하나 난수표로부터 선택한다. 5 행 3 열이라면 59이므로 ID 59, 119(59+60), 179, -- 학생을 표본 추출하면 된다.

•장점

- ①모집단을 대표하는 표본을 뽑을 수 있다.
- ②매번 난수를 찾는 번거로움이 없다.

•단점

- ①실제 표본 프레임을 구하는 것이 불가능한 경우가 많다.
- ②표본 응답자에게 반드시 해야 하므로 시간과 비용이 많이 든다.
- ③일련 번호에 따른 체계적 차이가 존재하면 조사가 왜곡된다.

계통 추출 방법이 이용되는 예를 들어보자.

**백화점 고객 만족도** 백화점 고객을 대상으로 만족도를 조사하는 경우 백화점 고객 리스트(카드 회원)를 이용하여 표본을 얻을 수 있으나 백화점 앞에서 조사하는 경우 표본을 뽑을 때 이 방법을 사용한다. 조사 일에 예상 고객 인원은 5,000 명이고 표본 수를 200 명이라 하면 25 번째, 50 번째, 75 번째, ...고객을 조사하면 된다. 물론 응답 거절이나 무응답으로 인하여 한-두 번호 오차는 있을 수 있다.

**전화 조사** 전화 번호 선택 시 일정 번호를 건너 뛰면 한다. 예를 들면 629-7622 이 선택 되면(Random Digit Dialing) 629-7722, 629-7822, ... 이런 식이다. 가정 전화와 회사 전화 번호가 어느 정도 그룹화 되어 있는 미국에서는 이런 방법이 가능하다. 물론 전화 여론 조사는 층화 추출(지역적 차이 고려)+계통 추출 방법이다.

(3)Stratified sampling (층화 추출법)

모집단을 인구학적 특성(예: 성별, 학년별, 직업별, 지역별)에 의해 그룹화 한 후 각 그룹에서 일정한 크기의 표본을 선택하여 전체 표본을 구성하는 방법으로 이 그룹을 층(strata)이라 한다. 설문 성향은 차이가 있으므로 각 그룹 내의 응답자들이 일정한 수만큼 표본으로 추출되어야 한다.

- ①모집단을 인구학적 특성에 의해 분류한다. 모집단을 응답 성향이 서로 다를 것이라 생각되는 인구학적 변인을 이용하여 모집단을 그룹(층)으로 나눈다.
- ②각 층에서 표본을 몇 개나 뽑을지는 층을 구성하는 크기에 따라 결정한다.
- ③각 그룹에 배당된 표본을 추출하는데 SRS 방법이나 계통 추출 방법을 사용한다.

○○대학교 학생(12,000 명) 만족도 조사를 위하여 200 명을 층화 추출 방법으로 뽑는다고 하자. 학교에 대한 학생 만족도는 학년별 차이가 있을 것이라 생각할 수 있으므로 층을 학년으로 구성하면 된다. ○○대학교 학생들의 학년별 비율을 조사하니 3:3:2:2 이다. 그러므로 1 학년, 2 학년은 60 명, 3, 4 학년은 각각 40 명을 뽑으면 된다. 각 층(학년)에 배당된 표본의 크기를 뽑을 때는 SRS 방법을 사용한다. 1 학년이 360 명이므로 1 학년에 일련 번호를 부여한 후 난수를 이용하여 60 명을 선출하면 된다. 다른 학년들도 동일한 방법으로 표본 응답자를 추출하면 된다.

학교 만족도에 성별 차이가 있다고 판단되면 ○○대학교 성비를 조사한다. 55%(남자):45%(여자)라면 6,600 명 남자 중 110 명을, 5,400 명의 여자 중 90 명을 선출하면 된다. 물론 각 성에서 표본을 뽑을 때는 SRS 방법이나 계통 추출방법을 사용한다.

모집단을 그룹화할 때 그룹을 나누는 변수를 하나만 사용할 필요는 없다. 그룹을 나누는 특성 변수가 2 개인 경우 Two-stage(2 단계) 층화 추출, 3 개 이상인 경우 Multi-stage (다단계) 층화 추출이라 한다. 예를 들어 성과 학년 모두를 이용하여 모집단을 그룹화 한다면 우선 성별에 의해 그룹을 나누고 각 그룹을 학년에 의해 다시 그룹화 하여 8 개의 그룹에 대해 표본을 추출하면 된다.

#### •장점

표본 추출 비용과 시간을 절약할 수 있다.

#### •단점

그룹화를 위하여 고려한 특성에 따라 그룹간 응답의 차이가 없다면 조사 비용이 증가한다.

OO기업에서 사내 직원 만족도를 조사한다고 하자. 사원 ID 를 이용하여 SRS 방법을 사용할 수 있으나 회사에 대한 만족도가 직급별 차이가 존재할 것이므로 직급을 층(그룹)으로 하여 층화 추출 방법을 사용하는 것이 적절하다. 기업의 규모가 크면 성별 혹은 입사 연수를 또 다른 층으로 사용하여 Multi-stage 층화 추출 방법을 사용하면 된다.

전화 여론 조사의 경우 응답자들을 지역에 의해 층화하고 전화 번호 선택은 계통 추출 방법을 이용하게 되므로 층화-계통 추출 방법을 이용하게 된다.

#### (4) Cluster sampling (군집 추출법)

모집단을 그룹화 한다는 면에서는 층화 추출 방법과 동일하지만 그룹간에는 응답의 차이가 없다는 점이 다르다. 군집 추출법에서 그룹은 군집이라 한다. 그룹 간 응답의 차이는

없으므로 한 그룹이 임의로 선택되면 (SRS 방법) 그 그룹에 속한 사람만을 대상으로 표본 크기만큼 추출하면 된다.

- ①모집단을 인구학적 특성에 의하여 군집으로 나눈다.
- ②난수에 의해 군집을 임의로 선택한다. 선택된 군집에 속한 모든 응답자를 표본으로 선택한다.
- ③안락 그룹에 속한 응답자를 모두 뽑으면 표본보다 많은 경우는 SRS 방법에 의해 군집에 속한 일부 응답자만 선택하면 된다.

○○대학교 학생(12,000 명) 가족에 대한 의식 조사를 위하여 표본을 200 명을 군집 추출 방법으로 뽑는다고 하자. 가족에 대한 인식은 단과 대학별 차이는 없다는 가정이 적절하다면 대학에 1-8 번 번호를 부여하고 난수표를 이용하여 8 개 단과 대학 중 한 개를 선택한다. 대학이 8 개 밖에 안되므로 종이에 대학 이름을 적고 그 중 하나를 뽑아도 된다. 경상 대학이 추출되었다면 경상 대학 학생 중 200 을 SRS 방법으로 뽑으면 된다.

#### ▪장점

- ①추출 비용과 시간을 절약할 수 있다.
- ②층화 추출 방법보다 비용과 시간을 더 줄일 수 있다.

#### ▪단점

나눈 그룹간 응답 성향의 차이가 있다면 조사 결과가 왜곡된다.

OO 기업에서 사내 식당에 대한 만족도 조사를 위하여 부서별 식당 이용의 차이가 없을 거라는 가설 하에 부서를 임의로 선택하여 그 부서 직원 전체를 대상으로 조사한다.

#### (5)어느 방법을 선택할 것인가?

SRS 방법은 가장 이상적인 방법이지만 표본 프레임의 작성이 어렵고 조사과정에서 비용과 시간이 너무 많이 든다. 군집 추출 방법은 비용이나 시간 면에서 효율적인 방법이지만 응답 성향의 차이가 없는 군집을 얻는 것은 현실적으로 불가능한 경우가 많다.

현실적으로 가장 많이 사용하는 방법은 층화 추출 방법이다. 각 층(그룹)에 할당된 표본을 뽑는 경우 SRS 방법이나 계통 추출 방법을 사용하므로 하나의 추출법이 단독으로 사용되는 것이 아니라 여러 방법이 결합되어 사용된다.

**(6) 한국 갤럽 연구소 표본 조사 방법**

규모 비례 확률 방법을 사용하여 전국 200 개 지역을 층화하고(층화 추출) 일련의 계통 추출 방법으로 가구 내 응답자를 선택한다. 좀더 상세히 살펴보면 표본 추출은 4 단계로 이루어진다.

① 6 개 도시(서울, 부산, 대구, 인천, 대전, 광주), 8 개도(경기, 강원, 충남·북, 경남·북, 전남·북) 12 개 층으로 나누고 도는 시, 읍, 면으로 나눈다.

② 6 개 도시, 시, 읍, 면의 모집단을 배열한 후 동(면은 리)을 뽑는다. (계통 추출) 이것이 1 차 표본 지역(primary sampling location)이다. 표본 지역 수는 표본 수에 따라 달라지나 표본의 크기가 1,500 이면 200 정도를 얻는다.

③ 실질 최종 표본 지역(actual final sampling location)인 반(부락)이 선택될 때까지 계통 표본 추출을 반복한다. 반은 20 가구, 부락은 20-80 가구로 구성한다.

④ 조사자가 선정 지역을 직접 방문하여 주민 명부에서 8 가구를 임의 선정한다. 응답자가 18 세 이상인 사람 중 생일이 가장 빠른 사람을 조사한다. 1 차 방문 시 그 사람을 만나지 못하는 경우 재 방문한다.

전화 조사의 경우에는 ②~③ 단계를 RDD 가 담당한다. 전화 번호 선택 시 계통 추출 방법이 사용된다.

**3.2.5. 비확률 표본 추출 방법 (non-probabilistic sampling method)**

앞 절에서 언급한 확률적 표본 방법을 적용하여 표본을 추출하려면 다음 이유로 현실적인 어려움에 봉착하게 된다.

- 표본 프레임을 작성할 수 없는 경우: 성폭력 피해자 실태 조사
- 표본 프레임을 얻을 수 있어도 조사 비용이 많이 드는 경우: 표본으로 추출된 특정 학생들의 ID 와 이름 리스트가 일일이 찾아 다니며 조사하는 경우.

이런 어려움을 해결하는 방법으로 비확률적 표본 추출 방법을 사용하게 된다. 비확률 표본 추출 방법을 사용할 때는 표본 오차 계산이 불가능하다.

**(1) Convenience sampling (편리 추출 방법)**

조사자가 자신의 편리한 방법으로 표본을 얻는 것을 편리 표본 추출 방법(convenience sampling)이라 한다. 때로는 우연 추출 방법(incidental sampling)이라고도 한다. 편리 추출

방법에서 가장 중요한 것은 객관적 타당성을 유지하는 것이다. 예를 들어 ○○대학교 학생 학교 만족도 조사를 위하여 도서관에서만 표본을 추출하는 경우보다는 각 단대 입구에서 조사한다면 이 조사는 객관적 타당성을 갖는다. 편리 추출 방법은 객관성만 유지할 수 있다면 가장 유용한 방법이다,

#### (2) Snow-ball 방법

성 폭력 피해자의 실태 조사, 에이즈 환자의 의식 조사와 같이 어떤 사람들이 응답 대상인지 모르는 경우 해당 사람들이 생길 때마다 조사하는 표본 추출 방법이다. 성 폭력 피해자가 경찰에 신고하거나 상담소를 찾는 경우 그 사람을 대상으로 설문 조사를 실시하면 된다.

#### (3) 할당 추출법 (quota sampling)

할당법(quota)이란 특정 변수(주로 인구학적 변인)의 구성비에 의해 표본을 선택하는 방법입니다. 예를 들어 표본 응답자 구성을 남자 51%, 여자 49%로 한다면 조사원이 조사 과정에서 그 비율을 맞추어 나가는 방법으로 비확률적(non-probabilistic) 표본 추출 방법입니다.

##### ▪장점

- ① 응답자 비율만 맞추면 되므로 조사가 간편하고 비용이나 시간적 측면에서 효율적이다.
- ② 모집단과 유사한 비율로 표본을 구성하면 신뢰도 높은 조사가 된다.

그러나 비율 맞추는데 급급하여 조사자의 편리성으로 인하여 조사 과정에서 편의가 발생할 가능성이 높고 비율 구성을 위한 변수(예: 성별) 선택이 잘못되면 조사의 신뢰성이 떨어진다.

#### (4) 판단 추출 방법 (judgmental sampling)

연구자가 모집단에 대한 사전 지식이 있는 경우 이를 이용하여 표본을 추출하는 방법이다. 저소득층 의식 조사를 하고자 할 때 저소득층인 사람들의 표본 프레임 작성은 거의 불가능하므로 조사자가 저소득층 밀집 지역인 ○○ 달동네를 표본 프레임으로 규정하고(가상 모집단의 일부) 그들 중 일부를 표본 추출하여 조사를 실시한다. 이 방법 역시 표본 프레임 규정에 대한 객관성 유지가 중요하다.

### 3.2.6. Comments

#### (1) Panel 구성

동일 조사를 여러 시점에 계속하는 경우 첫 번째 조사에 선택된 표본을 계속 이용하는 조사 방법이다. 여론 조사나 기업의 고객 만족도 조사를 여러 시점에서 성향, 변화 추이를 보고자 하는 경우 주로 사용한다. 현재 시청률 조사가 패널 조사 방법에 의존한다.

동일 응답자를 사용하므로 응답자의 성향의 변화를 알아보는 방법이지만 ①패널을 유지하는데 비용이 많이 들거나 현실적인 어려움이 있고 ②처음 표본 추출이 잘못되면 조사는 계속 왜곡되는 문제로 인하여 자주 사용되지 않는다.

시청률 조사의 경우는 **set-up box** 라는 장치를 이용하므로 계속 표본을 바꾸면 재설치, 이용 방법 설명 등 조사 비용이 증가하므로 패널 조사 방법을 사용한다. 한 번 잘못 추출하면 패널을 유지하기 힘들거나 조사가 계속 왜곡되므로 조사 기관은 자체 패널들을 스크린 하여 불성실한 패널은 대체하여 간다.

#### (2) Voluntary response sampling

온라인 조사와 같이 설문에 관심이 있는 사람들만 응답하거나 선물을 받고 싶어하는 사람들만 설문에 응답하게 되는 조사 방법의 표본을 자원 응답 표본(**voluntary response sampling**)이라 한다. 이 방법은 조사의 신뢰성이 현저히 떨어지므로 모집단을 대표할 수 있는 방법을 찾아야 한다. 예를 들어 사용자가 가장 많은 포털 사이트의 사용자 **DB** 를 표본 프레임으로 이용하면 된다. 선물은 응답 회수율 높이는 것이지 조사의 신뢰성을 높이지는 못한다.

#### (3) 출구 조사와 표본 추출 방법

출구 조사(**exit poll**)은 선거 결과를 사전에 예측하기 위하여 선거 당일 투표를 마치고 나오는 유권자를 대상으로 실시한다. 1997년 경기도 안양시 만안구 국회의원 보궐 선거에서 최초로 사용되었고 2000년 4월 13일 제 16대 총선에서 본격적으로 사용되기 시작하였다.

16대 총선에서는 선거구는 227개, 각 선거구는 50-60개 정도의 투표구가 있었다. 이 중 6개 정도를 임의로 선택하여(SRS) n 번째 사람(투표자 수를 예상하여 표본의 크기로 나누어)을 택하는 계통 추출 방법을 사용한다.

다음은 선거구 선택 방법을 정리한 것이다.

## ①최소자승법

지난 번 선거를 이용한다. j-번째 후보 i-번째 선거구 득표율  $p_{ij}$ , j-번째 후보의 득표율을  $p_j$  라면  $\sum_j (p_{ij} - p_j)^2$  을 최소화하는 6 개 지역을 선택한다. 당이 바뀐 경우, 새로운 선거구에는 적용 불가능하다.

## ②총화법

유사 투표구를 총화 하여 각 층에서 1-2 개씩 임의로 뽑아 투표구를 6 개 얻는다.

## ③순환식

시간대 별로 조사 투표구를 바꾼다. 투표구별 시간대 투표 성향이 다른 투표구가 혼재된 경우 조사가 왜곡될 가능성이 높다.

## ④대표지역 선정법

선거구의 특성을 잘 나타낼 것 같은 투표구를 조사자가 선택한다.

## 3.3. 척도

일정한 규칙에 의해 보기 문항들에 수치를 부여하는 도구를 척도(scale)라 한다. 척도를 얻을 수 있는 문항 보기 예는 “①매우 불만족”, “②불만족”, “③보통”, “④만족”, “⑤매우 만족”이다. 하나의 문항으로 응답자들의 만족 정도를 측정하는 경우 단일 척도 (문항)이라 하고 여러 개 척도 문항들의 합으로 임의의 개념을 측정하는 경우 이를 복합 척도 (문항)라 한다. 복합 척도에 의해 하나의 개념을 측정하는 경우 복합 척도를 구성하는 단일 척도 문항들의 신뢰도(reliability) 혹은 단일 차원성(내적 일치도: internal consistency)을 측정하게 된다.

- 복합 척도는 단일 척도로 제대로 측정할 수 없는 개념을 측정할 수 있고 복합 척도를 구성하고 단일 척도 문항 점수의 합으로 지표의 척도 값으로 사용한다.
- 척도는 질적인 의견을 양적인 값으로 변화시킴으로 다양한 통계 분석 방법 적용이 가능하다. 즉 분류형 변수(보기 문항)가 측정형 변수로 변환되는 것으로 각 문항 보기에 대해 비율만 계산할 수 있으나 척도 문항은 평균과 같은 하나의 대표값으로 요약될 수 있다.

- 단일 척도에 비해 복합 척도는 측정의 신뢰도를 높여 준다. 또한 단일 척도의 경우 1, 2, 3, 4, 5 점이라면 2 개 단일 척도 문항으로 구성된 복합 척도는 1, 2, 3, ..., 10 이 되므로 보다 측정형 변수의 성질에 가까워진다.

개개인을 등급화시키는 방법으로 등급간의 간격이 거의 동일하도록 조정하는데 사용되는 Thurston scale (서스톤 척도, differential rating: 차이 평가), 총합 평가(summation rating technique)이라 불리는 리커드 척도 (Likert scale), 척도를 구성하는 과정에서 문항들의 단일 차원성을 경험적으로 검증하도록 설계된 거트만 척도가 (Guttman scale, cumulative rating: 누적 평가) 흔히 사용된다. 리커드 척도가 가장 널리 사용되며 리커드 척도는 설문 조사 전에 점수(척도)가 정해지나 거트만 척도는 설문 조사 결과를 이용하여 척도가 정해진다. 리커드 척도 문항의 신뢰도 혹은 내적 일치도는 Cronbach  $\alpha$  값에 의해 판단되며 척도 문항의 그룹화는 요인 분석에 의한다. 이는 6 장에서 다룬다.

### 3.3.1. 서스톤 척도

L. L. Thurston 에 의해 제안된 척도로 각 척도 문항이 척도 상 어디에 위치할 것인가 평가자로 하여금 판단케 한 다음 대표적인 문항들로 척도 문항을 구성하는 방법이다. 각 문항에 대한 응답자의 평가 간 불일치도가 높으면 그 문항은 제외시키고 척도상의 각 점수를 대표할 수 있는 문항을 고른다. 척도를 구성하는 단계를 요약하면 다음과 같다.

- ①측정하고자 하는 항목(개념)을 명확히 규정한다.
- ②항목과 관련된 의견을 광범위하게 수집하고 의견을 100 개 정도 압축하고 카드에 간결하게 적는다.
- ③50-100 명의 평가자에게 제시하여 100 개 카드를 11 개 그룹으로 배열하되 5 개 부정적인 견해 (1 그룹이 가장 부정적), 1 개 중립적인 그룹, 5 개 긍정적인 그룹(11 그룹이 가장 긍정적)으로 나누게 한다.
- ④평가자의 판단에 근거하여 100 개 카드의 척도 점수를 계산한다. 카드의 척도 점수는 평가자의 점수의 중앙값을 포함한 그룹의 값이다. 예를 들어 A 카드 진술에 대해 50 명의 평가자가 3 그룹(3 점)에 10 명, 4 그룹(4 점) 10 명, 5 그룹(5 점)에 20 명, 6 그룹에 20 명이었다면 A 카드 척도 값은 5 점이다.

- ⑤평가자들간 척도 값의 차이가 심하거나 평가자들이 측정하려는 개념과 무관하다고 진술한 카드는 제외하고 각 값들이 등 간격이 되도록 20-30 개의 진술 카드를 선택하여 최종 설문지에 포함시킨다.

서스톤 척도는 문항 평가를 위하여 많은 시간과 노력이 필요해 거의 사용하지 않는다.

### 3.3.2. 리커트 척도

Rensis Likert 에 의해 제안된 척도로 다른 신뢰할 만한 조사자들이 사용하였던 척도 문항이나 자신의 경험이나 다양한 형태의 정보(신문, 서적, 발표 자료)를 이용하여 문항 보기를 구성한다. 척도 문항은 사실에 대한 판단보다는 가치를 묻는 것으로 간결하게 하며 5 점 척도(1, 2, 3, 4, 5 점 등간 척도)로 평가될 수 있도록 선택 보기가 구성된다.

- 측정하고자 하는 항목(개념)을 명확히 규정한다.
- 각 문항에 대해 5 점 척도로 구성된 응답 범주를 결정한다.

(예제)귀하는 OO 법안에 대해 어떤 의견을 가지고 계십니까?

- ①적극 반대    ②반대    ③중간 입장    ④찬성    ⑤적극 찬성

(예제)당신은 교양 세미나 과목에 대해 어떻게 생각하십니까?

- ①전혀 유익하지 않다    ②유익하지 않다    ③보통    ④유익하다    ⑤매우 유익하다

- 문항 내적 일치도(신뢰도)가 높은 유사 척도 문항을 합하여 하나의 척도 개념으로 판단한다. OO 대학 시설 만족도 조사를 위하여 여러 교내 시설에 대해 아래 척도 문항을 물은 후 유사 문항 (내적 일치도 높은)끼리 묶어 하나의 개념을 측정하게 된다.

(예제)귀하는 OO 시설에 대해 얼마나 만족하십니까?

- ①매우 불만족    ②불만족    ③보통    ④만족    ⑤매우 만족

장점으로는 서스톤 척도와는 달리 평가자를 필요로 하지 않기 때문에 척도 구성이 간단하고 편리하다. 또한 한 항목에 대한 응답의 범위에 따라 측정의 정밀성을 확보할 수 있다. 단점으로는 극단(1 점, 5 점)을 택하지 않는 응답자들이 있으면 마치 3 점 척도처럼 조사된다.

그리고 합계 점수가 뜻하는 바가 개념적으로 분명치 못하다. 즉, 점수의 단순한 합계에는 각 항목에 표현한 응답자의 태도의 강도가 묻혀버리기 때문이다. 예를 들어 교수 능력 만족도를 알아보기 위하여 강의 능력 만족도를 묻는 문항과 연구 능력 만족도를 묻는 문항을 구성했다고 가정하자. 강의 능력 3 점, 연구 능력 3 점을 응답한 사람과 강의 능력 5 점, 연구 능력 1 점을 응답한 사람은 각각 교수 능력에 대한 만족도는 6 점으로 동일하게 되는 문제가 발생한다.

### 3.3.3. 거트만 척도

Louis Guttman 에 의해 제안된 척도로 응답자 태도의 강도에 대한 연속적 증가유형을 측정하고자 하는 척도로 전제조건은 측정의 대상이 되는 척도가 단일 차원성을 지녀야 한다. 거트만 척도는 반응 중심 접근방법으로 문항과 응답자를 동시에 척도화한다. 척도를 구성하는 과정에서 문항들의 단일 차원성이 경험적으로 검증되도록 설계된 척도로 단일 차원적이고 누적적인데 이는 척도를 구성하는 여러 문항과 응답자가 갖는 총 평점의 관계가 누적되고 있기 때문이다.

구성 절차를 정리하면 (1)척도구성항목을 선정하고 (2)응답자의 응답을 scalogram 용지에 기입한 후 (3)재생계수(CR: coefficient of reproducibility)를 구한다. (4)그리고 구성항목을 조정하여 척도를 구성하면 된다,

scalogram 은 연구자가 거트만척도의 절차에 따라서 만든 척도가 완벽한 거트만 척도와 일치하는 정도, 즉 단일차원성, 누적성의 가정에 얼마나 부합하는지를 재생 계수(coefficient of reproducibility)를 통하여 그 정도를 파악한다. 재생계수 =  $1 - \frac{\text{응답의 오차수}}{\text{문항수} \times \text{응답자수}}$ 이다. 재생계수가 1 일 때 완벽한 척도구성가능성(scalability)을 갖는다. 보통 재생계수가 최소한 0.90 는 되어야 바람직한 거트만척도이다.

예를 들어 보자. 유사 개념을 측정하는 하는 문항 3 개가 있다고 하자. 단순성을 위하여 문항 보기는 “예”, “아니오” 로 구성되었다고 하자. 응답자 400 명을 선정하여 응답한 결과 “예”의 비율이 가장 높은 문항을 순서대로 배열한다.

	제일 높은 문항	그 다음 문항	제일 낮은 문항	지수 점수	척도 점수	빈도
척도 유형	+	+	+	3	3	60
	+	+	-	2	2	150
	+	-	-	1	1	130
	-	-	-	0	0	0
혼합 유형	-	+	-	1	2	15
	+	-	+	2	3	20
	-	-	+	1	0	15
	-	+	+	2	3	10

혼합 유형에 대해서는 지수 점수와 다른 척도 점수를 배정하고 재생 계수를 구한다. 재생 계수가 0.9 이상이면 배정한 척도 점수를 사용한다.

$$CR = (1200 - 60) / 1200 = 0.95$$

일단 문항이 척도화 되면 응답자를 서열화 한다. 장점으로는 (1)주로 질문이나 투표에 의한 개인의 태도를 측정하는데 유용하다. 다른 척도와는 달리 응답자 개인을 서열화하기 때문에 개인차에 대한 연구를 수행할 때 특히 유용하다. 단점으로는 (1)현실적으로 재생계수가 높으면서 위계적으로 구성된 문항을 만들고 배열하기가 매우 어려워 많은 문항을 척도화할 수 없다. (2)단일차원의 측정에는 문제가 없으나 다양한 특성을 포함하는 다차원적인 조사를 수행하기 위한 척도로는 구성이 불가능하다.

## [연습문제]

(1) ○○지역 성인 남녀를 대상으로 영화 보는 회수, 장르, 함께 보는 사람 등에 대한 실태와 영화관 만족도 설문 조사를 위하여 표본 200 명을 뽑으려고 한다. 아래 각 방법을 사용한다고 할 때, 표본 추출 과정을 기술하시오.

① SRS 방법

② 계통 추출 방법

③ 층화 추출법

④ 군집 추출법

⑤ ①~④의 문제점이 있으면 적고 현실적인 방법(비확률적 방법도 포함) 중 가장 좋은 추출 방법을 적으시오.

(2) 80 개 원의 면적을 계산하기 위하여 4 개만 추출하여 원의 면적을 추정하려고 한다. (페이지 63)

① SRS 방법에 의해 표본을 추출하고 원의 면적을 추정하시오.

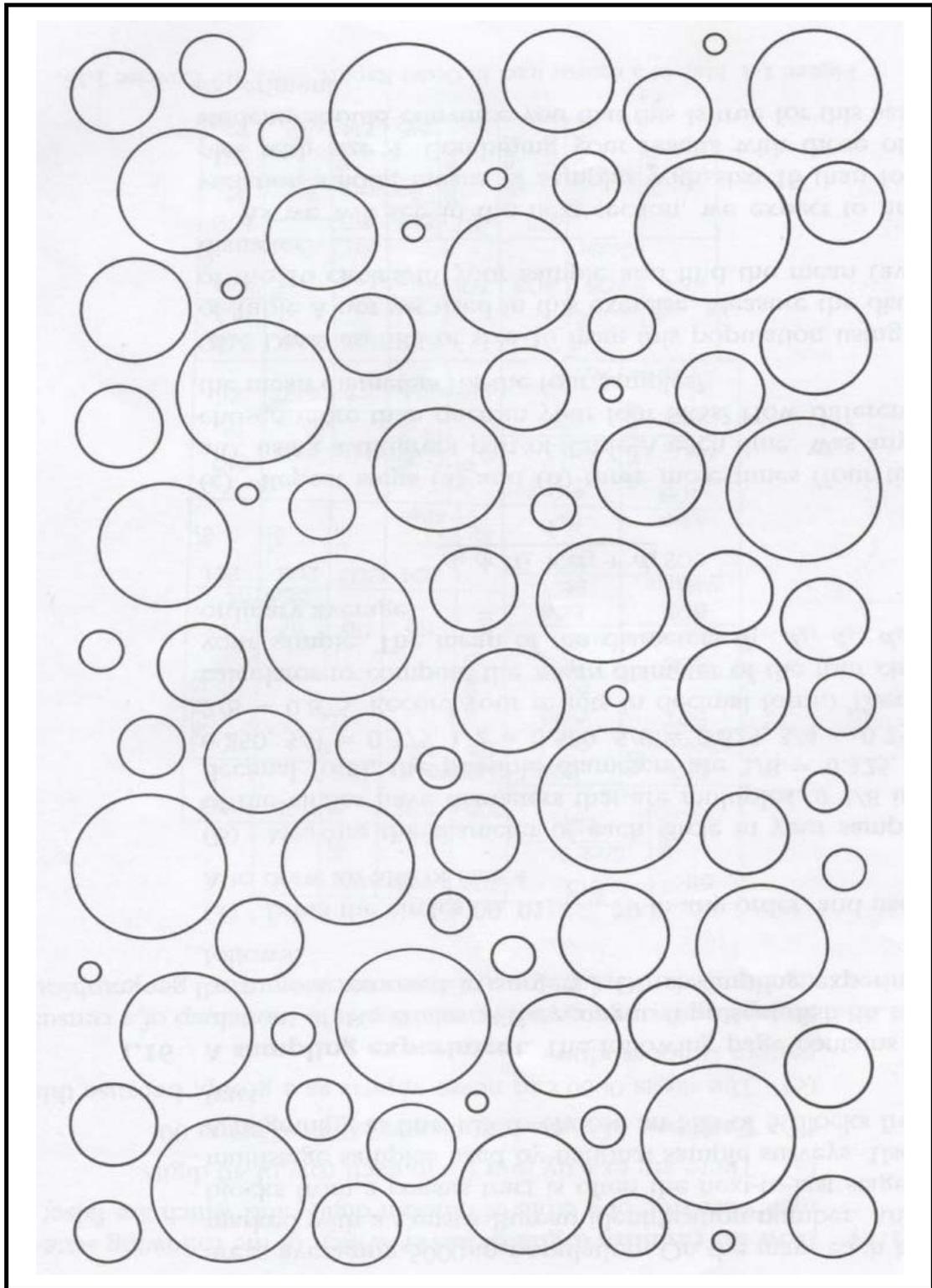
② 계통 추출 방법에 의해 표본을 추출하고 원의 면적을 추정하시오.

③ 층화 추출 방법에 의해 표본을 추출하고 원의 면적을 추정하시오.

④ 군집 추출 방법에 의해 표본을 추출하고 원의 면적을 추정하시오.

⑤ SRS 방법에 의해 표본 8 개를 추출하고 원의 면적을 추정하시오. ①의 결과와 비교하시오.

(3) 팀 프로젝트 설문 조사에 가장 이상적인 표본 추출 방법과 현실적인 방법을 제시하시오.



설문조사 <한남대학교 통계학과 권세혁교수>

설문조사 <한남대학교 통계학과 권세혁교수>