

제 14 장 표본조사

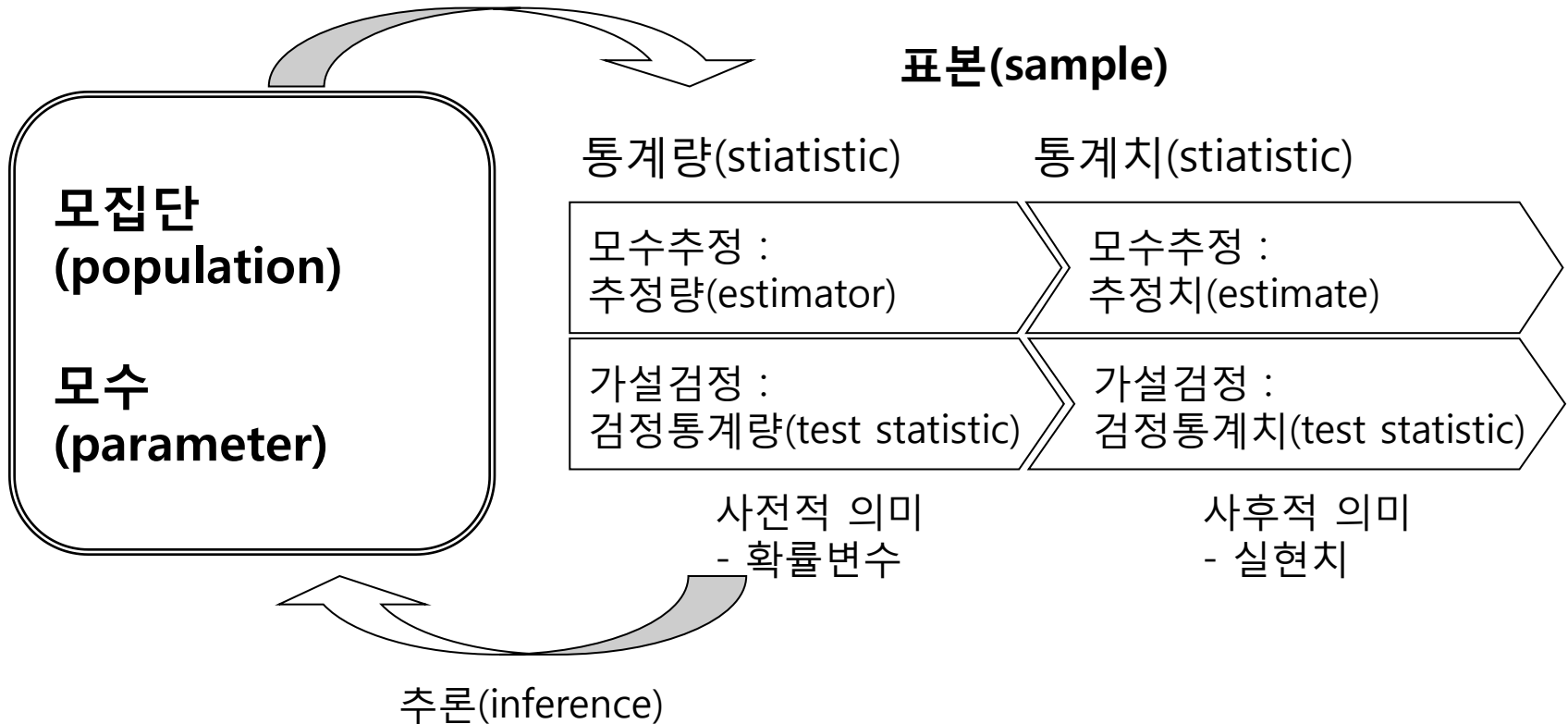
1. 모집단과 표본
2. 여론조사
3. 표본조사에서의 편의와 그 종류
4. 표본추출방법
5. 갤럽조사
6. 전화조사
7. 확률오차와 표준오차

1. 모집단과 표본

표본추출

표본추출(sampling)

표본(sample)



2. 여론조사

1993년의 LA 시장 선거: 여론조사와 실제 결과의 비교

LA Times Poll on LA Mayoral Election, LA Times, May 12, 1993

- among all registered voters: 민주당 Michael Woo 후보가 공화당 Richard Riordan 후보에 6% 포인트 앞서는 것으로 나타남
- among the likely voters: 공화당 Richard Riordan 후보가 민주당 Michael Woo 후보에 7% 포인트 앞서는 것으로 나타남
- 1993년 6월 8일의 실제 선거결과: 등록된 유권자의 44%만이 투표. 공화당 Richard Riordan 후보가 민주당 Michael Woo 후보에 8% 포인트 앞선 승리

2. 여론조사

다이제스트사 vs. 갤럽

1936년 미국 대통령 선거 결과의 예측

	루즈벨트의 득표율(%)	
실제선거결과	62	→ 실제 루즈벨트 승리
다이제스트사의 예측	43	→ 다이제스트사는 랜던 승리 예측
다이제스트사의 예측에 대한 갤럽의 예측	44	→ 갤럽은 다이제스트사가 랜던 승리 예측할 것으로 예측
갤럽의 예측	56	→ 갤럽은 루즈벨트 승리 예측

갤럽의 일방적 승리

- 다이제스트사의 잘못된 예측까지도 갤럽이 예측 (지피지기면 백전백승!)
- 다이제스트사의 표본추출방식에는 심각한 문제가 있었다.

3. 표본조사에서의 편의와 그 종류

표본추출편의 관련 두 가지 구절

- (i) “거리엔 괜찮은 사람들이 많은데 소개를 받으러 나온 자리엔 어디서 이런 사람만 나오는 거야” (유행가 구절)

- (ii) “서울이라고 못난이가 없을 리야 없겠지만 대처에서는 못난이들이 거리에 나와 행세를 하지 못하고, 시골에선 아무리 못난이라도 마음 놓고 나와 다니는 때문인지, 못난이는 시골에만 있는 것처럼 흔히 시골에서 잘 눈에 뜨인다.” (이태준 소설 ‘달밤’ 중)

3. 표본조사에서의 편의와 그 종류

편의(bias)

표본이 보이는 모집단과의 체계적인 차이를 편이라고 한다.

- 표본추출편의(sample selection bias): 확률적으로 표본이 모집단을 대표하지 못함
- 무응답편의(non-response bias): 응답하지 않는 사람들은 응답하는 사람들과 체계적으로 다를 수 있음
- 응답편의(response bias): 응답이 속마음과 다름
- 가구편의(household bias): 표본으로 뽑힐 확률은 가구원수가 많은 가구의 구성원이 가구원수가 적은 가구의 구성원보다 낮음

3. 표본조사에서의 편의와 그 종류

표본추출편의(sample selection bias)

다이제스트사의 표본추출방식

전화번호부나 클럽회원 명단에서 파악한 주소를 이용하여 여론조사

- 당시 전화나 클럽은 부의 상징
- 가난한 사람들은 표본에서 배제
- 가난한 사람에게 인기 있는 루즈벨트에게 불리한 예측치 나옴

- 표본에 포함시키고 표본으로부터 제외시키는 데 어떤 체계적인 경향이 개입되어 표본추출편의가 발생함
- 표본추출편의가 있는 경우 표본이 모집단을 대표하지 못함
- 표본추출편의 문제는 표본이 크다고 해서 해결되지 않음

3. 표본조사에서의 편의와 그 종류

표본추출편의(sample selection bias)



“무작위 추출”의 개념을 악용하는 아들!

3. 표본조사에서의 편의와 그 종류

상기하라! 돈 잃은 수많은 사람들의 소리 없는 아우성



돈을 딴 사람만 목소리가 큼니다.

3. 표본조사에서의 편의와 그 종류

무응답편의(non-response bias)

다이제스트사의 표본추출방식

설문지를 받은 1,000만 명 중 240만 명만이 응답

- 저소득층이나 고소득층이 응답을 거부하는 경향. 중산층 의견이 과도하게 반영됨
- 루즈벨트의 지지자들이 주로 응답하지 않음

- 응답하지 않은 사람들과 응답한 사람들 사이에 체계적인 차이가 있을 수 있다.
- 무응답자의 생각은 응답자와 다를 수 있다.
- 응답자가 모집단을 대표하지 못한다.
- 무응답비율이 높은 경우 무응답편의의 가능성을 고려해야 한다.

3. 표본조사에서의 편의와 그 종류

응답편의(response bias)

- 사람들의 응답은 설문지의 디자인, 문구, 혹은 질문의 어조나 조사원의 태도에 의해 영향을 받음
- 후보자의 이름 순서가 선거 결과에 큰 영향을 미치는 것이 대표적 사례임
- 응답 편의를 줄이려면 모든 응답자들이 가급적 같거나 비슷한 조사원과 면담하도록 하고 그 과정도 정해진 규칙에 따라 진행되도록 해야 한다.

3. 표본조사에서의 편의와 그 종류

가구편의(household bias)

- 가구별로 가구원 수에 큰 차이가 나고 가구당 한 구성원만 조사한다고 했을 때 표본으로 뽑힐 확률은 가구원 수가 많은 가구의 구성원이 가구원 수가 적은 가구의 구성원 보다 낮게 된다.
- 이에 대한 해결책으로는, (i) 표본으로 뽑힌 가구의 구성원 전원을 조사하는 방법, (ii) 가구원수에 비례하는 확률로 표본가구를 선정하는 방법 등이 있다

3. 표본조사에서의 편의와 그 종류

생존편의(survivorship bias)

- When estimating an overall rate of return from stock investment for the past 20 years:
- If you took a random sample of 100 stocks currently trading, this would be incorrect and overstate returns due to "survivorship bias".
- You would better sample 100 random stocks that were trading 20 years ago, as some might go bankrupt or merge under adverse circumstances.
- Ryu & Yoon (2013), "Relative performance of chaebol vs. non-chaebol in Korea over the past three decades: a value path approach" : 재벌성과를 측정하는데 있어 생존편의를 고려하기 위해 value path approach 이용

4. 표본추출방법

할당추출과 단순무작위추출

할당추출 (quota sampling)

모집단의 구성비에 따라 다양한 특성의 사람들을 뽑도록 미리 관측치수를 할당하여 표본추출

할당에 쓰인 조건 외의 나머지 특성에 관한 한 조사원의 자의가 개입되어 표본의 신뢰성이 떨어짐

단순무작위추출 (simple random sampling)

무작위 비복원 추출
-모집단에서 임의로 표본을 추출
-한번 뽑은 것은 복원하지 않음
-단순무작위표본
(simple random sample)

실제로 표본을 추출할 때 모집단 전체로부터 단순무작위추출하기는 매우 어려움
- 다단계군집추출이용

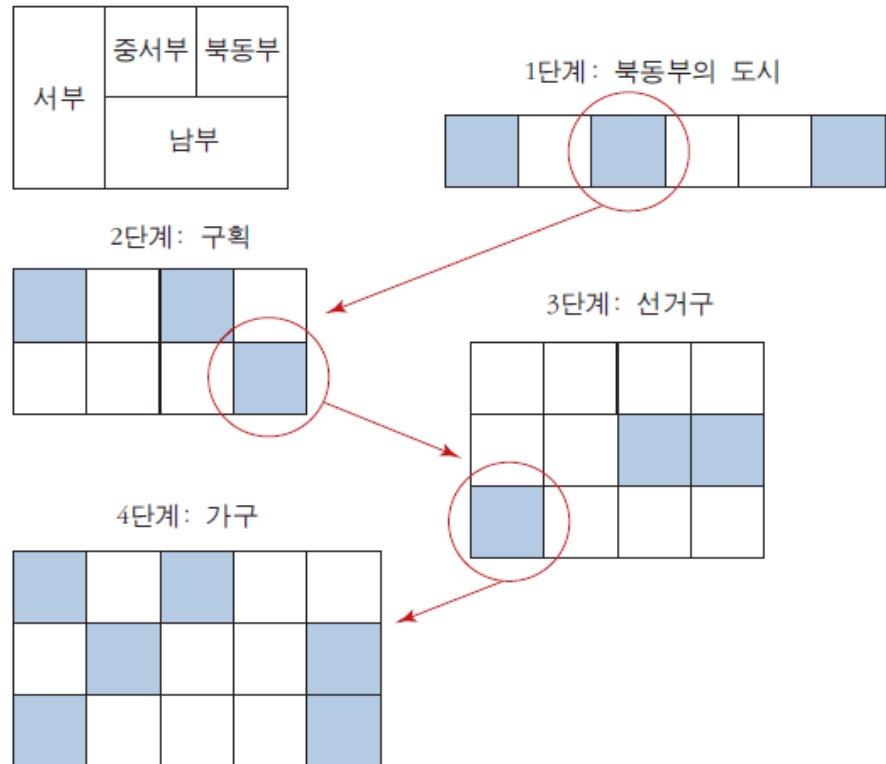
4. 표본추출방법

다단계군집추출(multistage cluster sampling)

갤럽의 표본추출방식

- 거주지를 바탕으로 해서 관측치수를 할당한다.
- 각 단계마다 확률적 방법을 써서 표본을 추출한다.

다단계군집추출 : 갤럽의 표본추출방법



4. 표본추출방법

단계마다 확률적 방법에 의한 표본추출

확률적 표본추출 방법의 특징

- 표본추출절차가 명확하고 이는 확률에 바탕을 둔다.
- 확률적 방법은 모집단을 이루는 개개의 구성원이 표본으로 선택될 확률을 계산할 수 있게 해준다.
- 조사원 스스로 누구와 면담할 지 결정할 수 없다.
- 제 3자에 의한 반복연구가 가능해지므로 연구결과의 통계적 유의성을 판단할 근거가 마련된다.

4. 표본추출방법

다단계군집추출에서 추출확률의 결정문제

표본추출 시 모집단에 속한 개개인이 표본으로 뽑힐 확률을 같게 만들어 주어야 바람직하다.

서울

	A			

부산

	B			

서울, 부산에서 3 개의 구획을 뽑는 경우 하나의 구획이 뽑힐 확률은

$$= 3/(20+10) = 1/10 \text{ 이다.}$$

속한 도시와 관계없이 A구획이 뽑힐 확률=B구획이 뽑힐 확률=1/10 이 성립하려면 1단계에서 도시를 뽑을 때

$$\text{서울이 뽑힐 확률} = 2/3$$

$$\text{부산이 뽑힐 확률} = 1/3$$

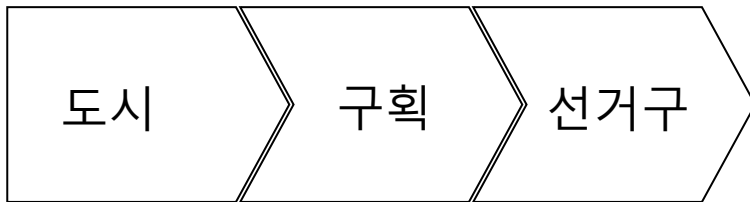
이 충족되어야 한다.

즉, 매 단계마다 최종 추출단위의 크기에 비례하여 추출확률을 조정해야 한다.

4. 표본추출방법

다단계군집추출에서 추출확률의 결정문제

매 단계에서 각 추출대상에 속한 가구수에 비례하여 추출확률을 정한다.



- 그 결과 하나의 가구가 표본에 뽑힐 확률은 그 가구가 속한 도시, 구획, 선거구와 관계없이 일정하게 정해진다.
- 추출확률을 가구원수가 아닌 가구수에 비례시키는 선에서 그침으로써 가구편의 (household bias)는 여전히 발생한다.
- 즉, 가구원수가 많은 가구의 구성원이 가구원수가 적은 가구의 구성원에 비해 표본으로 뽑힐 확률이 낮다.

5. 갤럽조사

갤럽조사

확률적 방법으로 표본을 추출하여 작은 크기의 표본을 가지고도 상당히 정확하게 결과를 예측한다.

- 편의를 최소화하려면 표본을 추출하는데 있어 객관적이고 공정한 확률적 방법을 사용해야 한다. 주관과 자의적인 판단은 편의를 가져오는 요인이 된다.
- 하지만 확률적 방법을 이용한다고 해서 모든 편의를 없앨 수 있는 것은 아니다.
 - 투표하지 않는 유권자
 - 아직 마음을 정하지 못한 유권자
 - 응답편의, 무응답편의 등
- 이를 해결하기 위해 갤럽은 여러 다양한 노력을 기울였다.

5. 갤럽조사

1984년 미국 대통령 선거 시 미국 갤럽이 사용한 설문지

SURVEY: A1813

DATE: October 25, 1984

No publication, reproduction, dissemination or other use of this questionnaire or any replies thereto, written or oral, is authorized by The Gallup Organization, Inc. Violators of this notice will be prosecuted to the fullest extent of the law.



Sponsored by leading Newspapers, Corporations and Agencies.

Copyright 1977 The Gallup Organization, Inc. Princeton, New Jersey 08540

SUGGESTED INTRODUCTION: I'm taking a GALLUP SURVEY. I'd like YOUR opinion on some topics of interest.

The Gallup Survey

Time started: _____

Time ended: _____

Length: _____

1. How much thought have you given to the coming November elections—quite a lot, or only a little?

- 1 Quite a lot
2 Some — (volunteered)
3 Little
Y None

2. Have you ever voted in this precinct or district?

- 1 Yes
2 No
Y Don't know

3. Where do people who live in this neighborhood go to vote?

- 1 Specify: _____
Y Don't know

4a. Are you NOW registered so that you can vote in the election this November?

- 1 Yes — (GO TO Q. 5)
2 No
3 Don't have to register — (GO TO Q. 5)
Y Don't know

4b. Do you plan to register so that you can vote in the November election?

- 1 Yes
2 No
3 Other: _____
Y None

5. Generally speaking, how much interest would you say you have in politics—a great deal, a fair amount, only a little, or no interest at all?

- 1 Great deal
2 Fair amount
3 Little
Y None

6. How often would you say you vote — always, nearly always, part of the time, or seldom?

- 1 Always
2 Nearly always
3 Part of the time
4 Seldom
5 Other: _____
Y Never vote

7. Suppose you were voting TODAY for president and vice president of the United States. Here is a Gallup Poll secret ballot listing the candidates for these offices. (TEAR OFF ATTACHED BALLOT AND HAND TO RESPONDENT.) Will you please MARK that secret ballot for the candidates you favor today — and then drop the folded ballot into the box. ([그림 14-2]의 투표용지를 건네줌)

INTERVIEWER: IF RESPONDENT HANDS BACK AND SAYS HE HASN'T MADE UP HIS MIND OR TO MARK IT, SAY: Well, would you please mark the candidates toward whom you lean as of today? IF RESPONDENT STILL CAN'T DECIDE OR REMARK THE BALLOT, PLEASE WRITE THAT BALLOT AND BE SURE TO DROP IT IN THE BOX.

8. Right now, how strongly do you feel about your chief executive officer, fairly strongly, or not strongly at all?

- 1 Very strongly
2 Fairly strongly
3 Not strongly at all
4 Didn't make choice
Y Don't know

9a. Do you, yourself, plan to vote in the election this November?

- 1 Yes
2 No — (GO TO Q. 10a)
Y Don't know

9b. How certain are you that you will vote — ABSOLUTE FAIRLY certain, or NOT certain?

- 1 Very strongly
2 Fairly strongly
Y Don't know

10a. If the elections for Congress were being held TODAY party would you like to see win in this Congressional District Democratic Party or the Republican Party?

- 1 Democratic — (GO TO Q. 11)
2 Republican
3 Other: _____
Y Undecided, refused

10b. As of today, do you lean more to the Democratic Party or the Republican Party?

- 1 Democratic
2 Republican
3 Other
4 Undecided
Y Refused

11. Here is a picture of a ladder. (HAND RESPONDENT) Suppose we say the top of the ladder (POINT) represents a person who definitely will vote in the November, and the bottom of the ladder (POINT) represents a person who definitely will not vote in it. How far up or down the ladder would you place (INTERVIEWER: CIRCLE NUMBER.)

- 10 9 8 7 6 5 4 3 2 1 0
Y Don't know

NOW, HERE ARE A FEW QUESTIONS SO THAT MY OFFICE CAN KEEP BACK OF THE CROSS-SECTION OF PEOPLE I'VE TALKED TO:

12. In politics, as of TODAY, do you consider yourself a Republican, Democrat, or Independent?

- 1 Republican
2 Democrat
3 Independent
4 Other: _____

13. In the election in November 1980—when Carter ran against Reagan and Anderson—did things come up which kept you from voting, or did you happen to vote? For whom?

- 1 Carter
2 Reagan
3 Anderson
4 Other
5 Voted, don't remember for whom
6 No, didn't vote
Y Don't remember if voted

14. Are you, or is your (husband/wife) a member of a labor union?

- 1 Yes, independent is
2 Yes, spouse is
3 Yes, both are
Y No, neither is

15. (HAND RESPONDENT CARD 2) Please tell me which of the categories on this card MOST NEARLY describes the kind of work the chief wage earner in your immediate family does. Just call off the number, please. (INTERVIEWER: IF THE CHIEF WAGE EARNER IS UNEMPLOYED, ASK WHAT TYPE OF WORK HE/SHE WOULD DO IF EMPLOYED.)

- 1 11
2 12
3 13
4 14
5 15
6 16 Other: _____
7 17 Can't say
8
9
10

16. What was the last grade or class you COMPLETED in school?

- 1 None or Grades 1-4
2 Grades 5, 6, 7
3 Grade 8
4 High school, incomplete (Grades 9-11)
5 High school, graduated (Grade 12)
6 Technical, trade, or business
7 College, university, incomplete
8 College, university, graduated

17. What is your religious preference—Protestant, Roman Catholic, Jewish, or an Orthodox church such as the Greek or Russian Orthodox Church?

- 1 Protestant
2 Roman Catholic
3 Jewish
4 Orthodox Church
5 Other: _____
Y None

18. How many persons 18 years and over are there now living in this household, including yourself? Include lodgers, servants, or other employees living in the household. (CIRCLE NUMBER)

1 2 3 4 5 6 7 8 9 or more

19. (HAND RESPONDENT CARD 3) From what nationality group or groups are you mainly descended? Just call off the number please.

- 1 11
2 12
3 13
4 14 Don't know (VOLUNTEERED)
5 or refused
6
7
8
9
10

20a. We are interested in finding out how often people are at home to watch TV or listen to the radio. Would you mind telling me whether or not you happened to be at home yesterday (last night/last Saturday) at this particular time? (INTERVIEWER: SEE INTERVIEWER'S BULLETIN FOR HANDLING THIS QUESTION.)

- 1 Yes, at home
2 No, not at home

20b. How about the day (night/Saturday) before at this time?

- 1 Yes, at home
2 No, not at home

20c. And how about the day (night/Saturday) before at this time?

- That was _____
1 Yes, at home
2 No, not at home

21. And what is your age? RECORD AGE: _____

22. CHECK WHETHER:

- 1 White man
2 White woman
3 Black man
4 Black woman
5 Other man (SPECIFY) _____
6 Other woman (SPECIFY) _____

So that my office can check my work in this interview if it wants to, may I have your name, address, and telephone number please?

NAME: _____

ADDRESS: _____

CITY: _____ STATE: _____ ZIP: _____

TELEPHONE: Area Code _____ Phone No. _____ Y No tel.

PLACE INTERVIEWER BADGE NUMBER HERE

I hereby attest that this is a true and honest interview.

(Interviewer's Signature) _____

Date of interview: _____

Time interview ended: _____

5. 갤럽조사

2007년 한국 대통령 선거 시 한국갤럽이 사용한 설문지

Gallup 0712

[17대 대선 관련 여론조사]

1014

연락하십니까? 여기는 한국갤럽조사연구소입니다. 19일에 있을 대통령 선거와 관련하여 국민 여러분의 의견을 알아보고 있으니 잠시만 협조해 주시면 감사하겠습니다.

- 지역 : 1. 서울 2. 부산 3. 대구 4. 인천 5. 광주 6. 대전 7. 울산 8. 경기 9. 강원 10. 충북 11. 충남 12. 전북 13. 전남 14. 경북 15. 경남 16. 제주 5*
- 지역크기 : 1. 대도시 2. 중소도시 3. 읍/면 7

SQ1) 실례지만, ○○님의 연세는 올해 민으로 어떻게 되십니까?
8-9
만 _____ 세 → **만 19세 이상 조사 가능**

SQ2) 성 별 : 1. 남자 2. 여자 10

문 1) ○○님께서는 19일에 있을 대통령 선거에 투표하시겠습니까? 아니면 투표하지 않으실 겁니까?
11
‘꼭 할 것이다’, ‘아마 할 것 같다’, ‘아마 하지 않을 것 같다’, ‘투표하지 않겠다’ 중에서 말씀해 주십시오.

1. 꼭 할 것이다
2. 아마 할 것 같다
3. 아마 하지 않을 것 같다
4. 투표하지 않겠다
5. 그 때 가봐서

문 2) 이번 대통령 선거에는 **대통합민주신당 ‘정동영’** 씨, **한나라당 ‘이명박’** 씨, **민주노동당 ‘권영길’** 씨, **민주당 ‘이인제’** 씨, **창조한국당 ‘문국현’** 씨, **민주연합 ‘정근모’** 씨, **경제공화당 ‘허경명’** 씨, **새시대정치연합 ‘전권’** 씨, **한국사회당 ‘글린’** 씨, **국민연대 ‘이수성’** 씨, **무소속 ‘이회창’** 씨가 출마했습니다. ○○님께서는 이중 누가 대통령이 되는 것이 조금이라도 더 좋다고 생각하십니까?

1. 정동영	5. 문국현	9. 전 권
2. 이명박	6. 문국현	10. 글 린
3. 권영길	7. 정근모	11. 이수성
4. 이인제	8. 허경명	12. 이회창

99. 없다/비슷하다/모르겠다/무응답

문2-1)문이 답하셨다면 여러 후보 중 누구에게 조금이라도 더 호감이 가십니까?
14-15

1. 정동영	5.	9. 전 권
2. 이명박	6. 문국현	10. 글 린
3. 권영길	7. 정근모	11. 이수성
4. 이인제	8. 허경명	12. 이회창

99. 없다/비슷하다/모르겠다/무응답

[문 3, 4는 문2, 2-1에서 응답된 시합을 □□에 넣어 질문할 것]

문 3 ○○님께서는 □□ 씨를 계속 지지하실 것 같습니까?
16
아니면 상황에 따라 다른 사람 지지도 바뀔 수도 있다고 생각하십니까?

1. 계속 지지할 것 같다
2. 다른 사람 지지도 바뀔 수도 있다

9. 모르겠다/무응답

문 4 ○○님께서는 지금 지지하시는 □□ 씨가 당선될 것이라고 생각하십니까?
17
아니면 당선되기 어려운 것이라고 생각하십니까?

1. 당선될 것이다
2. 당선되기 어려운 것이다

9. 모르겠다/무응답

문 5 ○○님께서는 지난 6월과 어제 저녁 실시된 대선 후보 TV 토론회를 보셨습니까?
18
(**보았다면**) 처음부터 끝까지 다 보셨습니까? 혹은 일부만 보셨습니까?

1. 처음부터 끝까지 다 보았다
2. 일부만 보았다
3. 못 보았다 → **문 7에 관**

문6-1) 그림, TV 토론회 가장 강한 후보는 누구라고 생각하십니까?
19

1. 정동영	4. 이인제
2. 이명박	5. 문국현
3. 권영길	6. 이회창

9. 없다/비슷하다/모르겠다/무응답

문6-2) ○○님께서는 이번 TV 토론회를 보시고 지지하는 후보가 바뀌거나 새로 생겼습니까? 아니면 변화가 없습니까?
20

1. 지지하는 후보가 바뀌거나 새로 생겼다
2. 지지 후보에 변화없다

9. 모르겠다/무응답

[모든 응답자에게 질문]

문 7) 현재 우리나라에는 **대통합민주신당, 한나라당, 민주노동당, 민주당, 국민중심당, 창조한국당, 참주인연합** 등의 정당에 있습니다. ○○님께서는 이중 어느 정당을 지지하십니까?
21
(9인 경우 : **그럼, 어느 정당에 조금이라도 더 호감이 가십니까?**)

1. 대통합민주신당
2. 한나라당
3. 민주노동당
4. 민주당
5. 국민중심당
6. 창조한국당
7. 참주인연합
8. 기타 정당
9. 없다/비슷하다/모르겠다/무응답

직업 ○○님의 직업은 무엇입니까?
22

1. 농/임/ 어업
2. 자 영 업
3. 불투 칼라 (판매/서비스직, 기능/숙련공, 일반직업직)
4. 화이트칼라 (사무/기술직, 경영/관리직, 전문/자유직)
5. 가정 주부
6. 학 생
7. 무 직
8. 기 타 (적용 것 : _____)

원래 ○○님 아버지의 고함은 어느 도입니까?
23

1. 서울	6. 대구/경북
2. 인천/경기	7. 부산/울산/경남
3. 강원	8. 제주
4. 대전/충청	9. 이북/기타
5. 광주/전라	

원래 ○○님께서는 학교를 어디까지 마치셨습니까?
24
(중퇴는 졸업에 포함되지 않음)

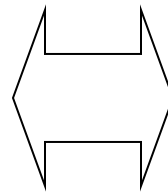
1. 초등학교이하	3. 고 졸
2. 중 졸	4. 대학 이상

조사에 응해 주신 것에 진심으로 감사드립니다!

5. 갤럽조사

투표하지 않는 유권자

선거결과를 예측할 때,
투표하지 않는 유권자는
모집단과 표본에서
제외하는 편이 좋다.



하지만 실제로는 투표에
참여하지 않을 거면서도
조사 시에는 투표할 거라고
답하는 경우가 많다.

투표할 것 같지 않은 유권자를 걸러내기 위한 갤럽의 설문지 문항

- ' 어디로 투표하러 가는가?'
- '지난번 선거에서 투표했는가?'

투표에 참여하지 않을 것 같은 사람들은 표본에서 제외하거나 그들의 응답에 낮은 가중치를 부여하면, 이는 선거결과를 예측하는데 있어 큰 도움이 된다. 22

5. 갤럽조사

아직 마음을 정하지 못한 유권자

아직 마음을 정하지 못한 유권자를 줄이기 위한 갤럽의 설문지 문항:

- '만약 오늘 당장 투표하게 된다면 누구를 뽑을 것인가?'
- 투표용지를 건네주어 실제 투표상황을 연출한다.

1984년 미국 대통령 선거 시 갤럽이 사용한 투표용지

<p><u>민주당</u></p> <p><input type="checkbox"/> 먼데일 페라로</p>	<p><u>공화당</u></p> <p><input type="checkbox"/> 레이건 부시</p>
---	--

5. 갤럽조사

자료의 점검과 가중치 부여

교육수준과 관련하여 아래와 같이 표본이 모집단에 비해 '고졸 이상' 학력자의 비율이 더 높은 경우

모집단	고졸미만(25%)	고졸이상(75%)
표본	고졸미만(20%)	고졸이상(80%)

- 모집단에서의 구성비보다 표본에서의 구성비가 더 큰 집단의 응답자에게는 이를 상쇄하기 위해 더 작은 가중치 부여. 위의 예에서 '고졸미만' 학력자에게 1의 가중치 준다면 '고졸 이상' 학력자에게는 $(75/80)/(25/20)=3/4$ 의 가중치 부여
- 가중치 조정 후 표본에서의 구성비 = $20:60 = 25:75 =$ 모집단 구성비

5. 갤럽조사

자료의 점검과 가중치 부여

가중치 부여

- 확률적 방법으로 뽑은 표본에 대해 사후적으로 '표본의 구성비=모집단의 구성비'가 되도록 집단별로 서로 다른 가중치 부여

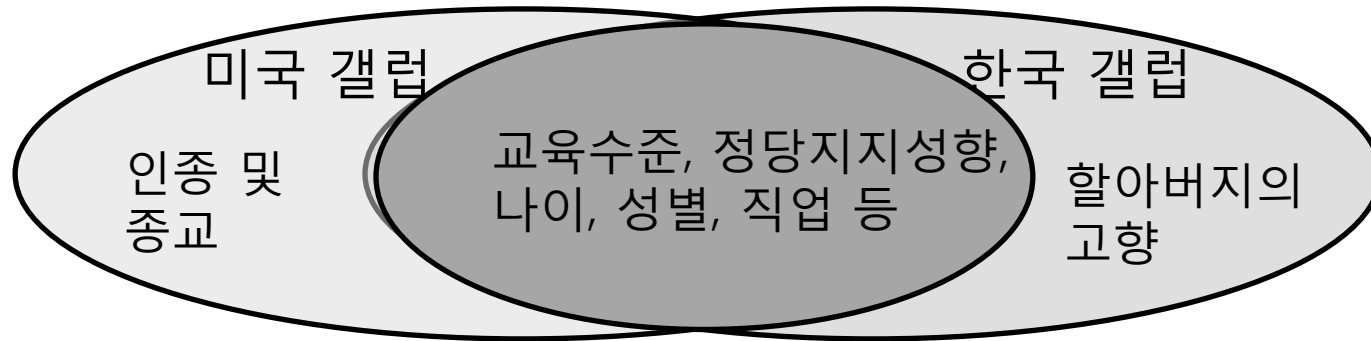
할당추출

- 표본이 모집단의 구성비 반영하도록 사전적으로 집단 별 표본크기를 할당

5. 갤럽조사

정치성향 파악하기 위한 질문 항목

정치성향에 영향을 미칠만한 기타 항목들도 설문에 들어간다.



몇 개의 같은 질문을 반복하여 응답자가 일관성 있게 답하는지 살펴본다.

- 일관성 없는 답이 나오면 그러한 응답자는 가중치를 낮게 주거나, 재조사를 실시한다.

6. 전화조사

전화조사

- 컴퓨터를 이용하여 전화번호를 단순 무작위 추출한다. 분석목적에 따라 전화번호를 지역별로 추출할 수도 있고 전국을 대상으로 추출할 수도 있다.
- 전화가 없는 사람이 전화가 있는 사람과 체계적으로 다르면 전화조사는 표본추출편의에 직면한다.
- 요즘은 집 전화와 휴대 전화를 어떻게 구분하여 처리할 것인가의 문제도 고려해야 한다.

6. 전화조사

전화조사

- 전화에 응답하지 않는 것도 편의를 가져올 수 있다. 일단 표본으로 추출된 전화번호에 대해서는 응답율을 높이기 위해 (i) 저녁, 주말에 조사한다든가, (ii) 전화를 받지 않는 경우 받을 때까지 반복적으로 전화한다.
- 전화조사는 면담조사에 비해 응답률이 높고, 비용이 적게 든다.
- 인터넷 조사는 비용이 저렴하고 조사기간이 짧다는 점에서 큰 장점이 있지만 응답자가 모집단을 대표하지 못하는 경우가 종종 발생할 수 있다.

7. 확률오차와 표준오차

확률오차

편의(bias)가 없더라도 확률오차 때문에 추정치는 참값과 다르다.
추정치는 참값으로부터 확률오차만큼 벗어난다.

예) 상자에 든 카드 중 1의 비율 추정

- 모집단: 0 과 1 이 적힌 카드가 많이 들어 있는 상자
- 표본: 상자로부터 무작위로 추출한 카드
- 추출편의, 무응답편의 등의 편의는 없지만 여전히 확률오차는 존재한다.
- (표본에서의 1의 비율)=(모집단에서의 1의 비율)+확률오차

7. 확률오차와 표준오차

표준오차

- 표준오차는 확률오차의 전형적인 크기
- 표본합의 표준오차, 표본평균의 표준오차, 표본비율의 표준오차는 표본합, 표본평균, 표본비율이 해당 기댓값으로부터 어느 정도나 떨어져 있는지 알려줌
- “수식어 없는 표준오차는 없다.” 표본합의 표준오차, 표본평균의 표준오차, 표본비율의 표준오차 등 구체적인 표준오차가 있을 뿐이다.