

1강 통계조사 개요

정보통계학과 이기재교수

학/습/목/차

1. 통계의 영광과 좌절

2. 통계조사

3. 표본조사의 특성

4. 표본조사의 기본용어

5. 표본추출법

영광 : 2012년 12월 제18대 대통령선거 예측조사

여론조사결과



박근혜(새누리당) 경합



득표율 50.10%



문재인(민주당) 경합



득표율 48.90%

실제개표결과



박근혜(새누리당) 당선



득표수 15,691,777 (51.6%)



문재인(민주당)



득표수 14,605,743 (48.0%)

영광 : 2012년 12월 제18대 대통령선거 예측조사

〈그림 1-1〉 제18대 대통령선거 예측값과 실제 개표 결과 비교



좌절 : 2012년 4월 제19대 국회의원선거 예측조사

■ 정당별 실제 의석 수

	새누리당	민주당	통합진보당
KBS	131-147	131-146	12-18
MBC	130-153	128-148	11-17
SBS	126-151	128-150	10-21
실제 의석 수	152	127	13

정보화시대의 통계

- 생활 속의 통계 예들

- ▶ 여론조사, 물가지수, 주가지수, TV시청률, 부동산 가격정보, 일기예보, 야구기록 등

통계

사회나 집단의 현상들을 수량화한 정보

통계의 이용

통계학적 개념에 의거하여 정보를 체계적으로 수집, 정리하고 그 결과에 따라 합리적인 의사결정을 내림

학/습/목/차

1. 통계의 영광과 좌절

2. 통계조사

3. 표본조사의 특성

4. 표본조사의 기본용어

5. 표본추출법

통계조사(survey)와 실험(experiment)

- 통계조사

- ▶ 복잡한 사회 또는 집단의 어떤 현상을 수량화 함으로써 객관적이고, 구체적인 특징을 파악하기 위한 일련의 과정

- 실험

- ▶ 연구자가 실험환경을 통제하고 조작을 가함으로써 특정 처리의 효과를 파악하는 과정

통계조사(survey)와 실험(experiment)

❖ 예제 : 흡연이 폐암에 미치는 영향



▶ 실험

처리그룹	매일 일정한 수준의 흡연 → 일정 시간 경과 후 폐의 상태 점검
대조그룹	금연

▶ 조사 : 폐암환자들의 흡연 정도 조사

■ 통계조사

▶ 전수조사

▶ 표본조사

우리나라의 대표적인 통계조사 예들

✚ 〈예 1.1〉 인구주택총조사

성, 연령 및 가구주와의 관계별 인구 - 시군구 - KOSIS

KOSIS 인구.가구 - 인구총조사 - 인구부문 - 총조사인구(2010) - 전수부문 - 성

통계설명자료 | 보도자료 | 스크랩 | 자료문의 | 도움말

조회기간: 부정기 ~ 2010 ~ 2010

수록기간 (부정기: 2010 ~ 2010)

단어검색 기본통계표

한글보기 소수점 계층 컬럼 구분 셀 단위 가중치 보기

상세보기 항목 행정구역별 1/2 성별 1/1 연령별 1/1

시계열선택 상세조건달기

1) 성, 연령 및 가구주와의 관계별 인구 - 시군구

숨김: 마우스 오른쪽 버튼을 클릭하면, 정렬기능을 사용할 수 있습니다.

단위: 명

행정구역별	성별	연령별	2010						
			일반가구원	가구주	가구주의 배우자	자녀	자녀의 배우자	가구주의 부모	배우
전국	계	합계	46,650,668	17,339,422	10,396,727	15,245,091	252,331	1,137,852	
		15세미만	7,771,133	113	0	7,168,180	0	0	
		15~19세	3,091,483	68,037	2,491	2,810,238	1,509	0	
		20~24세	2,452,384	384,415	50,452	1,776,904	13,129	0	
		25~29세	3,433,236	917,813	450,072	1,731,617	37,857	0	
		30~34세	3,666,529	1,442,504	1,113,750	859,304	59,226	0	
		35~39세	4,083,419	1,925,640	1,506,635	454,303	56,852	10	
		40~44세	4,113,159	2,159,231	1,571,931	234,849	37,877	2,590	
		45~49세	4,053,585	2,272,513	1,533,158	116,423	23,848	18,284	
		50~54세	3,777,701	2,170,673	1,397,213	60,229	13,444	54,329	
		55~59세	2,753,069	1,601,955	979,900	22,179	5,631	80,682	
		60~64세	2,171,350	1,285,517	708,416	8,106	2,051	113,155	
		65~69세	1,797,985	1,078,481	512,759	2,243	662	153,368	
		70~74세	1,544,641	938,820	355,443	516	244	200,458	
		75~79세	1,054,397	641,554	162,617	0	0	205,242	
		80~84세	561,758	314,741	42,982	0	0	166,685	
		85세 이상	324,839	137,415	8,908	0	1	143,049	

출처: 통계청 조사관리국 인구총조사과 +더보기

문의처: 통계청 042-481-373

※ 출처: 통계청 사회통계국 인구조사과

우리나라의 대표적인 통계조사 예들

❖ 〈예 1.2〉 대통령선거여론조사

	후보자	개표결과	여론조사	출처
13대 (1987)	노태우 김영삼 김대중	36.6% 28.0% 27.1%	34.5% 28.9% 27.9%	한국갤럽
14대 (1992)	김영삼 김대중 정주영	42.0% 33.8% 16.3%	39.6% 31.1% 15.7%	한국갤럽
15대 (1997)	김대중 이회창 이인제	40.3% 38.7% 19.2%	39.9% 38.9% 19.7%	MBC
16대 (2002)	노무현 이회창 권영길	48.9% 46.6% 3.9%	49.1% 46.8% 3.5%	KBS
제17대 (2007년)	이명박 정동영 이회창	48.7% 26.1% 15.07%	50.3% 26.0% 13.5%	KBS-MBC 공동
제18대 (2012년)	박근혜 문재인	51.6% 48.0%	50.1% 48.9%	방송3사 공동

우리나라의 대표적인 통계조사 예들

❖ <예 1.3> TV 시청률조사의 예 (2013년 1월 첫째 주)

1. AGB닐슨미디어 리서치사 발표 시청률

전국 시청률				수도권 시청률			
순위	채널	프로그램	시청률	순위	채널	프로그램	시청률
1	KBS2	주말연속극(내딸서영이)	39.5	1	KBS2	주말연속극(내딸서영이)	39.3
2	KBS1	일일연속극(힘내요미스터김)	28.2	2	KBS1	일일연속극(힘내요미스터김)	26.3
3	KBS2	개그콘서트	21.8	3	KBS2	개그콘서트	23.4
4	KBS1	KBS9시뉴스	20.8	4	KBS1	KBS9시뉴스	20.4
5	MBC	창사특별기획(마의)	17.5	5	MBC	무한도전	19.9
6	MBC	무한도전	17.3	6	MBC	창사특별기획(마의)	18.7
7	SBS	정글의법칙IN아마존	17	7	SBS	정글의법칙IN아마존	18.2
8	KBS1	한국인의밥상	16.5	8	SBS	일요일이 좋다(서바이벌오디션K팝스타2,런닝맨)	17.9
9	SBS	일요일이 좋다(서바이벌오디션K팝스타2,런닝맨)	16	9	MBC	주말특별기획드라마(백년의유산)	16.8

※ 출처: AGB닐슨 미디어리서치

우리나라의 대표적인 통계조사 예들

❖ 〈예 1.3〉 TV 시청률조사의 예 (2013년 1월 첫째 주)

2. TNmS사 발표 시청률

※ 출처: TNS미디어

전국 Top 20

1. 주말연속극<내딸서영이>	KBS2		40.6
2. 일일연속극<힘내요미스터김>	KBS1		30.4
3. TV소설<사랑아사랑아>	KBS2		18.0
4. 개그콘서트	KBS2		24.5
5. KBS뉴스광장 2부	KBS1		11.3
6. KBS뉴스광장 1부	KBS1		6.9
7. KBS뉴스9	KBS1		21.9
8. 남북의창	KBS1		11.8
9. 아침드라마<사랑했나봐>	MBC		14.2
10. 전국노래자랑	KBS1		15.4

서울수도권 Top 20

1. 주말연속극<내딸서영이>	KBS2		41.0
2. 일일연속극<힘내요미스터김>	KBS1		26.1
3. 개그콘서트	KBS2		25.7
4. 무한도전	MBC		21.8
5. 김병만의아마존정글의법칙	SBS		21.3
6. 아침드라마<사랑했나봐>	MBC		13.6
7. TV소설<사랑아사랑아>	KBS2		15.8
8. 도전1000곡	SBS		13.3
9. KBS뉴스광장 2부	KBS1		9.2
10. 일요일이 좋다	SBS		19.2

통계조사의 역사

■ 고대~현대에 이르기까지의 역사

고대

- 징병, 징세 목적의 호구, 호적조사
(예) BC 15세기, 구약성경 민수기의 인구조사

근대적 조사의
기원

- 19세기

현대 표본이론의
기초

- 1930년대

최초의
선거여론조사

- 1936년 미국 대통령선거

전수조사(complete enumeration survey)

일명 센서스(Census)

조사대상으로 하는 집단의 모든 개개의 단위를 조사하는 방법

- 예
 - ▶ 인구주택 총 조사, 농업 총 조사, 사업체 총 조사 등
- 사용목적
 - ▶ 어떤 정책 결정의 중요한 기초자료로 사용
 - ▶ 다른 표본조사의 기초자료로 활용
(전수조사 자료는 다른 표본조사의 추출틀로 사용됨)
- 전수조사의 어려운 점
 - ▶ 엄청난 조사 비용 및 시간 소요
 - ▶ 숙련된 많은 수의 조사원 필요

표본조사(sample survey)

전체 모집단 중에서 과학적인 추출방법에 따라 추출된
일부분을 대상으로 조사하여 얻어진 정보를 토대로
전체 모집단에 대한 특성을 추정하는 방법



학/습/목/차

1. 통계의 영광과 좌절

2. 통계조사

3. 표본조사의 특성

4. 표본조사의 기본용어

5. 표본추출법

표본조사의 장점

1

경제성

2

신속성

3

심도 있는 조사가 가능하다는 측면

4

조사의 정확성

5

숙명적 필요성

표본조사의 단점

1

모집단을 제대로 대표하지 못하는 표본을 사용할 경우 잘못된 통계를 만들게 됨

✦ 사례 : 1936년 Literary Digest 사의 미국 대통령 선거여론조사

- 전화번호부와 자동차 등록대장을 사용하여 1000만 여명
우편조사 : 230여명만 응답
- 랜돈 : 루즈벨트 = 55 : 45로 랜돈 승리 예측
- 실제 결과 : 루즈벨트 후보의 압도적 승리
→ 서민층이 과소 추출된 결과

2

희소한 부차모집단(소지역별 자료)의 특성까지 알기는 힘들

전수조사의 표본조사에 대한 역할

1

모집단 상황에 대한 기초자료와 추출틀로
사용됨

2

전수조사 자료는 모수 추정과정에서 중요한
자료로 이용됨

학/습/목/차

1. 통계의 영광과 좌절

2. 통계조사

3. 표본조사의 특성

4. 표본조사의 기본용어

5. 표본추출법

기본단위(elementary unit : 조사단위, 단위)

필요한 정보를 얻기 위해 조사를 할 때 조사 대상이 되는 최소의 요소

- 명확하게 정의되어야 하고, 구체적으로 확인할 수 있어야 함
 - ▶ 여론조사 : 개인
 - ▶ 가계조사 : 가구
 - ▶ 작물 생산량 조사 : 일정 면적의 경지

모집단(population)

조사목적에 의하여 규정되는 모든 기본단위들의 모임

- 유한모집단, 무한모집단
- 명확한 정의가 필요
 - ▶ 예) 농가 – 300평 이상의 경지에 농사를 짓는 가구
- 목표모집단, 조사가능 모집단
 - ▶ 목표모집단 : 조사목적에 의해 개념상 규정된 모집단
 - ▶ 조사가능 모집단 : 표본을 추출하기 위해 규정된 모집단

추출단위(sampling unit)

표본 추출을 위해 설정한 기본단위들의 모임

- 모든 기본단위가 누락, 중복되지 않고 추출단위에 포함되게 해야 함

예) 여론조사

- ➔ 개인을 추출하는 경우 : 추출단위 = 기본단위 = 개인
- ➔ 가구를 추출하는 경우 : 추출단위 = 가구, 기본단위 = 유권자

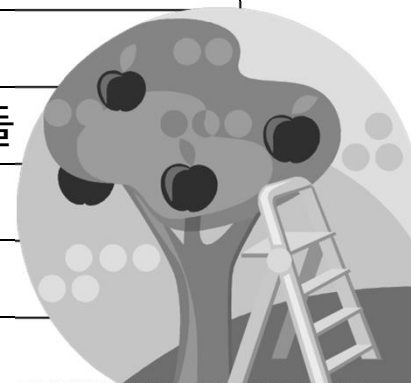
추출틀(sampling frame)

모집단에 속하는 모든 추출단위의 목록

- ▶ 개인, 가구, 사업체 등의 명부, 문서철, 지도 등
- ▶ 모집단의 모든 추출단위를 누락 없이 그리고 중복 없이 포함
예) 1936년 미국의 여론조사 실패 : 불완전한 추출틀이 주요 원인
- ▶ 추출틀의 작성 가능성을 염두에 두고 추출틀 결정

❖ <예 1.4> 경북지역 사과나무 한 그루 당 평균수확량 조사

기본단위	사과나무 그루
목표모집단	경북에 있는 모든 사과나무들
조사가능모집단	경북의 등록된 과수원에 있는 사과나무들
추출단위	과수원
추출틀	경북지역 과수원 목록



학/습/목/차

1. 통계의 영광과 좌절

2. 통계조사

3. 표본조사의 특성

4. 표본조사의 기본용어

5. 표본추출법

표본(sample)

모집단의 특성치를 알아보기 위하여
모집단의 기본단위 중에서 추출된 부분집합

- 표본설계(design of sample survey) 시의 고려사항
 - ▶ 모집단 특성에 대한 대표성 지니면서도 경제적 추출법
 - ▶ 표본오차를 목표수준 이내로 유지하면서 비용을 최소화하는 표본 크기 결정
 - ▶ 효율적인 추정 방식

확률추출법(probability sampling)

추출단위에 대하여 사전에 정해진 추출확률에 따라 표본을 추출하는 방법

- 단순임의추출법, 층화추출법, 계통추출법, 집락추출법
- 표본이 모집단의 특성이나 구조를 잘 반영하여 조사결과를 모집단으로 일반화 가능
- 추정량의 통계적 오차를 확률적으로 계산 가능함

비확률추출법(non-probability sampling)

- ▶ 추출단위들이 표본으로 추출될 확률을 객관적으로 나타낼 수 없음
- ▶ 대부분 조사자의 주관적 판단에 의해 표본을 추출
- ▶ 전형법(typical sampling method),
할당법(quota sampling method)
- ▶ 한계 : 추정값의 정확성 평가 불가, 과학적인 조사방법이 아님

❖ <예 1.5> 예이츠(Yates)의 실험

- ▶ 실제 1200개 돌의 무게 = 54.15 그램
- ▶ 비확률추출법 적용 : 추정 평균 = 66.34 그램
 - 12명에게 모집단을 잘 대표한다고 생각하는 크기 20개의 표본을 전형법으로 뽑는 시행을 3번씩 반복하도록 함
 - 각각의 조사자가 얻은 표본평균의 평균을 구함

비확률추출법(non-probability sampling)

❖ <예 1.6> 미국 대통령선거 여론조사의 예

연도	여론조사 결과	실제선거 결과
1936	루우즈벨트(55.7%)	루우즈벨트(62.5%)
1940	루우즈벨트(52.0%)	루우즈벨트(55.0%)
1944	루우즈벨트(51.5%)	루우즈벨트(52.3%)
1948	듀이(49.5%)	트루먼(49.5%)
1952	아이젠하워(51.0%)	아이젠하워(55.4%)
1956	아이젠하워(59.5%)	아이젠하워(57.8%)
1960	케네디(51.0%)	케네디(50.1%)
1964	존슨(64.0%)	존슨(61.3%)
1968	닉슨(43.0%)	닉슨(43.5%)
1972	닉슨(62.0%)	닉슨(61.8%)
1976	카터(48.0%)	카터(50.0%)
1980	레이건(47.0%)	레이건(50.8%)
1984	레이건(59.0%)	레이건(59.2%)
1988	부시(56.0%)	부시(53.4%)
1992	클린턴(49.0%)	클린턴(43.0%)
1996	클린턴(48.0%)	클린턴(49.2%)
2000	W.부시(47.0%)	W.부시(47.9%)
2004	W.부시(49.0%)	W.부시(50.7%)
2008	오바마(53.0%)	오바마(52.9%)
2012	오바마(49.0%)	오바마(51.0%)



출처: 한국갤럽조사연구소