1.1 통계와 통계학

제주권역
대학 e-러닝 지원센터

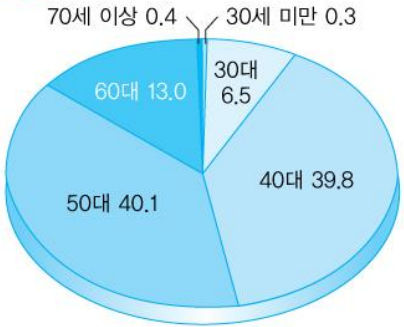
통계(statistic)

통계란 특정집단을 대상으로 한 조사나 실험에 의하여
구한 결과에 대한 요약된 형태의 표현이다.

경제통계	물가, 실업률, GDP
인구통계	인구의 출생, 결혼, 사망
농업통계	농산물의 생산과 소비
사회조사 분석통계	선호도 조사, 의식조사
실험결과 분석통계	임상실험결과, 오염도 측정

2006년 제4회 지방선거 당선자 분석표

연령별



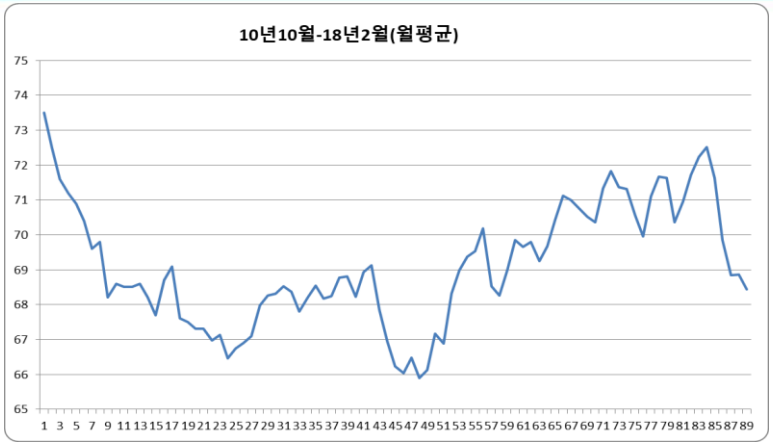
학력별

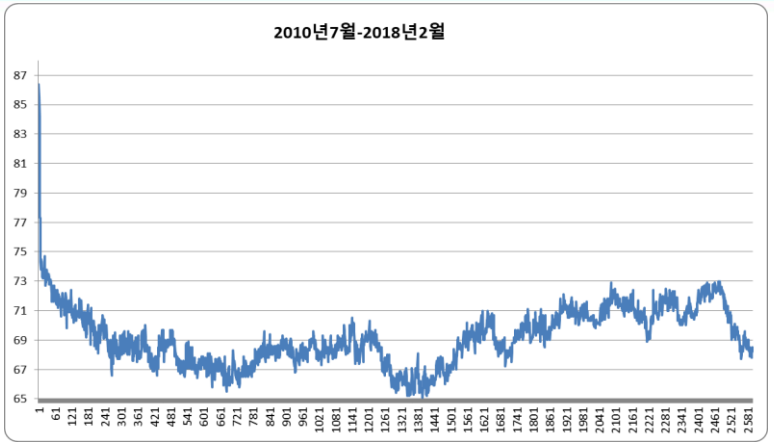
연령	단위(%)
무학(독학 포함)	0.3
초졸	3.4
중졸	4.9
고졸	17.4
대졸	54.2
대학원 졸	14.5
기타	5.3

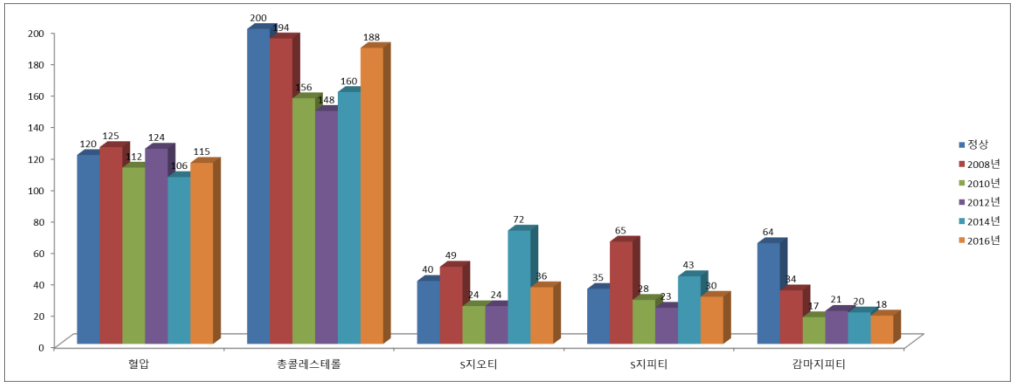
투표율(중앙선관위 최종 집계 · %)

구분	17대 총선	2006 지방선거
서울	62.2	49.2
부산	61.9	48.1
대구	59.3	48.3
인천	57.4	44.2
광주	60.2	46.3
대전	58.9	49.5
울산	62.0	52.8
경기	59.7	46.2
강원	59.7	58.4
충북	58.2	54.7
충남	56.0	55.7
전북	61.2	57.9
전남	63.4	64.2
경북	61.5	61.2
경남	62.3	57.8
제주	61.1	67.3
합계	60.6	51.3

자료 : 문화일보, 2006. 6. 2.


10년10월-18년2월(월평균)


2010년7월-2018년2월


혈압
총콜레스테롤
sGOT
sGPT
감마GPT
정상
2008년
2010년
2012년
2014년
2016년


Median Starting Salary
10th
25th
Median
75th
90th
Chemical Engineering
Computer Engineering
Electrical Engineering
Aerospace Engineering
Economics
Physics
Computer Science
Industrial Engineering
Mechanical Engineering
Math
Physician Assistant
Civil Engineering
Construction
Finance
MIS
Philosophy
International Relations
Chemistry
Marketing
Geology
Political Science
Accounting
Architecture
Information Technology (IT)
Business Management
Agriculture
History
Communications
Film
Nursing
Journalism
Geography
Art History
Biology
English
Forestry
Anthropology
Health Care Administration
Psychology
Graphic Design
Sociology
Hospitality & Tourism
Drama
Criminal Justice
Nutrition
Music
Interior Design
Spanish
Religion
Education
0
50000
100000
150000
200000
Salary (USD)

Rank	Undergraduate Major	Starting Median	Mid-Career Median	% Change
#1	Chemical Engineering	$63,200	$107,000	69.3%
#2	Computer Engineering	$61,400	$105,000	71.0%
#3	Electrical Engineering	$60,900	$103,000	69.1%
#4	Aerospace Engineering	$57,700	$101,000	75.0%
#5	Economics	$50,100	$98,600	96.8%
#6	Physics	$50,300	$97,300	93.4%
#7	Computer Science	$55,900	$95,500	70.8%
#8	Industrial Engineering	$57,700	$94,700	64.1%
#9	Mechanical Engineering	$57,900	$93,600	61.7%
#10	Math	$45,400	$92,400	103.5%

#11	Physician Assistant	$74,300	$91,700	23.4%
#12	Civil Engineering	$53,900	$90,500	67.9%
#13	Construction	$53,700	$88,900	65.5%
#14	Finance	$47,900	$88,300	84.3%
#15	Management Information Systems (MIS)	$49,200	$82,300	67.3%
#16	Philosophy	$39,900	$81,200	103.5%
#17	International Relations	$40,900	$80,900	97.8%
#18	Chemistry	$42,600	$79,900	87.6%
#19	Marketing	$40,800	$79,600	95.1%
#20	Geology	$43,500	$79,500	82.8%

기술통계(descriptive statistic)란 자료를 요약한 기초적인 통계를 말한다.

숫자표현	**평균, 표준편차, 중위수, 최빈값**
그림표현	**막대그림표, 원그림표, 꺾은선 그림표**

*(참고)정보, 통계 및 지표

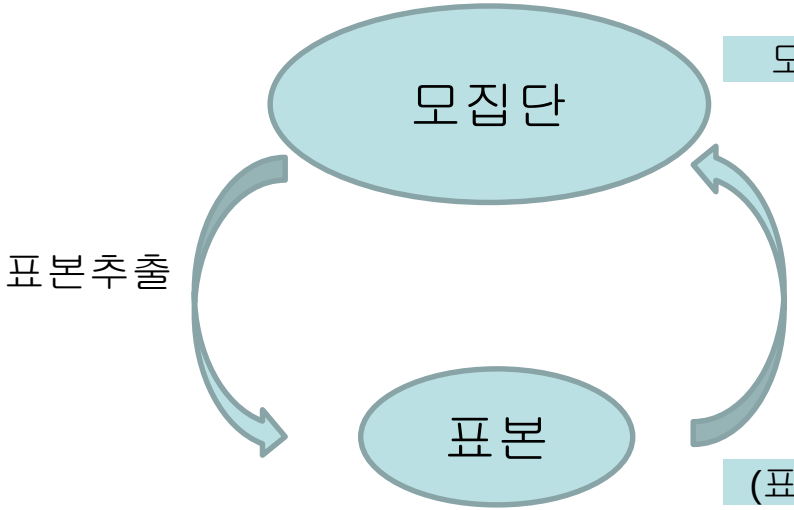
정보(information)	집단이 아닌 개별적인 사실을 파악한 것
통계(statistic)	이 정보 중 목적에 따라 필요한 것만을 집계하여 집단의 상태를 집약적으로 나타낸 숫자
지표(indicator)	자료의 이동방향 또는 추세를 제시하기 위하여 통계의 비율이나 증감률의 형태로 체계적으로 정리한 것

통계학(statistics)

통계학이란 불확실한 현상을 대상으로 자료를 수집하고 정리하며,
이 자료가 수집된 대상에 대하여 적절한 모형을 설정하고
추정(estimation), 검정(testing) 및 예측(forecasting)을 하는 학문이다.

수리통계학	통계학의 기본적인 이론을 다룸(확률론, 추론 등)
응용통계학	수리통계학에서 정립된 이론을 바탕으로 실제 자료 분석에 응용하는 방법을 연구(표본론, 회귀분석, 시계열분석 등)
기술통계학	모집단에 대한 추론이나 어떤 결론을 도출함이 없이 수집된 정보를 간단명료하고 유용하게 정리하는 문제를 다룸
추리통계학	표본을 기초로 하여 모집단의 특성을 추정하고 일반화하며 또한 예측하는 문제를 다룸

1-1-2. 통계학의 기본 용어



(1)모집단(population):연구자의 관심대상이 되는 모든 개체의 집합
(2)표본(sample):모집단에서 조사대상으로 채택된 일부 집단(부분집합)
(3)추론(inference):표본의 정보를 바탕으로 모집단에 관한 의사결정,추정,예측을 하는 것
(4)신뢰도(reliability):추론에 대한 신뢰성의 측도
(5)변수(variable):관측 때마다 서로 다른 값을 취하는 것으로 변수를 표현하는 방법으로 일반적으로 X,Y,Z와 같은 영문자로 표현
(6)모수(parameter):모집단의 수량적인 특성(모평균,모분산,모집단 표준편차..)
(7)(표본)통계량(statistic):표본의 수량적인 특성(표본평균,표본분산,표본표준편차..) 표본에 담긴 정보를 요약하는 공식
(8)추정량(estimator):모수를 추정하는 공식을 나타내는 통계량
(9)추정치(estimate):추정량(공식)에 실제의 관찰값을 넣어 계산한 통계량의 값
(10)표본오차(sampling error):모집단을 조사하지 않고 표본조사의 결과만을 가지고 모집단의 특성을 추정할 때 발생하는 오차
(11)비표본오차(non-sampling error):관찰오류, 누락, 오기 등 표본추출과정에서 오류로 인하여 발생하는 오차

1-2 자료수집

1-2-1. 통계분석과 자료수집

통계분석(Statistical analysis)

-특정 집단을 대상으로 자료를 수집해 대상집단에 대한 정보를 구하고,
적절한 통계분석 방법을 이용해 의사결정을 하는 과정
-이러한 의사결정을 통계적 추론이라고 하는데 통계적 추론에는 추정과 가설검정이 있음

추정(Estimation)

"대상집단의 특성값(모수)이 무엇일까?"를 추측

가설검정(Hypothesis test)

대상집단에 대해 특정한 가설을 설정한 후 그 가설의 채택여부를 결정

통계분석의 5단계

통계분석이란 수리통계학과 응용통계학에서 정리된 이론을 이용해
실제 자료로부터 유용한 정보를 구하는 과정이다.

자료의 수집

수집된 자료의 요약·정리

모수의 추정

검정

모형 분석

자료수집방법

- 조사 : 전수조사 / 표본조사
- 실험 : 대상집단의 일부에 처리를 가하고 결과를 관측해 자료를 수집하는 방법
  -treatment group(처치집단/치료집단) vs. control group(대조군/통제집단/비교집단)

사례1

여론조사기관에서 1997년 12월에 국회에서 의결한 금융실명제 보완 입법에 대한 국민들의 지지율을 조사했다.

사례2

한 제약회사에서 새로 개발된 AIDS 치료제의 효과를 분석하는 실험을 실시했다.

- AIDS 감염환자 20명 선정→치료제 투약→치료효과 측정
- AIDS 감염환자 20명 선정→ treatment group(10명)→치료제 투약→치료효과 측정
  ↘ control group(10명)→치료제 투입 안함

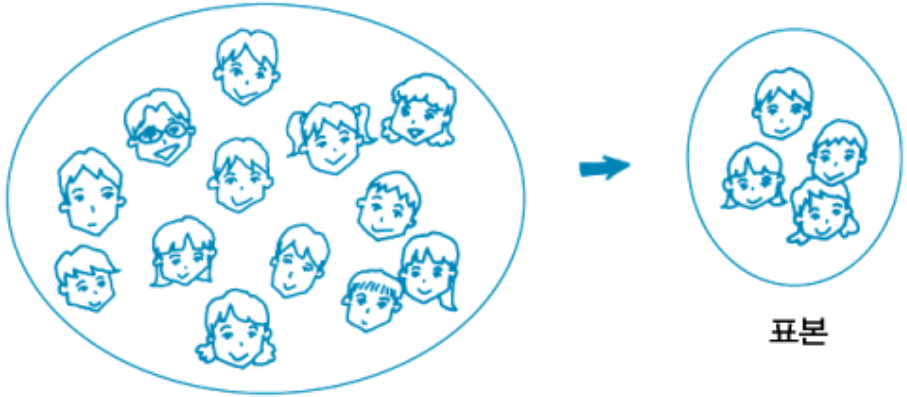
1-2-2. 표본조사

총조사(Census)

대상집단 모두를 조사하는 방법. '인구 및 주택 총조사'가 그 대표적인 예

표본조사

대상집단의 일부를 관측해 그 대상집단 전체에 대한 정보를 구하는 과정.
총조사에 비해 시간과 비용을 절약할 수 있음



표본조사에서 이용되는 용어

모집단(population)	조사하고자 하는 대상집단전체
원소(element)	모집단을 구성하는 구성원소
표본(sample)	조사하기 위하여 뽑힌 모집단의 일부
모수(parameter)	표본관측에 의하여 구하고자 하는 모집단의 특성값

사례1

- 모집단 : 전국의 유권자
- 원소 : 유권자 개개인
- 표본 : 조사대상자로 뽑힌 유권자 1000명
- 모수 : 금융실명제 보완입법에 대한 지지율

표본조사의 유의사항

1. 표본이 합리적으로 추출되었는가(연령별, 지역별,성별, 학력별 등)
2. 질문의 형식이 특정사항을 선호하도록 유도되어서는 안 된다
3. 조사방법(면접, 우편, 전화 등)에 따라서도 조사결과에 차이가 있을 수 있다
   - •무응답자에 대한 처리
   - •전화조사 시간대
4. 표본조사 시점에 대한 고려를 해야 한다
5. 나타난 결과에 대한 절대적인 신뢰를 해서는 안 된다

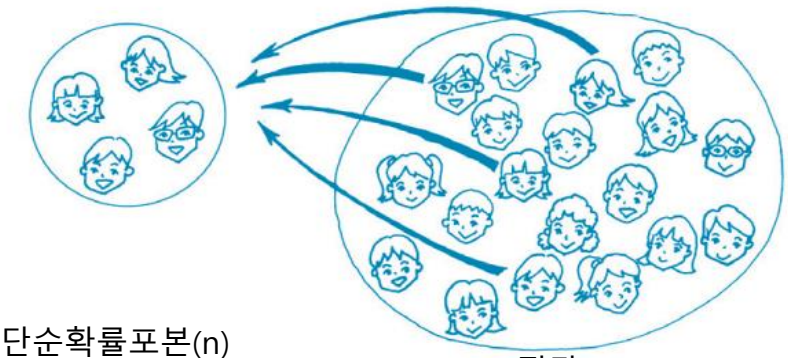
표본조사결과 발표 시에 포함되어야 할 사항

- 모집단의 정의
- 표본의 크기
- 조사 방법 (면접조사, 우편조사, 전화조사 등)
- 조사기간
- 표본추출방법

표본추출방법- (1)단순랜덤추출법(simple random sampling)

- 확률추출법 : 단순랜덤추출, 계통추출, 층화추출, 집락추출
- 비확률추출법 : 할당추출법(quota sampling)

그림 2-1 단순랜덤추출법



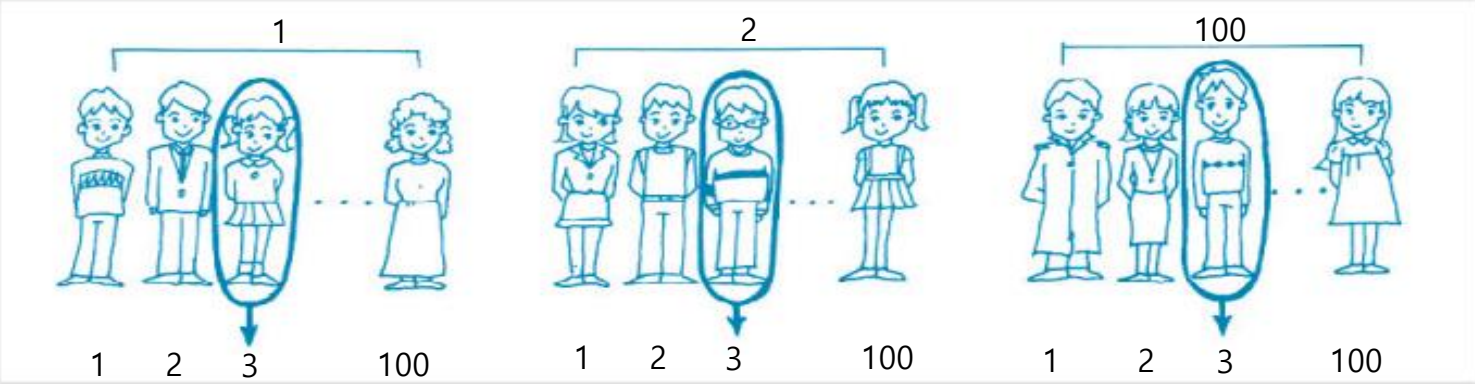
- N개의 모집단에서 n개의 표본이 추출될 가능성을 동일하게 해 주는 표본추출 방식
- 난수표 이용
  - 0부터 9까지의 정수가 동일 비율로 포함된 표
  - 예를 들어 50명으로 구성된 모집단에서 10명을 추출한다면 8열 1행 마지막 두자리 숫자부터 시작

[표 8] 난수표

Line/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
7	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
10	85475	36857	53342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
11	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
12	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
13	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
14	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
16	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
17	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
18	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
19	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
21	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
22	54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
23	32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
24	29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953
26	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
27	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280
28	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253

표본추출방법- (2)계통추출법(systematic sampling)

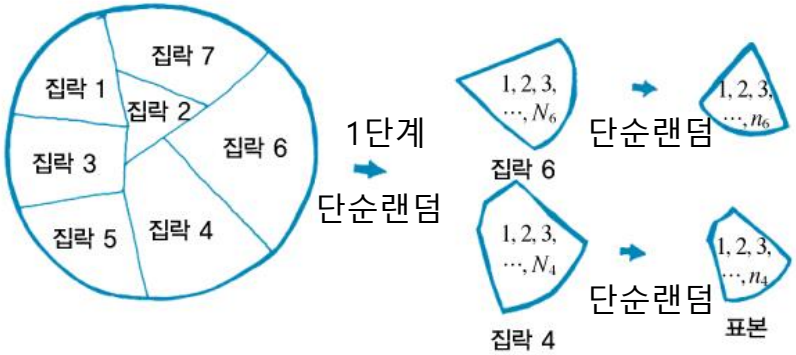
그림 2-2 계통추출법



- k개의 구간으로 구분 : k≤N/n(예, N=10,000, n=100)
- 첫 구간에서 임의의 번호(예:3)을 추출
- 그 점에서 매 k번째 떨어진 간격에 위치한 단위를 추출
- 품질관리(QC) sampling에 많이 이용

표본추출방법- (3)집락추출법(cluster random sampling)

그림 2-3 집락추출법

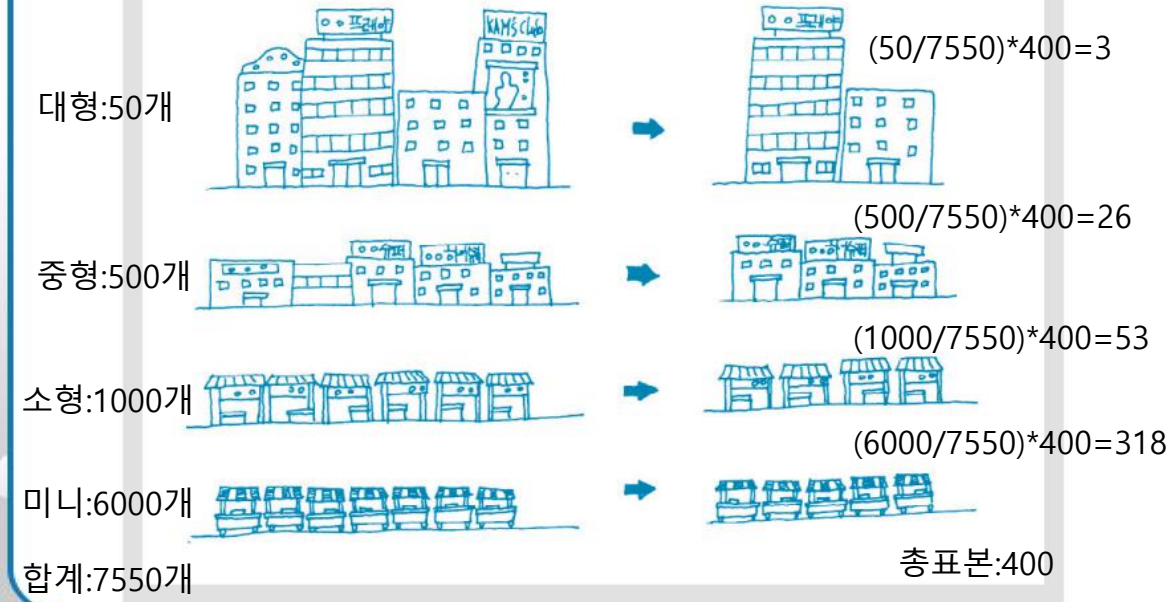


- 모집단이 몇 개의 집락으로 결합
- 1단계 집락추출
  - 먼저, 집락을 단순램덤으로 추출
  - 다음으로, 각 집락에서 표본을 단순랜덤으로 추출
- 2단계 집락추출
  - 먼저, 집락을 단순램덤으로 추출
  - 다음으로, 1차로 추출된 집락에서 하위 집락을 단순램덤으로 추출
  - 마지막으로, 각 집락에서 표본을 단순랜덤으로 추출

- (예)서울 25개 자치구에서 300가구 추출
  - 1단계 집락추출 : 25개 중 5개 구 추출 → 각 구에서 60가구 추출
  - 2단계 집락추출 : 25개 구 중 5개 수 추출 → 각 구에서 4개 동 추출 → 각 동에서 15가구 추출

표본추출방법- (4)층화추출법(stratified random sampling)

그림 2-4 층화추출법



- 표본이 각 계층을 골고루 대표할 수 있도록 추출
- 모집단을 상호배타적인 층(strata)으로 구분
- 각 층에서 단순램덤추출