

2022학년도 2학기

경영 통계학

담당교수: 백수정



학습 목표

1. 데이터와 표본을 기술하는 기본용어들을 이해함
2. 숫자형 데이터와 범주형 데이터의 차이점을 이해함
3. 시계열과 횡단면 데이터의 차이를 알아봄
4. 데이터의 측정 수준 및 코딩하는 방법을 이해함
5. 표본과 모집단에 대해 올바른 용어를 이해함
6. 일반적인 표본추출 방법과 그 수행 방법을 이해함
7. 일상적으로 발표되는 데이터의 출처를 알아봄
8. 서베이 설계, 서베이 유형, 응답 척도에 대해 기본 요소들을 이해함

변수와 데이터

데이터 용어

- 관측치(observation): 우리가 분석하려는 개체(사람, 기업, 지역 등)의 집합에서 어떤 하나의 개체를 말함.
- 변수(variable): 객체 또는 개체의 어떤 특징을 말함.예: 근로자의 임금, 구매 상품 청구금액.
- 데이터세트(data set): 모든 변수의 모든 값으로 이루어져 있으며, 우리가 선택한 모든 관측치를 말함.

변수와 데이터

하나의 데이터세트에 여러 변수가 있을 수 있다.

데이터의 형태와 변수의 개수에 따라 분석의 주제와 방법이 달라질 수 있다.

본 교과서에서는

- 단일변량(univariate; 변수 한 개) 데이터세트
- 이변량(bivariate; 변수 두 개) 데이터세트
- 다변량(multivariate; 변수 세 개 이상) 데이터세트를 다룬다.

변수와 데이터

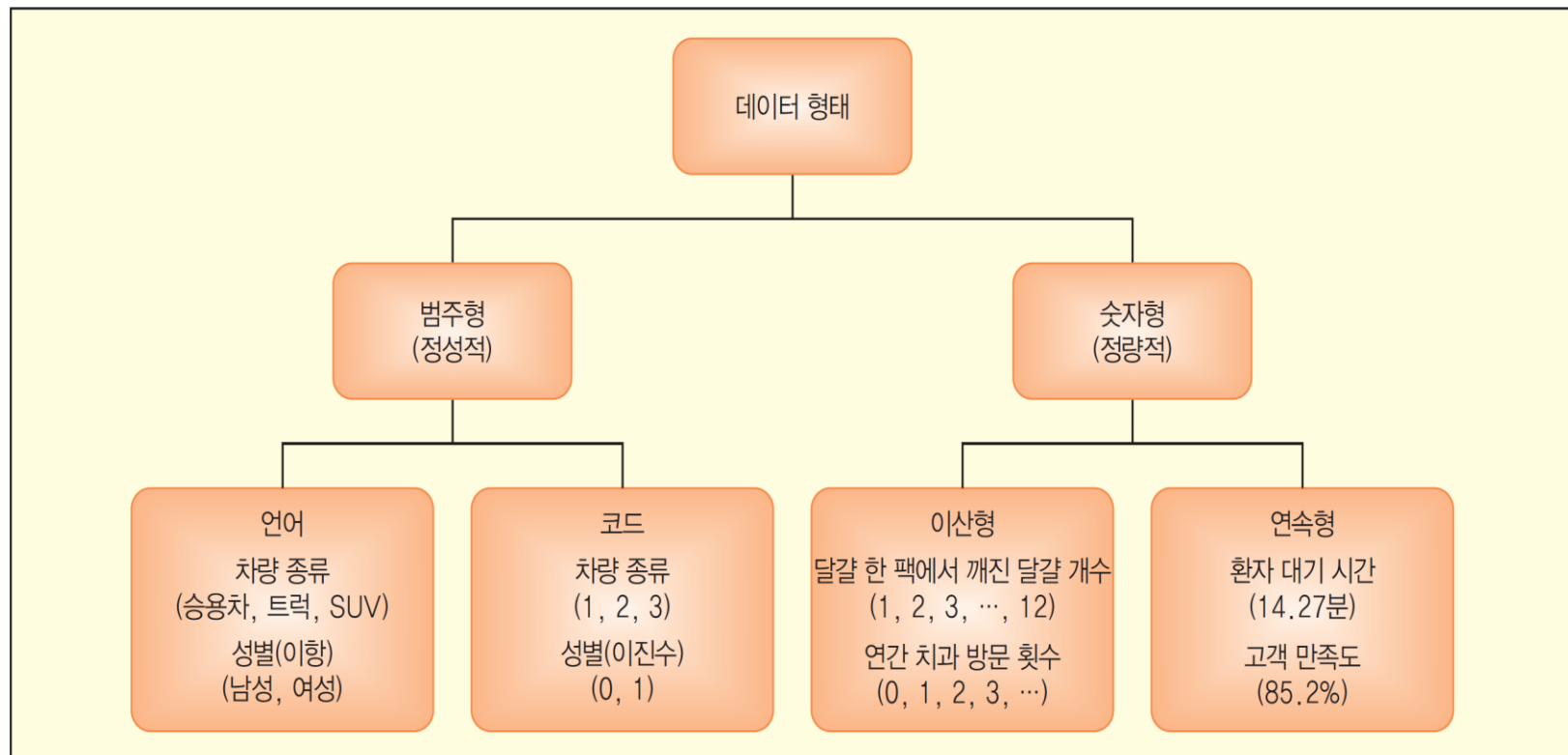
표 2.2: 변수 개수와 주요 작업

데이터세트	변수	예	주요 작업
단일변량	한 개	소득	히스토그램, 기본 통계량
이변량	두 개	소득, 나이	산포도, 상관성
다변량	세 개 이상	소득, 나이, 성별	회귀분석

변수와 데이터

그림 2.1

데이터 유형과 예



시계열 데이터(time series data)

표본의 모든 관측치들이 동일 간격(년, 월, 일)으로 서로 다른 시점에서 측정된다.

관측치 간의 시간을 주기(periodicity)라고 하며, 여기에는 연간, 분기, 월, 주, 일, 시간 등

예

- **거시경제 시계열 데이터:** 국민소득(GDP, 소비, 투자), 경제지표(소비자물가지수, 실업률, S&P 500지수), 통화 데이터(M1, 국채 수익률, 소비자 신용, 연방 부채) 등
- **미시경제 시계열 데이터:** 기업의 매출액, 시장점유율, 부채비율, 재고회전율, 제품 품질등급 등

횡단면 데이터(cross-sectional data)

각 관측치들의 서로 다른 개별 단위(예: 사람)를 같은 시점에서 보여줌
(예; 월간 비자카드 사용액)

횡단면 데이터의 경우 우리의 관심사항은:

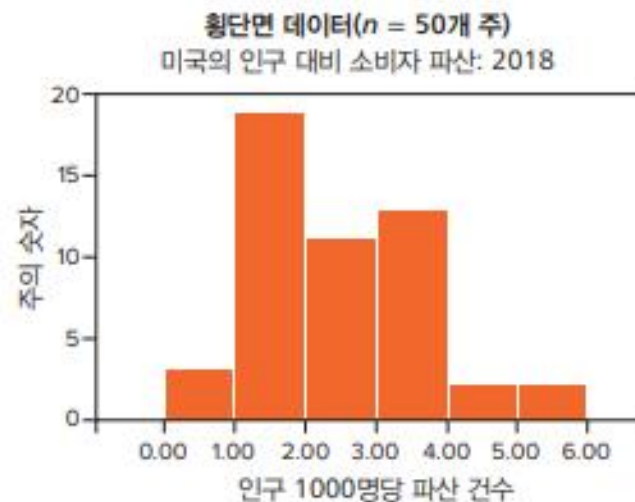
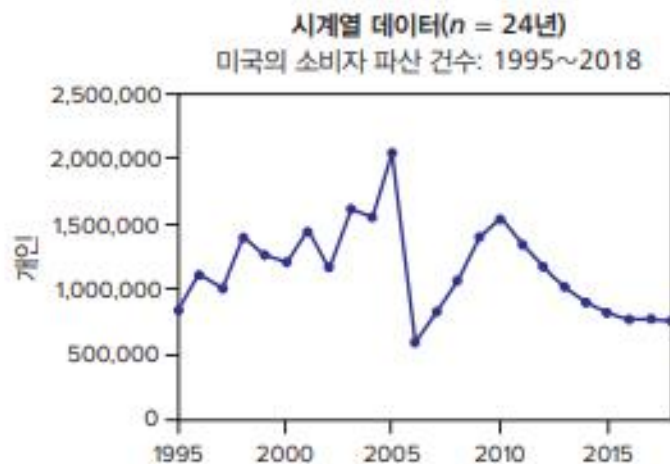
- 관측치들 간의 차이점(예: Subway 20개 프랜차이즈의 미수금).
- 관측치 사이의 상호관계(예: 그림 2.2에서 보듯이 Subway 20개 프랜차이즈의 미수금과 매출액의 상관성 여부).

횡단면과 시계열을 결합하여 횡단면 및 시계열 합동(pooled) 데이터를 만들 수 있음

시계열 및 횡단면

그림 2.2

시계열과 횡단면 데이터 사례

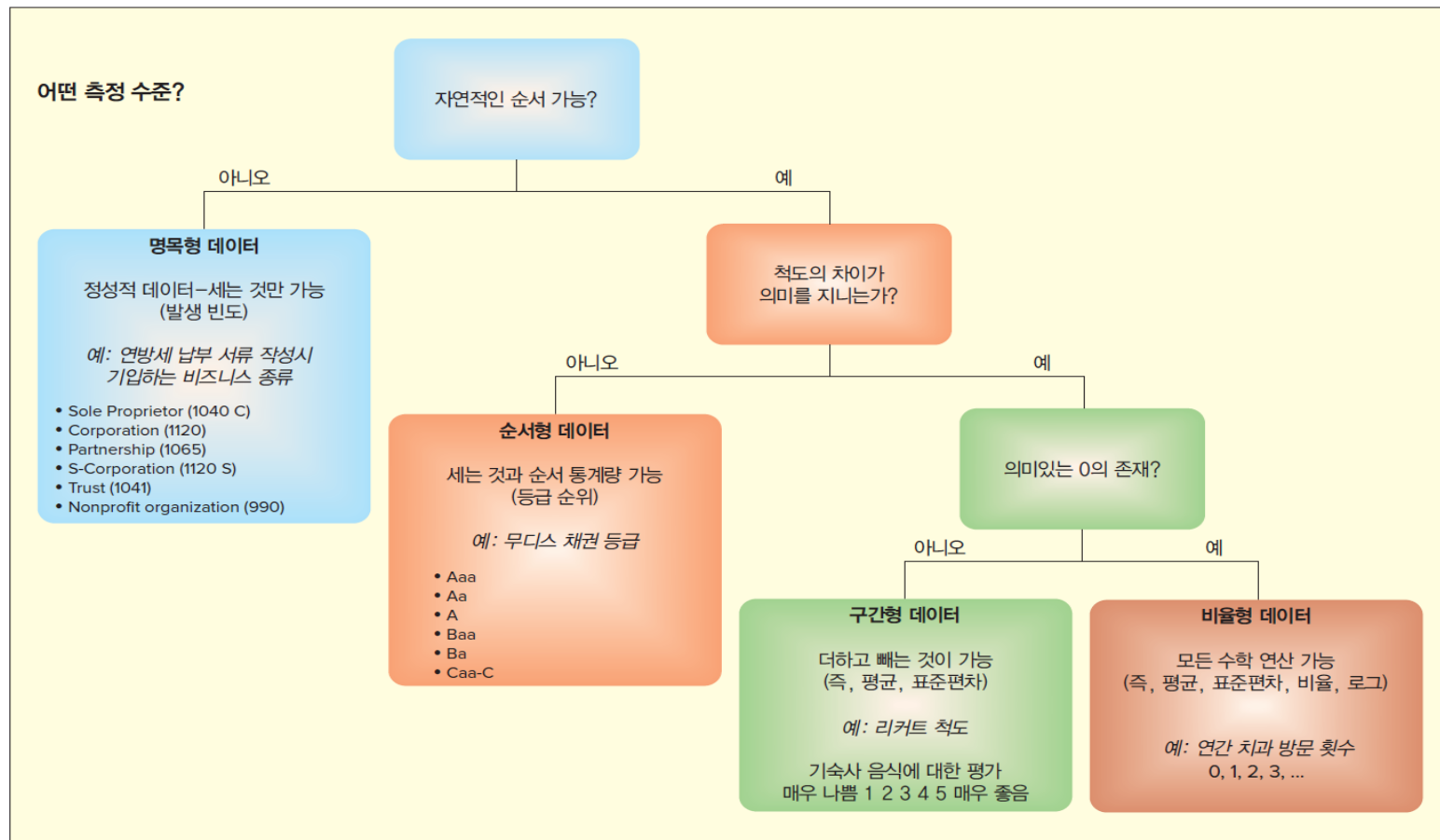


출처: American Bankruptcy Institute (www.abiworld.org).

측정 수준

그림 2.3

측정 수준의 결정



측정 수준

측정 수준	특징	예
명목척도 명목형(nominal)	범주(구분)만 가능	성별, 학년, 거주지, 선수 등번호 등
서열척도 서열(순서)형(ordinal)	범주(구분)+순서 거리는 분명한 의미가 없음	석차, 좋아하는 정도 등
등간척도 등간형(interval)	범주(구분)+순서+구간 거리가 의미를 지님	온도(섭씨 57°)
비율척도 비율형(ratio)	범주(구분)+순서+구간 +절대값 0	지불가능금액 (2,170만 달러)

측정수준: 1. 명목척도

- 명목형 데이터(nominal data)는 단지 범주만을 구별.
- 명목형 데이터는 “정성적”, “범주형”, “분류형” 데이터와 동일한 의미이며, 숫자로 코딩될 수 있다(예: 1 = Apple, 2 = Samsung, 3 = Dell, 4 = HP).
- 적용 가능한 수학적 계산은 세는 것(예: 빈도수), 그리고 최빈값(mode)과 같은 간단한 통계량 정도이다.

사례:

지난 달에 보험료를 청구한 적이 있습니까?

1. 예 2. 아니오

측정수준: 2. 서열척도

- 순서데이터코드는 등급으로 표시될 수 있다
(예를 들어 1=자주, 2=가끔, 3=드물게, 4=전혀)
- 데이터 코드 간의 간격은 의미가 없다(가령 1과 2, 또는 2와 3, 또는 3과 4 사이의 '간격'이 특별한 의미를 갖지 않는다).
- 순서형 데이터는 특히 사회과학, 마케팅, 인사관리 연구 등에서 일반적이다. 순서형 데이터에 대해서는 유용한 통계적 검정이 많다.

사례

MS Access를 얼마나 자주 이용합니까?

1. 자주 2. 가끔 3. 드물게 4. 전혀

측정수준: 3. 등간척도

- 순서를 매길 수 있을 뿐만 아니라 척도 포인트 사이의 간격이 의미를 지닌다(예: 60°F 와 70°F 의 간격은 20°F 와 30°F 의 간격과 동일).
- 숫자 간의 간격은 *거리*를 의미하기 때문에 평균 계산과 같은 수학적 계산을 할 수 있다.
- 0이 갖는 의미는 임의적이며, 따라서 비율은 의미가 없다(예: 60°F 가 30°F 보다 2배 따뜻한 것은 아님).

사례:

어제 기온은 얼마였습니까? (화씨로)

측정수준: 4. 비율척도

- 비율형 데이터(ratio data)는 다른 세 가지 타입 데이터의 성질을 모두 보유하고 있을 뿐만 아니라 의미있는 0(측정량이 없는 상태)을 추가적으로 보유
- 0의 존재로 인해 데이터 값의 비율이 의미를 가짐
(예: 순익 2,000만 달러는 1,000만 달러의 두 배이다)
- 비율형 데이터라고 해서 0이 반드시 관찰될 필요는 없다. 0은 절대적 기준 척도

사례

책의 무게가 얼마였습니까?

측정수준: 리커트 척도

- 구간형 데이터의 특별한 경우로서 서베이 연구에서 자주 사용
- 리커트 척도의 세밀함은 응답문항을 몇 단계(대개 5 또는 7)로 하느냐에 달려 있음

“대학에 진학할 고등학교 학생들은 외국어 공부를 필수로 해야 한다.” (하나를 고르시오)

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
매우 그렇다	대체로 그렇다	그저 그렇다	대체로 아니다	매우 아니다

리커트(이산 척도 포인트 사용)

매우 나쁨 1 2 3 4 5 6 7 매우 좋음

리커트(슬라이더 사용)

매우 나쁨 _____ ▼ _____ 매우 좋음

측정수준: 리커트 척도

- 응답문항을 홀수(대개 5 혹은 7)로 하면 자연스럽게 중앙점(‘보통’)이 생김
- 응답자로 하여금 어느 한쪽 편에 서게 만들기 위해 중앙점을 없애기도 함
- 리커트 데이터는 숫자로 코드화할 수 있는데 간격이 동일하다면 어떤 숫자를 이용해도 됨

리커트 코딩: 1에서 5까지 척도

5 = 많은 도움을 줄 것이다
4 = 조금 도움을 줄 것이다
3 = 영향을 미치지 않는다
2 = 조금 해를 끼칠 것이다
1 = 많은 해를 끼칠 것이다

리커트 코딩: -2에서 +2까지 척도

+2 = 많은 도움을 줄 것이다
+1 = 조금 도움을 줄 것이다
0 = 영향을 미치지 않는다
-1 = 조금 해를 끼칠 것이다
-2 = 많은 해를 끼칠 것이다

데이터 유형을 파악하기 위한 절차

질문	“Yes”인 경우
Q1. 의미 있는 0의 존재?	비율형 데이터(통계적 계산 허용)
Q2. 척도 포인트 사이의 간격이 의미를 지니는가?	구간형 데이터(일반적 통계 가능, 예: 평균, 표준편차)
Q3. 척도 포인트가 순서를 나타내는가?	순서형 데이터(특정 형태의 비모수적 통계 검정에 국한됨)
Q4. 이산형 범주인가?	명목형 데이터(세는 것만 가능, 예: 최빈값)

표본추출의 개념

모집단이나 표본이냐?

모집단(population)은 우리가 관심을 두는 모든 항목으로서 유한일 수도 있고(예: 어떤 비행편의 모든 탑승객), 사실상 무한일 수도 있다(예: 생산 설비에서 출하되는 모든 콜라)

표본(sample)은 모집단 중 우리가 실제로 분석하는 부분집합

표본추출의 개념

표본 또는 센서스?

표본(sample)은 모집단에서 일부 항목들만 뽑아 그것을 살펴보는 아이디어

센서스(census; 총조사)는 어떤 모집단에 속해 있는 모든 항목들을 조사하는 것
하지만 미국의 센서스(10년마다 실시)가 미국내 모든 사람들을 다 조사하는 것은 아님

- 인구이동, 불법 이주민, 미응답, 예산제약, 사실확인의 한계 등

표본추출의 개념

표 2.4

표본인가, 센서스인가?

표본이 선호되는 상황	센서스가 선호되는 상황
<p>무한대의 모집단</p> <p>모집단이 무한대이거나 크기가 확정되어 있지 않은 경우 센서스는 불가능함(어떤 볼트 조립라인이 생산을 계속하고 있거나 어떤 의사가 환자를 계속 진찰하는 경우).</p> <p>파괴적 테스트</p> <p>어떤 것에 대한 측정이 그것을 파괴하거나 가치를 떨어뜨릴 수 있음 (배터리 수명 측정, 자동차 충돌 실험).</p> <p>시의적절한 결과</p> <p>표본추출이 센서스에 비해 결과를 보다 신속히 얻을 수 있음(채소의 살모넬라균 오염 체크).</p> <p>정확성</p> <p>한정된 자원을 센서스에 제한적으로 투입하기보다는 경험 있는 인력을 고용하고 조사요원 교육과 데이터 안전장치의 개선에 예산을 사용하는 것이 더 바람직할 수 있음.</p> <p>비용</p> <p>센서스를 시행하는 것이 가능하더라도 시간 또는 경제적 비용이 예산을 초과할 수 있음.</p> <p>민감한 정보</p> <p>어떤 큰 조직의 성취률에 대해 조사할 때 전체 직원보다는 소수 표본 사원에 대해 심층면접을 하는 것이 더 많은 정보를 얻을 수 있고 비밀유지도 개선될 수 있음.</p>	<p>소규모 모집단</p> <p>모집단의 규모가 작은 경우 데이터 수집에 많은 비용이 들지 않기 때문에 표본추출을 할 이유가 거의 없음.</p> <p>대규모 표본</p> <p>필요한 표본의 크기가 모집단의 크기에 육박할 경우 차라리 센서스를 수행하는 편이 나음.</p> <p>데이터베이스의 존재</p> <p>데이터가 입력되어 있는 경우라면 모든 사례들을 조사할 수 있음.</p> <p>법률적 제한</p> <p>은행은 영업이 끝날 때마다 '모든' 은행원의 서랍 속에 있는 현금을 세야 함. 미국 의회는 2000년 인구 총조사에서 표본추출 방식을 금지함.</p>

표본추출의 개념

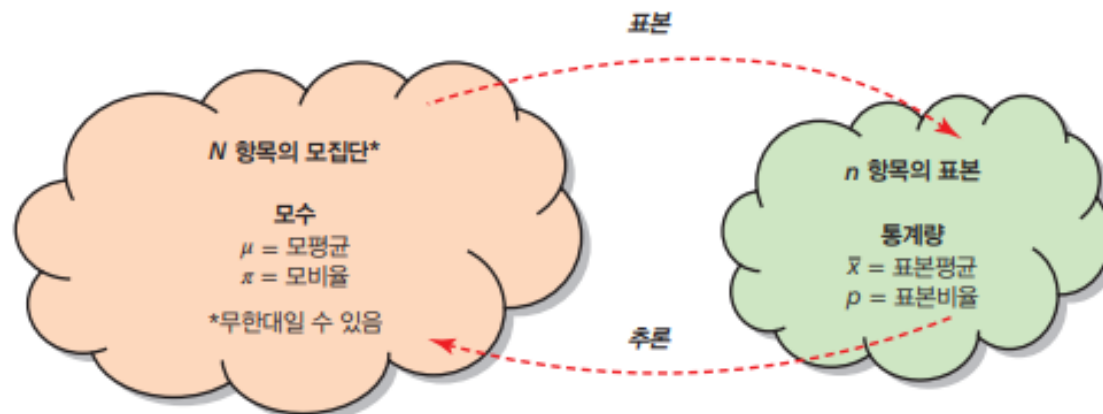
모수와 통계량

모수(parameter): 집단에 대한 측정 또는 특징(예: 평균 또는 비율). 우리가 모집단 전체를 관찰할 수 있는 경우가 드물기 때문에 대개 모수의 값은 알려져 있지 않음. 대개 그리스문자로 표기(μ 또는 π)

통계량(statistic)은 어떤 표본으로 계산한 측정치(예: 평균 또는 비율). 대개 로마문자로 표기(\bar{x} 또는 p)

그림 2.4

모집단과 표본



표본추출의 개념

- 모집단에서 추출된 n 개의 표본으로부터 통계량(statistics)을 계산하고 이를 모집단의 모수(parameters)를 추정하는데 이용
- 혼동을 피하기 위해 모수와 해당 통계량에는 서로 다른 기호를 사용
- 모집단 평균은 μ (그리스문자 뮤의 소문자), 표본평균에는 m 를 사용

표본추출의 개념

목표 모집단(target population)

- 표본 통계량이 모집단 모수를 잘 추정하기 위해서는 모집단을 신중하게 규정하고 표본을 과학적으로 추출함으로써 표본이 모집단을 대표해야 한다.
- 목표 모집단(target population)은 우리가 관심을 갖고 있는 모집단을 말한다(예: 미국 휘발유 가격).
- 표본추출틀(sampling frame)은 표본을 추출하는 대상 그룹을 말한다(예: 115,000개 주유소).
- 표본추출틀이 목표 모집단과 다르다면 표본 추정치가 의미를 갖기 어렵다.

표본추출의 방법

단순임의표본	항목의 선택 시 난수를 사용(예: 비자 카드 소지자)
계통표본	매 k 번째 항목을 선택(예: 레스토랑 고객)
층화표본	규정된 계층 내에서 임의로 선택(예: 나이, 직업, 성별)
군집표본	계층이 지역(예: 우편번호)이라는 점을 제외하고는 층화표본과 유사

표본추출의 방법: 1. 단순임의표본(simple random sample)

- 일반적으로 모집단의 크기를 N 으로, 표본의 크기를 n 으로 표시한다.
- 단순임의표본에서는 모집단에 있는 N 개의 모든 항목들이 n 개의 표본으로 선택될 가능성이 서로 동일하다.
- 이를 위한 한 가지 방법은 포커칩에 N 개의 데이터 값을 적은 다음 이를 바구니에 넣고 완전히 섞은 다음 n 개의 칩을 뽑는 것이다.

표본추출의 방법: 2. 복원 및 비복원 추출

- 기술의 발전으로 임의표본은 쉽게 만들 수 있다.
- 비복원추출(sampling without replacement)은 어떤 항목이 일단 표본으로 선택되면 다시는 표본으로 고려될 수 없는 경우를 의미한다.
- 복원추출(sampling with replacement)은 임의의 숫자가 한 번 이상 나올 수 있는 것을 의미한다.

표본추출의 방법: 2. 복원 및 비복원 추출

비복원추출은 어떤 상황에서 문제를 일으키는가?

- 모집단이 유한한 상황에서 표본의 크기가 모집단에 가까운 경우

주의:

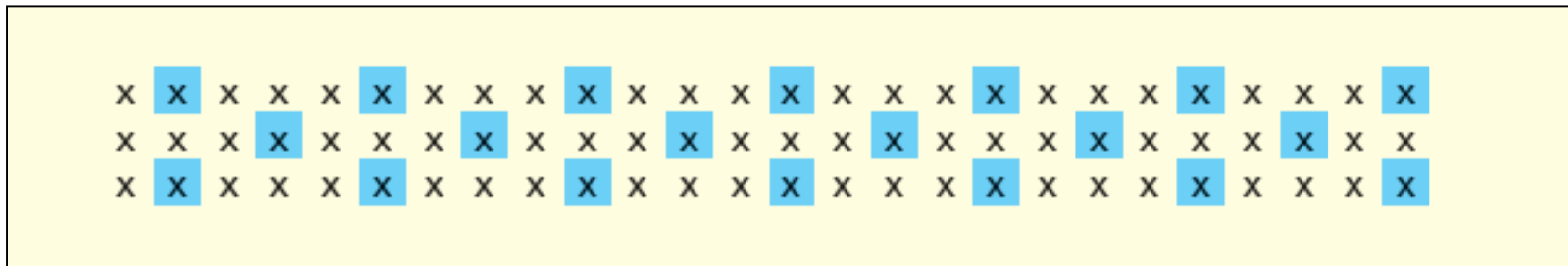
일반적인 기준은 어떤 유한한 모집단에 대해 만약 표본이 모집단의 5% 이하이면(즉 $n/N \leq 0.05$), 해당 모집단은 사실상 무한한(effectively infinite) 것으로 간주됨
반대로 말하면 어떤 모집단이 표본의 20배보다 크면(즉 $N/n \geq 20$) 해당 모집단은 사실상 무한함

표본추출의 방법: 3. 계통추출(systematic sampling)

무작위 추출의 한 방법으로서 주어진 리스트 중 처음 k 개의 항목에서 임의로 하나를 고르고 그 다음부터는 매 k 번째 항목을 계속 추출해 나가는 방식

아래는 두 번째 항목에서 시작하여 매 네 번째 항목을 추출함으로써 총 $N = 78$ 개 항목에서 $n = 20$ 개 항목의 표본을 만드는 것을 보여줌

- $N/n = 78/20 \approx 4$ (주기)임을 주의할 것.



표본추출의 방법: 4. 층화추출(stratified sampling)

- 모집단에 대한 사전 정보를 활용할 경우 표본의 효율성을 높일 수 있음.
특히 모집단이 비교적 성질이 동일하고 크기가 알려진 소그룹, 즉 계층 또는 층(strata)들로 분리될 수 있을 때 이 방법 적용 가능
- 각 계층 내에서 원하는 크기의 표본을 단순 임의로 추출할 수 있음
- 또는 전체 모집단을 대상으로 임의표본을 추출한 다음 적절한 가중치를 이용하여 각 계층별 추정치를 결합시킬 수 있음

예: 성별 인종 등

표본추출의 방법: 5. 군집 표본

- 층들이 지리적 영역으로 이루어진 경우
- 군집추출을 한 단계(one-stage)만으로 수행하는 경우에는 임의로 k 개의 하위지역들(또는 군집들)을 골라 거기에 속해 있는 모든 항목들로 표본을 만든다.
- 2단계(two-stage) 군집추출의 경우에는 일차적으로 k 개의 하위지역들(또는 군집들)을 임의로 고른 다음, 선택된 각 군집에서 다시 임의표본을 추출하는 방식이다.

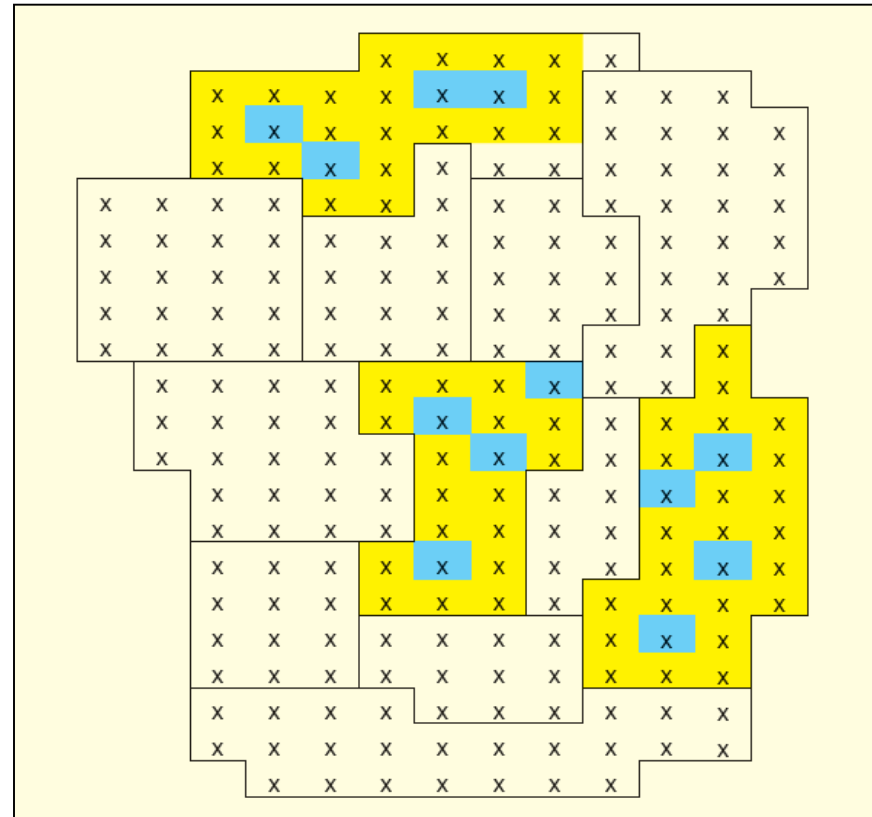
표본추출의 방법: 5. 군집 표본

군집표본

임의로 선정된 3개의
클러스터로부터 4개의 원소가
선택된 예 (2단계 군집표본추출)

그림 2.7

2단계 군집추출: 임의로 3개의
군집을 뽑은 다음, 각 군집에서
임의로 4개의 항목을 뽑음



표본추출의 방법: 비임의추출 방법

판단표본	“전형적인” 항목의 선택에 전문가적 지식을 활용 (예: 인터뷰 대상 사원)
편의표본	이용 가능한 표본을 사용 (예: 점심을 같이 하는 동료에게 질문)
포커스 그룹	대표적 사람들로 구성된 패널과의 심도 있는 대화

표본추출의 방법: 판단표본(judgement sample)

- 비확률적 추출 방법으로서 표본을 추출하는 사람의 전문지식에 의해 모집단을 대표하는 항목들 선택
- 잠재의식적인 편향이 발생할 수 있다(즉, 선택에서 *비임의성*을 야기)
- 쿼터(quota) 표본추출은 판단추출의 한 형태로서 각 범주(예: 남/여)에서 몇 명을 고를지를 면접원이 선택하는 방식

표본추출의 방법: 편의표본(convenience sampling)

- 어떤 표본이든 간편하기만 하면 그것을 취하는 것. 유일한 장점은 빠르다는 것

사례

- 공항의 안전문제에 대한 기사를 쓰면서 여행을 자주하는 동료기자 인터뷰
- 기업의 책임자가 부서장에게 직원들의 업무외 웹서핑이 얼마나 만연했는지 질문

표본추출의 방법: 포커스 그룹(focus group)

- 어떤 이슈에 대해 자유토론을 하거나 아이디어를 모으기 위한 목적으로 모집단을 대표할 수 있는 사람들로 구성된 패널
- 비록 임의추출은 아니지만 포커스 그룹은 기업과 학술연구 등에서 널리 활용됨. 전문가들의 통찰력은 단순한 “숫자” 이상이기 때문

기타 데이터 모으는 방법

- 비즈니스 분야에서는 기술의 발전으로 데이터를 모으는 여러가지 다른 방법들이 사용되고 있음
- 소매점이나 편의점, 레스토랑, 주유소 등에서는 POS(판매시점정보관리) 시스템을 통해 판매 데이터를 실시간으로 집계할 수 있음
- 기업들은 고객들이 사용하는 포인트 적립 카드를 통해 고객의 정보와 해당 고객의 구매 패턴을 연결시킴

표본크기

- 표본을 얼마나 크게 할지는 측정 대상이 기본적으로 어느 정도의 변동성을 갖고 있는지, 그리고 추정을 얼마나 정확히 하기를 원하는지에 달려 있음
- 예를 들어 음료수 Mountain Dew의 카페인 함유량의 경우 공장에서 만들어져 캔이나 병마다 함유량이 매우 일정하기 때문에 표본의 크기가 작아도 평균을 구하는 데 충분함
- 그러나 차에 포함된 카페인은 고객이 얼마나 오래 우려내는가에 따라 달라지기 때문에 더 큰 표본이 필요
- 조사의 목적, 표본추출의 비용, 예산, 시간제약 등도 표본의 크기를 결정하는 데 고려해야 할 요인들임

오차 또는 편향의 원천

- 표본추출에서 **편의(bias)**이라는 용어는 분석하고자 하는 모집단의 모수를 체계적으로 과대 또는 과소평가하는 경향을 말한다.
- **오차(error)**라는 용어는 일반적으로 모집단 모수를 부정확하게 추정하게 만드는 표본 방법론 상의 문제점들을 가리킨다.
- 서베이를 아무리 신중하게 수행해도 오차가 발생할 가능성이 존재한다.

오차 또는 편향의 원천

오차 유형	특징
비응답 편향	응답자와 비응답자가 다름.
선택 편향	자발적 참여자가 전체를 대표하지 않음.
응답 오차	응답자가 틀린 정보를 줌.
포함 오차	모집단이나 틀을 잘못 규정
측정 오차	질문 내용이 편향적이거나 불명확
면접원 오차	면접원에 의해 응답이 영향을 받음.
표본추출 오차	임의적이고 피할 수 없음.

데이터 출처

통계학 코스의 목적 중의 하나는 필요한 데이터를 잘 찾도록 도와주는 것
다행히 요즘은 많은 훌륭한 데이터들을 도서관이나 인터넷, 개인적인 구입 등을
통해 쉽게 이용할 수 있음

데이터 유형	예
미국 고용 관련 데이터	U.S. Bureau of Labor Statistics
미국 경제 데이터	<i>Economic Report of the President</i>
연감	<i>World Almanac, Time Almanac</i>
정기간행물	<i>Economist, Bloomberg Businessweek, Fortune,</i>
색인	<i>The New York Times, The Wall Street Journal</i>
데이터베이스	Compustat, Citibase, U.S. Census
세계 데이터	<i>CIA World Factbook</i>
웹	Google, Yahoo!, MSN

인터넷 데이터 출처

인터넷 데이터 출처

출처

Bureau of Economic Analysis
Bureau of Justice Statistics
Bureau of Labor Statistics
Central Intelligence Agency
Economic Report of the President
Environmental Protection Agency
Federal Reserve System
Food and Drug Administration
National Agricultural Statistics Service
National Center for Education Statistics
National Center for Health Statistics
Statistics Canada
U.N. Department of Economic and Social Affairs
U.S. Census Bureau
U.S. Federal Statistics
World Bank
World Demographics
World Health Organization

웹사이트

www.bea.gov
www.bjs.gov
www.bls.gov
www.cia.gov
www.gpo.gov/erp
www.epa.gov
www.federalreserve.gov
www.fda.gov
www.nass.usda.gov
nces.ed.gov
www.cdc.gov/nchs
www.statcan.gc.ca
www.un.org/depts/unsd
www.census.gov
www.usa.gov/statistics
www.worldbank.org
www.demographia.com
www.who.int/en

서베이

서베이 연구의 기본 절차

- Step 1: 연구의 목적을 세운다.
- Step 2: (시간, 돈, 인원 등) 예산을 짠다.
- Step 3: (목표 모집단, 틀, 표본크기 등) 연구를 설계한다.
- Step 4: 서베이 유형과 관리방법을 정한다.
- Step 5: (설문지 등) 데이터 수집 도구를 설계한다.
- Step 6: 서베이 도구를 사전 테스트하고 필요시 수정한다.
- Step 7: 서베이를 수행한다(필요시 후속 조치).
- Step 8: 데이터를 코드화하여 분석한다.

서베이 유형

서베이 유형	특징
우편	현재 주소가 있는 리스트가 필요함(사람들이 자주 이동하기 때문). 보통 응답률이 낮으며, 무응답 편향(nonresponse bias)이 예상됨. 우편번호 리스트는 특정한 계층을 목표로 삼을 수 있게 도와주지만, 참여를 유도하기 위해 표지에 서베이의 목적을 분명히 밝혀야 하며, (쿠폰과 같은) 인센티브를 제공해야 함(웹서베이가 비용의 측면에서 더 나을 수 있음)
이메일	기업들은 E메일을 통해 등록된 이용자들에 대해 쉽게 서베이를 할 수 있음. 무응답편의가 문제가 될 수 있으나, E메일을 등록한 사람들은 이미 표적모집단이라고 할 수 있으며 자사의 제품이나 서비스에 호의적인 사람들임(물론 E메일을 등록하지 않는 사람들은 예외).
인터넷	많은 판매자들은 Web에 기반한 조사방법을 제공함('web survey tools'를 검색해보라). 이 경우 자신의 이해와 관련한 문제(예를 들어 자동차 소유자의 만족도)에 대해 타깃 그룹을 잘 설정해야 함. (너무 바쁘다든지 하는 이유로) 무응답편의가 예상되며, (사기 혹은 스팸 등의) 불신의 문제도 있을 수 있음. 상품을 탈 수 있는 기회 등을 제공하는 것도 응답을 증가시킬 수 있는 방법

서베이 유형

서베이 유형	특징
전화	전화기 저편에서 ‘간단한 서베이’에 응해주시겠냐는 질문을 받아본 적이 있을 것임 그러나 응했던 적이 있는가? 당신과 같은 기존 고객을 타깃으로 할 때조차도 무응답 편 의가 문제가 됨(정치 관련 서베이같이) 크게 문제될 게 없는 전화조사도 (자동녹음 전화 로 인한) 불신, 발신자 거르기, 보이스메일 등으로 인해 성공하는 경우가 많지 않음
면접	면접은 비용과 시간이 많이 소요되지만 표본 크기가 작아도 결과의 질을 높일 수 있어 가치가 있음. 면접을 통하면 복잡하고 민감한 주제에 대해 정보를 얻을 수 있음. 면접원 을 잘 훈련시켜서 면접을 신중하게 수행하도록 해야 함(추가 비용).

서베이 가이드라인

계획	서베이의 목적이 무엇인가? 진정으로 알고 싶은 것이 무엇인가? 참여 인력들의 전문지식은? 외부에서 조달해야 하는 가장 중요한 기술은? 어느 정도의 정밀도가 필요한가? 예산을 어떻게 쓰는 것이 최선인가
설계	높은 응답률과 유용한 데이터를 얻기 위해서는 서베이를 설계하는 데 시간과 돈을 투자해야 한다. 관련 책자와 자료들을 활용하여 불필요한 오류를 막아야 한다.
품질	준비에 신경을 써야 한다. 고급인쇄와 선전활동은 품질에 대한 사람들의 기대를 높인다. 지저분한 설문지는 무시된다. 어떤 서베이(예: 인터넷)는 특별한 소프트웨어가 필요할 수도 있다.
예비조사	자신에게는 질문내용이 명확하더라도 다른 사람에게는 그렇지 않을 수 있다. 설문지를 친구나 동료에게 사전에 조사해볼 수 있다. 그러나 최선의 방법은 자신과는 아무 이해관계가 없는 소수의 순진한 응답자들을 대상으로 조사하는 것이다.
거두기	서베이의 목적을 분명히 명시하고, 감사 기념품(예: 할인쿠폰, 무료 선물)을 제공하거나 또는 추천(예: 어떤 믿을 만한 전문가 그룹으로부터) 등으로 포장하면 응답률이 높아질 것이다.
전문지식	데이터 수집과 표 작성 등을 스스로 수행하는 경우에도 초기 단계에서 외부(또는 내부) 전문가의 도움을 받는 것을 고려하라. 문제가 발생할 때까지 아무 조치를 취하지 않는 것보다 초기에 자문을 받는 것이 비용면에서 더 효과적이다.

설문지 디자인

- 설문지 여백을 많이 둔다.
- 짧고 명료한 지시문으로 시작한다.
- 설문의 목적을 서술한다.
- 익명성을 보장한다.
- 완성된 설문의 제출방법을 알려준다.
- 자연스런 순서에 따라 섹션을 나눈다.
- 응답자가 해당되지 않는 문항은 건너뛰게 한다.
(예를 들어 “7번문항에 아니오라고 대답했다면 문항 15번으로 바로 가시오”)
- 예비설문에 따라 필요하면 수정한다.
- 설문지를 가능한 한 짧게 한다.

질문 유형

- 개방형
- 빈칸 채우기
- 체크하기
- 순위 선택
- 그림
- 리커트 척도

질문 문장

질문을 어떻게 하느냐에 따라 응답이 크게 영향을 받는다.

예를 들어:

1. 세금인하에 찬성하는가?
2. 세금인하가 고속도로 유지보수를 소홀히 하는 것을 의미할 경우 찬성하는가?
3. 세금인하가 교사와 경찰의 해고를 의미할 경우 찬성하는가?

가능한 응답을 모두 포괄하는지 확인하라.

- 예를 들어, 결혼했는가? ☐예 ☐아니오

구간이 겹치거나 범부가 불분명하면 문제된다.

- 아버지의 나이는? ☐35-45 ☐45-55 ☐55-65 ☐65 이상
- 아버지가 45세이거나 돌아가셨다면 어디에 표시를 하나?

데이터 품질

- 대개 서베이 응답은 숫자로 코드화된다(예: 1 = 남성, 2 = 여성).
- 대개 결측치(missing value)는 특정한 문자로 표시(예: 빈칸, 마침표, 별표).
- 어떤 설문 응답지가 너무 많은 오류와 빈칸으로 되어 있다면 응답지 전체를 뺄 수도 있음
- 여타 선별 이슈들은 복수응답, 단답형 질문에 대한 비정상적 답변, “범위” 응답, 일관되지 못한 응답 등
- 자신의 데이터 코딩 원칙을 잘 기록해 놓아야 함

서베이 소프트웨어

- 설문 작성과정을 자동화하는 소프트웨어가 많아 여러가지 질문 포맷을 사용할 수 있으며, 질문들을 건너뛰어 다음 섹션으로 가게 만들 수 있고, 설문지 편집을 다양하게 만들어 볼 수 있음
- 요즘은 대부분의 설문이 온라인 상에서 이루어지기 때문에 소프트웨어 역시 이런 부분을 반영하여 응답자의 익명성을 보장한다거나 설문에 두 번 응답하는 것을 방지하는 기능을 제공함
- 설문의 디자인, 작성과 관리에는 세심한 준비가 필요하며, 올바른 정보를 얻고 높은 응답률을 달성하기 위해서는 계획을 세워야 함을 명심할 것