

제2장 자료의 수집과 측정



제2장 자료의 수집과 측정

목차

2.1 자료의 수집

2.2 관측 연구

2.3 실험 연구

2.4 측 정

2.5 자료의 척도

2.6 자료의 분류



2.1 자료의 수집- 통계학과 자료과학



2.1 자료의 수집-자료의 구조/ 변수/ 개체



↑
자료

	변수 1	변수 2	...
개체 1			
개체 2			
...			



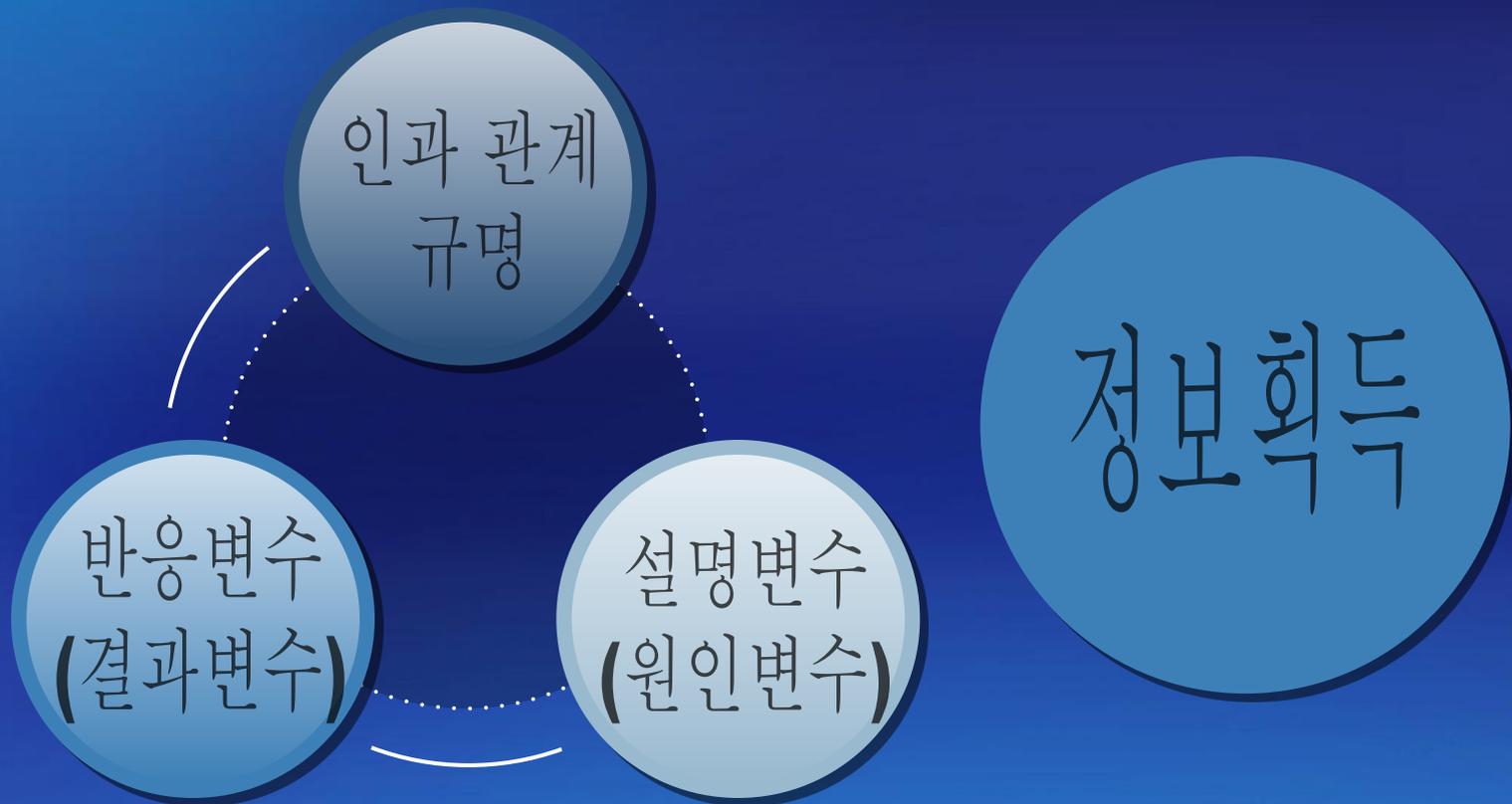
- 반응변수(종속변수, 목적변수, 결과변수)
- 설명변수(독립변수, 입력변수, 원인변수)



- 관측개체 ← 관측연구
- 실험개체 ← 실험연구

[그림 2.1] 자료의 구조, 변수, 개체

2.1 자료의 수집-연구의 목적



2.1 자료의 수집 - (예제 2.1) 관측연구의 예



2.1 자료의 수집 - (예제 2.2) 실험연구의 예



2.1 자료의 수집 - 관측연구와 실험연구의 비교



처리

실험연구에서 실험개체에 연구자가 특정한 실험조건을 가하는 것

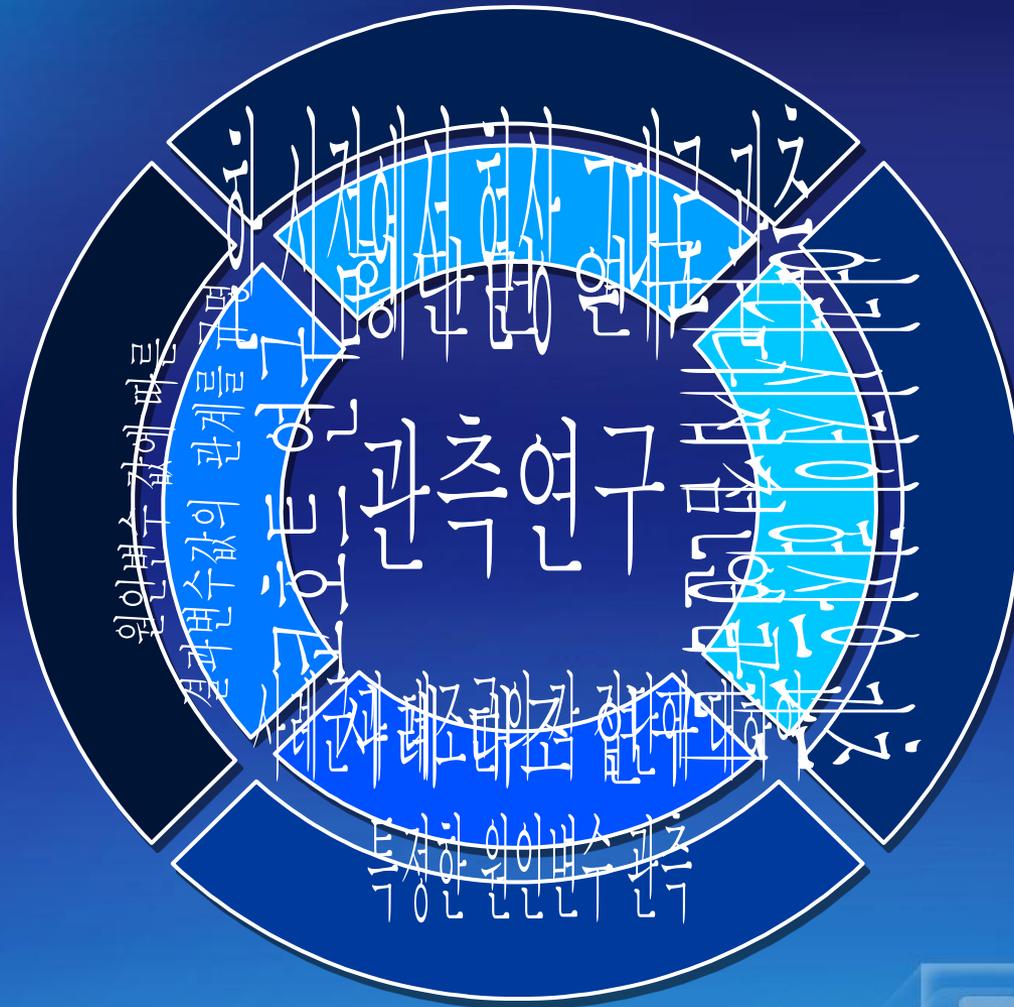
잠재변수

반응변수에 중요한 영향을 주나 설명변수나 연구에 포함되지 않는 변수

교락

반응변수에 영향을 주는 두 개 이상의 변수 효과를 구별하지 못할 때

2.2 관측연구- 관측연구의 종류



2.2 관측연구-(예제 2.3) 횡단면 연구의 예

시청률

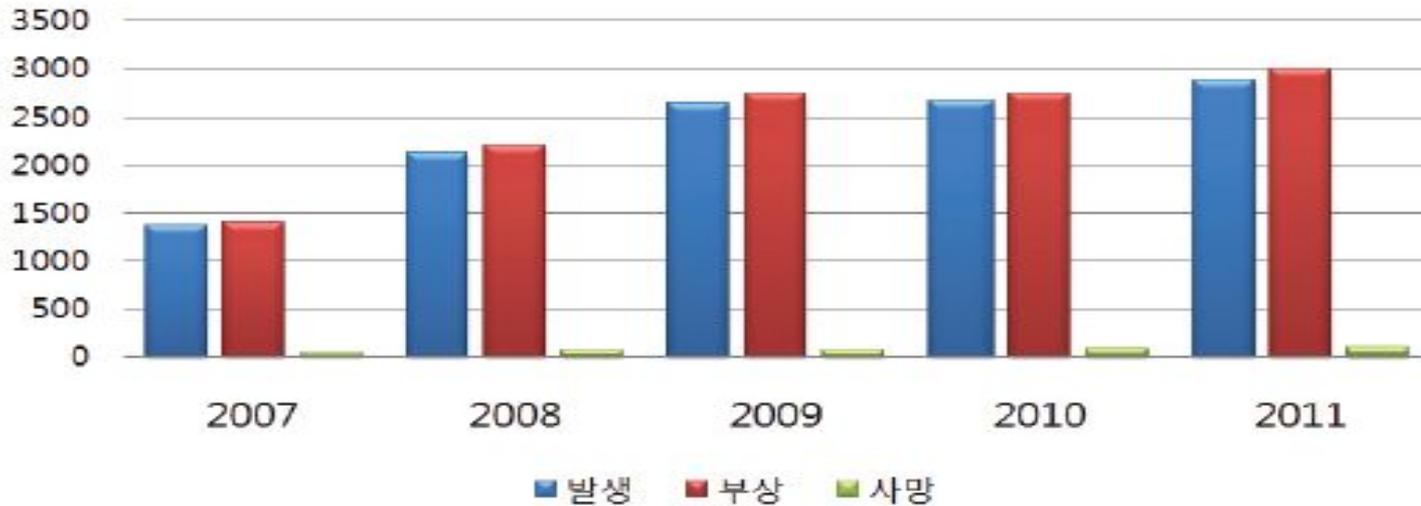
일일시청률		주간시청률	역대시청률	전체 편성표	
종합	드라마	예능	2013년 08월 23일(목) 방송 기준		
프로그램명			채널	시청률	
1	별도 달도 따줄게		KBS1		31.90%
2	KBS 뉴스9		KBS1		23.20%
3	각시탈		KBS1		22.30%
4	천사의 선택		KBS1		16.80%
5	아랑사또전		KBS1		16.20%
6	해피투게더 시즌3		KBS1		15.90%
7	순간포착 세상에 이런일이		KBS1		14.70%
8	KBS 뉴스광장		KBS1		13.70%
9	한국인의 밥상		KBS1		13.60%
10	인간극장		KBS1		13.40%

[그림 2.3] 횡단면 연구의 예: 시청률 조사

TV
시청률 조사

- 개체 : 3,000가구 패널
- 변수 : 각 가구가 시청한 프로그램들
- 반응/ 설명변수 : 구별 없음

2.2 관측연구-(예제 2.4) 다시점 연구의 예



[그림 2.4] 다시점 연구의 예: 자전거 사고 통계

자전거 사고

개체 : 11,689건의 자전거 사고

변수 : 발생한 사망자수와 부상자수

반응/설명변수 : 구별 없음

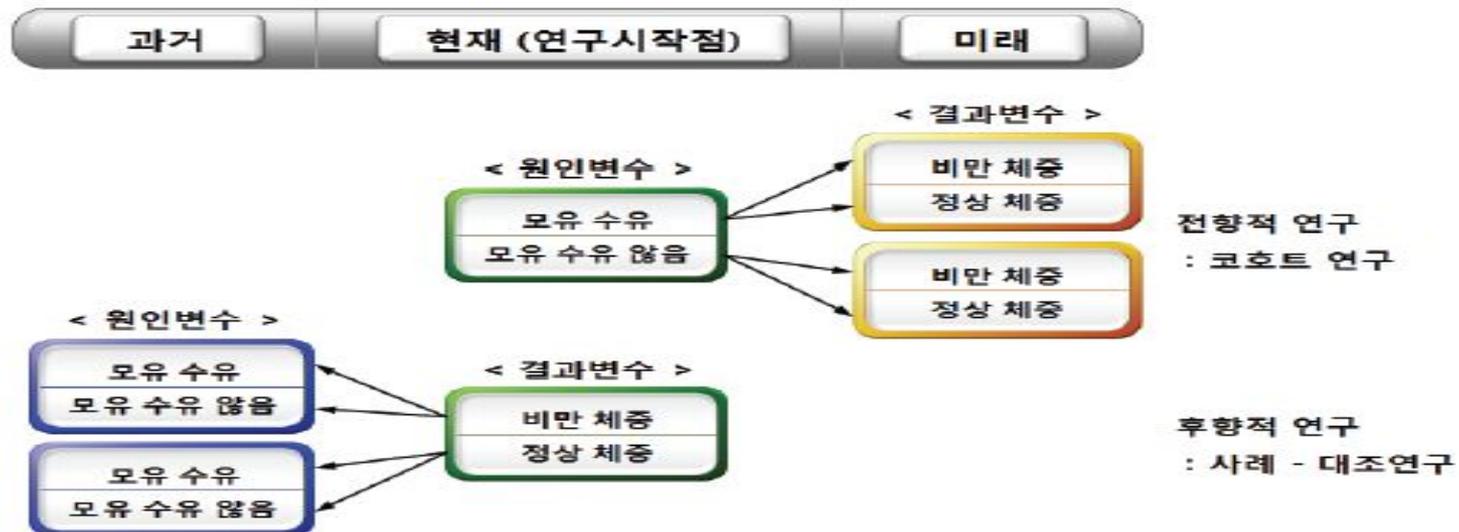
2.2 관측연구-사례 대조연구와 코호트연구

사례 대조연구

현재의 결과변수값에 따라 사례군과 대조군으로 구분한 후 각 집단에 대하여 과거에 특정한 원인변수가 있었는지를 알아보는 관측연구

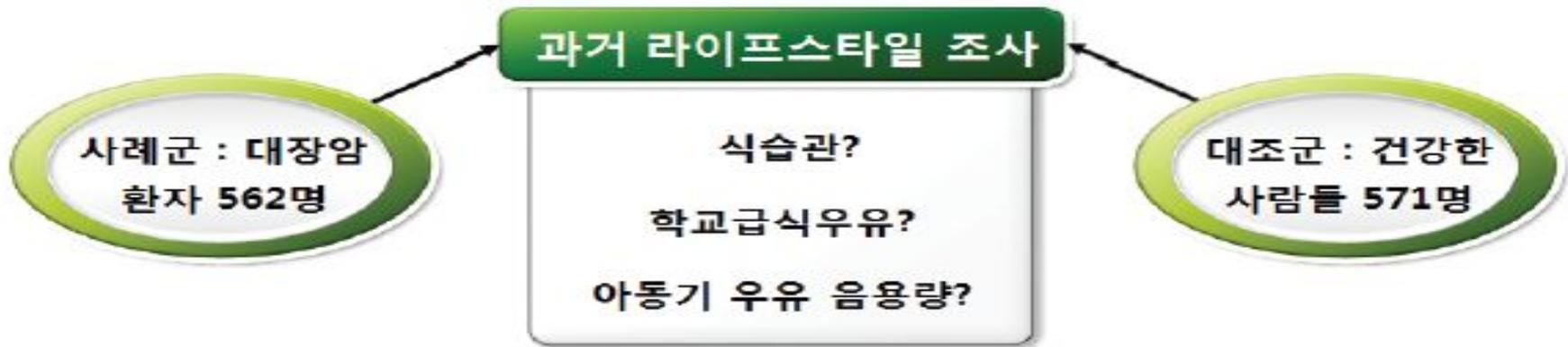
코호트 연구

원인변수값을 미리 측정한 후 일정한 기간을 추적하고 결과변수값을 측정하여 두 변수 간의 관계를 규명하는 관측연구

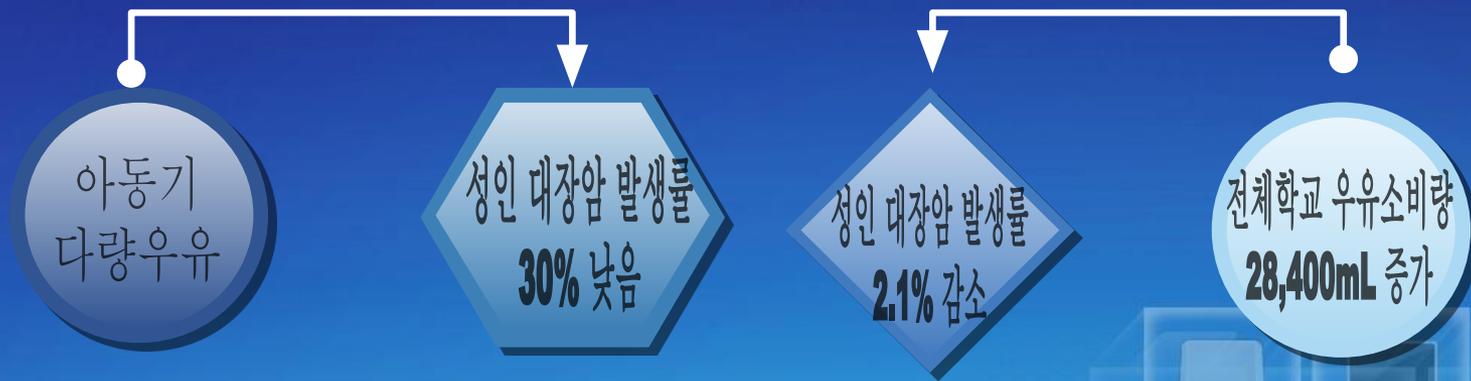


[그림 2.5] 원인변수, 결과변수 및 시간별로 구분한 사례-대조 연구와 코호트 연구의 비교

2.2 관측연구-(예제2.5) 사례-대조연구의 예



[그림 2.6] 후향적 연구의 예: 사례-대조 연구



2.2 관측연구-(예제 2.6) 코호트 연구의 예



[그림 2.7] 전향적 연구의 예: 코호트 연구

2.3 실험연구- 임의화 비교 실험

변수 사이의 인과 관계를 밝혀내기 위해서는 연구자가 개입하여 교각의 가능성이 없도록 할 수 있는 실험연구가 더 좋음

임의화 비교실험

각 실험 개체는 처리를 하는 처리군과 처리를 하지 않는 대조군에 랜덤하게 배치된 후에 랜덤한 순서로 실험이 실시되게 함으로써 실험의 임의화(랜덤화)를 이룸



2.3 실험연구-(예제 2.7) 임의화 비교실험의 예



[그림 2.8] 임의화 비교실험의 예: 위약, 이중눈가림 실험

이중눈가림 실험

실험대상자와 실험연구자 모두가 어떤 처치를 받았는지를 모르는 실험

2.3 실험연구- 실험계획

실험계획법

해결하려는 문제에 대해 최소 횟수의 실험으로 수집한 자료로 최대의 유용한 정보를 얻을 수 있는 실험을 계획하는 통계학의 응용분야

완전 임의화 계획

모든 실험개체를 각 처리집단에 임의 할당하여 균질한 실험환경을 만드는 실험설계

임의화 블록 계획

비슷한 특성을 가진 실험개체들끼리 블록으로 묶은 후에 블록 내의 실험개체들이 처리조합에 임의로 할당되게 하는 계획

2.3 실험연구-(예제 2.8) 완전 임의화 계획의 예



[그림 2.9] 완전 임의화 계획의 예: 처리가 두 개인 경우

성적비교

개체: 수강학생

결과변수: 평가점수 원인변수: 강의방법

처리: 강의실 수업, 온라인 수업

2.3 실험연구-(예제 2.9) 완전 임의화 계획의 예

	올리브 오일	포도씨 오일	해바라기 오일
10g	①⑧⑦	⑩⑬⑧	⑫⑨⑦
30g	②⑬⑤	⑮⑥③	⑭⑪④

[그림 2.10] 완전 임의화 계획의 예: 처리가 여섯 개인 경우

맛의 차이 비교

개체 : 접시에 담긴 시금치 샐러드

결과변수 : 맛의 평가점수 원인변수 : 오일 종류와 호두의 양

처리: 6개, 3회 반복/처리

2.3 실험연구-(예제 2.10) 임의화 블록 계획의 예



[그림 2.11] 임의화 블록 계획의 예: 두 개의 블록과 두 개의 처리

2.4 측정

개체가 가지는
특성을 숫자로
표현

도구를 사용하고
단위를 선택하여
측정을 기록

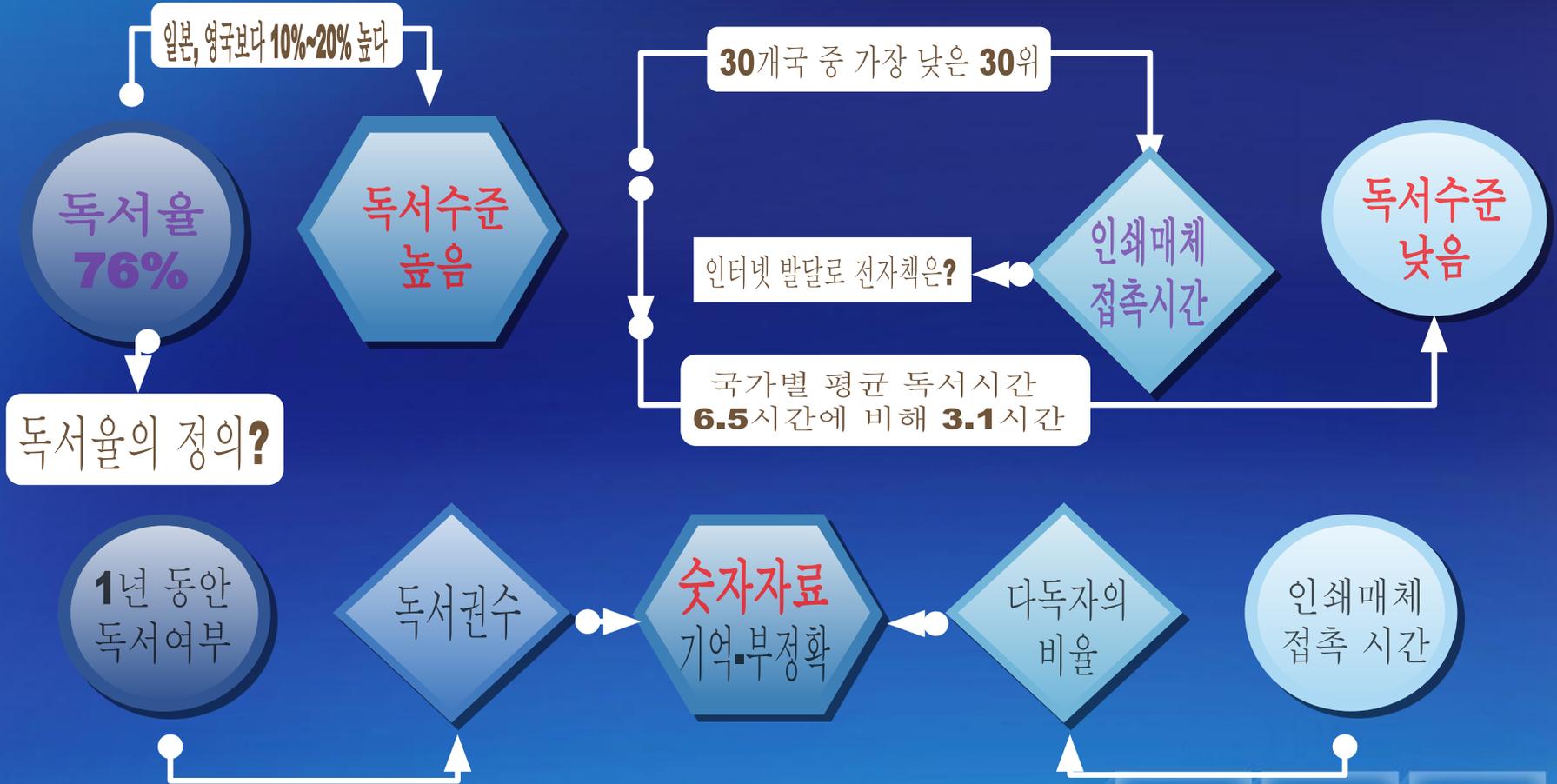
줄자- 허리둘레 **28**인치
저울- 돼지고기 **600g**
IQ검사/ 취업률 등

측정



2.4 측정- 변수의 정의 / 숫자자료 관측

국가지식경쟁력을 나타내는 독서수준



2.4 측정 - 연구목적에 적합한 자료를 얻기 위한 조건

연구목적에 적합한 자료를
얻기 위한 조건

명확한 변수의 정의

교통사고 사망자수
측정

- 24시간 이내 사망자수?
- 1년 이내 사망자수?

타당성있는
측도로 측정

학생들 지능검사

- 영어로 된 시험이라면?

대학평가

- 취업률?
- 취업자수?

정확하게 측정

측정값

실제 값+ 편의+ 임의 오차
편의↓ 임의오차↓



2.5 자료의 척도-명목/순서/구간/비척도

명목척도

개체의 특성을 구별하기
위해서 숫자 부여

학번/ 지역코드
혈액형

순서척도

개체의 특성 구별
우열관계 혹은 대소관계

학점: **5/ 10/ 15/ ...**
선호도: **1/2/3/ ...**

자료의 척도

구간척도

순서척도가 가지는 순서 정보
+ 두 측정값의 차이에 의미

온도 : **10°C와 20 °C**
IQ : **120과 130**

비척도

가장 높은 수준의 척도로서 측정값들의 구별
서열, 차이에 더불어 비의 의미있는 **0** 존재

체중: **0kg/ 40kg/ 90kg**
점수: **0점/ 30점/ 90점**



2.6 자료의 분류-질적자료/ 양적자료



2.6 자료의 분류-이산형자료와 연속형자료

