

# 제 1장

## 통계학의 기초



고려대학교 경영대학 박 광태

# 통계학

## ◆ 통계학의 중요성

- ◆ 불확실성하의 의사결정문제와 직접 연관된 학문

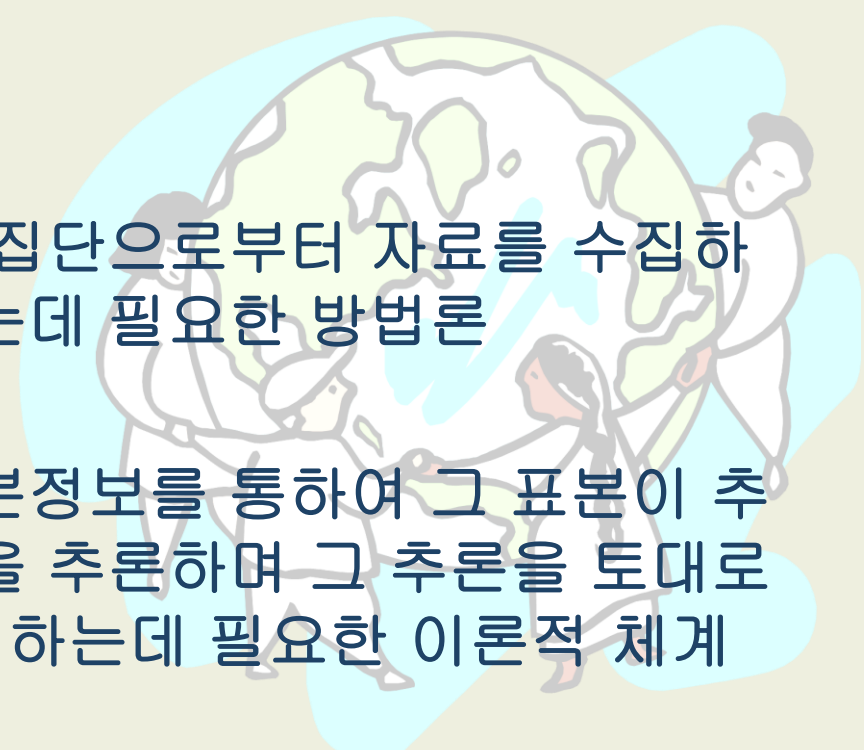
## ◆ 통계학의 정의

### ◆ 기술통계학

- ◆ 분석의 대상이 되는 집단으로부터 자료를 수집하고 요약하고 정리하는데 필요한 방법론

### ◆ 추측통계학

- ◆ 표본을 추출하고 표본정보를 통하여 그 표본이 추출된 모집단의 특성을 추론하며 그 추론을 토대로 의사결정대안을 제시하는데 필요한 이론적 체계



# 모집단과 표본

## ▶ 모집단

- ◆ 모집단이란 통계분석의 대상이 되는 모든 개체들의 집합

## ▶ 표본

- ◆ 표본이란 모집단으로부터 임의로 추출된 모집단의 부분집합



# 모수와 표본통계량

## ◆ 모수

- ◆ 모집단의 특성을 나타내 주는 수치로 모평균, 모분산, 모비율이 쓰임

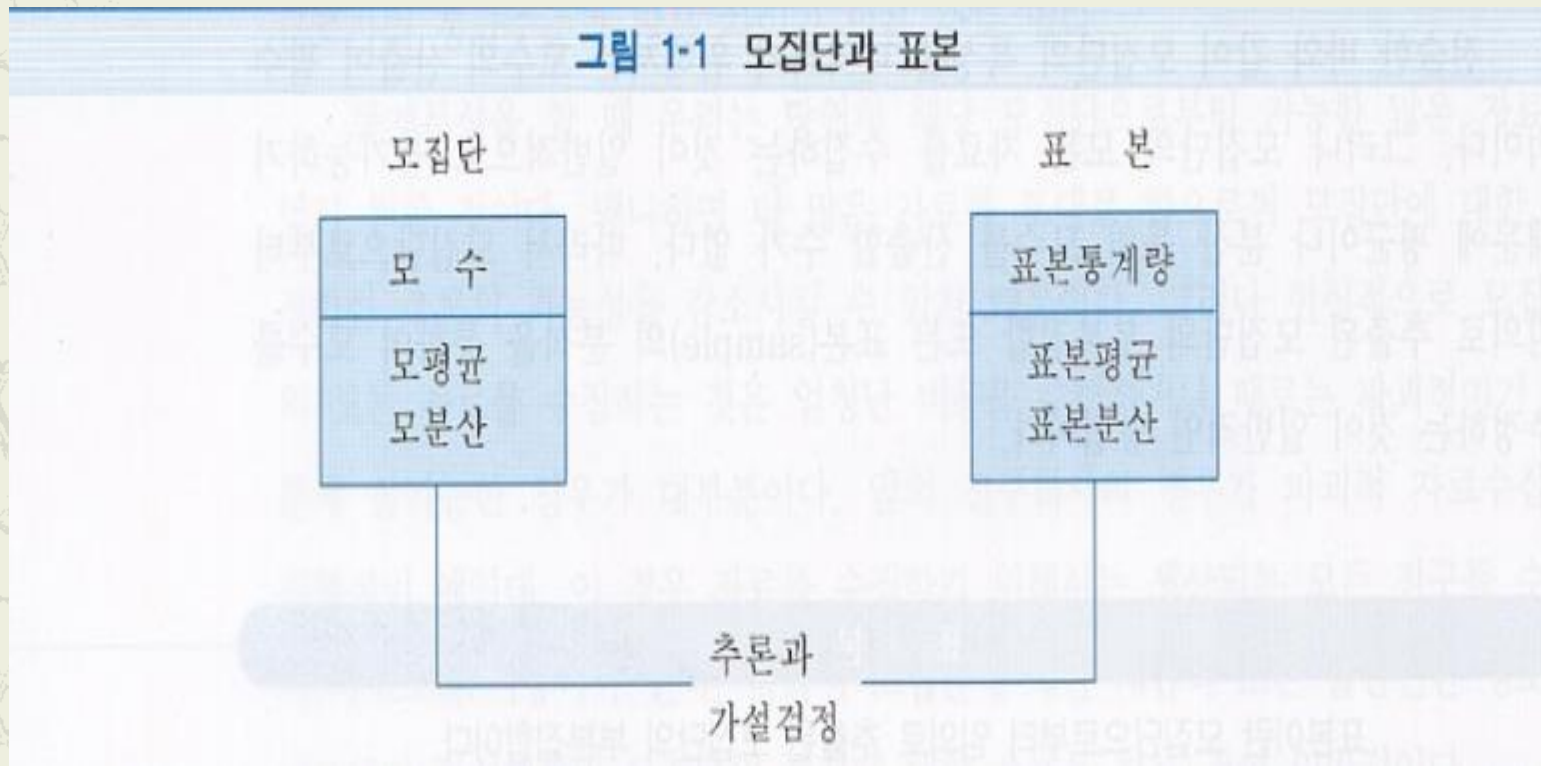
## ◆ 표본통계량

- ◆ 표본의 특성을 나타내는 척도로 표본평균, 표본분산, 표본비율이 쓰임



# 모수와 표본통계량

▶ 그림 1-1(p.8)



# 표본추출법

## ◆ 무작위 추출

- ◆ 모집단을 구성하는 개체들이 선택될 확률이 동일하도록 추출
- ◆ 어떤 한 개체의 선택이 다른 개체의 선택에 영향을 미치지 않도록 추출

## ◆ 계통추출

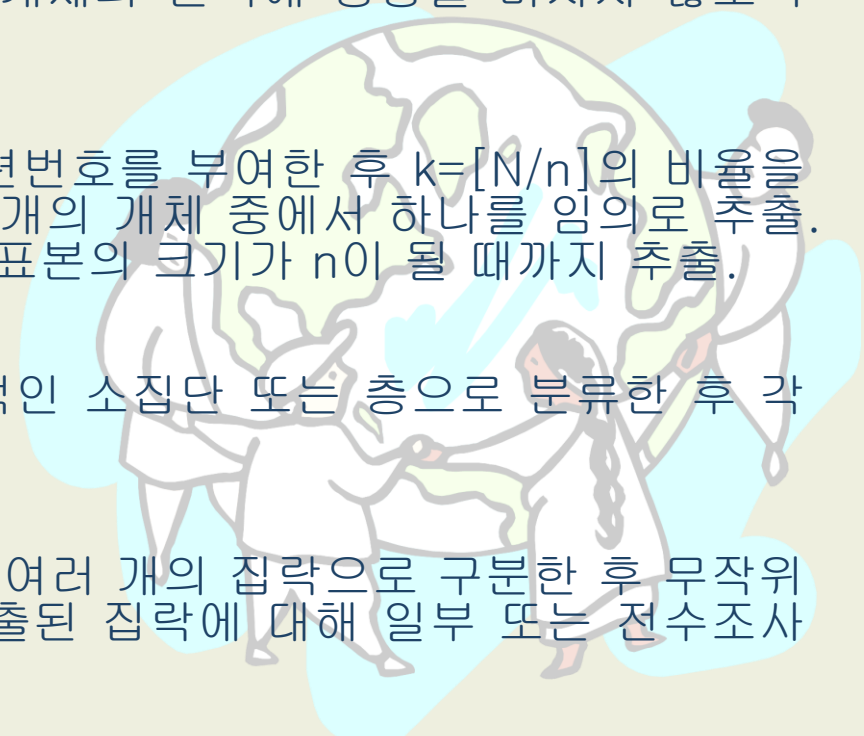
- ◆ 각 개체에 1에서 N까지의 일련번호를 부여한 후  $k = [N/n]$ 의 비율을 구함. 일련번호에 따라 처음 k개의 개체 중에서 하나를 임의로 추출. 그리고 나서 매 k번째 자료를 표본의 크기가 n이 될 때까지 추출.

## ◆ 층화추출

- ◆ 모집단을 두 개 이상의 동질적인 소집단 또는 층으로 분류한 후 각 층으로부터 무작위 추출

## ◆ 집락추출

- ◆ 모집단을 일정한 기준에 따라 여러 개의 집락으로 구분한 후 무작위 추출로 일부 집락을 추출. 추출된 집락에 대해 일부 또는 전수조사를 함.



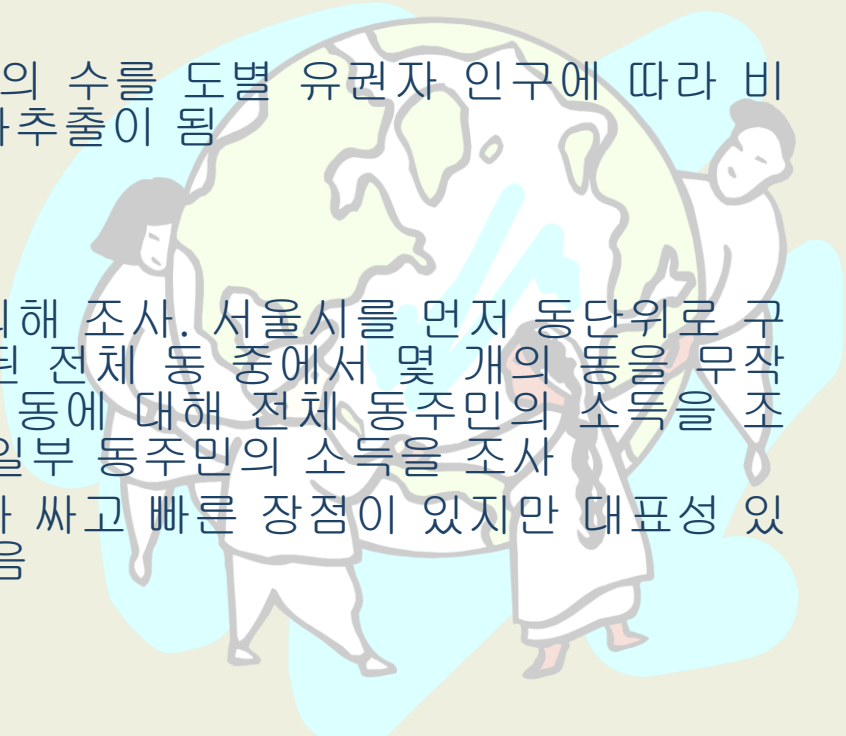
# 표본추출법

## ◆ 층화추출

- ◆ 대선후보에 대한 선호도를 조사하기 위해 400명을 층화추출. 유권자를 도별로 나누어 8개의 층을 구성하고 각 도에서 50명의 유권자를 무작위로 추출
- ◆ 만일 각 도에서 추출되는 표본의 수를 도별 유권자 인구에 따라 비례적으로 결정하면 비례적 층화추출이 됨

## ◆ 집락추출

- ◆ 서울시의 소득을 집락추출에 의해 조사. 서울시를 먼저 동단위로 구분하여 집락을 구성하고 구성된 전체 동 중에서 몇 개의 동을 무작위로 추출. 마지막으로 추출된 동에 대해 전체 동주민의 소득을 조사하든지 또는 무작위 추출로 일부 동주민의 소득을 조사
- ◆ 다른 표본추출법에 비해 경비가 싸고 빠른 장점이 있지만 대표성 있는 표본의 추출에는 문제가 있음



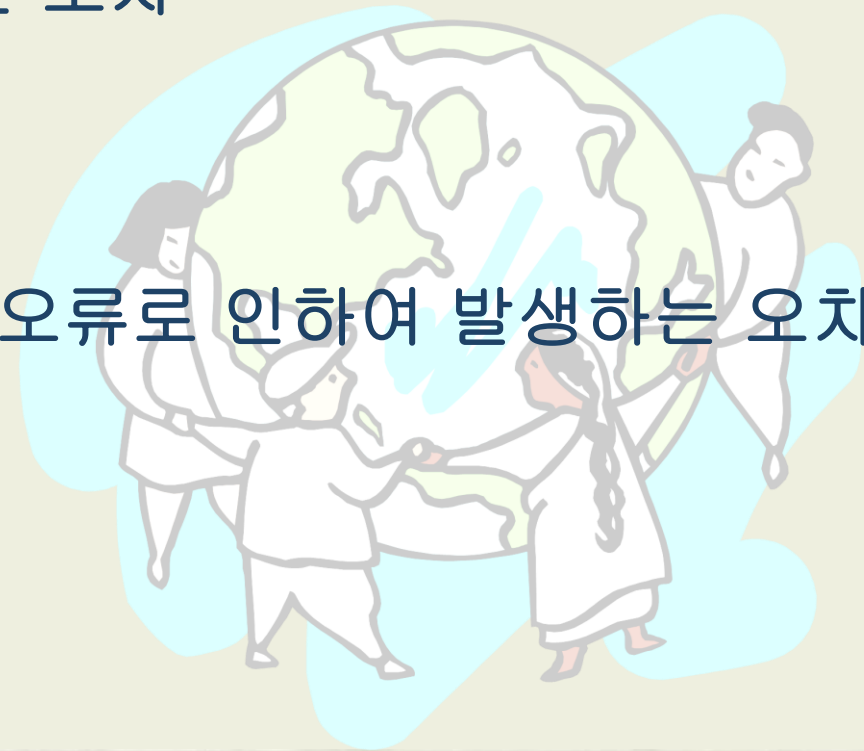
# 표본오차와 비표본오차

## ▶ 표본오차

- ◆ 모집단의 일부분인 표본에 의해서 전체의 특성을 파악하려는 데서 오는 오차

## ▶ 비표본오차

- ◆ 표본추출과정에서의 오류로 인하여 발생하는 오차





## 제 2장

# 기술통계 1: 표와 그래프



고려대학교 경영대학 박 광태

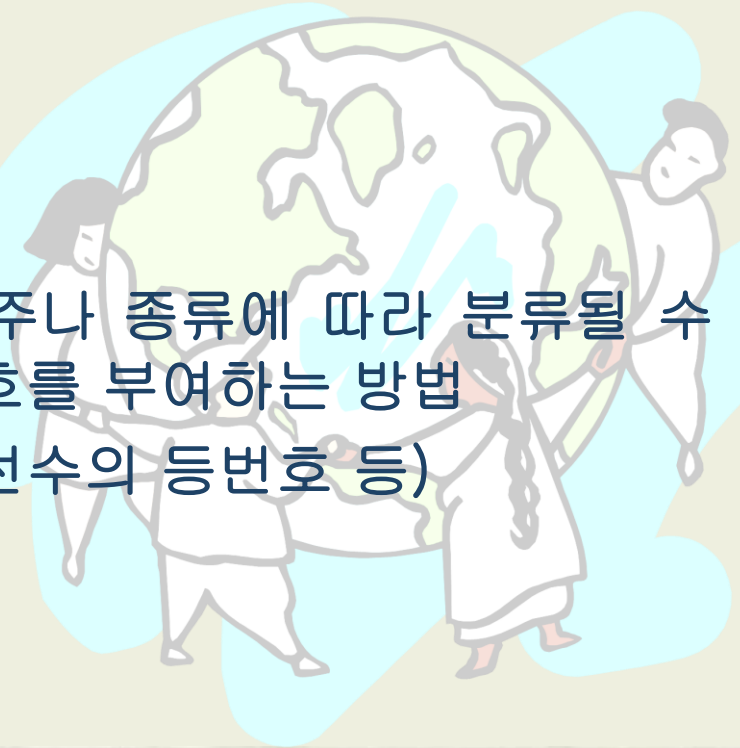
# 자료의 형태

## ◆ 자료의 형태

- ◆ 질적자료 : 명목자료, 서열자료 → 이산형자료
- ◆ 수치자료 : 구간자료, 비율자료 → 이산형자료, 연속형자료

## ◆ 명목척도(nominal scale)

- ◆ 측정대상이 그들이 속한 범주나 종류에 따라 분류될 수 있도록 측정대상에 수치나 부호를 부여하는 방법  
(학년, 주민등록번호, 운동선수의 등번호 등)
- ◆ =, ≠ 이 가능



# 자료의 형태

## ◆ 서열척도(ordinal scale)

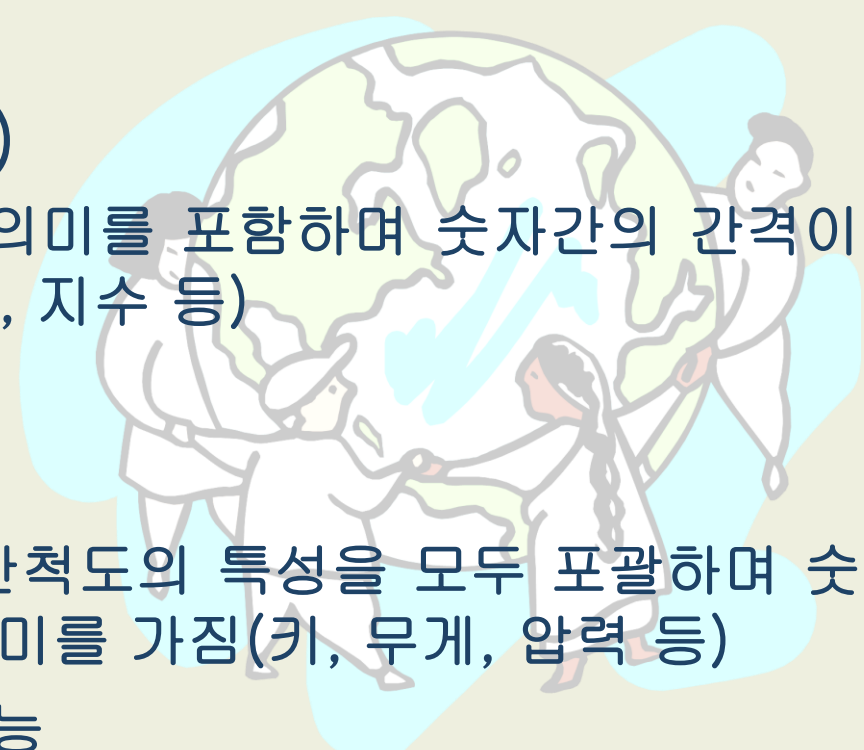
- ◆ 개체간의 서열관계를 나타내어 주는 척도(제품선호도, 국가간의 GNP순위, 올림픽 금메달 순위 등)
- ◆  $=, \neq, \leq, \geq$  이 가능

## ◆ 구간척도(interval scale)

- ◆ 명목척도와 서열척도의 의미를 포함하며 숫자간의 간격이 산술적 의미를 가짐(온도, 지수 등)
- ◆  $=, \neq, \leq, \geq$  이 가능

## ◆ 비율척도(ratio scale)

- ◆ 명목척도, 서열척도, 구간척도의 특성을 모두 포괄하며 숫자간의 비율이 산술적 의미를 가짐(키, 무게, 압력 등)
- ◆  $=, \neq, \leq, \geq, +, -, \times, \div$  이 가능



# 표본추출법

## ▶ 표본추출법

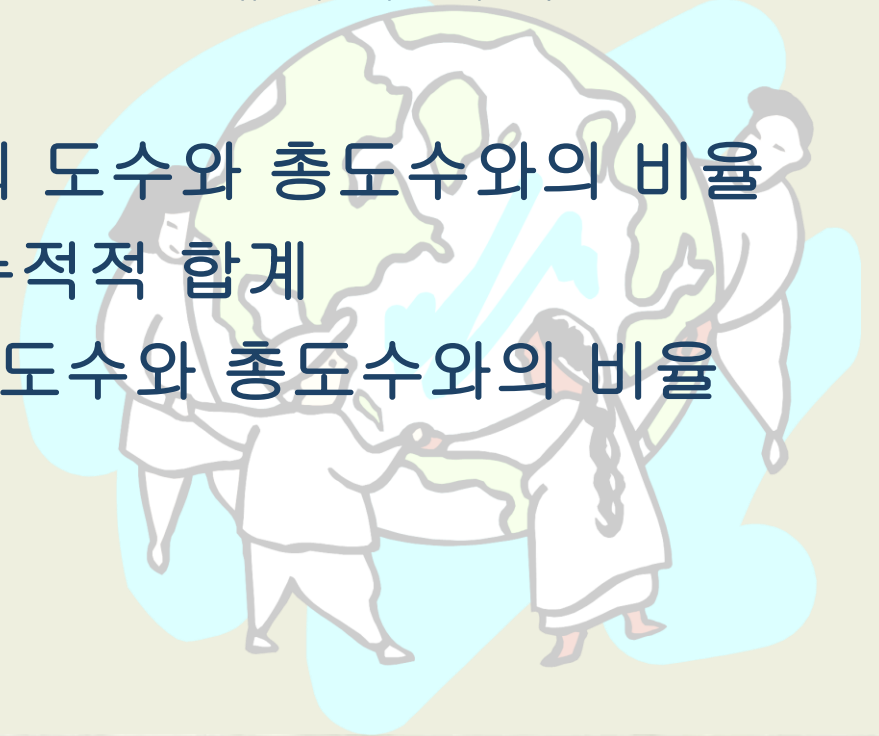
- ◆ 전수조사(census) : 대상모집단의 모든 구성원소를 전부 조사 → 예) 인구센서스
- ◆ 표본조사 : 모집단으로부터 추출된 일부만을 조사 → 예) 여론조사



# 도수분포표

## ◆ 도수분포표

- ◆ 수집된 자료를 제한된 수의 구간으로 나누고 각 구간에 속한 자료의 빈도를 헤아려 기록한 표
- ◆ 상대도수 : 각 계급의 도수와 총도수와의 비율
- ◆ 누적도수 : 도수의 누적적 합계
- ◆ 누적상대도수 : 누적도수와 총도수와의 비율



# 도수분포표

## ◆ 도수분포표 작성

- ◆ 단계 1 : 범위 ( $x_{max} - x_{min}$ )를 구함
- ◆ 단계 2 : 계급의 수를 결정
- ◆ 단계 3 : 계급의 폭을 결정 (범위/계급의 수)
- ◆ 단계 4 : 계급의 경계값을 결정
- ◆ 단계 5 : 각 계급의 도수를 구함



# 도수분포표

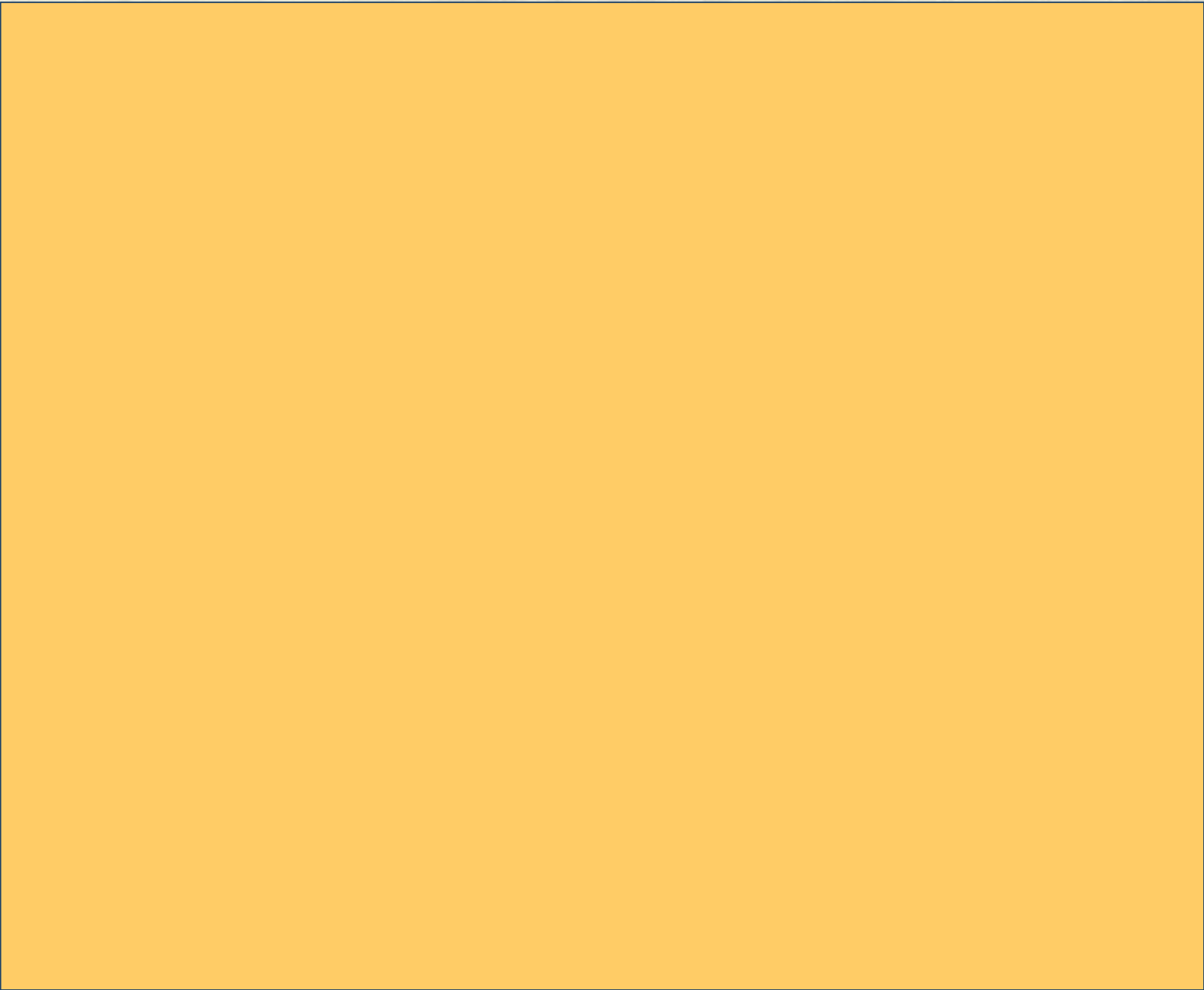
## ▶ 도수분포표 작성(p.32의 예제2-1)

### 예제 2-1

표 2-1의 자료는 천마회사 사원의 월급자료를 보여준다. 이 회사의 경영자는 자사 사원들의 월급에 대한 분포현황을 알고 싶어한다. 자신이 이 회사의 인사담당이라 생각하고, 도수, 상대도수, 누적도수, 누적상대도수가 모두 나타난 도수분포표를 앞에서 설명한 다섯 단계를 이용하여 작성해 보시오.

표 2-1 천마회사 사원의 임금(월급) (단위: 천 원)

1,050	820	1,120	1,080	1,190	1,150	840	890
880	1,070	940	925	825	1,495	1,260	920
1,060	880	1,100	1,085	1,240	1,025	1,050	1,150
1,090	1,315	945	900	1,125	1,160	1,400	940
1,200	1,120	1,175	1,150	955	900	800	1,050
950	1,525	1,325	1,360	1,050	1,200	1,100	1,400
1,000	1,075	1,225	1,500	1,280	975	1,120	910
1,270	1,140	1,120	1,050	1,315	1,100	1,140	1,150
800	1,250	1,150	1,205	1,080	1,375	1,030	1,100
1,230	1,515	920	1,320	1,430	1,350	1,050	1,380

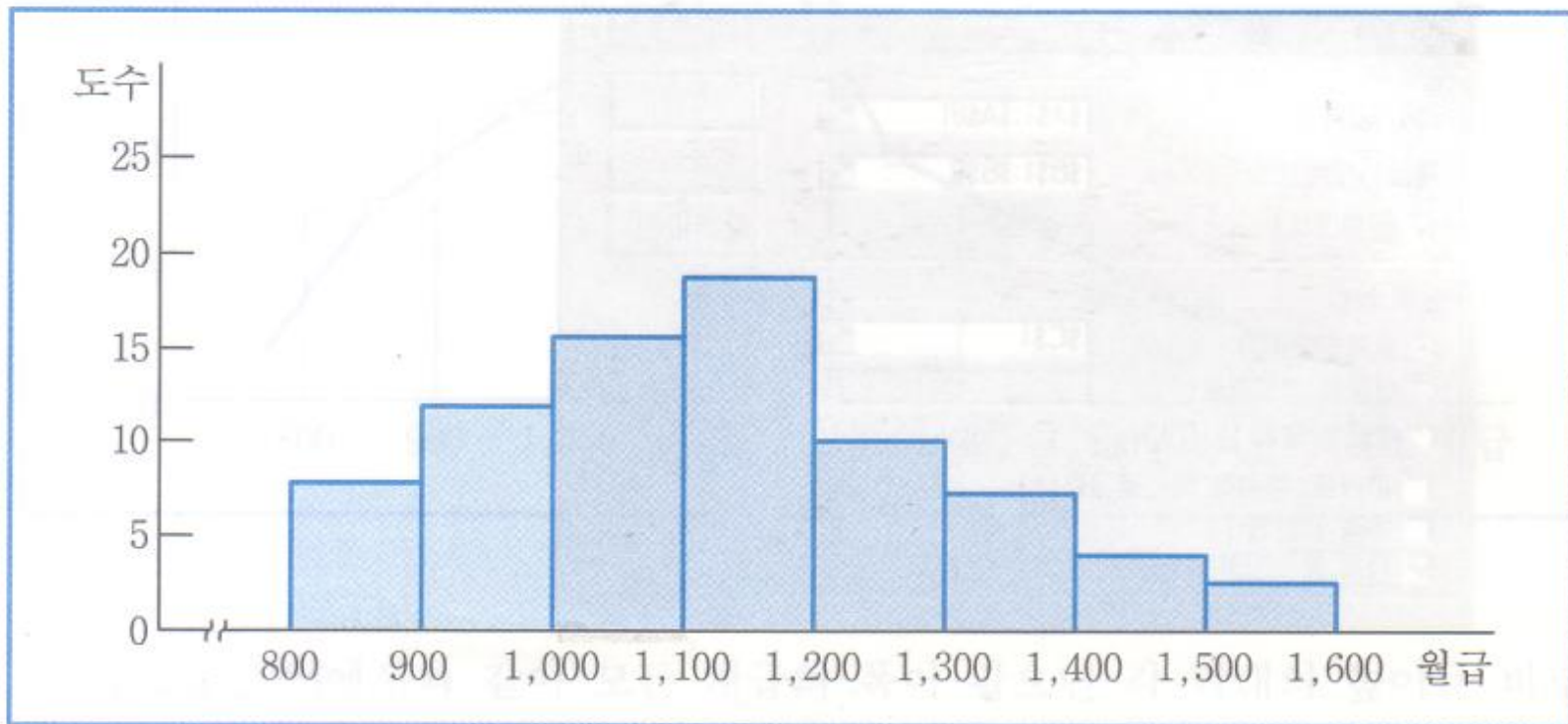




# 그래프

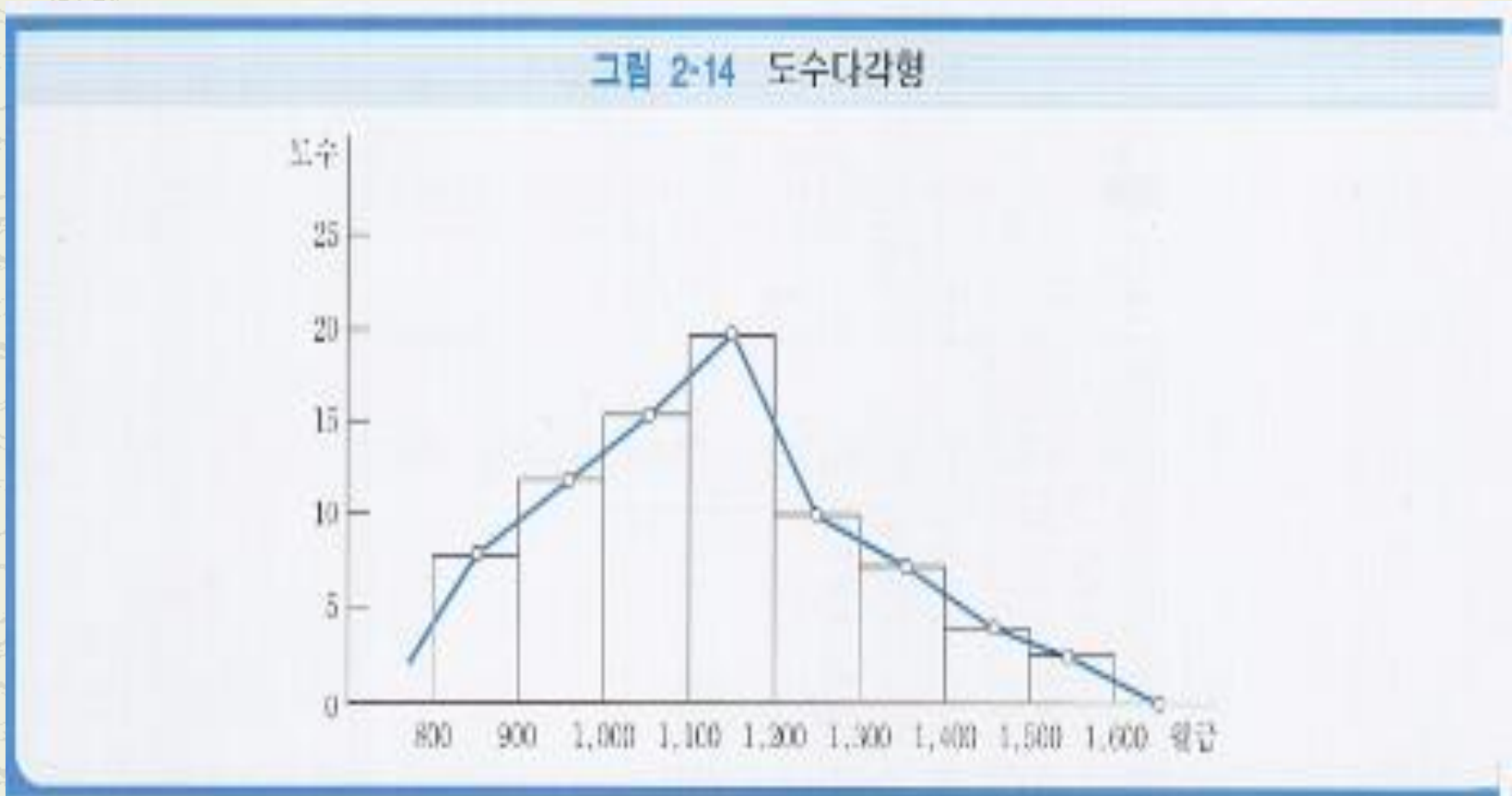
## ▶ 막대그림표 (p. 34의 그림 2-5)

그림 2-5 표 2-2에 대한 막대그림표



# 그래프

▶ 도수다각형 (p. 40의 그림 2-14)



# 그래프

▶ 줄기-잎 그림 (p. 45의 그림 2-21)

표 2-4 삼진기업 종업원의 나이자료

43	34	21	26	57	64	51	30	38	60
19	26	41	47	58	50	42	32	25	18
36	27	38	37	24	49	56	51	20	29
32	51	60	52	45	44	33	25	29	27
31	35	39	19	20	30	46	47	32	24
29	22	20	37	34	29	60	55	30	41

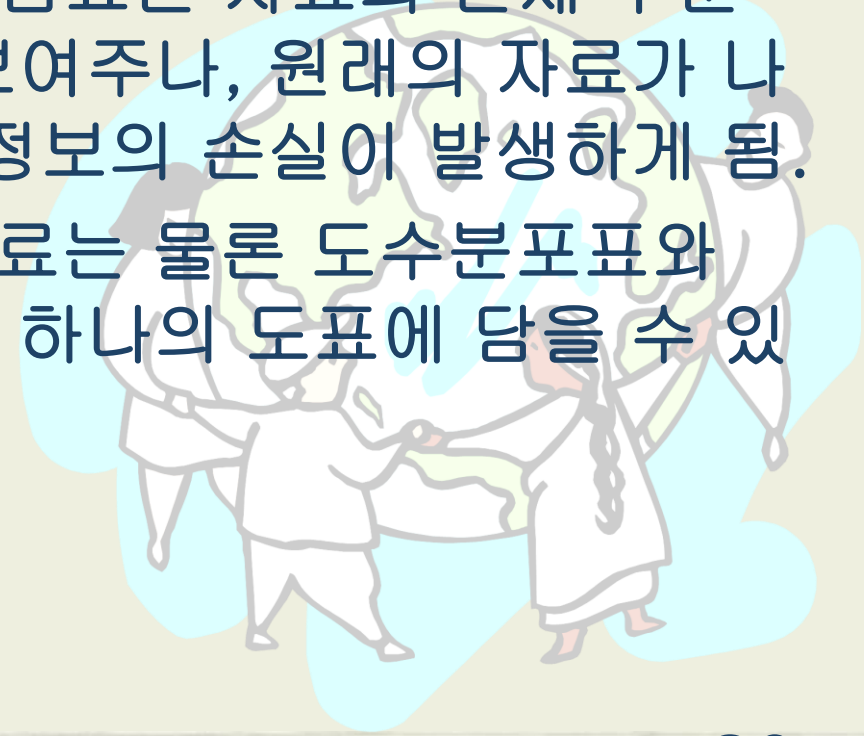
그림 2-21 표 2-4에 대한 줄기-잎 그림

1	8	9	9															
2	0	0	0	1	2	4	4	5	5	6	6	7	7	9	9	9		
3	0	0	0	1	2	2	2	3	4	4	5	6	7	7	8	8	9	9
4	1	1	2	3	4	5	6	7	7	9								
5	0	1	1	2	5	6	7	8										
6	0	0	0	1	4													

# 즐거-읽 그림

## ▶ 즐거-읽 그림

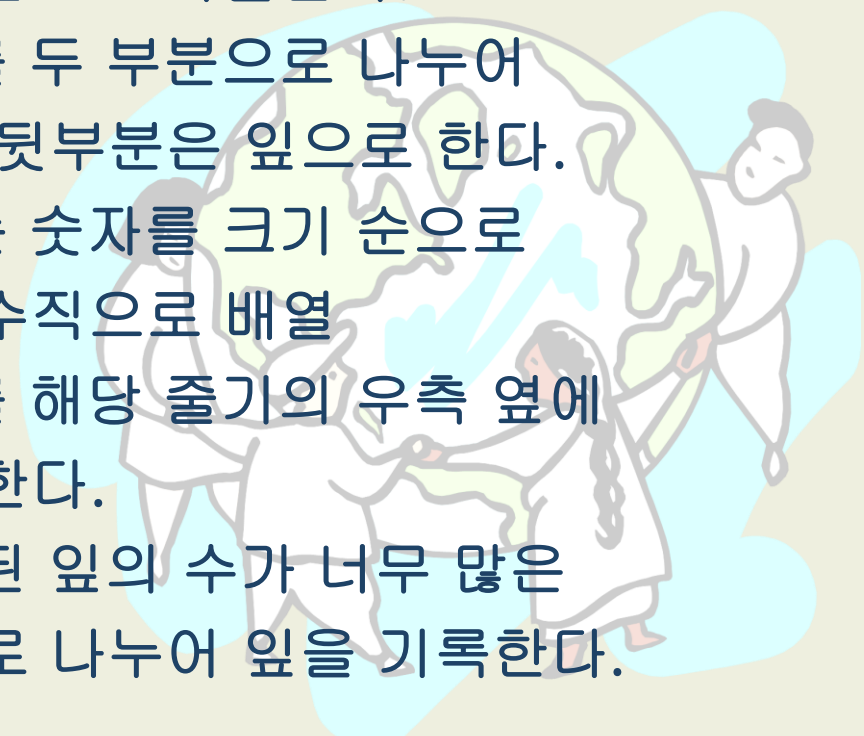
- ◆ 도수분포표와 막대그림표는 자료의 전체적 분포를 일목요연하게 보여주나, 원래의 자료가 나타나지 않기 때문에 정보의 손실이 발생하게 됨.
- ◆ 즐거-읽 그림은 원자료는 물론 도수분포표와 막대그림표의 장점을 하나의 도표에 담을 수 있는 방법임.



# 즐거-읽 그림

## ◆ 즐거-읽 그림 작성방법

- ◆ 단계 1 : 원자료를 크기 순으로 나열한다.
- ◆ 단계 2 : 원자료의 숫자를 두 부분으로 나누어 앞부분은 즐거, 뒷부분은 읽으로 한다.
- ◆ 단계 3 : 즐거에 해당하는 숫자를 크기 순으로 위에서 아래로 수직으로 배열
- ◆ 단계 4 : 원자료의 수치를 해당 즐거의 우측 옆에 앞 부분만 기록한다.
- ◆ 단계 5 : 한 즐거에 기록된 읽의 수가 너무 많은 경우에는 두 줄로 나누어 읽을 기록한다.



## 제 3장

# 기술통계 2: 요약특성치



고려대학교 경영대학 박 광태

# 중심경향도 (1)

## ▶ (산술)평균(mean)

### ◆ 모평균

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

### ◆ 표본평균

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

## 중심경향도 (2)

- ◆ 중앙치(median) :
  - ◆ 자료를 크기 순으로 배열했을 때 중앙에 있는 값
- ◆ 최빈치(mode) :
  - ◆ 가장 빈도가 높은 자료
- ◆ 1사분위수와 3사분위수의 위치

$$Q1의 위치 = \frac{[Me의 위치] + 1}{2}$$

$$Q3의 위치 = [Me의 위치] + \frac{[Me의 위치] + 1}{2}$$



# 중심경향도 (2)

## 예제 3-3(p.54)

### 예제 3-3

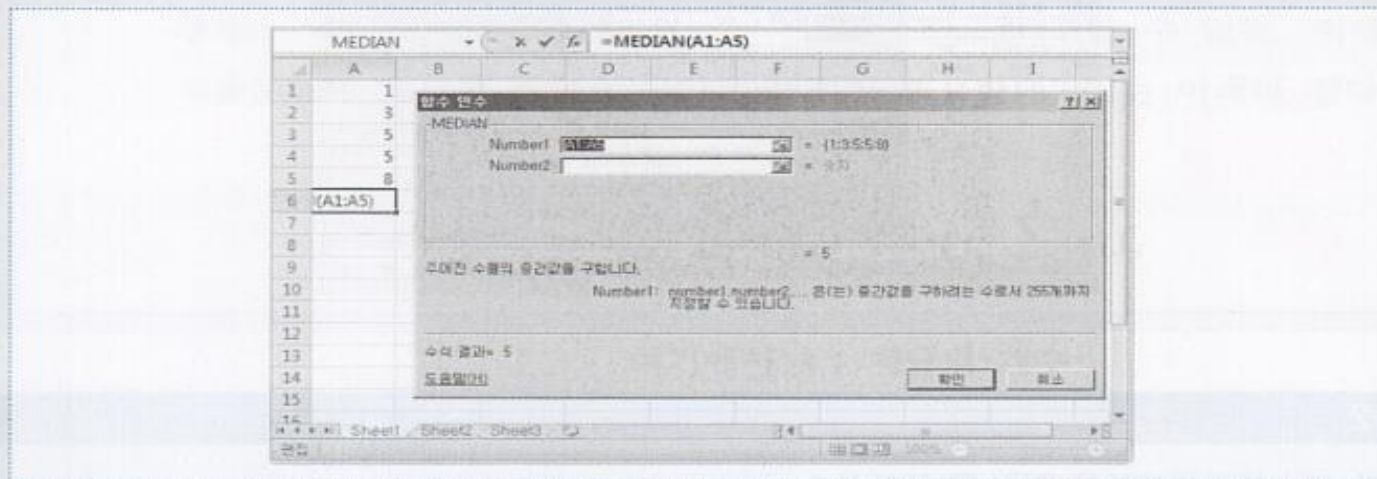
다음의 자료에 대한 중앙치를 구하시오.

(a) 1, 3, 5, 5, 8

(b) 1, 1, 2, 4, 5, 8

Excel 함수를 이용할 경우 (a)와 (b)에 대해 각각 그림 3-2와 같이 '수식-통계함수'에서 =MEDIAN(A1:A5), =MEDIAN(A1:A6)를 지정하면 각각 5와 3을 얻을 수 있다.

그림 3-1 엑셀함수를 이용한 산술평균 계산



# 중심경향도 (2)

## 예제 3-4(p.55)

### 예제 3-4

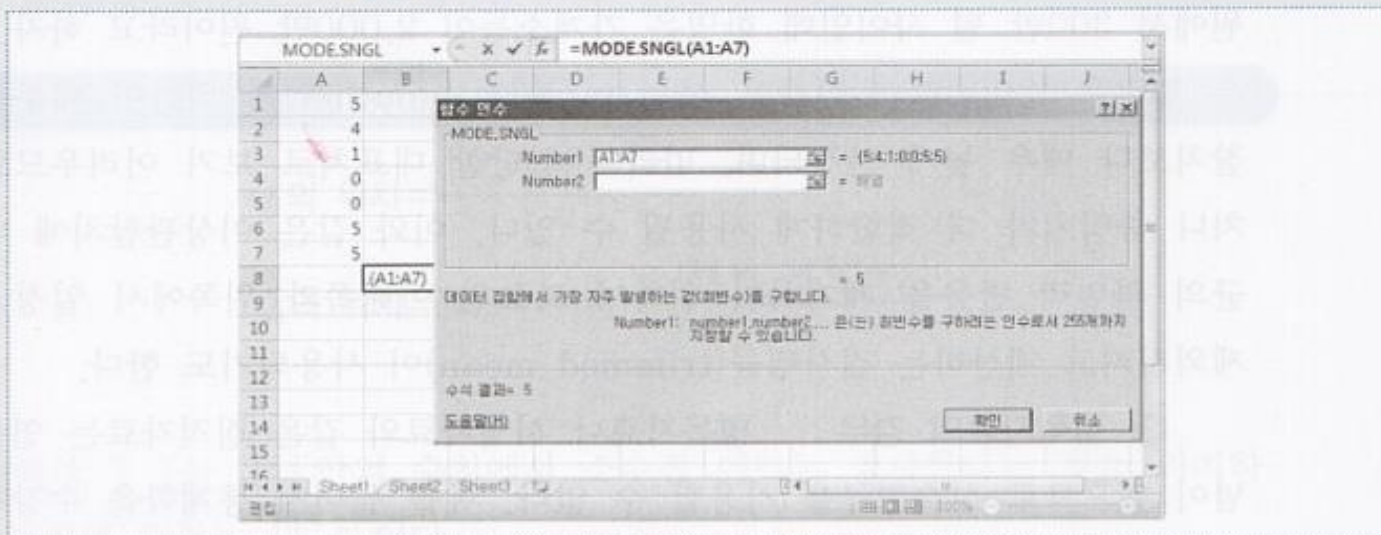
다음 자료의 최빈치를 구하시오.

(a) 5, 4, 1, 0, 0, 5, 5

(b) 0, 0, 4, 2, 5, 2

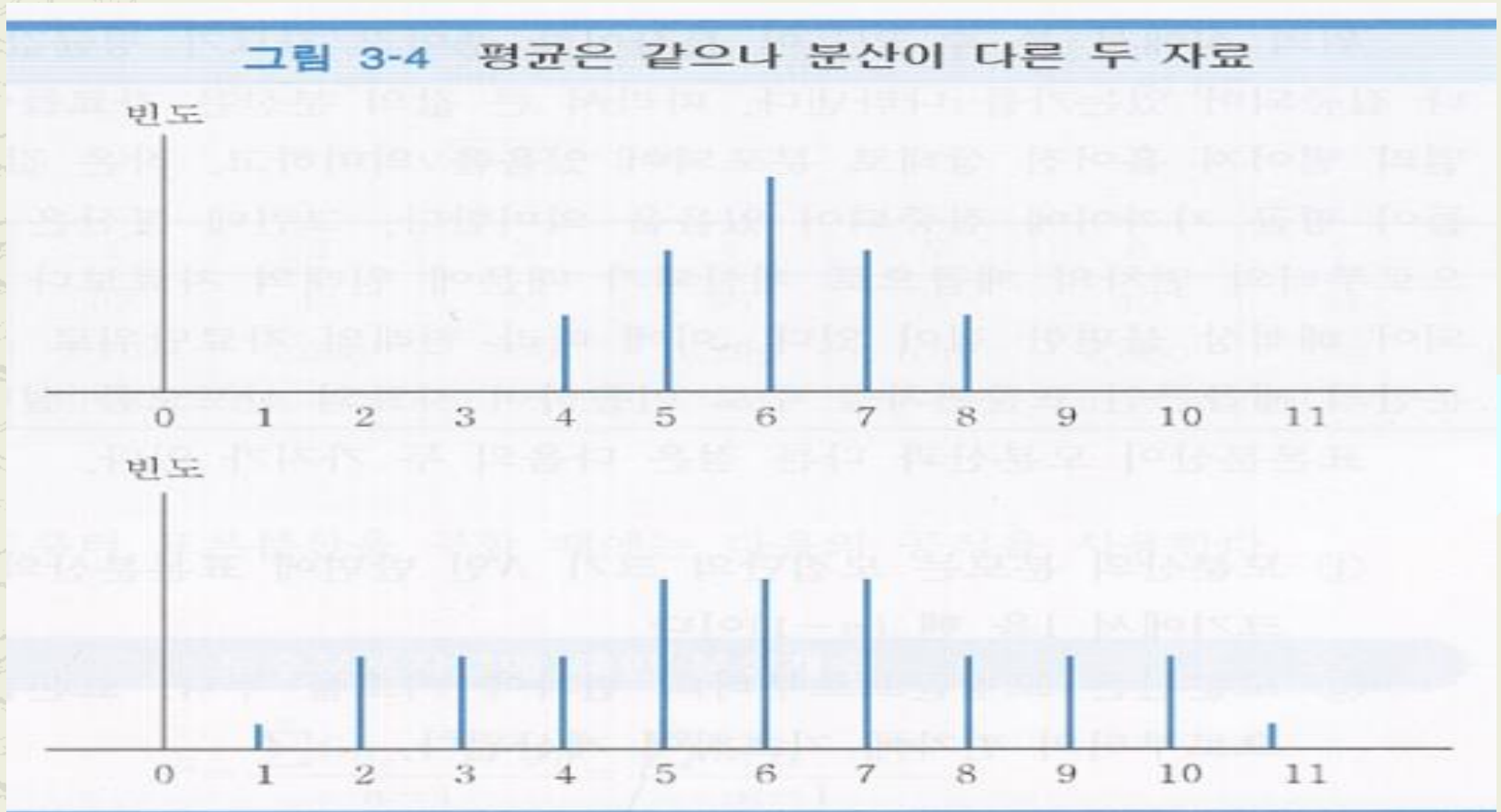
Excel 함수를 이용할 경우 그림 3-3과 같이 '수식-통계함수'에서 =MODE.SNGL(A1:A7)을 지정하면 (a)의 최빈치 5를 구할 수 있다.

그림 3-3 엑셀함수를 이용한 최빈치 계산



# 산포도

◆ 그림 3-4(p.59)



# 산포도 (1)

## ◆ 분산(variance)

### ◆ 모분산

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

### ◆ 표본분산

$$s^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$

# 산포도 (1-1)

## ▶ 예제 3-6(pp.60-61)

### 예제 3-6

한 도시에서 7일 동안의 교통사고 발생건수는 다음과 같다.

일 자	1	2	3	4	5	6	7
건 수	52	44	30	36	26	40	31

이 표본자료의 분산과 표준편차를 구하라.



# 산포도 (1-1)

▶ 예제 3-6(pp.60-61)



# 산포도 (2)

## 범위(range):

- ◆ 최대값과 최소값의 차이

## 평균절대편차(MAD)

$$MAD = \frac{\sum |x_i - \mu|}{N}$$

## 변동계수(CV):

- ◆ 단위가 다른 두 자료군의 산포도를 비교하는데 적합  
(상대적 표준편차라 불림)

예제) A회사: 주식가격 평균 50, 표준편차 4

B회사: 주식가격 평균 20, 표준편차 3

$$CV_A = 8\% < CV_B = 15\%$$

$$CV = \frac{\text{표준편차}}{\text{평균}} \times 100$$