

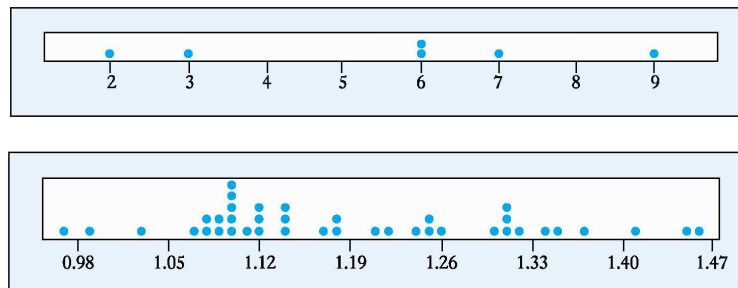
II. 기술통계학(descriptive Statistics)

도표/그래프	수리
<ul style="list-style-type: none"> - 점도표 / - 선그래프 - 막대그래프 - 줄기와 잎 그림 - 상자그림 - 도수분포표 - 히스토그램 - 도수분포다각형 - 누적도수분포다각형 	<ul style="list-style-type: none"> - 중심경향값 : 평균, 중앙값, 최빈값 - 산포도 : 범위, 분산, 표준편차 - 상대적 위치 : 사분위수, 백분위수, Z-값

1. 도표/그래프를 이용한 자료의 정리

1-1. 점도표(Dot Diagram)

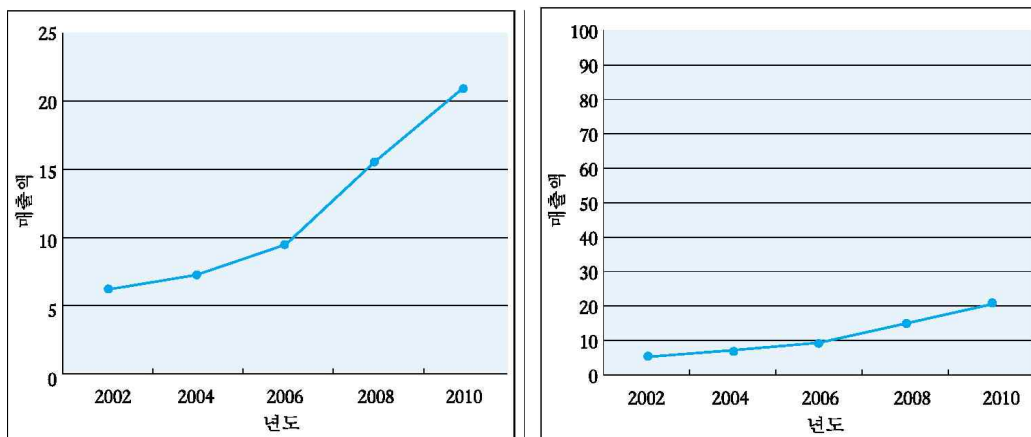
① 모든 자료를 수직선 위에 점으로 표시한 것



1-2. 선 그래프(Line Graph)

① 시계열 자료(시간간격으로 기록된 자료)를 정리

② 미래에 대하여 예측 가능



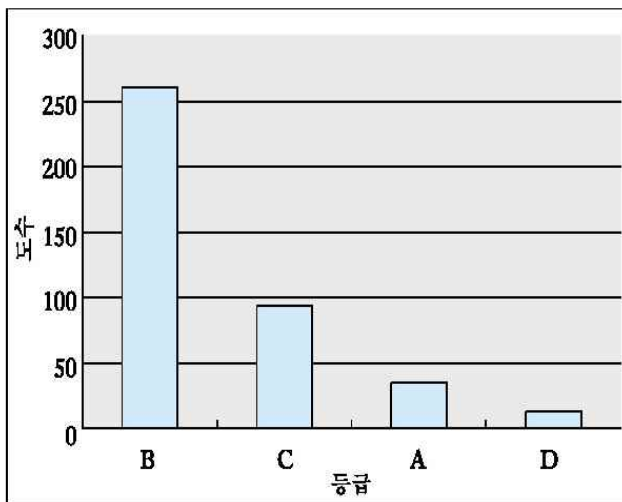
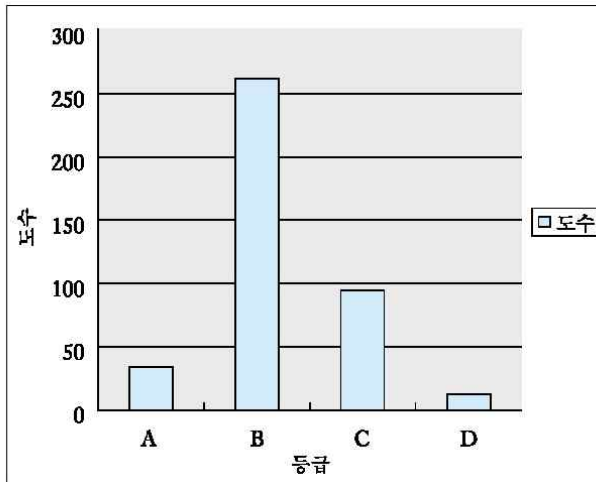
1-3. 막대그래프(Bar Graph)

- ① 범주형(질적) 자료의 도수분포 상황을 쉽게 알아보기 위하여 수직선 상의 상대도수와 같은 높이의 막대를 그린 그래프

- ② 점도표와 같이 자료의 양이 많지 않을 때 사용

【예】 공교육에 관한 조사에서 400명의 교장에게 교육의 질에 관하여 등급을 부여하게 하였다.

등급	도수
A	35
B	260
C	93
D	12
합계	400



참고 _____ (pareto graph)

: 가장 큰 값부터 순서대로 정렬한 막대그래프

1-4. 줄기와 잎그림

- ① 세로선의 왼쪽을 줄기, 오른쪽을 잎으로 하여 자료를 정리
- ② 자료의 값을 그대로 보존하여 자료의 분포 형태를 보여줌

【예】 A 동물원에 있는 동물의 평균 수명 정리표

물소	16	양	15	야생말	26	미니 나귀	20	염소	15
풍산개	15	토끼	6	개미핥기	10	박쥐	17	호랑이	25
돌고래	25	샴	21	하마	40	시타 통가	15	기린	30
고슴도치	8	망구스	8	하이 에나	25	표범	23	사랑새	18
코끼리	63	흑염소	12	타조	35	나무늘보	10	수달	21

동물의 평균 수명

줄기	잎
0	8 6 8
1	6 5 5 2 0 7 5 0 5 8
2	5 1 6 5 0 3 5 1
3	5 0
4	0
5	
6	3

- ❶ 십의 자리의 수를 작은 수부터 크기순으로 세로로 나열한다.
- ❷ ❶에서 나열한 수의 오른쪽에 세로선을 긋는다.
- ❸ 각 십의 자리의 수에 해당하는 일의 자리의 수를 세로선의 오른쪽에 순서대로 가로로 나열한다.

1-5. 상자그림

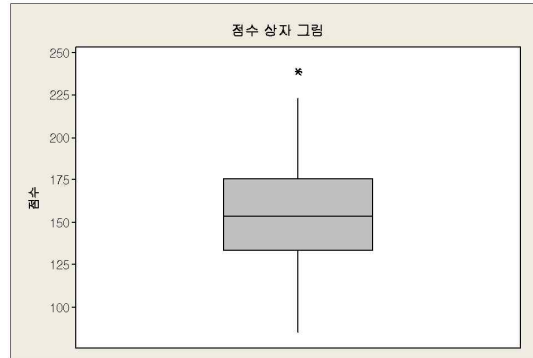
- ① 자료의 분포에 대한 정보를 사분위수를 이용하여 나타내 주는 그림
- ② 수염(whisker), 이상점(outlier)

【예】 80명 학생의
한 학기 시험
총점 자료

$$Q_3 = 175.50$$

$$\text{중앙값} = 153.5$$

$$Q_1 = 133.25$$



1-6. 도수분포표

- ① 자료를 몇 개의 계급으로 나누고 각 계급의 도수를 조사하여 나타낸 표
- ② 관련 용어

㉠ 변량 : 어떤 집단의 특성을 수량화한 것

㉡ 계급 : 변량을 일정한 간격으로 나눈 구간

㉢ 계급의 크기 : 구간의 너비

㉣ 계급값 : 각 계급을 대표하는 값, 그 계급의 가운데 값

$$(\text{계급값}) = \frac{(\text{계급의 양 끝값의 합})}{2}$$

㉤ 도수 : 각 계급에 속하는 자료의 수

㉥ 계급구간 : 각 계급의 폭, 두 개의 연속된 계급상한들의 크기

$$(\text{계급구간}) = \frac{(\text{자료의 최댓값} - \text{자료의 최솟값})}{\text{계급의 수}}$$

㉦ 계급의 수 : 보통 6~20개 정도가 바람직

【예】 학생 40명의 한 달 동안 용돈 지출 금액(만원)

33	8	11	22	12	9	8	33	32	30	22	12	24	
21	26	12	15	37	30	16	18	9	28	18	25	37	
16	25	28	25	16	24	31	25	28	26	15	12	35	38

1단계 : 범위(R)=(최댓값-최솟값) 계산

2단계 : 계급의 수 결정: 스타제스에 의한 방법

$$K = 1 + \frac{\log n}{\log 2} \quad (\text{단, } K \text{는 계급의 수, } n \text{은 전체도수})$$

3단계 : 계급구간 결정 $C = \frac{R}{K}$

4단계 : 도수 산정

1단계 : $R=38-8=30$

$$\begin{aligned} 2\text{단계} : K &= 1 + \frac{\log 40}{\log 2} = 1 + \frac{\log 4 + \log 10}{\log 2} \\ &= 1 + \frac{0.6021 + 1}{0.3010} = 1 + 5.32 \\ &= 6.32 \\ &\rightarrow 7 \end{aligned}$$

$$3\text{단계} : C = \frac{30}{7} = 4.29 \rightarrow 5$$

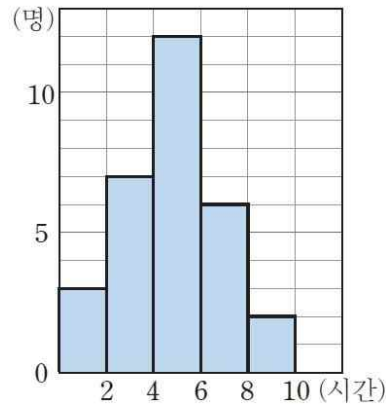
각 계급은 5~9, 10~14, 15~19, 20~24, 25~29, 30~34, 35~39

계급	도수
5~9	4
10~14	5
15~19	7
20~24	5
25~29	9
30~34	6
34~39	4
합 계	40

1-7. 히스토그램(Histogram)

- ① 변수가 양적이고 연속변수일 때, 직교좌표의 횡축 상에 도수분포표의 각 계급을 나타내는 구간을 밀변으로 하고 계급에 대응하는 도수를 높이로 하는 직사각형을 모든 계급에 대해 만들어 연결한 그래프
- ② 각 직사각형에서 가로의 길이인 계급의 크기가 일정하므로 직사각형의 넓이는 세로의 길이인 각 계급의 도수에 정비례
- ③ 두 집단 이상을 비교하는 것이 용이하지 않음.

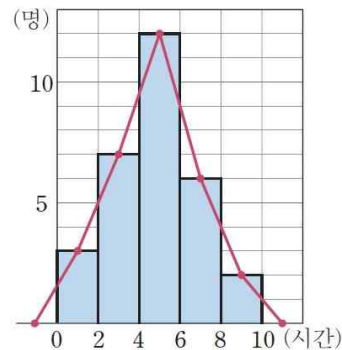
- ❶ 가로축에 각 계급의 양 끝 값을 써넣는다.
- ❷ 세로축에 도수를 써넣는다.
- ❸ 각 계급을 가로로 하고, 도수를 세로로 하는 직사각형을 차례대로 그린다.



1-8. 도수분포다각형

① 2개 이상의 자료의 분포 상태를 동시에 나타내어 비교하는 데 편리

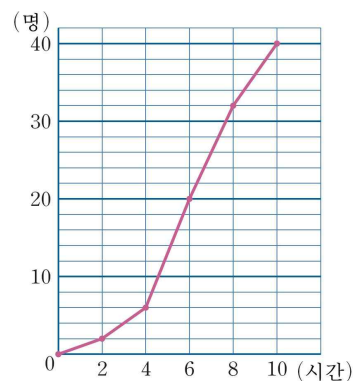
- ❶ 히스토그램의 각 직사각형에서
윗변의 중앙에 점을 찍는다.
- ❷ 히스토그램의 양 끝에 도수가
0이고, 크기가 같은 계급이
하나씩 더 있는 것으로 생각하고
그 중앙에 점을 찍는다.
- ❸ ❶과 ❷에서 찍은 점을 차례대로
선분으로 연결한다.



1-9. 누적도수분포다각형

① 특정한 어떤 점수까지 누적된 도수나 백분율을 알고자 할 때

- ❶ 가로축에 각 계급의 양 끝값을,
세로축에 누적도수를 써넣는다.
- ❷ 처음 계급의 작은 쪽의 끝값과
0을 순서쌍으로 하는 점을
찍는다.
- ❸ 각 계급의 큰 쪽의 끝값과
그 계급의 누적도수를
순서쌍으로 하는 점을 찍는다.
- ❹ 각 점을 차례대로 선분으로
연결한다.



【참고】 그래프 사용과 해석 시 유의점

- ① 그래프 도식의 일반 규칙인 세로축이 가로축 길이의 $2/3 \sim 3/4$ 이 되게 하였는지 확인한다.
- ② 막대그래프나 히스토그램의 경우 막대의 폭이 동일한지 확인한다.
- ③ 그래프가 시각적으로 집단 간의 차이의 유무와 변화의 폭을 급변동 혹은 완만하게 그렸을 때 수적 차이와 수적 변동 폭을 고려하여야 한다.