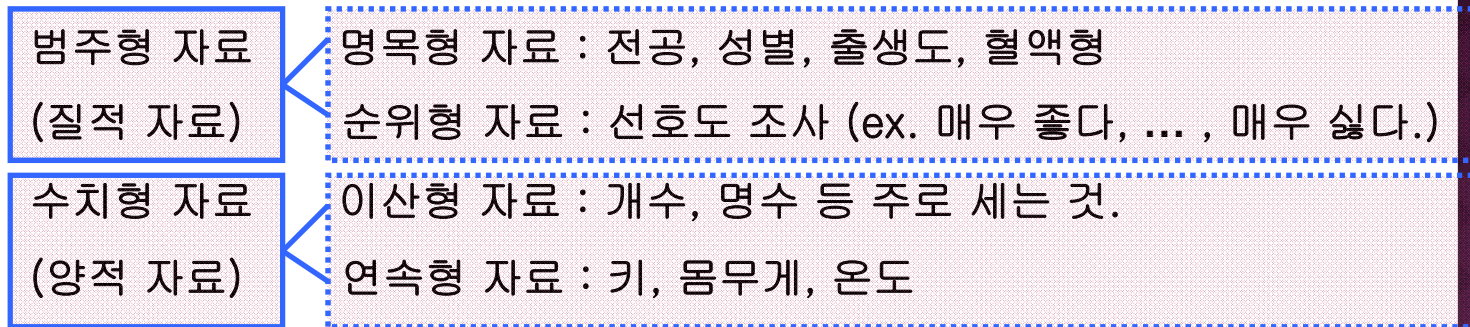


확률 및 통계

제2주 자료의 종류, 자료의 표현

hylee@silla.ac.kr

자료의 종류



❖ 변수 : 관심의 대상이 되는 특성.

(변수의 종류는 자료의 종류에 대응된다.)

범주형 자료의 도표화 - 1

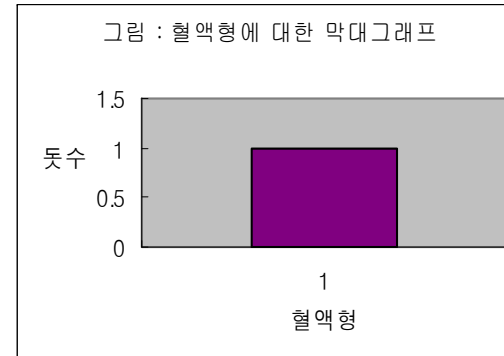
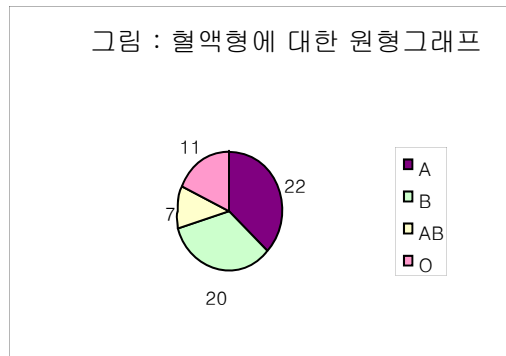
❖ **덧수 분포표** : 범주, 덧수, 상대덧수

❖ **원형 그래프 (Pie Chart)**

: 원의 중심각을 상대 덧수에 비례하도록 나누어 그린 그림.

❖ **막대 그래프 (Bar Chart)**

: 막대의 높이가 상대덧수 (혹은 덧수)에 비례하도록 그린 그림이다.



*(비교) 각 범주간의 덧수를 비교하는 경우 → **막대 그래프**

전체적인 분포에 관심이 있는 경우 → **원형 그래프**

범주형 자료의 도표화 - 2

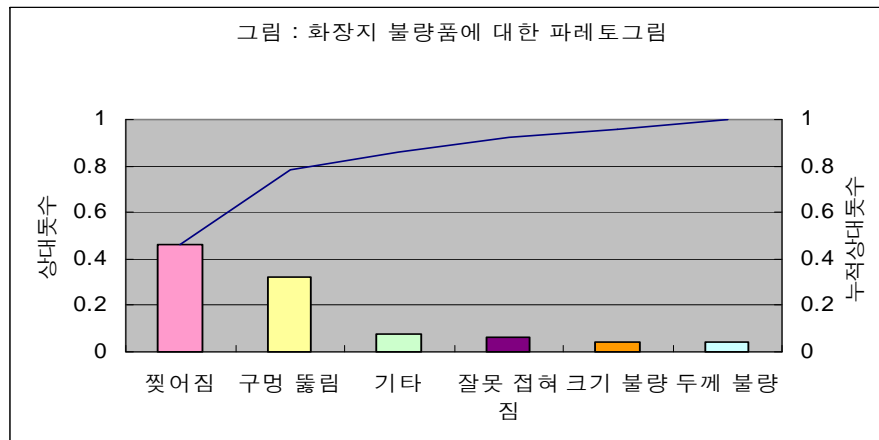
❖ 파레토 그림 (Pareto Diagram)

: 전체 부의 80%를 대략 20%의 사람이 소유

20% 사람 : 중요한 소수 (vital few)

80% 사람 : 사소한 다수 (trivial many)

- 명목형 자료에서 ‘중요한 소수’를 찾는 데 파레토 그림이 유용하게 쓰인다.



예제

이산형 자료의 도표화

- 취하는 값의 개수가 적은 경우 : 범주형 자료의 방법 이용
- 취하는 값의 개수가 많은 경우 : 연속형 자료의 방법 이용

연속형 자료의 도표화 - 1

- ◎ 점도표 (dot diagram) : 표본의 크기가 작은 경우에 유용, 정보 손실이 없는 방법

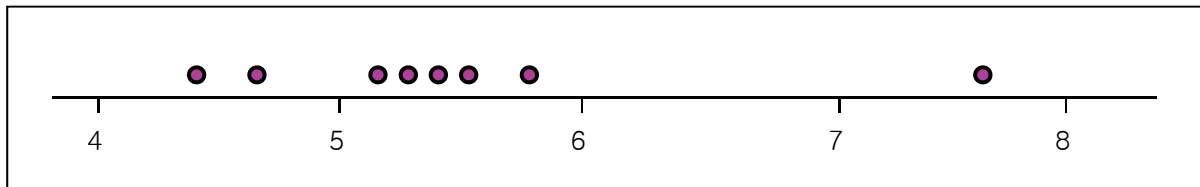


그림 : GNP에 대한 교육비 지출 비율

연속형 자료의 도표화 - 2

◎ 숫수분포표

- 범위 : 최대값 - 최소값
- 계급의 개수 : 5-15 ; 전체적인 분포를 얻을 수 있도록 정한다.
- 계급구간의 폭 : 범위 ÷ (계급의 숫수)의 값보다 약간 크게 정한다.

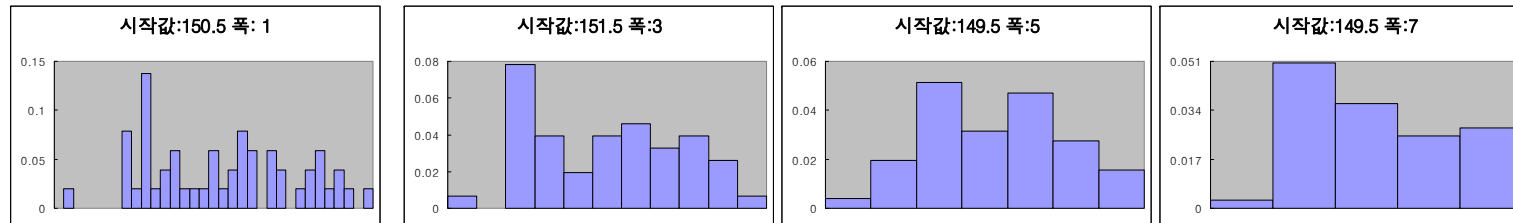


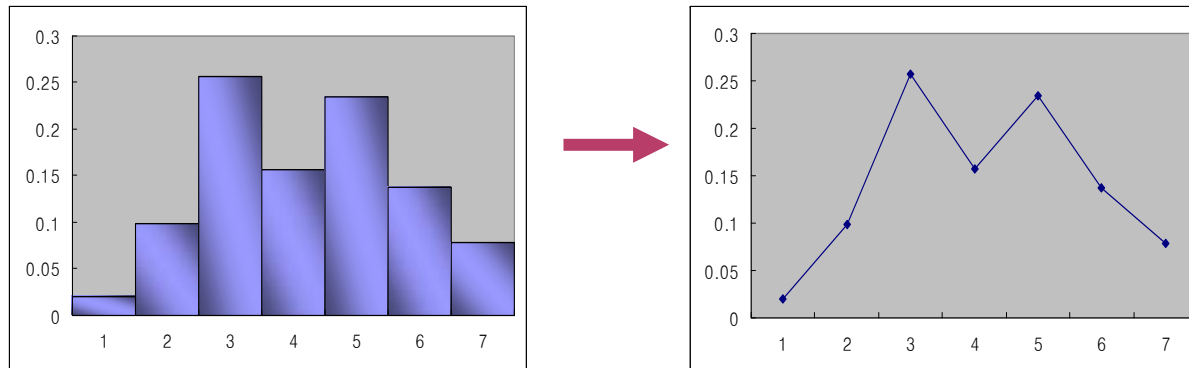
그림 : 계급구간의 폭과 계급구간의 시작값의 변화에 따른 히스토그램의 변화

- ◎ 히스토그램 (Histogram) : 숫수분포표에 의하여 막대그래프와 같은 형태로 그리되 전체 막대의 넓이의 합이 1이 되도록 그린다.
- 계급구간의 막대의 높이 = 상대숫수 ÷ 계급구간의 폭

연속형 자료의 도표화 - 3

◎ 돛수다각형 (Frequency Polygon)

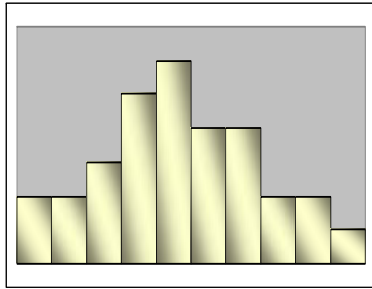
- 히스토그램에서 각 계급구간의 막대 상단의 중앙점을 직선으로 연결하여 그린다.
- 장점: 20이상의 분포를 비교하는 데 유용하다.



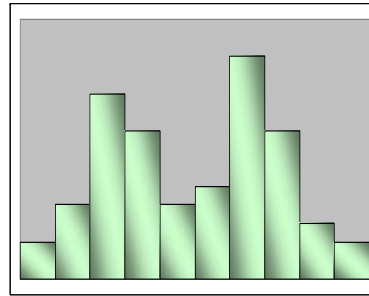
◎ 줄기-잎 그림 (Stem-leaf plot)

- 자료가 갖는 정보를 잃어버리지 않고, 그 분포를 비교적 쉽게 파악하는 방법
- 주로 2자리 숫자로 이루어진 자료에 많이 쓰인다.

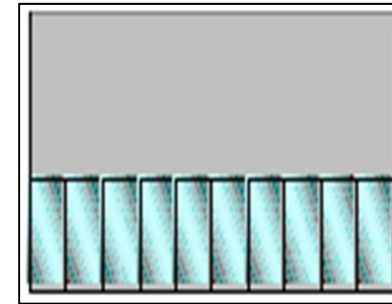
분포의 모양



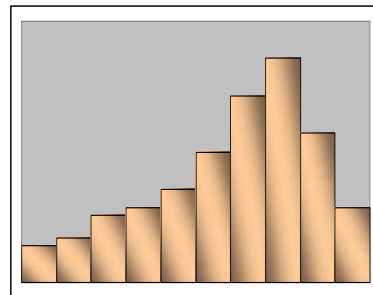
(a) 종모양



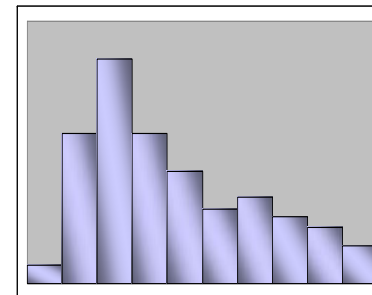
(b) 이봉형



(c) 균일형



(d) 오른쪽으로 편중



(e) 왼쪽으로 편중

Thank You!

