

1장 정보의 시각화

P. 42

1. 일상에서의 통계

- 예제 : 자연과 농의 통신에 자료비역을 비교하여 경위에 따라 여름방학 알바를 하지 않아도 된다는 결론을 유도해 볼 수 있음.
- 예제 : 소득통계 관련 통계청의 가계동향조사 결과를 어떻게 받아들여야 하는가에 따라 사회와 정치에 대한 이해가 달라짐.

* 통계란? : 주어진 데이터에서 정보를 추출하는 것.

- 데이터 : 수집된 사실, 숫자, 기록 등등
- 정보 : 특정 의미가 부여된 데이터

P. 43

2. 통계 학습 목적

- 통계 산출방법 학습
- 통계 활용방법 학습

* 통계를 배워야 하는 이유

- 세상에서 벌어지는 일들을 이해하고 설명하기 위해.
- 사실관계를 정확하게 판단하기 위해.
- 통계지식을 전공영역에서 활용하기 위해.

3. 정보전달 방법 선택

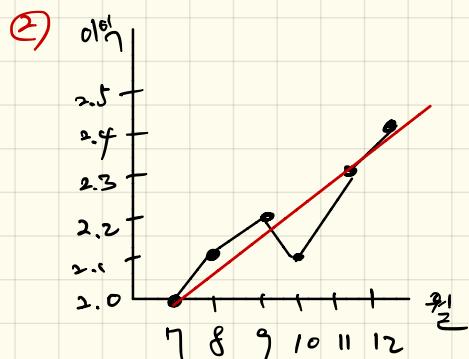
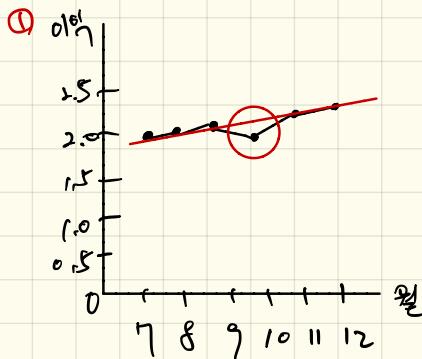
- 동일한 데이터에서 다른 정보, >경우에 따라서는 심지어 서로 상반되는 정보를 만들어낼 수 있다.
- 예제 : 어떤 회사의 월별 실적
(단위: 억 원)

월	7	8	9	10	11	12
이익	2.0	2.1	2.2	2.1	2.3	2.4

이런 정보를 확인할 수 있나? 아마도 아래와 같이
쉽게 가능성이 높다.

“월별 이익이 완만하지만 꾸준히 증가한다.”

정말 이것 봄일까? 아래 두 그래프를 비교해 보자.



회사의 CEO라면 경위에 따라 두 그래프 중
하나를 선택해서 사용할 것이다.

경우 1 : 회사가 꾸준히 성장하고 있음을
증명하고 싶을 때 그래프 ① 사용

경우 2 : 회사의 실적을 과장하여 홍여주고자
할 때 그래프 ② 사용

4. 다른 내용 : 정로 시작화 4방향 네 가지

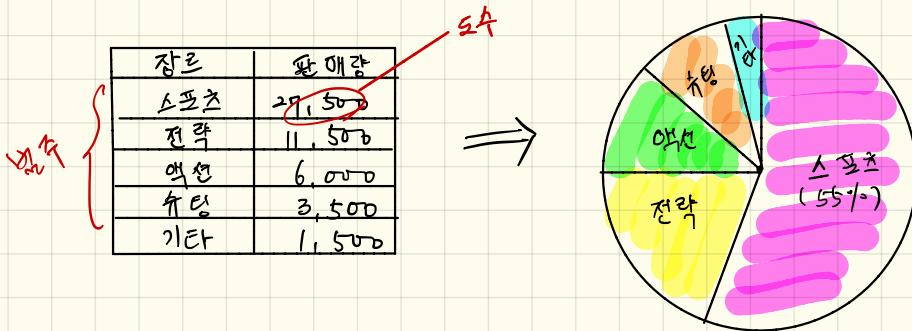
- ① 파이차트 (원그래프)
- ② 막대그래프
- ③ 히스토그램
- ④ (꺽은) 선그래프

* 알아두어야 할 개념 두 가지

- ① 별주 : 동일한 성질을 갖는 부류 또는 범위
- ② 도수 : 특정 벌주에 속하는 개체의 개수

5. 파이차트

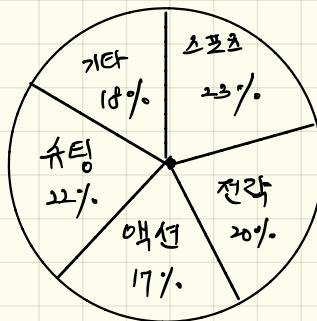
예제 어떤 기업에서 판매한 제품의 장르별 판매량



• 파이차트의 장단점

장점	단점
각 제품이 차지하는 비율을 한눈에 볼 수 있게 도와줌.	항목별 축자들의 차이가 크지 않을 때 별 의미 없음.

예제: 범주별 차이가 크지 않으면 파이차트에서 각 범주별로 차지하는 영역의 크기가 잘 구분되지 않음.



6. 막대그래프

p.50

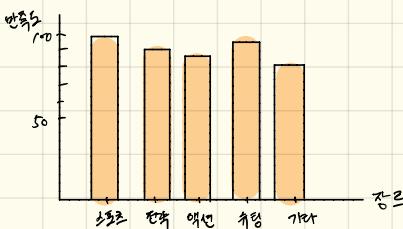
예제 : 게임 장르별 소비자 만족도

장르	만족도(%)
스포츠	99
전략	90
액션	85
슈팅	95
기타	80

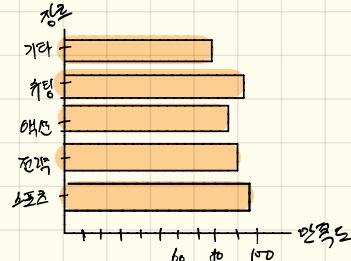
주의 사항:

- 파이차트 사용 적절치 않음: 범주별로 차지하는 상대적 비율을 비교하는 것이 아니기 때문
- 막대그래프 활용이 넓다 적절함
- 막대그래프: 헤치아 도수(또는 퍼센트 등)를 x축과 y축에 지정하고 범주별로 도수(또는 퍼센트 등) 크기만큼 막대를 그리는 그래프

주의 사항: 막대들의 폭이 일정해야 함



(수직) 막대그래프



(수평) 막대그래프

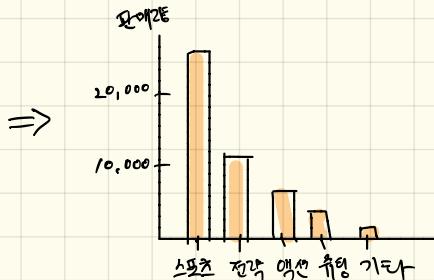
• 막대 그래프의 장점

i) 범주별 상대적 비율이 아니라 범주별 미세한 차이를 명확히 전달할 수 있음.

ii) 여러 종류의 데이터를 함께 나타내어 비교할 수 있도록 도와줄 수 있음.

iii) 예제 : 사이차트를 막대그래프로 나타낼 수 있음.

장르	판매량
스포츠	27,500
전략	11,500
액션	6,000
슈팅	3,500
기타	1,500



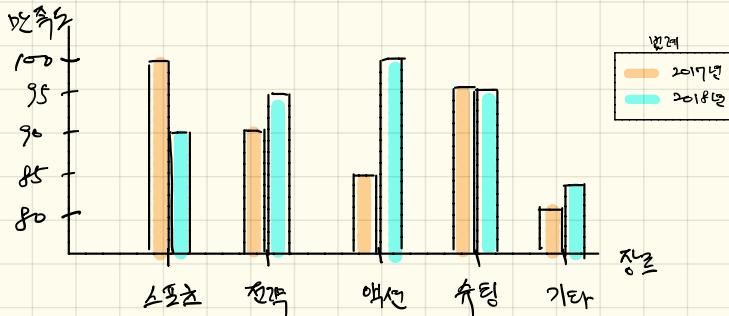
ii) 예제

장르	판속도(%)
스포츠	99
전략	90
액션	85
슈팅	95
기타	80

(2017년)

장르	판속도(%)
스포츠	90
전략	95
액션	99
슈팅	95
기타	85

(2018년)



7. 히스토그램

p.58

- 데이터 종류 두 가지

i) 범주적 데이터 (질적 데이터) :

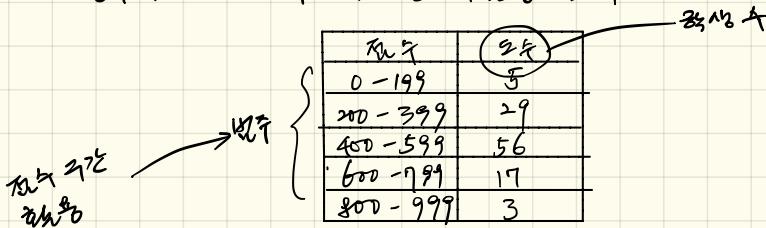
- 범주의 성질이나 특성에 따라 나뉘어진 데이터
- 히스토그램으로 나타낼 수 있음.

ii) 수치적 데이터 (양적 데이터) :

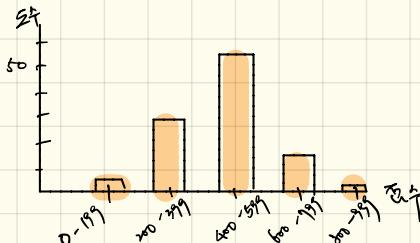
- 나이, 길이, 시간, 천수 등 숫자의 범위에 따라 나뉘어진 데이터

p.61

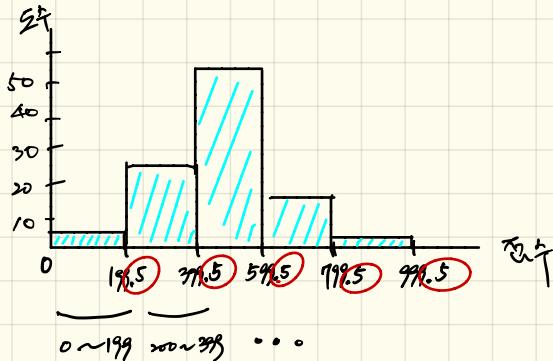
- 예제 : 시험점수 구간별 수험생 수자



- 막대그래프 활용은 주제가 맞음 : 구간별 연속성이 제대로 반영되지 못함



- 구간의 연속성을 반영하기 위해 리스트그래프를 사용할 수 있음



주의사항 : 구간 구분을 199, 399, ... 등 대신에
199.5, 399.5, ... 등으로 하는 이유는
모든 정수가 하나의 구조에 포함되도록
하기 위함임.

* 도수밀도

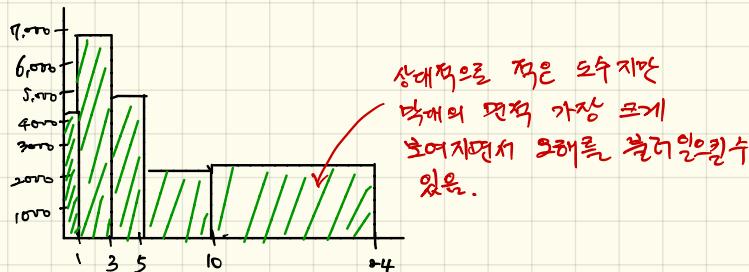
- 리스트그래프에 사용되는 막대들의 폭은 적절하게 지정하면 된다.
- 하지만 구간별 크기가 다른 경우 막대의 높이는 도수의 크기에 비례하도록 해야 한다.
이를 위해 도수밀도를 계산하여 막대의 높이로 사용해야 함

$$\text{도수밀도} = \frac{\text{도수}}{\text{구간의 크기}} = \frac{\text{도수}}{\text{막대의 폭}}$$
$$= \text{막대 높이}$$

예제: 사설들의 하루 개인 시간

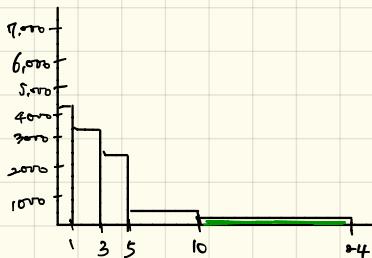
제일 시간	오늘
0 - 1	4,300
1 - 3	6,900
3 - 5	4,900
5 - 10	2,000
10 - 24	2,100

- 도수밀도를 사용하지 않은 히스토그램



- 해설: 도수밀도를 막대의 높이로 사용해야 함.
즉, 막대의 넓이가 도수만 있자해야 함.

제일 시간	오늘	구간 크기	도수밀도
0 - 1	4,300	1	4,300
1 - 3	6,900	2	3,450
3 - 5	4,900	2	2,450
5 - 10	2,000	5	400
10 - 24	2,100	10	150



8. 선그래프

- 막대그래프의 상단 중심부를 선으로 연결하여 액 벤주를 비교하는 그래프

- 특징**

(i) 시계열 데이터의 변화정도를 전달할 때 유용함.

(ii) 주소별 도수의 특별한 규칙도는 선그래프를 활용해야 정확한 정보를 전달할 수 있음.

(iii) 복수의 데이터들의 변화를 한 눈에 전달하고자 할 때 막대우리프보다 효과적임.

- 주요개념 : 시계열 데이터**

- 주가, 지지도, 센서데이터, 기록 추세 데이터 등 시스템으로 구성된 데이터
- 선그래프의 일종인 시계열 그래프에 적합한 데이터

- 선그래프 예제 ① : 시계열 그래프**

- 시계열 데이터를 분석하기 위한 그림
- 예제 : 실질 국내총생산(GDP) 성장을 추이

2010년 연세기각기준, 계절조정계열 전기 대비 증감률



*주의 : 교재와 다른 내용과 차이.
예제만 다른.

9.74

• ② 누적도수 그래프

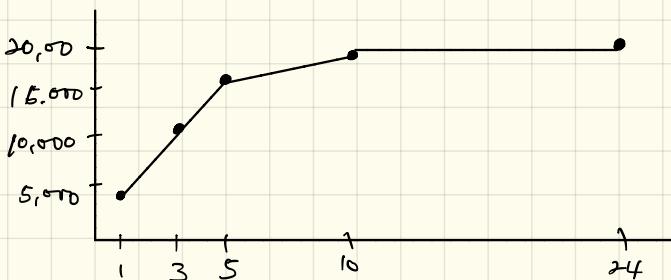
- 누적도수 현행을 전달할 때 유용함.

1회	도수	상한시간	누적도수
0-1	4,300	1	4,300
1-3	6,900	3	11,200
3-5	4,900	5	16,100
5-10	2,000	10	18,100
10-24	2,100	24	20,200

질문 : 최대 3시간 기한하는 사람의 수?

$$\text{답} : 4,300 + 6,900 = 11,200 \text{ 명}$$

\Rightarrow 아래 모양의 누적도수 그래프를 이용하여
누적도수의 차이를 한 눈에 확인 가능

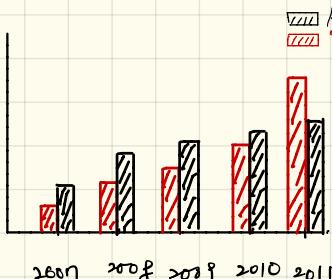


- 주의사항 : 막대그래프 사용은 적절하지 못함.

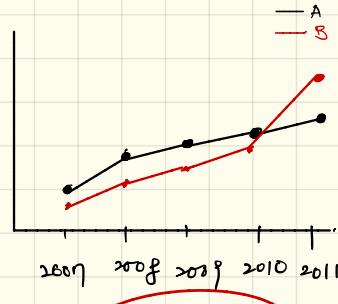
▶ 79

- 선 그래프 사용 예제 ③

- 복수의 시계열 데이터 비교
- 막대그래프처럼 선그래프를 이용하여 복수의 데이터를 하나의 그래프로 나타낼 수 있음.
- 예제: A, B 두 회사의 연도별 수익 비교



(막대 그래프)



(선 그래프)

B 회사의 성장을 강조하고자 하면 선 그래프를 사용하는 것이 보다 효과적일 수 있음.