

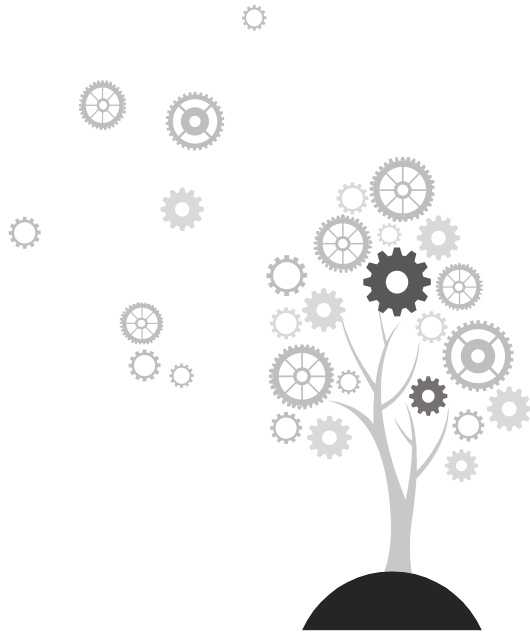
Contents

Chapter 1 서론

1.1 통계학이란 무엇인가?

1.2 모집단과 표본

1.3 통계학의 목표



01 통계학이란 무엇인가?

통계학은 자료의 수집과정을 설계하고, 자료를 요약하고 해석하여 결론을 끌어내거나 일반화하는 전체적인 원리와 방법론을 제공하여 준다.

02 모집단과 표본

← 원수조사를 보통 하지 않기 때문에,
보통 표본 자료를 활용하여 통계를 실시한다.

- 모집단 (Population)

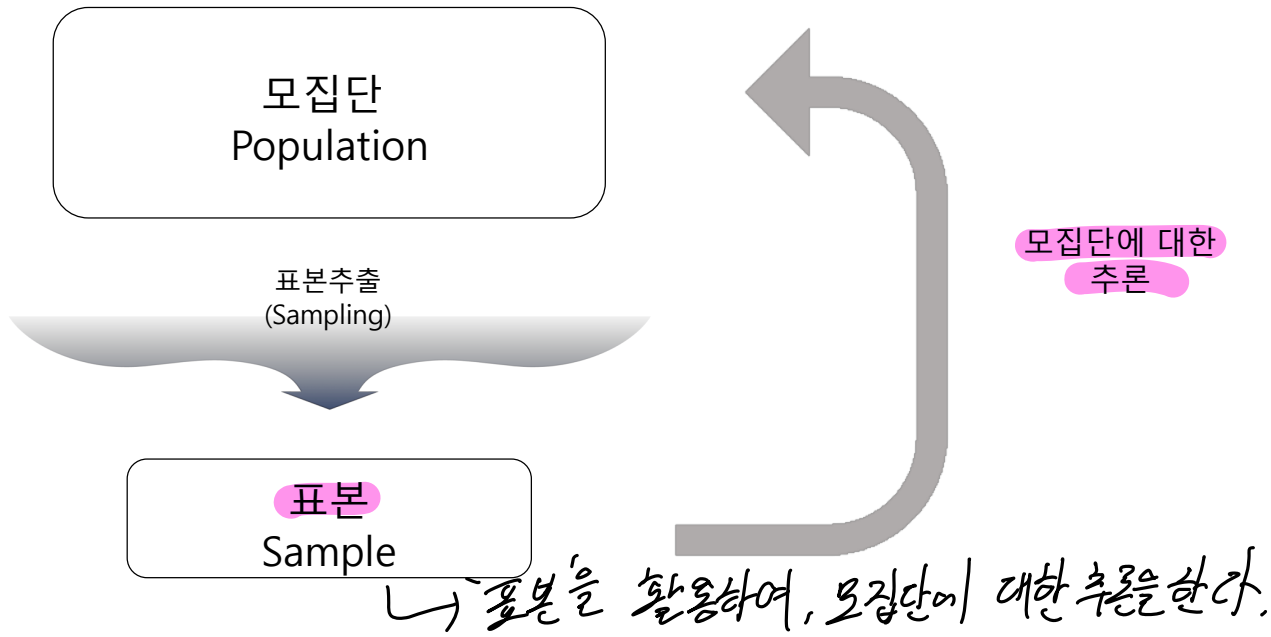
: 얻고자 하는 정보와 관련 있는 모든 개체로 부터 얻을 수 있는 모든 관측값들의 집합

- 표본 (Sample)

: 모집단의 일부분, 원하는 정보를 얻기 위해 수행한 관측과정을 통하여 실제로 얻어진 관측결과의 집합

Chapter 1 서론

02 모집단과 표본



Chapter 1 서론

03 통계학의 목표

1. 관측값으로부터 효율적인 추론을 할 수 있도록 표본추출의 과정과 범위를 설계한다.

↓ 표본 자료를 분석하여

표본에 포함되어 있는 정보를 분석하여 모집단에 관하여 추론을 한다. 이때 추론에 수반되는 불확실성도 측정한다.

↑ 통계의 목표.

※ 자료 = 관측값

Contents

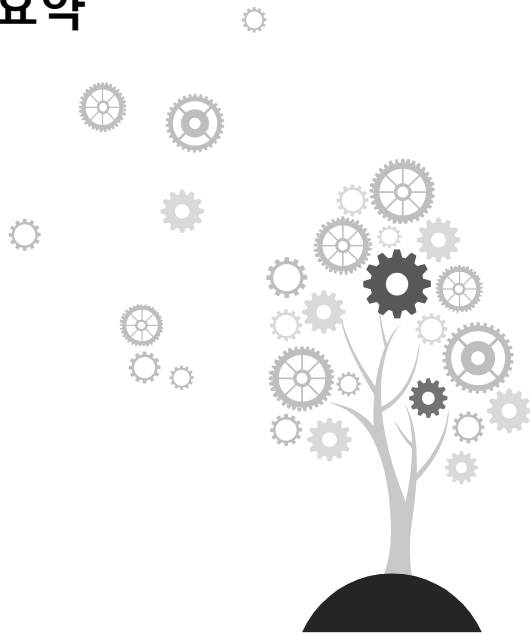
Chapter 2 표와 그림을 통한 자료의 요약

2.2 자료의 형태

2.3 범주형 자료의 요약

2.4 이산형 자료의 요약

2.5 연속형 자료의 요약



Chapter 2 표와 그림을 통한 자료의 요약

01 자료의 형태

- 수치자료(numerical data)

: 관측값이 수치로 측정되는 자료 (= 양적자료(quantitative data))

→ 관측된 수치 자료가 될 수 있는 경우

① 연속형자료 : 관측가능한 값이 연속적인 경우 ex) 키, 몸무게, ...

② 이산형자료 : 관측가능한 값이 셀 수 있는 경우 ex) 자동차 사고 건수, 불량품의 개수, ...

→ 관측된 수치 자료가 될 수 없는 경우.

- 범주형자료(categorical data)

: 관측결과가 몇 개의 범주 또는 항목의 형태로 나타나는 자료 (= 질적자료(qualitative data))

① 순위형자료(ordinal data) : 범주 순서에 의미가 있는 경우 ex) 선호도

② 명목형자료(nominal data) : 범주 순서에 의미가 없는 경우 ex) 성별, 혈액형, 지역, ... , 학번

Chapter 2 표와 그림을 통한 자료의 요약

01 자료의 형태

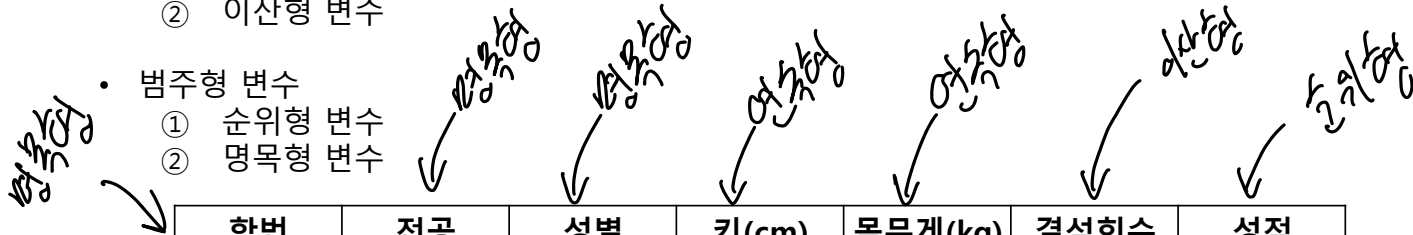
- 변수 (variable) : 관측되어지는 특성

- 수치형 변수

- ① 연속형 변수
- ② 이산형 변수

- 범주형 변수

- ① 순위형 변수
- ② 명목형 변수



학번	전공	성별	키(cm)	몸무게(kg)	결석회수	성적
1214xxxx	통계학과	남	178.4	xx.xxx	2	A+
1215xxxx	경영학과	남	180.2	xx.xxx	1	B0
1213xxxx	교육학	여	163.1	xx.xxx	3	C+
...						

Chapter 2 표와 그림을 통한 자료의 요약

02 범주형 자료의 요약

범주형 자료를 주로 요약할 때,
사용하는 것.

- 도수분포표(frequency table)

- ① 도수 : 각 범주에 속하는 관측값의 개수
- ② 상대도수 : 도수를 자료의 전체 개수로 나눈 비율
- ③ 누적상대도수 : 배열된 각 범주에서 그 범주를 포함한 이 전 범주들의 상대도수들의 합
- ④ 도수분포표 : 각 범주와 그 범주에 대응하는 도수와 (누적)상대도수를 나열한 표

예제 1 한 회사에서 새로 개발한 자동차의 외형에 대하여 고객 150명을 임의로 뽑아 선호도를 조사하였다. 150명 중에서 71명은 좋다고 답하고, 42명은 그저 그렇다고 답하고, 28명은 싫다고 답하고, 9명은 답을 하지 않았다. 조사에 대한 도수분포표를 작성하라.

범주	도수	상대도수	누적상대도수
합		/	

Chapter 2 표와 그림을 통한 자료의 요약

02 범주형 자료의 요약

범주형 자료를 원형 그래프로 요약할 때, 사용하는 것.

- 원형그래프(Pie chart)

- ① 원을 상대도수에 비례하여 중심각을 나누어 마치 파이의 조각을 나눈 것 같은 형태를 가진 그림
- ② 각 범주나 일부의 범주가 전체에서 차지하는 비율 파악에 용이
- ③ 수치적인 내용파악이 어려워 추가로 기입하기도 함

예제 2 어느 고등학교에서 한 학급을 임의로 선택하여 그 학급 60명에 대한 혈액형 검사를 실시한 결과가 다음과 같다. 이 자료에 대한 원형그래프를 그려라

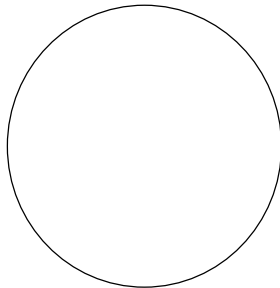
B	A	O	O	A	B	A	B	AB	O	B	A	A	B	A
A	O	A	B	A	O	AB	A	A	A	A	AB	B	B	B
AB	B	A	AB	A	B	B	O	B	O	B	A	O	B	O
AB	O	B	A	AB	A	B	A	B	A	B	O	B	A	A

Chapter 2 표와 그림을 통한 자료의 요약

02 범주형 자료의 요약

예제 2 어느 고등학교에서 한 학급을 임의로 선택하여 그 학급 60명에 대한 혈액형 검사를 실시한 결과가 다음과 같다. 이 자료에 대한 원형그래프를 그려라

범주	도수	상대도수
A		
B		
AB		
O		
합		



Chapter 2 표와 그림을 통한 자료의 요약

02 범주형 자료의 요약

- 막대그래프(Bar chart)

- ① 각 범주에서 도수 (상대도수)의 크기를 막대로 그려 나타냄
- ② 도수를 그리거나 상대도수를 그리는 것에 상관 없이 그림의 모양은 같음
- ③ 각 범주간의 도수 비교에 용이
- ④ 범주가 전체에서 차지하는 비율을 파악하기 어려움

예제 3 예제2에서 주어진 자료에 대한 막대그래프를 그려라

도수 또는 상대도수

상대도수를
사용할 것을
권장함!!!

Chapter 2 표와 그림을 통한 자료의 요약

02 범주형 자료의 요약

- 파레토 그림(Pareto diagram)

- ① 막대그래프의 일종으로, 상대도수의 크기가 큰 순서로 범주를 왼쪽부터 오른쪽으로 배열하여 그림
- ② 각 범주의 누적상대도수를 막대 위 중앙에 표시하고 그 점들을 연결
- ③ 명목형 자료에서 다수를 점유하고 있는 중요한 소수의 범주를 찾는 데 용이
- ④ 누적상대도수를 통해서 상대도수가 증가하는 정도와 큰 도수의 범주들이 차지하는 비율 파악

예제 4. 어느 화장지 회사에서 생산된 미용 화장지 500장을 임의로 뽑아 불량품을 조사하여 불량률의 종류를 알아보았더니, 구멍 뚫림이 16개, 잘못 접혀짐이 3개, 찢어짐이 23개, 크기 불량 2개, 두께 불량 2개, 그리고 4종류의 불량률 각각 1개씩 있었다. 불량률이 1개인 불량률의 종류를 기타의 항목으로 묶어 하나의 범주로 하여 불량률의 종류에 대한 파레토그림을 그려라.

범주형 자료 요약 -

- 2수 분포도
- 막대 그래프
- 원형 그래프
- 파레토 그림

전부 상대도수를
표시하고 있음.

세로축에는 '상대도수'가 표시됨.

Chapter 2 표와 그림을 통한 자료의 요약

03 이산형 자료의 요약

별첨형 자료 (특히, 순위형 자료)

• 관측값의 종류가 적은 경우 : 순위형 자료를 요약하는 기법이용

• 관측값의 종류가 많은 경우 : 연속형 자료를 요약하는 기법이용

Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

- 점도표 (Dot diagram)
 - ① 수평선을 긋고 눈금을 표시한 후 각 관측값에 해당하는 위치에 점으로 표시
 - ② 관측값의 개수가 적은 경우 이용 (20~25개)

예제 6 각 국가의 교육비 대 GNP의 비율이 표에 주어져 있다. 교육비 지출 비율에 대한 점도표를 그려라

국가	GNP대비 교육비 지출 비율	국가	GNP대비 교육비 지출 비율
한국	4.4	캐나다	7.6
일본	4.7	영국	5.2
미국	5.3	이탈리아	5.4
타이완	5.7	말레이시아	5.5

Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

- 도수분포표

- ① 연속형 자료에서 모든 관측값을 포함하는 범위를 몇 개의 구간으로 나누어 각 구간에 포함되는 관측값의 개수를 세어서 작성
- ② 계급(class) : 나누어진 구간의 각 부분
- ③ 계급구간 (class interval) : 각 계급에 포함되는 값의 범위
- ④ 계급구간의 폭 : 각 계급 구간의 크기

- 연속형 자료를 위한 도수분포표의 작성방법

- ① 자료의 범위 : 자료에서 최댓값과 최솟값을 찾아 자료의 범위(range: 최댓값-최솟값)을 구한다
- ② 계급구간의 폭 : 계급의 개수가 5개~15개 정도 되도록 대략 정하고 자료의 범위를 구간의 개수로 나누어 얻은 값보다 조금 큰 값을 계급구간의 폭으로 정한다.
- ③ 계급구간 : 모든 관측값을 포함하도록 각 계급구간의 경계점을 구한다
- ④ 도수 : 각 계급구간에 속하는 관측값의 개수를 세어 계급의 도수를 구한다.
- ⑤ 상대도수 : 각 계급의 도수를 전체 관측값의 개수로 나누어 계급의 상대도수를 구한다.

범위 구하기

계급 개수를
5개로 하는 것을
권함.

계급의 개수

Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

예제 7. 다음의 자료는 어느 대학 통계학과 신입생 51명의 키를 센티미터 단위로 기록한 것이다.
이 자료에 대한 도수분포표를 만들어라

181	161	170	160	158	169	162	179	183	178	171	177	163
158	160	160	158	174	160	163	167	165	163	173	178	170
167	177	176	170	152	158	160	160	159	180	169	162	178
173	173	171	171	170	160	167	168	166	164	174	180	

Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

- 히스토그램 (Histogram)

- ① 도수분포표로부터 계급구간과 막대높이가 주어짐으로써 그릴 수 있음

$$\text{막대높이} = \frac{\text{상대도수}}{\text{계급구간의 폭}}$$

자료의 범위 ÷ 계급의 개수 = 계급구간의 폭.

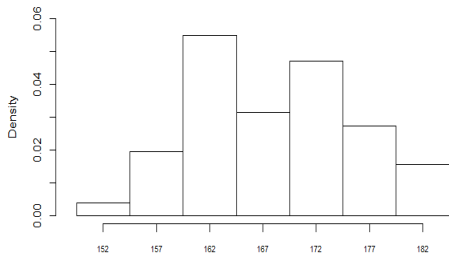
- ② 히스토그램의 전체 면적은 1

- ③ 계급구간에 따른 도수변화의 경향 (자료의 분포) 파악 가능

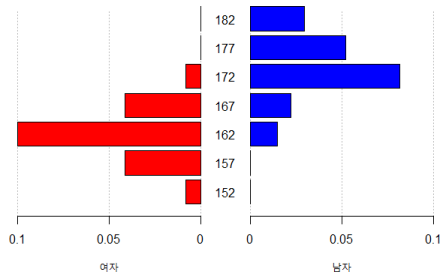
핵심!!!

예제 8. 예제 7의 통계학과 신입생의 키에 관한 자료에 대한 히스토그램을 그려라

통계학과 신입생 키에 대한 히스토그램



성별로 구분된 통계학과 신입생의 키

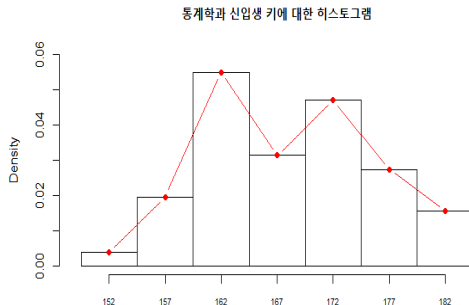


Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

- 도수다각형 (Frequency Polygon)
 - ① 히스토그램에서 각 계급구간의 막대 상단의 중앙점을 직선으로 연결하여 그림
 - ② 히스토그램과 같이 그리기도 하고 도수다각형만 그리기도 함
 - ③ 자료의 분포에 대한 상태파악에 용이
 - ④ 여러 개의 자료를 비교할 때 하나의 좌표에 여러 종류의 선으로 비교가능

예제 9. 예제 7의 통계학과 신입생의 키에 관한 자료에 대한 도수다각형을 그려라



Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

- 줄기-잎 그림 (Stem-and-leaf Plot)

: 관측값의 앞 단위와 뒷 단위를 정한 후, 앞 단위는 줄기로 하고 뒷 단위는 잎으로 하여 그림

- 줄기-잎 그림 작성방법

- ① 관측값을 보고 앞 단위와 뒷 단위를 정한다
- ② 앞 단위를 줄기로 하여 순서대로 세로로 배열하고 그 옆에 수직선을 그린다
- ③ 뒷 단위를 잎으로 하여 해당하는 관측값을 앞 단위 오른쪽에 가로로 기입한다
- ④ 각 줄기에서 앞 부분의 값을 작은 숫자가 왼쪽에 오도록 크기 순서로 재배열한다

예제 11. 예제 7의 통계학과 신입생의 키에 관한 자료에 대한 줄기-잎그림을 그려라.

```
15 | 288889
16 | 0000000122333456777899
17 | 00001111333446778889
18 | 0013
```

※한 줄기를 세분화할 수 있다.

ex) 15^- | ...
 15^+ | ...

줄기-잎 그림을 그리기 편,
관측값들을 작은 값부터 sort 한다.

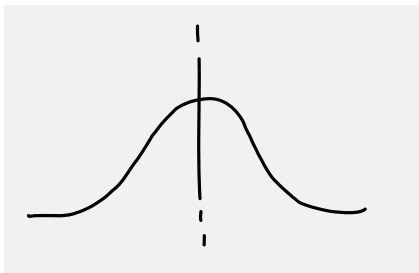
Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

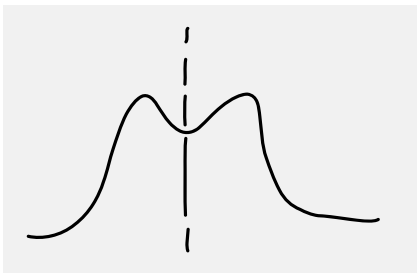
- 분포의 모양

- ① 대칭형 분포 (symmetric distribution)

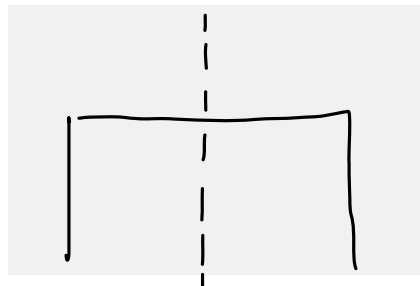
- : 분포의 좌우 모양이 중앙을 기준으로 대칭인 분포



종모양



이봉형

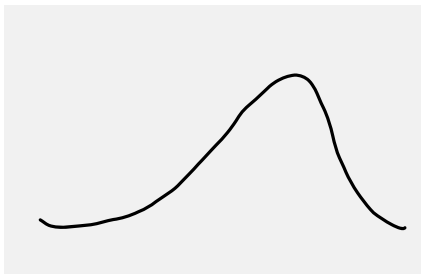


균일형

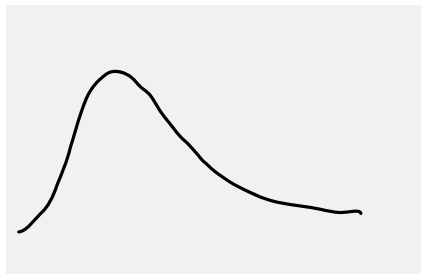
Chapter 2 표와 그림을 통한 자료의 요약

04 연속형 자료의 요약

- 분포의 모양
- ② 편중된 분포 (skewed distribution)
 - : 분포의 모양이 왼쪽이나 오른쪽으로 편중되어 있는 분포



오른쪽으로 편중 (skewed to the left)



왼쪽으로 편중 (skewed to the right)