

일반통계학 제2장 자료의 요약

Jinseog Kim

Dongguk University

jskim1986@gmail.com

2018-03-08

Contents

1	자료의 정리	2
1.1	자료의 종류	2
1.2	자료의 측정척도	2
1.3	자료의 정리: 도수분포표	3
1.4	도수분포표를 이용한 자료의 정리: 예제	4
1.5	그래프를 이용한 자료의 정리	6
1.6	막대 그래프(bar chart) : 예제	7
1.7	줄기잎그림	9
1.8	줄기잎그림 예제	10

1 자료의 정리

1.1 자료의 종류

- 질적자료(qualitative data)/양적자료(quantitative data)
- 질적자료: 관측결과가 숫자가 아닌 범주나 속성으로 분류되는 자료, 범주형자료(categorical data)라고도 함
 - 성별(남, 여), 색상(붉은색, 노란색, 파란색), 종교(불교, 천주교, 기독교), 혈액형(O, A, B, AB) 등
- 양적자료: 관측결과가 숫자로 표현되는 수치적 자료
 - 무게, 연령, 길이, 성적, ...
 - 이산형: 관측값이 셀 수 있는 값
 - 연속형: 관측값이 구간에서의 자료

1.2 자료의 측정척도

1. 명목척도(nominal scale)

- 성별, 혈액형, 자동차 회사 등
- 통계분석의 편의를 위하여 숫자로 나타내기도 함 (단 이경우, 숫자로서의 의미는 없음)
- 빈도(frequency) 분석

2. 순위척도(ordinal scale)

- ▣ 고객의 만족도나 상품의 선호도, 성적 A, B, C, D, F
- ▣ 상대적 순위

3. 구간척도(interval scale)

- ▣ 크기의 순서, 관측치 간의 차이에 대한 분석
- ▣ 비율(ratio)에 대한 분석은 의미가 없음
- ▣ 평균과 분산

4. 비율척도(ratio scale)

- ▣ 비율계산이 가능한 척도
- ▣ 몸무게, 키,...

1.3 자료의 정리: 도수분포표

▣ 도수분포표

- ▣ 질적자료(범주형) 또는 이산형, 연속형 자료를 표를 이용하여 정리하는 방법
- ▣ 계급(class): 자료의 값을 그룹화 한 것
- ▣ 도수(frequency): 각 계급에 속한 자료의 수
- ▣ 상대도수(relative frequency): 도수를 전체 자료의 수로 나눈 값

1.4 도수분포표를 이용한 자료의 정리: 예제

▣ 40명의 통계학 성적

71 64 70 78 58 49 79 65 83 69
62 47 60 70 66 79 63 60 67 92
59 67 66 78 63 62 69 56 66 81
68 86 70 66 82 72 73 54 70 71

```

x <- c(
  71, 64, 70, 78, 58, 49, 79, 65, 83, 69,
  62, 47, 60, 70, 66, 79, 63, 60, 67, 92,
  59, 67, 66, 78, 63, 62, 69, 56, 66, 81,
  68, 86, 70, 66, 82, 72, 73, 54, 70, 71)
xc <- cut(x, c(40, 50, 60, 70, 80, 90, 100))
xt <- table(xc)
xt <- data.frame(xt)
xt$rf <- round(xt[,2]/sum(xt[,2]), 2)
names(xt) <- c("계급", "도수", "상대도수")
kable(xt)

```

계급	도수	상대도수
(40,50]	2	0.05
(50,60]	6	0.15
(60,70]	19	0.48
(70,80]	8	0.20
(80,90]	4	0.10
(90,100]	1	0.02

1.5 그래프를 이용한 자료의 정리

▣ 범주형 자료

- ▣ 막대 그래프(bar chart)

- ▣ 원그래프(pie chart)

▣ 연속형 자료

- ▣ 줄기-잎 그림(stem and leaf plot)

- ▣ 히스토그램(histogram): 도수분포표를 막대그래프로 나타낸 것

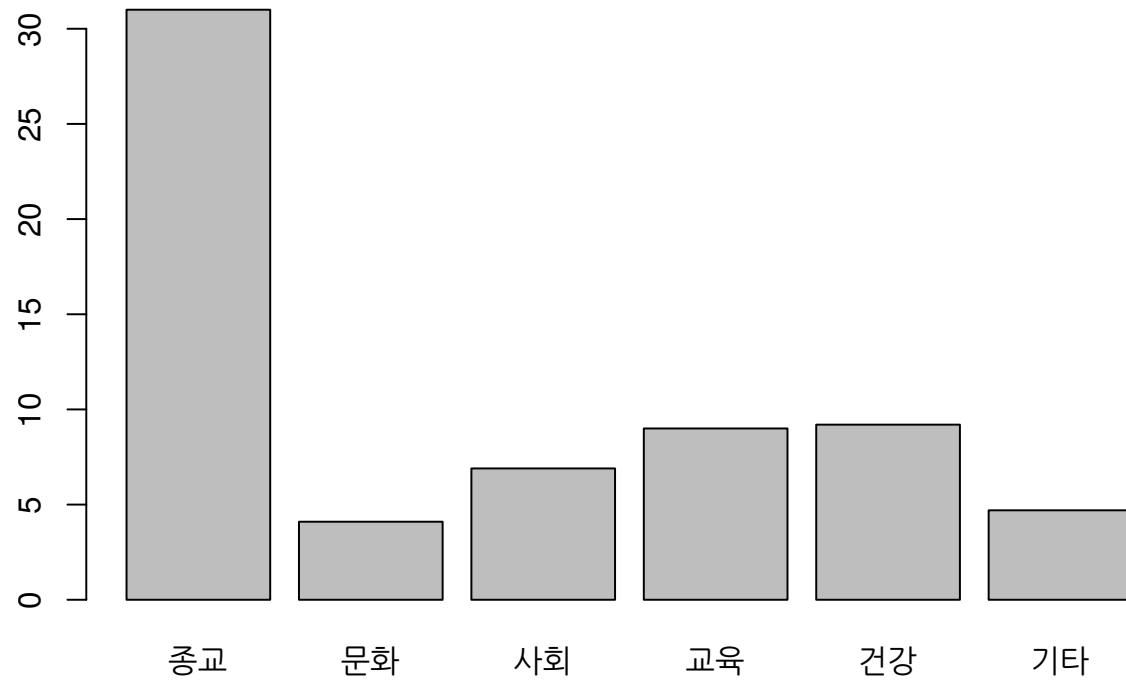
- ▣ 선그래프(line chart)

1.6 막대 그래프(bar chart) : 예제

▣ 다음 자료는 어느 사회단체에서 한 해 동안 받은 기부금의 내역이다

```
## 종교 문화 사회 교육 건강 기타  
## 31.0 4.1 6.9 9.0 9.2 4.7
```

```
y <- c(31.0, 4.1, 6.9, 9.0, 9.2, 4.7)  
names(y) <- c("종교", "문화", "사회", "교육", "건강", "기타")  
barplot(y)
```



1.7 줄기잎그림

- ▣ 도수분포표는 관측값들을 몇개의 구간(계급)으로 나누어 정리하기 때문에 정보 손실 발생
- ▣ 관측값들의 처음 몇 자리는 줄기로 정하고 나머지 부분을 잎으로 결정
- ▣ 줄기 - 잎 그림 작성순서
 1. 자료의 자리수를 고려하여 줄기와 잎을 결정
 2. 줄기를 크기순으로 나열, 옆에 수직선을 그은 후, 각 줄기에 해당되는 자료의 잎 부분을 오른쪽 나열
 3. 각 줄기에서 잎의 값을 크기순으로 나열

1.8 줄기잎그림 예제

▣ 통계학 수강인원 25명에 대한 성적이 다음과 같다. 줄기 - 잎 그림을 작성하시오.

```
x <- c(
  64, 84, 82, 81, 68, 85, 76, 89, 93,
  77, 66, 64, 86, 74, 64, 70, 53, 98,
  59, 79, 57, 59, 65, 67, 80)
stem(x, scale=1)
```

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 5 | 3799
## 6 | 4445678
## 7 | 04679
## 8 | 0124569
## 9 | 38
```