

3강. 탐색적 자료분석[EDA] 시각화 2

◆ 담당교수 : 이태림

들어가기

■ 주요용어

용어	해설
히스토그램	연속형 관찰값의 구간별 도수를 상대적인 막대의 길이로 나타낸 그래프
상자그림	사분위수와 중앙값으로 상자를 만들고 최대 최소에 선을 연결한 그래프
줄기 잎 그림	수치로 된 자료를 줄기와 잎으로 분류하여 자료의 분포개형을 파악하는 그림

연습문제

1. 다음 줄기그림의 박스안의 자료의 실제값은?

- ① 18
- ② 380
- ③ 80
- ④ 8

1	5	8		
2	3	8	8	
3	1	4	6	8
4	2	4	5	
5	0			

정답 : ③

해설 : 자료를 줄기와 잎으로 분해하여 작성된 줄기 잎의 원자료는 줄기부분과 잎부분을 결합해야 한다.

2. 줄기그림과 히스토그램의 차이점은?

- ① 줄기그림이 줄기의 수를 임의로 정할 수 있다.
- ② 잎으로 원자료를 표시해줌으로써 히스토그램 보다 통계적 정보의 손실이 적다.
- ③ 자료의 분포개형을 좀 더 자세히 파악할 수 있다.
- ④ 자료의 분포양상을 하나의 통계량으로 나타낼 수 있다.

정답 : ②

해설 : 자료를 줄기와 잎으로 분해하여 표시하므로 히스토그램은 도수로 카운트되는 데 반해 줄기그림은 잎에 실제 자료값을 담고 있다.

3. 상자그림을 그리기 위해 필요한 수치는?

- ① (최소값, 최빈치, 평균, 중앙치, 최대값)
- ② (최소값, 아래사분위수, 중앙치, 위사분위수, 최대값)
- ③ (최소값, 최대값, mid값, spread값, 중앙치)
- ④ (mid H, mid E, mid M, mid D, mid C)

정답 : ②

해설 : 상자그림은 다섯수치요약(최소값, 아래사분위수, 중앙치, 위사분위수, 최대값)으로 상자를 작성하고 최대 최소에 선으로 연결하여 작성하는 그래프이다.



정리하기

1. 줄기그림: 수치로 된 자료를 줄기와 잎으로 분류하여 자료의 분포개형을 파악하는 그림

2. 줄기그림의 작성방법

- ① 자료의 줄기부분을 선택하고 나머지 부분을 잎으로 정한다.
- ② 줄기값을 크기순으로 세로로 나열하고 그 옆에 수직선을 긋는다.
- ③ 각 줄기에 해당되는 자료의 잎부분을 줄기의 오른쪽에 가로로 나열한다.
- ④ 각 줄기에서 잎의 값을 크기순으로 재배열한다.
- ⑤ 줄기열 왼쪽에 각 줄기의 도수를 세어 기록한다.
- ⑥ 누적도수를 그림에 써넣는다.

3. 히스토그램과 줄기그림의 공통점과 차이점

-공통점

외양적인 테두리가 동일하다는 것으로 자료의 분포개형을 파악하는데 중요한 정보

를 제공한다. 각 구간에 속하는 자료점의 도수에 비례하는 막대기둥을 갖고 있다.

-차이점

- ① 줄기그림은 구간내의 자료들의 도수가 아닌 숫자로 구별되어 있으므로 도수로 나타낸 히스토그램이 정보의 손실을 가져온다.
- ② 히스토그램은 구간의 폭을 정수배로 늘리려면 원자료에 의존하지 않으면 안 되는데 줄기그림에서는 작성된 원 줄기그림을 이용하여 좀 더 효율적으로 쉽게 줄기수를 조절할 수 있다.
- ③ 줄기그림은 구간폭이 정수이어야 하기 때문에 히스토그램처럼 구간폭을 임의로 정할 수 없다.

4. 줄기그림에서 관찰할 수 있는 자료의 특징

- ① 자료가 몇 개의 그룹으로 나누어지는가?
- ② 분포의 중앙이 어디인가? 즉 자료가 어느 줄기에 집중적으로 몰려있는가로 집중도가 높은 구간을 알 수 있다.
- ③ 좌우대칭을 평가하여 자료가 어느 한쪽으로 몰려 있는가를 파악한다.
- ④ 자료의 퍼짐정도를 등을 비교한다.
- ⑤ 특이점의 존재여부를 파악한다.

5. 상자그림의 작성 : 다섯수치요약을 자료로 하여 좀 더 시각적으로 분포의 특징을 파악하기 쉽게 나타내는 방법이다. 즉 다섯수치요약을 그래프화한 것이다.

<작성방법>

- ① 다섯수치 요약을 구한다. (Min, HL, Med, HU, Max)
- ② 위 아래 사분위수(H)로 네모상자의 양끝을 잡는다.
- ③ 중위수(M)로 상자내에 "+" 또는 세로막대로 표시한다.
- ④ 최소값(Min)과 최대값(Max)을 점으로 연결한다.

참고자료

- 1. 허명희 (2014). 데이터 시각화, 자유아카데미.
- 2. 허명희 (2014). R을 활용한 탐색적 자료 분석, 자유아카데미.