

12강. 다변량 자료의 시각화 이해 1

◆ 담당교수 : 이태림

들어가기

■ 주요용어

용어	해설
별그림	다변량을 동시에 별의 형태로 나타낸 그림
산점도 행렬	다변량 변수를 2개씩 변수 쌍의 산점도 행렬로 나타낸 그림
평행좌표 플롯	다변량 변수들을 각각의 평행선상에 타점한 점들을 연결하여 그린 그래프
조건부플롯	세 변수 중 한 변수의 조건화 상에서 나머지 두변수의 관계

연습문제

1. 다변량 자료를 표현하는 단순한 방법으로 세 변수의 축으로 변수값의 크기를 나타내어 삼각의 형태로 상대적인 비교를 가능하게 하는 방법은?

- ① 스파게티그래프
- ② 별그림
- ③ 파이조각그림
- ④ 바이올린그림

정답 : ②

해설 : 별그림: 다변량을 동시에 별의 형태로 나타낸 그림

2. 피셔의 붓꽃자료로 *setosa*, *versicolor*, *Virginica*의 3품종을 시각적으로 분류할 수 있는 방법은?

- ① 산점도 행렬
- ② 평행좌표플롯

- ③ 주성분분석
- ④ 등고선도?

정답 : ①

해설 : 산점도 행렬: 다변량 변수를 2개씩 변수 쌍의 산점도 행렬로 나타낸 그림

3. 조건부 플롯에서 조건화 변수가 연속형 변수일 때 범주화하기 위한 R 함수는?

- ① conditioning plot()
- ② equal.count()
- ③ 평행좌표 플롯
- ④ 자료의 재표현

정답 : ②

해설 : Z를 몇 개의 구간으로 나누어 범주화하기 위한 R function

정리하기

1. 다변량 자료의 시각화는 현실적으로 그래프로 작성하는데 3차원을 넘는 시각화 정 보는 처리하기가 어렵다.
2. 별그림(star plot)은 다변량 자료를 표현하는 단순한 방법으로 세 변수의 축으로 변수값의 크기를 나타내어 삼각별의 형태로 상대적인 비교가 가능하다.
3. 파이조각그림(pie segments plot)으로 각 변수는 파이조각의 면적으로 나타내진다.
4. 산점도 행렬(scatterplot matrix)이란 2개의 변수 쌍으로 작성된 산점도를 변수의 개수만큼 P×P행렬에 체계적으로 배열해놓은 그래프이다.
5. 조건부 플롯(conditioning plot)이란 3개 변수 X, Y와 Z의 관계를 탐색하는 과정에서 변수 X와 Y의 관계를 Z에 조건화하여 검토하는 그래프이다.

참고자료

1. 허명희 (2014). 데이터 시각화, 자유아카데미.
2. 허명희 (2014). R을 활용한 탐색적 자료분석, 자유아카데미.