**특이값: 정상적이라고 생각되는 데이터의 분포 범위 밖에 위 치하는 값 = 이상치**

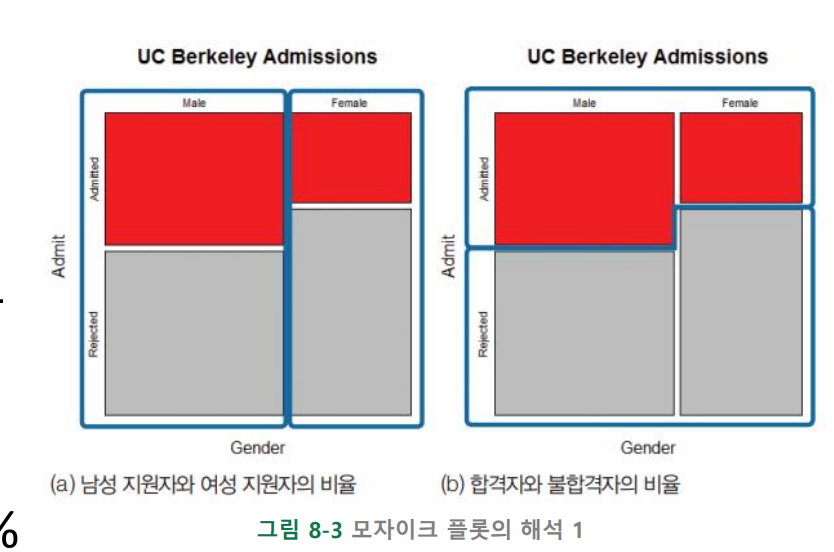
**특이값은 입력 오류에 의해 발생하기도 하고, 일반인의 몸무게 자료에 씨름선수의 몸무게가 합쳐진 경우처럼 실제로 특이한 값일 수도 있음**

<보고 뭐인지 알게만>

**모자이크 플롯 : 다중변수 범주형 데이터에 대해 각 변수의 그룹별 비율을 면적으로 표시하여 정보를 전달**

**버블차트 : 앞에서 배운 산점도 위에 버블의 크기로 정보를 표시하는 시각화 방법**

**트리맵 : 사각타일의 형태로 구성되어 있으며, 각 타일의 크기와 색깔로 데이터의 크기를 나타냄**



**<파란색 박스가 의미하는 것은?>**

**ggplot: 주로 미적인 그래프를 그릴 때 사용 어렵고 ggplot2 패키지 설치 필요함**

**차원 축소: 고차원 데이터를 2,3 차원 데이터로 축소하는 기법을 말하는데, 2,3 차원으로 축소된 데이터로 산점도를 작성하여 데이터 분포를 확인하면 고차원상의 데이터 분포를 추정 가능**

**회귀분석: 통계학에서 사용하는 자료 분석 방법 중 하나로 여러 자료들 간의 관계성을 수학적으로 추정 및 설명하는 것을 말함**

**독립변수(independent variable): 주식시세에 영향을 미치는 요인들(기업의 매출액, 원유가격, 국제정세, 정부정책 발표) = 설명변수**

**종속변수(dependent variable): 독립변수의 영향에 따라 값이 결정되는 주식시세 = 반응변수**

텍스트이(가) 표시된 사진

자동 생성된 설명

**머신러닝: 방대한 데이터를 컴퓨터가 스스로 분석하고 학습하여 유용한 정보를 얻어내거나 미래를 예측하기 위한 예측모델을 만들어내는 기술 ex) 군집화와 분류**

**군집화(clustering): 주어진 대상 데이터들을 유사성이 높은 것끼리 묶어 주는 기술**

**->묶음들을 군집(cluster), 범주(category), 그룹(group), 클래스(class) 등 다양한 용어로 부름**

**분류(classification): 그룹group, class의 형태로 알려진 데이터들이 있을 때 그룹을 모르는 어떤 데이터에 대해 어느 그룹에 속하는지를 예측하는 기술**

텍스트, 영수증이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명