<Dataset 분석 보고서>

정보융합학부 2018204002 박정재

#1

1. 기본 데이터 셋인 trees를 불러온 후 다음과 같은 3가지 조건을 각각 저장해줍니다.
2. tree1은 “Girth”(직경)열의 평균을 colMeans로 구한 후 “Girth”(직경)열의 값들 중 그들의 평균보다 큰 값들을 조건으로 지정합니다.
3. tree2는 “Height”(높이)열의 값들 중 80보다 큰 값들을 조건으로 지정합니다.
4. tree3는 “Volume”(부피)열의 값들 중 50보다 큰 값들을 조건으로 지정합니다.
5. 마지막으로, sum을 이용하여 이 3가지 조건들을 모두 만족시키는 벚나무의 개수를 구하면 3개가 나옵니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

#2

1. tips 데이터 셋을 불러오기 위해 “reshape2”패키지를 설치합니다.
2. tips를 불러 자료구조를 확인합니다.
3. head를 이용하여 처음 6개의 자료를 추출합니다.

테이블이(가) 표시된 사진

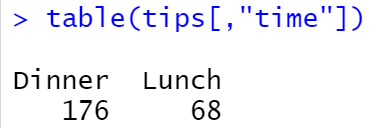
자동 생성된 설명

1. table을 이용하여 “day”열의 속성들 각각의 개수를 구해 요일별로 팁을 받는 빈도를 구합니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

1. (4)와 마찬가지로 table을 이용하여 “time”열의 속성들 각각의 개수를 구해 시간별로 팁을 받는 빈도를 구합니다.



1. (5)의 table을 number라는 변수에 저장하고, number에서 “Dinner”와 ”Lunch”의 수를 각각 DinnerNum, LunchNum이라는 변수에 저장합니다. 그리고, subset을 이용하여 “time”열이 각각 “Dinner”와 “Lunch”일 때의 데이터 셋을 Dinner, Lunch라는 변수에 저장합니다. 그 후 Dinner와 Lunch를 colSums를 이용하여 각각의 “total\_bill”, “tip”, “size”열의 합들을 구하고 그것을 각각 DinnerNum과 LunchNum으로 나눠 주어 저녁과 점심때의 해당 평균들을 구합니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

1. “tip”열의 값들을 “total\_bill”열로 나눈 값들을 percent에 저장합니다. 그 후 mean을 이용하여 계산하면 손님들이 결제금액 대비 약 16%의 팁을 준다는 것을 알 수 있습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명