

HW1-2. 국회의원 표결 분석

안상호

2020 년 4 월 17 일

과제를 수행하기 전, 반드시 “백터 이해하기.docx” 파일을 읽어보세요.

1. Homework 개요

얼마전 4 월 15 일 대한민국 제 21 대 국회의원 선거로 새로운 20 대 국회의 구성원들이 선출 되었습니다. 이로써 제 20 대 국회는 5 월 29 일에 그 임기가 끝나게 됩니다.

여러분들은 이번 Homework 를 통해서 국회의원 표결 데이터를 분석하는 방법에 대한 기초적인 방법에 대해서 배우고, 이를 20 대 국회에서 표결한 안건들에 적용해보게 될 것입니다.

1.1. 데이터 소스

- virtualData.csv
 - 가상의 국회의원들의 표결 데이터
 - 2 열은 소속 정당의 이름(가상)
 - 3~10 열은 각 의안 별 투표 내역

1.2. 문제

(기본) 문제는 총 7 문제로, 1 ~ 5 번까지는 예제로 주어진 데이터를 사용하여 진행됩니다
(응용) 6 ~ 7 은 직접 선택 및 수집한 데이터로 진행됩니다.

문제 1

- 파일을 읽어 들인 후, 데이터프레임의 행 이름을 각각 매칭되는 의원들의 이름으로 수정한 다음, 의원들의 이름이 담겨 있는 첫 번째 열을 제거한 데이터프레임을 출력

문제 2

- dataframe 의 각 의원 별 표결 정보에서 “찬성”은 1 로 “반대”는 -1 로 “기권”은 0 으로 <NA>는 0 으로 변경

참고 : 기존 데이터 프레임을 사용해 데이터를 변환하면 문제가 생길 수 있습니다. 새로운 데이터 프레임을 만들고 값을 새로 할당하는 방법을 사용해주세요. (물론 문제가 발생하지 않는다면 기존 데이터 프레임을 사용하셔도 상관없습니다.)

문제 3

- 각 의원 별 표결 정보를 내적해서 나온 결과를 저장한 데이터 프레임
 - 숫자가 클수록 의원간 투표가 비슷한 분포를 보여준다는 뜻
 - 숫자가 작을수록 의원간 투표가 다른 분포를 보여준다는 뜻

	강병원	강석호	강효상	고용진	곽상도	권성동	금태섭	김경수	김경협	김기선	김두관	김민기	김병기	김삼화	김석기	김성수
강병원	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
강석호	1	3	-1	1	-3	1	1	1	1	-1	1	1	1	1	1	1
강효상	-3	-1	3	-3	1	1	-3	-3	-3	3	-3	-3	-3	-3	-1	-3
고용진	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
곽상도	-1	-3	1	-1	3	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1
권성동	-1	1	1	-1	-1	3	-1	-1	-1	1	-1	-1	-1	-1	1	-1
금태섭	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김경수	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김경협	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김기선	-3	-1	3	-3	1	1	-3	-3	-3	3	-3	-3	-3	-3	-1	-3
김두관	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김민기	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김병기	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김삼화	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3
김석기	1	1	-1	1	-1	1	1	1	1	-1	1	1	1	1	1	1
김성수	3	1	-3	3	-1	-1	3	3	3	-3	3	3	3	3	1	3

참고 : 다음과 비슷한 형태의 데이터 프레임이 완성되어야 함

문제 4

- 문제 3 을 통해 만들어진 데이터 프레임을 분석해보세요.
 - 예) 표결 결과가 가장 크게 차이나는 경우는 박○○ 의원과 오○○ 의원을 비교했을 때 이다. 등

문제 5

- **문제 2** 에서 데이터 프레임을 만든 방법을 이용해 정당 별 표결 평균 및 분산을 구해보세요. 평균과 분산은 8 차원 벡터형태 이어야 합니다.

문제 6

- **의안 정보 시스템** 에서 원하는 표결 정보 10 개를 받아 의원 간 표결 비교표를 만들고 분석해보세요. 또, 정당 별 평균 및 분산도 구해보세요.
 - 의안 정보 시스템 [웹페이지 주소](#)
 - 페이지에 접속하고 원하는 의안명을 클릭하신 후, 의원성명 탭을 클릭합니다. 검색된 의원 명, 소속 정당, 표결 결과를 모두 긁어 하나의 의안정보 당 csv 파일 한 개를 만듭니다(드래그로 선택 후 엑셀로 가공). 즉, 총 10 개의 csv 파일이 생성되어야 합니다.
- 10 개의 표결정보 데이터(csv 파일)를 하나의 테이블로 만들어서 분석을 진행합니다. 방법은 2 가지가 있습니다. 2 가지중 한가지를 사용하시면 됩니다
 1. 다음 설명은 %in% 를 사용해 문제를 해결하는 방법입니다.
 - 추가할 데이터의 각 행의 이름을 조사해 기본 데이터프레임에 있는 이름인지확인
 - 기본 데이터 프레임에 있는 이름이라면, 기본 데이터프레임에 새로운 열을 만들고 해당 이름에 해당하는 행에 그 이름에 해당하는 정보(표결 정보)를 추가
 - 기본 데이터 프레임에 없는 이름이라면, 기본 데이터프레임에 새로운 열을 만든 뒤, 그 이름에 해당하는 새로운 행을 만들고 이름에 해당하는 정보(당명, 표결 정보)를 추가
 2. merge 계열 함수를 사용하시면 문제를 해결할 수 있습니다.
 - merge 계열 함수를 사용하지 않고, %in% 를 사용해 문제를 해결하시면 가산점이 있습니다.
 - outer join 을 하기 위해서는 merge(x=a, y=b, by = c("key1", "key2"), all = T) 함수를 사용하시면 됩니다. x 와 y 는 결합할 데이터 프레임들, by 는 기준 열, all 은 outer join 의 유무 입니다.

문제 7

- 문제 6 에서 선택한 10 개의 표결 정보들을 선택한 이유를 서술하고 분석한 내용들을 토대로 나름대로의 분석결과를 도출하세요.

2. 필수사항

- 문제 해결에서 새로 작성하는 변수, 열 이름 등은 모두 저장하는 데이터의 의미를 내포한 이름을 가질 것
- R 소스코드 작성시 코드에 대한 설명을 성실히 작성할 것
- 최소 라인 수는 30 줄 이상이 되어야함. 이때, 공백, 변수 값 출력, 주석은 라인수에서 제외

3. 제출방식

- R 마크다운을 사용해 작성할 것
- 제출은 **HW1-2_학번_이름.Rmd** 와 **HW1-2_학번_이름.docx** 형식으로 제출할 것

4. 유의사항

- 조교는 R 을 처음 다뤄본다 가정하고 코드에 대한 설명을 볼 것이므로 최대한 자세하게 작성할 것.
- 수업시간에 다루지 않은 것을 사용해도 되지만 추가점수는 없음.