

통계계산프로그래밍 과제 1

마감: 2021년 10월 11일(월)

파트1.(60점)

1. 피보나치의 수열(Fibonacci sequence)은 수학에서 잘 알려진 수열 중 하나이다. 처음 두 개의 구성요소는 [1,1]로 정의된다. 그 다음의 구성요소는 바로 이전 두 개의 구성 요소들의 합으로서 정의된다. 다음과 같은 방법으로 변형된 피보나치 수열을 생성할 수 있도록 R코드를 작성하시오.

1) 처음 두 구성 요소들의 값을 모두 2로 변경한다.

2) 연속된 두 구성 요소들의 합이라고 정의된 규칙을 차로 변경한다. 예를 들어 세 번째 구성 요소의 값은 두 번째 구성 요소가 가지는 값에서 첫 번째 구성 요소가 가지는 값을 뺀 것이다.

2. 2021년 9월 현재, 대한민국 국회 의석수 현황은 총 297명으로, A, B, C, D, E, F, G, H 당의 지역구·비례대표 수가 각각 (158, 11), (85, 19), (1, 5), (0, 3), (0, 3), (0, 1), (0, 1), (6, 4) 명이다. 이 중 30명을 비복원 임의추출하여 특별위원회를 구성하는 경우 위원회의 정당 및 지역구·비례대표 구성표를 8*2 표로 제시하라.

* `sample()` 함수를 사용. 출력은 임의적이다.

3. 파스칼의 삼각형: 조합 수 ${}_nC_k$ 가 담긴 리스트 L을 제시하라 ($k=0, \dots, n; n=1, \dots, 10$). 즉,

L[[1]] : 1 1

L[[2]] : 1 2 1

L[[3]] : 1 3 3 1

⋮

L[[10]] : 1 10 45 120 ⋯ 45 10 1

힌트: ${}_nC_k = {}_{n-1}C_{k-1} + {}_{n-1}C_k$, for $k=1, \dots, n-1$.

파트2.(40점) 메이저리그(Lahman 패키지)의 Salaries, Batting, Pitching)의 2015년 데이터에 대하여 다음 질문들에 답하라.

4. 선수들의 연봉의 10분위수, 즉 0%, 10%, 20%, ..., 90%, 100% 분위수를 구하라. 이들은 수열로서 어떤 패턴을 보이는가? 수열의 로그변환은 어떤 패턴을 보이는가?

5. 메이저리그 팀들의 팀별 중간연봉의 그래프를 막대그림으로 제시하라. 최대 중간연봉 대 최소 중간연봉의 비를 산출하라.

6. 추신수 선수(playerID "choosh01")가 속한 Texas Rangers의 선수들의 연봉을 막대그림으로 표출하라. 추신수 선수의 연봉은 소속 팀의 중간연봉의 몇 배인가?

7. 총 연봉 중 투수연봉이 차지하는 비율은 얼마인가? (* 투수의 정의는 Pitching 데이터에 이름이 있는 선수로 한다. 다른 정의를 사용하는 경우 세부적 사항을 명확히 기술할 것).