

IRD - przykładowe kolokwium

Zadanie 1

Napisz funkcję o nazwie `LiczSumyKolumn`, która dla dowolnej macierzy będzie zwracała wektor sum kolumn i będzie jednocześnie wyświetlała w konsoli komunikat w następującym formacie:

```
wektor_sum_kolumn <- LiczSumyKolumn(macierz)
```

```
## Sumy w kolumnach: -2.308092 -3.619076 -4.416957 -2.00685 -1.189115
```

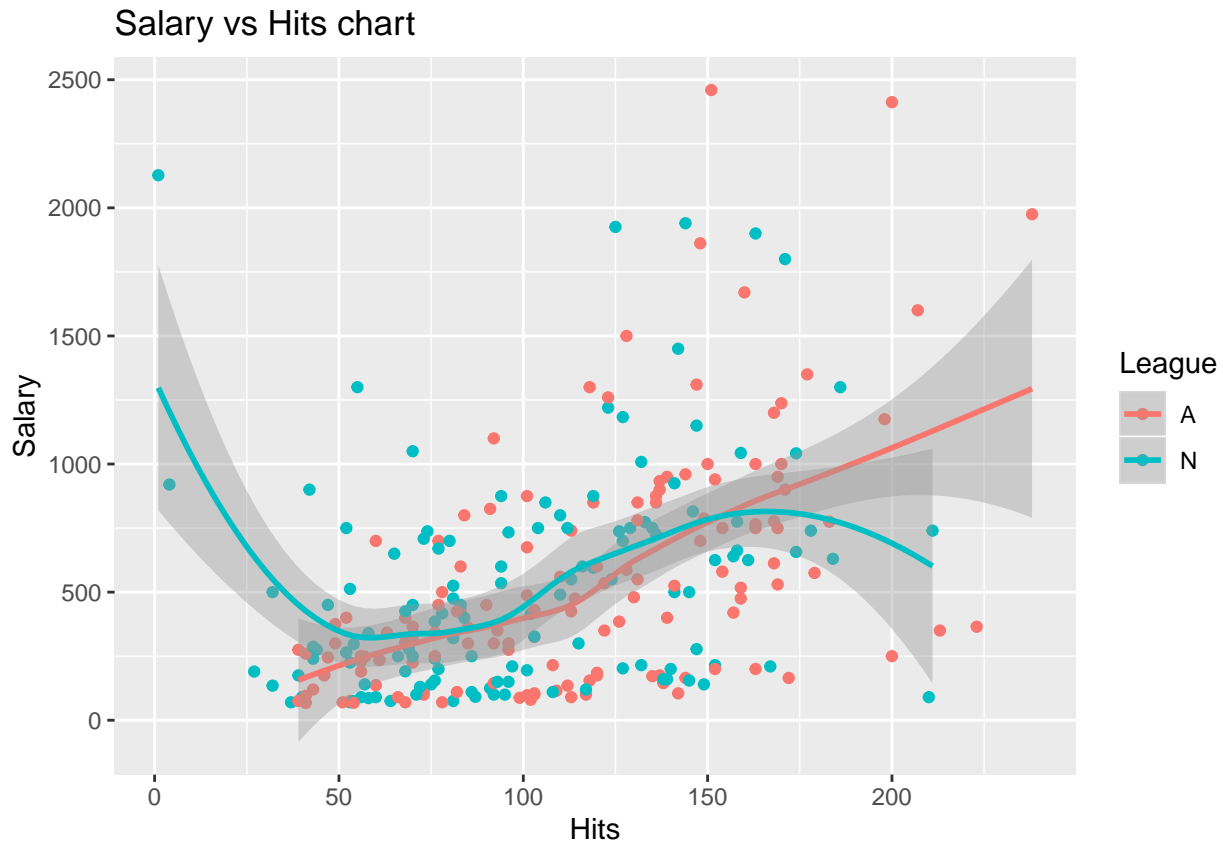
Ustaw ziarno na twój numer indeksu. Stwórz macierz o wymiarach 5x5 z liczbami z rozkładu normalnego zaokrąglonymi do 3 miejsc po przecinku i zastosuj na niej napisaną funkcję. Wyświetl w konsoli wszystkie otrzymane sumy, które są mniejsze od 0.

Zadanie 2

- Wczytaj zbiór `FITNESS`. Napisz liczbę zmiennych i obserwacji w zbiorze.
- Wyświetl medianę, dominantę, minimum i maksimum zmiennej `Age`.
- Zamień zmienną `RunTime` na zmienną binarną, która przyjmuje wartość „1”, dla wartości większych lub równych średniej i 0 w przeciwnym przypadku oraz nadaj tej zmiennej typ czynnikowy.
- Stwórz nową kolumnę o nazwie `Diff`, która będzie różnicą pomiędzy zmiennymi `MaxPulse` i `RunPulse`.

Zadanie 3

- Wczytaj dane `HITTERS` z pakietu `ISLR`. Napisz, ile jest braków danych po czym je usuń.
- Wybierz ze zbioru kolumny `Hits`, `Years`, `Salary`, `League` oraz `Division` i pogrupuj zbiór według zmiennych `League` i `Division` oraz policz średnią w grupach dla pozostałych zmiennych. Zinterpretuj przykładową otrzymaną wartość dla zmiennej `Salary`.
- Narysuj wykres identyczny jak poniższy.



Zadanie 4

- Wczytaj dane COLLEGE z pakietu ISLR. Ustaw ziarno na twój numer indeksu. Podziel losowo zbiór danych na zbiory treningowy i uczący w proporcji 60:40.
- Na zbiorze treningowym zbuduj las losowy i drzewo decyzyjne, które będą klasyfikować studentów względem tego, czy uczęszczali do college'u prywatnego czy publicznego. Zmienna objaśniana: **Private**, pozostałe zmienne to zmienne objaśniające.
- Napisz przykładową regułę decyzyjną otrzymaną za pomocą drzewa.
- Porównaj zbudowane modele na zbiorze testowym za pomocą macierzy klasyfikacji. Napisz, ile wynosi Accuracy i Missclassification error rate dla poszczególnych modeli oraz napisz co te statystyki oznaczają. Który model lepiej klasyfikuje studentów?