

INDUKOWANE REGUŁY DECYZYJNE

Zadanie 1

(6 pkt) 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: 2.82 -1.05 2.13 -2.39 7.5
```

(4 pkt) 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

(3 pkt) Wczytaj zbiór *FITNESS*. Napisz liczbę zmiennych i obserwacji w zbiorze.

(2 pkt) Wyświetl medianę, dominantę, minimum i maksimum zmiennej *RunTime*.

(3 pkt) 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 czynnika.

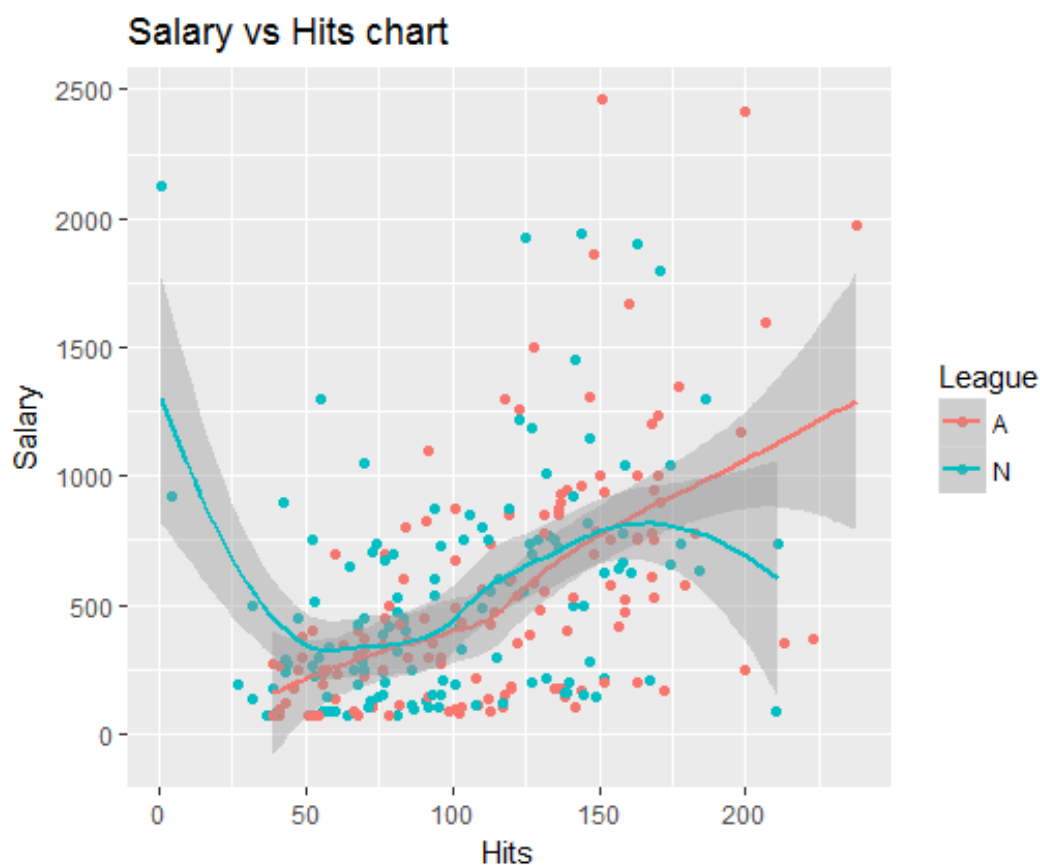
(2 pkt) Stwórz nową kolumnę o nazwie *Diff*, która będzie różnicą pomiędzy zmiennymi *MaxPulse* i *RunPulse*.

Zadanie 3

(1 pkt) Wczytaj dane *HITTERS* z pakietu *ISLR*. Napisz, ile jest braków danych po czym je usuń.

(5 pkt) Wybierz ze zbioru kolumny *Hits*, *Years*, *Salary*, *League* oraz *NewLeague* i pogrupuj zbiór według zmiennych *League* i *Division* oraz policz średnią w grupach dla pozostałych zmiennych. Zinterpretuj przykładową otrzymaną wartość dla zmiennej *Salary*.

(4 pkt) Narysuj wykres identyczny jak poniższy i oraz (1 pkt) wyeksportuj go do pliku jpg.



Zadanie 4

(2 pkt) 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.

(4 pkt) 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.

(1 pkt) Napisz przykładową regułę decyzyjną otrzymaną za pomocą drzewa.

(3 pkt) Porównaj zbudowane modele na zbiorze testowym za pomocą macierzy klasyfikacji. Napisz, ile wynosi Accuracy i Missclassification error dla poszczególnych modeli oraz napisz co te statystyki oznaczają. Który model lepiej klasyfikuje studentów?