

Modelowanie Statystyczne w Zarządzaniu Wierzytelnościami Masowymi

Laboratorium 1. Lista zadań

Zadanie 1

Stwórz strukturę danych `x <- data.table(U = runif(10))`. Następnie napisz funkcję `goodBadProp`, która przyjmuje jako argumenty strukturę `tab` (`data.table`) oraz liczbę `p` z przedziału $(0, 1)$. W ciele funkcji wykonaj operacje:

- dodaj do struktury `tab` kolumnę `GoodBad`, która przyjmuje wartości `1`, gdy $U > p$, a `0` w przeciwnym przypadku,
- jako wynik funkcji zwróć strukturę `data.table` z wyliczonymi częstościami `good`ów i `bad`ów.
- zadbaj o to, by kod funkcji nie zależał od wprost od `U` (nazwy kolumny w `x`)

Zadanie 2

Stwórz tabelę o nazwie `rndNumbers` (klasa `data.table`) o 100k wierszach i kolumnach:

- `U` z wartościami rozkładu jednostajnego na $(0, 1)$
- `Z` z wartościami rozkładu normalnego
- `E` z wartościami rozkładu eksponencjalnego
- `G` z wartościami $0, 1$
- `P` z wartościami z rozkładu Poissona o średniej 2

Wyznacz (korzystając z funkcjonalności `data.table`) statystyki opisowe wybranych 2 kolumn. Wyznacz (korzystając z funkcjonalności `data.table`) statystyki opisowe kolumn `U`, `Z`, `E` w rozbiciu względem kolumny `G`. Wyznacz (korzystając z funkcjonalności `data.table`) statystyki opisowe kolumn `U`, `Z`, `E` w rozbiciu względem czy `P` jest większe od swojej średniej.

Zadanie 3

Do tabeli `rndNumbers` z zadania 2 dodaj kolumnę `Id` przypisując do niej `.I`. Stwórz tabelę `rndNumbers2` jako podzbiór tabeli `rndNumbers` dla `Id <= 10`. Stwórz tabelę `tab` (10 wierszy) z kolumnami:

- `Id` o wartościach z wektora `c(1, 2, 3)` przypisanymi losowo
- `Id2` o wartościach z wektora `c(6, 7, 8)` przypisanymi losowo

Nałóż klucz `Id` na tabelę `rndNumbers2`. Nałóż klucz `Id` na tabelę `tab`. Wykonaj joiny tabel `tab` i `rndNumbers2` w różnej kolejności (np. `tab[rndNumbers2]`, lub `rndNumbers2[tab]`). Co otrzymujesz w wyniku (po jakiej kolumnie wykonywany jest join; wymuś joina po drugiej kolumnie)? Do czego służą parametry `nomatch`, `allow.cartesian`?

Zadanie 4

Stwórzcie `data.table` `x` o 1mln wierszy i dwóch kolumnach z liczbami z rozkładu jednostajnego i normalnego. Zapiszcie tabelę `x` do pliku `csv`. Zweryfikujcie czas zapisu przy użyciu standardowych funkcji zapisu do `csv` np. `write.table`, `write.csv` oraz `fwrite` (`data.table`). Wczytajcie utworzony plik używając funkcji `read.table` oraz `fread` porównując czas wczytywania danych. Zwróćcie uwagę na ilość zajmowanego miejsca przez zmienną w R oraz wielkość pliku `csv`.