

Wprowadzenie do programu R - zadania

Kurs: Analiza i wizualizacja danych w R

prowadzący: Piotr Ćwiakowski

Zadanie 1.

- 1) Stwórz fikcyjną bazę danych ze wzrostem i wagą dla 10 osób.
- 2) Stwórz zmienną płeć.
- 3) Za pomocą odpowiednich komend policz dla wszystkich i zapisz w tej samej tabeli indeks BMI.
- 4) Wyniki zaokrąglaj do dwóch liczb po przecinku.
- 5) Obliczenia zapisz w zewnętrznym pliku (*csv*).

Zadanie 2.

Wykonaj kod generujący bazę pacjenci. Następnie:

- 1) Przetłumacz nazwy zmiennych na angielski (albo dowolny inny język).
- 2) Na ten sam język konwertuj etykiety zmiennych typu **factor** w bazie danych.
- 3) Jeśli jakies zmienne nominalne nie są typy **factor**, to je konwertuj na ten typ.
- 4) Stwórz w bazie odpowiedniki tekstowe i numeryczne zmiennych typu **factor**.
- 5) Przypisz do nazw wierszy numery id i skasuj tą zmienną z bazy.
- 6) Korzystając z kodu z zadania pierwszego, policz indeks BMI dla pacjentów z bazy.
- 7) Stwórz zmienną logiczną, która przyjmuje wartość TRUE jeśli dana osoba jest wyższa niż przeciętna w próbie i FALSE w przeciwnym wypadku.
- 8) Konwertuj zmienną na **factor** przyjmując dowolne etykiety.

Zadanie 3.

Wczytaj poprawnie bazy danych:

- *free_time.csv*
- *free_time 2.csv*
- *free_time 3.txt*
- *free_time 4.csv*

Zwróć uwagę na separator kolumn, sposób kodowania braków danych, separator dziesiętny i rozszerzenie pliku.

Dodatkowo: Jeśli potrafisz, wczytaj do R dowolną metodą zbiór *free_time 2.xlsx*

Zadanie 4.

Wczytaj bazę *diamonds* (z pakietu *ggplot2*). Następnie:

- 1) wyświetl wiersze od 200-tnego do 300-tnego (wszystkie kolumny),
- 2) zapisz do odrębnego zbioru danych diamenty, które mają jakoś cięcia "Ideal",
- 3) zapisz do odrębnego zbioru danych diamenty o przeciętnej cenie poniżej mediany.

Zadanie 5. (dodatkowe)

Wczytaj bazę `Pima.tr` z biblioteki `MASS`.

- 1) Narysuj wykres rozrzutu zmiennej `bmi` oraz zmiennej `ped`.
- 2) Narysuj wykres słupkowy dla zmiennej `type`.
- 3) Narysuj histogram zmiennej `age`.

Zadanie 6. (dodatkowe)

Wczytaj bazę `czas.wolny.csv`. Wykonaj następujące polecenia:

- 1) Narysuj wykres punktowy zmiennej `wiek` oraz `sen`.
- 2) Narysuj wykres słupkowy zmiennej `wyksztalcenie`.
- 3) Narysuj wykres pudełkowy zmiennych `zatrudnienie` oraz `czas_wolny`.
- 4) Narysuj histogram zmiennej `podrozowanie`.