

Zaawansowane programowanie w R

mgr Piotr Cwiakowski, dr Piotr Wójcik

Zajęcia 5-6

Informacje wstępne

1. Pracę domową 5-6 należy przesłać do *środy, 29 listopada 2017 r. do godz. 24.00* na adres pcwiakowski@wne.uw.edu.pl, w tytule wpisując imię i nazwisko, numer indeksu i hasło “Praca domowa - programowanie”.
2. Do wiadomości załączamy jedynie plik z naszym kodem (rozszerzenie .R).
3. Na początku pliku umieszczamy w komentarzu krótką informację o tym co udało nam się zrobić, a co nie. Wykonanie wszystkich zadań na dane zajęcia daje 2% punktów końcowych.
4. W razie pytań i wątpliwości piszemy na adres pcwiakowski@wne.uw.edu.pl

Powodzenia!

Zajęcia 5.

W ćwiczeniach 5.1.-5.5. wykorzystaj jedną z reprezentantek rodziny funkcji apply.

Ćwiczenie 5.1.

Na podstawie bazy `state.x77` znajdziemy średnią, medianę, max i min dla: liczby ludności, dochodu, oczekiwanej długości życia oraz powierzchni we wszystkich stanach. Wyniki zwrócimy w jednej, przejrzystej tabeli.

Ćwiczenie 5.2.

Przypuśćmy, że mamy bazę danych złożoną z informacji o płci badanej osoby oraz wzroście. Proszę obliczyć średnią w zależności od płci (funkcją `tapply`). Kod generujący dane:

```
Wzrost <- rnorm(100, mean=170, sd=10)
Kobieta <- factor(floor(2*runif(100)))
d <- data.frame(Kobieta, Wzrost)
```

Ćwiczenie 5.3.

Utwórz listę zawierającą 100 wektorów: `c(1)`, `c(1,2)`, `c(1,2,3)`, \dots , `c(1,\dots,100)`

Ćwiczenie 5.4.

Wygeneruj 10 zbiorów z parami liczb x i y . Następnie dla każdego zbioru policz równanie regresji liniowej i wynik zapisz do listy.

Ćwiczenie 5.5.

Wyniki regresji z punktu piątego pobierz z każdego elementu listy funkcją `coef()` i zapisz do macierzy - w kolejnych wierszach wyniki kolejnych modeli.

Ćwiczenie 5.6.

Za pomocą funkcji `switch` napisz formułę, która na podstawie zmiennej `name` stwierdzi, czy imię jest męskie, czy żeńskie (na podstawie ostatniej litery z wyjątkiem imion Kuba, Bonawentura, Barnaba). Następnie zwektoryzuj funkcję. Przykładowy ciąg imion:

```
imiona <- c("anTek", "BASia", "czaREK", "Daria", "EWELINA", "filip")
```

Podpowiedź: Funkcję `switch` można zagnieżdżać. Podpowiedź: Aby dowiedzieć, się jaki jest ostatni znak w danym ciągu wykorzystaj funkcję: `substr()`

Zajęcia 6.

Ćwiczenie 6.1.

Wykorzystując funkcje `parS/Lapply()` z pakietu `parallel` przeprowadź estymację modelu regresji na próbach składających się z 1000 obserwacji (1-1000, 1001-2000, 2001-3000, itd.) zapisując wyniki każdej estymacji.

UWAGA! dane możesz podzielić na listę mniejszych podzbiorów za pomocą funkcji `split()`

```
dane_lista_po1000 <- split(dane,  
                           rep(1:1000, each = 1000))
```

Ćwiczenie 6.2.

Analogicznie jak w ćwiczeniu 6.1. przeprowadź estymację powyższego modelu na próbach składających się z 1000 obserwacji (1-1000, 1001-2000, 2001-3000, itd.) korzystając z funkcji `foreach()` i wykorzystując przetwarzanie równoległe.

Ćwiczenie 6.3.

Napisz funkcję, która będzie estymowała powyższy model (i zapisywała wszystkie wyniki estymacji) na kolejnych rozłącznych podpróbach składających się z `n` obserwacji, gdzie `n` będzie jedynym argumentem funkcji (ostatnia próba może mieć mniej obserwacji niż `n`, jeśli `1e6` nie jest podzielne przez `n` bez reszty). Razem z wynikami estymacji zapisuj wielkość próbki, na której model jest szacowany.

Ćwiczenie 6.4.* (za dodatkowy punkt)

Porównaj efektywność czasową poszczególnych podejść z ćwiczeń 6.1, 6.2 i 6.3 - możesz w tym celu wykorzystać mniejszy zbiór danych - np. pierwsze 100000 obserwacji wykorzystaj funkcję `benchmark()` lub `microbenchmark()`