

Program kursu „Podstawy języka R”

Dzień 1 (6 h)

Moduł 1. Podstawowe algorytmy i struktury danych w R (6 h)

Prowadzący: Prof. UEK dr hab. Paweł Lula

1. Instalacja i konfiguracja pakietu
2. Korzystanie z bibliotek (pakietów)
3. R-Studio
4. Podstawy pracy z pakietem R
 - zmienne i operator przypisania
 - zmienne typów prostych (logiczne, numeryczne, łańcuchowe)
 - operatory i funkcje wbudowane do działań na wartościach typów prostych
 - wartość NULL
 - braki danych
 - zapis i odczyt danych z plików binarnych
5. Wektory
 - tworzenie wektorów
 - indeksowanie wektorów
 - operatory do działań na wektorach
 - funkcje do działań na wektorach
 - selekcja elementów wektora
 - wektorowa reprezentacja zbiorów i operacje na zbiorach
6. Macierze i tablice
 - tworzenie macierzy
 - indeksowanie elementów macierzy
 - selekcja elementów macierzy
 - operacje na macierzach
 - etykietowanie wierszy i kolumn macierzy
 - tablice wielowymiarowe
7. Reprezentacja wartości nominalnych i porządkowych
 - tworzenie wektorów przechowujących wartości nominalne i porządkowe
 - operacje na wartościach nominalnych i porządkowych
8. Ramki danych
 - tworzenie ramek danych
 - etykietowanie wartości w ramkach danych

- selekcja danych przechowywanych w ramkach (wybór wierszy i kolumn)
- zapis i odczyt ramek danych z plików tekstowych csv (wymiana danych z pakietem Excel)

9. Listy

- tworzenie list
- dostęp do elementów listy
- modyfikowanie list
- operacje na listach
- iteracyjne przetwarzanie elementów listy

10. Funkcje

- definiowanie obiektów typu funkcja
- wartość zwracana przez funkcję
- parametry funkcji

11. Skrypty i podstawy programowania

- komentarze
- operacje wejścia/wyjścia
- instrukcja if
- pętle for, while, repeat
- wyjątki i ich obsługa

Dzień 2 (6 h)

Moduł 2. Przygotowanie, wizualizacja oraz elementarna analiza danych w R (6 h)

Prowadzący: Przemysław Jaśko

1. Import/eksport danych w R:

- import/eksport danych do R ze schowka systemu Windows, z plików csv, z plików tekstowych o określonej strukturze, z plików xlsx,
- import/eksport danych z repozytoriów z poziomu R: m.in. pakiet SmarterPoland (Bank Danych Lokalnych GUS), pakiet eurostat, dane z rynków finansowych: m.in. pakiet quantmod, funkcje własne pobierające archiwa zip z notowaniami giełdowymi z repozytoriów takich jak BOSSA, stooq,
- import danych z tabel html,
- zapisywanie i odczytywanie danych w postaci plików binarnych R (RData).

2. Wstępne przetwarzanie i przygotowanie danych do analizy:

- narzędzia do przetwarzania i czyszczenia danych z pakietu base,
- narzędzia do przetwarzania i czyszczenia danych z pakietów dplyr (m.in. funkcje select, filter, mutate, arrange, summarise), tidyR (gather, spread),
- elementarne metody postępowania z brakami danych.

3. Wizualizacja danych w R:

- wybrane narzędzia wizualizacji danych z pakietu graphics,
- wybrane narzędzia wizualizacji danych z pakietu ggplot2.

4. Elementarna analiza danych w R:

- podstawowe statystyki opisowe dla zbioru danych,
- elementarne wprowadzenie dotyczące generowania liczb pseudolosowych,
- podstawowe działania na macierzach w R,
- elementarne metody numeryczne poszukiwania miejsc zerowych i pierwiastków funkcji,
- symulacje danych z modelu KMNRL oraz UMNRL (autokorelacja składników losowych, heteroskedastyczność wariancji) oraz procedura oszacowania parametrów metodą KMNK oraz (E)UMNK oraz postępowanie w ramach weryfikacji założeń modeli.

Dzień 3 (4 + 2 h)

Moduł 3. Wprowadzenie do wnioskowania statystycznego z R (4 h)

Prowadzący: Prof. UEK dr hab. Daniel Kosiorowski

1. Wybrane metody estymacji rozkładu prawdopodobieństwa jednowymiarowej zmiennej losowej oraz jej charakterystyk liczbowych.
 - Estymacja parametrów rozkładu zmiennej losowej za pomocą metody największej wiarygodności
 - Estymacja przedziałowa wybranych charakterystyk liczbowych populacji
 - Estymacja jądrowa gęstości prawdopodobieństwa zmiennej losowej 1D
2. Symulacyjne badanie własności statystycznych wybranych estymatorów charakterystyk liczbowych populacji
 - rozkłady z próby dla: średniej przyciętej, mediany, odchylenia standardowego, mediany odchyłeń absolutnych od mediany (MAD), rozstępu międzykwartylowego (IQR), popularnego współczynnika asymetrii, kurtozy.
 - elementarne wprowadzenie do metody bootstrap
3. Stosowanie wybranych statystycznych testów istotności na przykładach z zakresu ekonomii.
 - rodzina testów Studenta: tzn. m. in. oryginalny test Studenta *vs.* test Studenta z poprawką Welsha, jedno i dwuczynnikowa analiza wariancji, test Hotellinga T^2 , sprawdziany spełnienia założeń stosowalności testów Studenta.
 - rodzina testów Wilcoxona
 - wybrane testy zgodności i normalności
4. Elementy wielowymiarowej analizy statystycznej.
 - nieparametryczna estymacja rozkładu prawdopodobieństwa wektora losowego 2D
 - regresja nieparametryczna: regresja jądrowa i regresja k- najbliższych sąsiadów
 - nieparametryczna estymacja warunkowego rozkładu prawdopodobieństwa
 - macierz kowariancji i wektor średnich *vs.* estymator minimalnego wyznacznika macierzy kowariancji (MCD)

Moduł 4. Elementy uczenia maszynowego w R (2 h)

Prowadzący: Prof. UEK dr hab. Paweł Lula

1. Analiza skupień
 - macierz odległości pomiędzy obiektami
 - klasyfikacja metod analizy skupień
 - metody aglomeracyjne

- dendrogram i jego interpretacja
 - podobieństwo wyników klasyfikacji
 - metoda k-średnich
2. Klasyfikacja wzorcowa
- Klasyfikatory liniowe
 - Drzewa decyzyjne
 - Lasy losowe

Literatura

- [1] P. BIECEK, *Przewodnik po pakiecie R*, Oficyna Wydawnicza GiS, 2017.
- [2] M. GĄGOLEWSKI, *Programowanie w języku R: analiza danych, obliczenia, symulacje*, Wydawnictwo Naukowe PWN SA, 2014.
- [3] C. KLEIBER, A. ZEILEIS, *Applied econometrics with R*, Springer Science & Business Media, 2008.
- [4] D. KOSIOROWSKI, *Wstęp do statystyki odpornej: kurs z wykorzystaniem środowiska R*, Wydawnictwo Uniwersytetu Ekonomicznego, 2012.
- [5] J. VERZANI, *Using R for introductory statistics*, CRC Press, 2014.
- [6] M. WALESIAK, E. GATNAR, *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, 2009.