

Analiza danych w R (kod: K-ANALIZA-R)

Opis i cel kursu

Umiejętność pozyskiwania i analizowania danych jest w dzisiejszych czasach kluczowa, zarówno w obszarze nauki jak i biznesu. Wyciąganie trafnych wniosków ze zbieranych na bieżąco danych, a także danych historycznych, pozwala osiągnąć postęp przy zarządzaniu firmą jak i prowadzeniu badań naukowych.

Nie każda instytucja potrafi wykorzystać potencjał, jaki kryje się w danych. Aby analizować duże ilości, często niepełnych informacji konieczne jest zastosowanie metod statystycznych i znajomość odpowiednich technologii. Te umiejętności zdobędziesz na naszym kursie!

R jest jednym z najpopularniejszych języków do analizy danych dzięki wieloletniemu rozwojowi, bogatemu zapleczu narzędzi i zaangażowanemu środowisku specjalistów z niego korzystających.

Olbrzymie repozytorium gotowych rozwiązań dla R, czyli CRAN, pakiety tidyverse, caret czy ggplot2 i inne stanowią dojrzały ekosystem gotowych do użycia modułów, a założenia przyświecające twórcom języka R pozwalają na sprawne posługiwanie się metodami statystycznymi i efektywną wizualizację danych.

Na naszym kursie można poszerzyć swoją wiedzę o zagadnienia związane z analizą danych i poznać narzędzia wykorzystywane w tym celu.

Od uczestników oczekiwana jest podstawowa znajomość matematyki i obsługi komputera. Jeśli chcesz się dodatkowo przygotować, możesz skorzystać z naszej oferty - bootcampa Specjalista MS Excel lub Analiza danych. Nie musisz być profesjonalnym programistą ani mieć doświadczenia w analizie danych.

Program

1. Wprowadzenie do R i RStudio

- R jako język programowania i RStudio jako wygodny interfejs użytkownika
- Typy danych w R: skalar, wektor, macierz, ramka danych, factor, lista
- R jako kalkulator: operatory arytmetyczne, porównawcze i logiczne, rodzaje poleceń: wyrażenia vs przypisania
- Przetwarzanie danych: wybór elementów wektora/macierzy na podstawie indeksów, tworzenie nowych kolumn w ramce danych, wybór wierszy z ramki danych na podstawie warunków
- Podstawowa analiza danych numerycznych (obliczanie średniej, sumy, itp.) i tekstowych (tabela częstości)

2. Statystyka opisowa

- Rodzaje danych statystycznych: ilościowe vs jakościowe
- Obliczanie i interpretacja statystyk opisowych dotyczących wartości przeciętnej, zróżnicowania oraz kształtu rozkładu zmiennej ilościowej
- Nieparametryczne miary rozkładu: dominanta/moda, mediana, kwantyle, decyle, kwantyle
- Podsumowanie rozkładu zmiennej jakościowej – tabele częstości

3. Wnioskowanie statystyczne

- Popularne rozkłady prawdopodobieństwa, m.in. jednostajny, normalny, t-Studenta, Chi-kwadrat, F-Snedecora
- Generowanie liczb (pseudo)losowych, określanie ziarna generatora
- Estymacja punktowa vs. estymacja przedziałowa: przedział ufności, poziom istotności, wartość p
- Testowanie hipotez dotyczących średniej i wariancji w jednej próbie

Zapytaj o szczegóły

tel. 22 63 64 164
akademia@alx.pl

Najbliższe terminy

2024-10-28 (Zdalnie)
2024-10-28 (Warszawa)

- Testowanie hipotez dotyczących porównania średniej i wariancji w dwóch próbach zależnych i dwóch próbach niezależnych
 - Analiza korelacji liniowej Pearsona i Spearmana dla zmiennych ilościowych
 - Analiza zależności dla zmiennych jakościowych (test chi-kwadrat)
 - Testowanie hipotez dotyczących średniej i wariancji w więcej niż dwóch próbach niezależnych (jednoczynnikowa ANOVA), testy porównań wielokrotnych
 - Testy nieparametryczne do porównywania rozkładów: test Wilcoxon-Manna-Whitney'a, test Kruskala-Wallis
4. **Analiza regresji liniowej w R**
- Założenia modelu regresji liniowej, interpretacja wyników regresji (parametry, dopasowanie, istotność zmiennych i modelu)
 - Jakościowe zmienne objaśniające w modelu
 - Wykrywanie obserwacji nietypowych
 - Metody doboru zmiennych
 - Diagnostyka modelu, testowanie spełnienia założeń: testowanie normalności reszt, testowanie homoskedastyczności, testowanie braku autokorelacji
 - Postępowanie w przypadku niespełnienia założeń
 - Regresja logistyczna
 - Prognozowanie na podstawie stworzonego modelu, sprawdzanie jakości prognozy, podział na próbę uczącą i testową
5. **Wizualizacja danych z wykorzystaniem pakietu ggplot2**
- Przegląd typów wykresów
 - Dodawanie i dostosowywanie elementów wykresu
 - Wizualizacja danych w grupach i podgrupach
 - Tworzenie wykresów z wykorzystaniem kreatora (pakiet esquisse)
 - Wykresy interaktywne (pakiet ggplotly)
 - Przykładowe pakiety wspomagające tworzenie wykresów
 - Zestawianie kilku wykresów w jednym oknie graficznym
 - Eksport wykresów do różnych formatów plików
6. **Zarządzanie danymi w R**
- Sortowanie i filtrowanie baz danych wg. wybranych kryteriów - pakiet dplyr
 - Praca ze zmiennymi tekstowymi - pakiety stringr i glue
 - Rekodowanie zmiennych jakościowych (np. łączenie kilku poziomów w jeden) - pakiet forcats
 - Zamiana zmiennej ilościowej na jakościową (podział na przedziały) – cut()
 - Praca z danymi w formacie daty - pakiet lubridate
 - Łączenie kilku baz danych według pola z „kluczem” – funkcje *_join()
7. **Podstawy programowania w R**
- Struktury danych w R – praca z ramkami danych (data frames) oraz z listami - operator \$, funkcja class()
 - „Potoki” poleceń - operator %>%
 - Tworzenie własnych funkcji
 - Pętle i przetwarzanie warunkowe – for(), if ... else, ifelse()
 - Korzystanie z wartości domyślnych dla argumentów funkcji
 - Jednoczesne przekazywanie wielu argumentów do funkcji już istniejących – mechanizm „...”
 - Automatyzacja powtarzalnych obliczeń – pakiet purrr
 - Podstawowa obsługa błędów – funkcje stop() i tryCatch(); possibly()
 - Debugging – lokalizowanie błędów w kodzie i poszukiwanie ich przyczyn
 - Praca z plikami zewnętrznymi
 - Import/eksport danych do/z R z/do innych formatów (np. pliki tekstowe, csv, Excel, SPSS)- pakiety foreign i readxl

Zapytaj o szczegóły

tel. 22 63 64 164

akademia@alx.pl

Najbliższe terminy

2024-10-28 (Zdalnie)

2024-10-28 (Warszawa)

8. Współpraca R z pakietem MS Office

- Tworzenie raportów z analiz w R w postaci dokumentów MS Word - pakiet officer
- Stosowanie własnych szablonów raportów w MS Word
- Zarządzanie zawartością plików Excela z poziomu R - pakiet openxlsx
- Wykorzystanie wcześniejszych elementów szkolenia do utworzenia rozbudowanego raportu z wykonanych analiz

9. Gdzie szukać pomocy

- Korzystanie z plików pomocy
- Strony WWW autorów pakietów

Zapytaj o szczegóły

tel. 22 63 64 164
akademia@alx.pl

Przeznaczenie i wymagania

Kurs dla osób, które chcą poznać język R i jego zastosowania w analizie danych.

Wymagana podstawowa znajomość matematyki i obsługi komputera.

Najbliższe terminy

2024-10-28 (Zdalnie)
2024-10-28 (Warszawa)

Certyfikaty

Uczestnicy szkolenia otrzymują imienne certyfikaty sygnowane przez ALX.