

## 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 oraz przetwarzanie danych w R

- 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)
- tworzenie prostych własnych funkcji
- proste pętle i przetwarzanie warunkowe: for(), if(), ifelse()
- inne przydatne funkcje do zarządzania danymi: class(), str(), ls(), rm(), identical(), dim(), names(), tail(), head()
- instalowanie i używanie dodatkowych pakietów
- import/eksport danych do/z R z/do innych formatów (np. pliki tekstowe, csv, Excel)
- przetwarzanie danych z wykorzystaniem funkcjonalności pakietów data.table dplyr, chaining itp.

#### 2. Statystyka opisowa i wizualizacja danych w R

- 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, kwartyle, decyle, kwantyle

### Zapytaj o szczegóły

tel. 22 63 64 164  
akademia@alx.pl

### Najbliższe terminy

2019-12-16 (Warszawa)

- Podsumowanie rozkładu zmiennej jakościowej – tabele częstości
- Analiza graficzna rozkładu zmiennej ilościowej – histogram i wykres ramkowy
- Analiza graficzna zmiennej jakościowej – wykres słupkowy, kołowy
- Dostosowywanie opcji graficznych w R (marginesy),
- Dodawanie nowych elementów do wykresów (legenda, punkty, linie)
- Zestawianie kilku wykresów w jednym oknie graficznym
- Wykorzystanie dostępnych palet kolorów, tworzenie własnych palet
- Automatyczne zapisywanie wykresów jako plików graficznych (jpg, gif, png, pdf)
- Wizualizacja danych z wykorzystaniem pakietu ggplot2
- Analiza i wizualizacja danych w podgrupach

### 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
- 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 znaków, test Wilcoxon-Manna-Whitney'a, test rangowanych znaków Wilcoxon, test Kruskala-Wallisa
- Testy dla wskaźnika struktury w jednej próbie oraz dwóch i więcej próbach zależnych i niezależnych: test chi-kwadrat, test McNemara, test Cochra

### 4. Analiza regresji liniowej w R

- założenia modelu regresji liniowej, interpretacja wyników regresji (parametry, dopasowanie, istotność zmiennych i modelu),
- jakościowe zmienne objaśniające w modelu – rozkodowanie na zmienne zero-jedynkowe,
- wykrywanie obserwacji nietypowych,
- testowanie hipotez prostych i złożonych, metoda od ogólnego do szczególnego, metody automatycznego doboru zmiennych
- diagnostyka modelu, testowanie spełnienia założeń: testowanie normalności reszt, testowanie współliniowości, testowanie poprawności formy funkcyjnej, testowanie stabilności parametrów, testowanie homoskedastyczności, testowanie braku autokorelacji,
- postępowanie w przypadku niespełnienia założeń
- zlogarytmowana zmienna zależna – interpretacja wyników modelu (elastyczności, semi-elastyczności)
- prognozowanie na podstawie oszacowanego modelu, sprawdzanie jakości prognozy, podział na próbę uczącą i testową

### 5. Współpraca R z pakietem MS Office

- Tworzenie raportów z analiz w R w postaci dokumentów MS Word i MS PowerPoint – pakiet ReporteRs
- Zarządzanie zawartością plików Excela z poziomu R – pakiet XLConnect
- Wykonywanie poleceń R z poziomu Excela – pakiet BERT, Basic Excel R Toolkit

### Zapytaj o szczegóły

tel. 22 63 64 164  
akademia@alx.pl

### Najbliższe terminy

2019-12-16 (Warszawa)

- Integracja R i Excela – pakiet RExcel
- Wykorzystanie wcześniejszych elementów szkolenia do utworzenia rozbudowanego raportu z wykonanych analiz

## **Przeznaczenie i wymagania**

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

Wymagana podstawowa znajomość matematyki i obsługi komputera.

## **Certyfikaty**

Uczestnicy szkolenia otrzymują imienne certyfikaty sygnowane przez ALX.

## **Zapytaj o szczegóły**

tel. 22 63 64 164

akademia@alx.pl

## **Najbliższe terminy**

2019-12-16 (Warszawa)