

```
174 coef(model)
175 coef(model)[1]
176 exp(coef(model)[1]) # exponentiate the log of price, to convert to original units
177 ggplot(diamondSample, aes(x = carat, y = price)) +
178   geom_point(colour = "blue") +
179   geom_smooth(method = "lm", colour = "red", size = 2) +
180   scale_x_log10() +
181   scale_y_log10()
182
183 *** Regression Diagnostics
184 # easy to get regression diagnostic plots. The same as with
```

R dla badaczy - poziom podstawowy

ptbrio szkolenie, 5 i 6 listopada 2019 r.



```
see Q-Q plot to see linearity which means residuals are normally distributed
c(2, 2) # set up for multiple plots on the same figure.
col = "blue"
c(1, 1) # plot layout to single plot on a 1x1
```

Marek Młodożeniec
Sotrender

NOWOŚĆ

Tematyka szkolenia

Pakiety statystyczne wykorzystywane w agencjach badawczych rzadko łączą możliwości analizy danych z funkcją tworzenia estetycznych i w pełni konfigurowalnych wykresów – każdy badacz i analityk zna frustrację, która towarzyszy przeklejaniu tabel wynikowych do Excela lub PowerPointa, a potem formatowaniu każdego wykresu oddzielnie. Nawet po zakończeniu pracy nad raportem, każde powtórzenie pomiaru wymaga wykonania szeregu czasochłonnych czynności, zanim powstanie raport z kolejnej edycji badania. Wszystko to wydłuża czas pracy, utrudnia jej automatyzację i zwiększa ryzyko błędów.

Język programowania R oferuje niemal nieograniczone możliwości analizy i wizualizacji danych. Jest elastyczny i wszechstronny – umożliwia analizę danych o dowolnej strukturze, w dokładnie taki sposób, jaki jest nam potrzebny. Umożliwia zarówno wykonywanie doraźnych analiz eksploracyjnych, jak i świetnie się nadaje do automatyzacji zadań powtarzalnych. Pozwala na tworzenie estetycznych, interaktywnych wykresów. Każdy badacz słyszał o możliwościach R, ale panuje powszechna obawa przed jego używaniem, ze względu na konieczność nauki programowania.

Zamierzam pokazać, że korzystanie z pakietu R wcale nie jest trudne, i umożliwić bezproblemowe przejście przez pierwszy etap nauki. Uczestnicy szkolenia zostaną wyposażeni w podstawowe umiejętności i szereg praktycznych podpowiedzi, które umożliwią im stopniowe zwiększanie biegłości w korzystaniu z R w toku codziennej pracy z danymi.

Szczegółowy zakres szkolenia

- Prezentacja zalet, możliwości i zastosowań R
- Instalacja narzędzi i zapoznanie z interfejsem RStudio
- Pakiety: ich rola, możliwości, instalacja i ładowanie
- Typy zmiennych, konwersje między typami
- Przetwarzanie zmiennych tekstowych, logicznych i numerycznych
- Tworzenie i przekształcanie macierzy i ramek danych
- Przeszukiwanie obiektów wielowymiarowych i odwoływanie się po adresie do ich elementów
- Selekcja przypadków i zmiennych
- Programowanie w potokach: operator `%>%`
- Opis statystyczny jednej zmiennej: wyliczanie parametrów i opis rozkładu
- Definiowanie własnych funkcji statystycznych
- Wczytywanie danych z zewnętrznych źródeł
- Przekształcanie struktury danych z formatu długiego w format szeroki i na odwrót
- Łączenie baz danych wierszami i kolumnami
- Rekodowanie zmiennych i tworzenie nowych zmiennych jako funkcji zmiennych istniejących
- Wizualizacja, opis i testowanie istotności współzależności między zmiennymi:
 - tabele kontyngencji
 - porównania średnich w podgrupach
 - analiza korelacji
 - model regresji liniowej

Do kogo kierowane jest szkolenie

Kurs podstawowy kierowany jest do badaczy ilościowych i analityków danych, którzy chcą zacząć wykorzystywać R do analizy danych z badań ankietowych i innych źródeł. Osoby mające doświadczenie w pracy z R mają szansę odświeżenia i uporządkowania posiadanej wiedzy. Warsztat stanowi wprowadzenie do posługiwania się R - sprawne wykorzystywanie go w pracy badawczej będzie wymagało dalszego szlifowania umiejętności po zakończeniu kursu.

Korzyści jakie odniosą uczestnicy

Po ukończeniu tego szkolenia uczestnik będzie potrafił:

- Instalować R i RStudio, oraz sprawnie pozyskiwać i uruchamiać nowe pakiety
- Korzystać z funkcjonalności RStudio, wspomagających programowanie w R

- Korzystać z pomocy i dokumentacji wbudowanej w język R
- Posługiwać się szeroką paletą funkcji do przetwarzania zmiennych numerycznych, logicznych i tekstowych
- Tworzyć i przekształcać wielowymiarowe struktury danych, takie jak macierze i ramki danych
- Przeszukiwać ww. struktury danych i odwoływać się do ich zawartości
- Wykorzystywać operator `%>%` do tworzenia potoków operacji na danych
- Definiować własne funkcje statystyczne
- Wczytywać dane z zewnętrznych źródeł: pliki tekstowe, arkusze Excel, bazy SPSS, internet
- Zmieniać format danych z formatu baz relacyjnych („długiego”) na częściej stosowany w badaniach format „szeroki”
- Łączyć bazy danych wierszami (dodawanie przypadków) i kolumnami (dodawanie zmiennych)
- Rekodować zmienne i dokonywać podstawowych operacji czyszczenia danych
- Przeprowadzać podstawowe analizy statystyczne:
 - opis parametryczny jednej zmiennej
 - tabele rozkładu jednej zmiennej
 - tabele kontyngencji
 - porównania między średnimi (test t, ANOVA)
 - analiza korelacji
 - model regresji liniowej
- Eksportować dane, wyniki analiz i wykresy do zewnętrznych katalogów

Kluczowe pojęcia

Pojęcia z zakresu programowania: obiekt, zmienna, typ zmiennej, konwersja typów, konkatenacja, generator liczb pseudolosowych, funkcja, argument funkcji, argument domyślny, macierz, ramka danych, adres, format długi i szeroki, klucz, pętla, inkrementacja, plik CSV (comma-separated values).

Pojęcia statystyczne: średnia, mediana, kwartył, minimum, maksimum, rozstęp, wariancja, odchylenie standardowe, odchylenie ćwiartkowe, zmienna kategoryjna, zmienna ciągła, rozkład zmiennej, rekodowanie zmiennej, liczebność, częstość, rozkład brzegowy, rozkład warunkowy, rozkład łączny, tabela kontyngencji, niezależność stochastyczna, istotność statystyczna, test t, analiza wariancji, korelacja, współczynnik korelacji, regresja liniowa, histogram, wykres pudełkowy, wykres rozrzutu.

Forma szkolenia

Szkolenie będzie miało formę warsztatową. Zdecydowaną większość czasu spędzimy, pisząc i testując kod na własnych komputerach. Moduły tematyczne składające się z teorii i przykładów będą przedzielane ćwiczeniami indywidualnymi, służącymi utrwaleniu materiału.

Stopień zaawansowania szkolenia:

Kurs podstawowy przeznaczony jest dla osób bez, lub z niewielkim doświadczeniem w posługiwaniu się R. Podstawowa wiedza na temat stosowanych w badaniach analiz, jak np. tabele krzyżowe, korelacje czy testy istotności będzie pomocna, ale zagadnienia te w razie potrzeby zostaną objaśnione. Doświadczenie w programowaniu w innych językach nie jest wymagane.

Prowadzi:



Marek Młodożeniec

sotrender

Analityk danych z 15-letnim doświadczeniem w badaniach. Badacz rynku (MDM Dendrite, Millward Brown), specjalista R&D (Gemius, 4P Research Mix), Data Scientist (OPI PIB, Stratosfera by Deloitte), obecnie zajmuje się analizą ruchu w sieci w firmie Sotrender. Szkolenia

dla badaczy oraz zajęcia ze studentami z zakresu statystyki, programowania i wizualizacji danych prowadzi od 2008 roku.

Tytuł: R dla badaczy – poziom podstawowy

Data: 5 i 6 listopada 2019 r.

Miejsce: PTBRiO, Szarotki 11, Warszawa

Koszt do 15 października: 1499 zł, później 1699 zł netto

Kontakt: szkolenia@ptbriio.pl, 22 648 44 92

>> Więcej <<

>> Zgłoś się <<