
Analiza danych behawioralnych przy użyciu języków R i Python

*Wykorzystanie danych zorientowanych
na użytkownika w celu uzyskania
realnych wyników biznesowych*

Florent Buisson

przekład: Jacek Janusz

Spis treści

Wstęp	xi
-------------	----

Część I. Zrozumienie zachowań

1. Koncepcja przyczynowo-behawioralna stosowana w analizie danych	3
Dlaczego do wyjaśniania ludzkich zachowań należy zastosować analizę przyczynową?	4
Różne rodzaje analizy	4
Istoty ludzkie są skomplikowane	5
Zakłócenia, czyli ukryte niebezpieczeństwa rozwiązywania problemów za pomocą regresji	8
Dane	9
Dlaczego korelacja nie jest związkiem przyczynowym? Rola czynnika zakłócającego	9
Zbyt wiele zmiennych może zepsuć zabawę	11
Podsumowanie	17
2. Zrozumienie danych behawioralnych	19
Podstawowy model ludzkiego zachowania	20
Cechy osobowe	21
Poznanie i emocje	23
Intencje	24
Działania	25
Zachowania biznesowe	26
Jak połączyć ze sobą zachowania i dane?	28
Zdefiniowanie sposobu myślenia pozwalającego osiągnąć integralność behawioralną	28
Nieufność i weryfikacja	29
Identyfikacja kategorii	30
Dostrajanie zmiennych behawioralnych	32
Zrozumienie kontekstu	33
Podsumowanie	36

Część II. Diagramy przyczynowe i usuwanie czynników zakłócających

3. Wprowadzenie do diagramów przyczynowych	39
Diagramy przyczynowe i koncepcja przyczynowo-behawioralna.....	40
Diagramy przyczynowe reprezentują zachowania	41
Diagramy przyczynowe reprezentują dane	42
Podstawowe struktury diagramów przyczynowych	46
Łańcuchy.....	47
Rozgałęzienia	50
Zderzacze	52
Typowe przekształcenia diagramów przyczynowych	53
Dzielenie (dezagregacja) zmiennych.....	54
Agregacja zmiennych	55
Co z cyklami?.....	57
Ścieżki	60
Podsumowanie	61
4. Tworzenie diagramów przyczynowych od podstaw	63
Opis problemu biznesowego i konfiguracji danych	64
Dane i pakiety.....	64
Zrozumienie głównej relacji.....	65
Identyfikacja zmiennych, które mogą zostać uwzględnione w diagramie przyczynowym.....	67
Działania.....	68
Intencje	70
Poznanie i emocje	72
Cechy osobowe.....	72
Zachowania biznesowe	74
Trendy czasowe	75
Walidacja obserwowalnych zmiennych w oparciu o dane.....	76
Relacje między zmiennymi numerycznymi.....	77
Relacje między zmiennymi skategoryzowanymi.....	80
Relacje między zmiennymi numerycznymi a skategoryzowanymi	82
Iteracyjne rozbudowywanie diagramu przyczynowego	84
Identyfikacja pośredników dla zmiennych nieobserwowalnych.....	84
Identyfikacja dalszych przyczyn	86
Iteracje.....	86
Uproszczenie diagramu przyczynowego.....	87
Podsumowanie	88

5. Używanie diagramów przyczynowych do usuwania czynników zakłócających z analiz danych ..	89
Problem biznesowy: sprzedaż lodów i wody butelkowanej	90
Rozłączne kryterium ustalania przyczyny	92
Definicja	92
Blok pierwszy	93
Blok drugi	95
Kryterium tylnej furtki	95
Definicje	96
Blok pierwszy	98
Blok drugi	99
Podsumowanie	101

Część III. Profesjonalna analiza danych

6. Rozwiązywanie problemu brakujących danych	105
Dane i pakiety	107
Wizualizacja brakujących danych	108
Ilość brakujących danych	111
Korelacja braków danych	113
Rozpoznawanie brakujących danych	118
Przyczyny braków danych – klasyfikacja Rubina	121
Rozpoznawanie zmiennych MCAR	123
Rozpoznawanie zmiennych MAR	124
Rozpoznawanie zmiennych MNAR	127
Brak danych jako skala	129
Obsługiwanie braku danych	132
Wprowadzenie do imputacji wielokrotnej	133
Domyślna metoda imputacji: predykcyjne dopasowanie średniej	136
Od PMM do imputacji z rozkładem normalnym (tylko język R)	138
Dodawanie zmiennych pomocniczych	140
Skalowanie liczby uzupełnianych zbiorów danych	142
Podsumowanie	142
7. Ocenianie niepewności za pomocą metody bootstrap	145
Wprowadzenie do metody bootstrap: „odpytywanie” samego siebie	146
Pakiety	146
Problem biznesowy: niewielki zbiór danych z wartościami odstającymi	146
Bootstrapowy przedział ufności dla średniej z próbki danych	148

Bootstrapowe przedziały ufności w przypadku doraźnych statystyk	153
Wykorzystanie metody bootstrap w analizie regresji	155
Kiedy należy używać metody bootstrap?	159
Warunki wystarczające do zastosowania tradycyjnych metod szacowania wartości centralnej	160
Warunki wystarczające do wyznaczenia zwykłego przedziału ufności	160
Ustalanie liczby prób bootstrapowych	163
Optymalizacja metody bootstrap w R i Pythonie	164
Język R – pakiet boot.	164
Optymalizacja dostępna w Pythonie.	167
Podsumowanie	168

Część IV. Projektowanie i analizowanie eksperymentów

8. Projektowanie eksperymentów – podstawy.	171
Planowanie eksperymentu – teoria zmiany	172
Cel biznesowy i wskaźnik docelowy	173
Interwencja	175
Logika behawioralna	177
Dane i pakiety.	179
Ustalenie randomizacji i wielkości/mocy próby	180
Randomizacja.	180
Wielkość próby i analiza mocy	183
Analizowanie i interpretowanie wyników eksperymentów	198
Podsumowanie	201
9. Randomizacja warstwowa	203
Planowanie eksperymentu	205
Cel biznesowy i wskaźnik docelowy	205
Zdefiniowanie interwencji	207
Logika behawioralna	208
Dane i pakiety.	208
Określenie losowego przypisania i wielkości/mocy próby.	209
Losowe przypisanie	209
Analiza mocy za pomocą symulacji bootstrapowych.	218
Analizowanie i interpretowanie wyników eksperymentu	225
Oszacowanie współczynnika ITT w przypadku interwencji zachęcającej.	225
Wyznaczanie wskaźnika CACE w przypadku interwencji obowiązkowej.	227
Podsumowanie	233

10. Randomizacja klastrowa i modelowanie hierarchiczne	235
Zaplanowanie eksperymentu	236
Cel biznesowy i wskaźnik docelowy	236
Definicja interwencji	236
Logika behawioralna	238
Dane i pakiety	238
Wprowadzenie do modelowania hierarchicznego	239
Kod języka R	240
Kod języka Python	243
Określanie losowego przypisania i wielkości/mocy próby	244
Przypisanie losowe	245
Analiza mocy	247
Analiza eksperymentu	255
Podsumowanie	256

Część V. Użycie zaawansowanych narzędzi w analizie danych behawioralnych

11. Wprowadzenie do moderacji	261
Dane i pakiety	261
Behawioralne odmiany moderacji	262
Segmentacja	262
Interakcje	268
Nieliniowości	270
Jak stosować moderację?	272
W jakich przypadkach należy stosować moderację?	273
Wiele moderatorów	283
Walidacja moderacji za pomocą metody bootstrap	288
Interpretacja poszczególnych współczynników	291
Podsumowanie	296
12. Mediacja i zmienne instrumentalne	299
Mediacja	300
Zrozumienie mechanizmów przyczynowych	300
Zniekształcenia pojawiające się podczas ustalania przyczyn	301
Identyfikacja mediacji	303
Mierzenie mediacji	304
Zmienne instrumentalne	309
Dane	309

Zrozumienie i zastosowanie zmiennych instrumentalnych	310
Pomiar	313
Stosowanie zmiennych instrumentalnych – najczęściej zadawane pytania. . .	316
Podsumowanie	317
Bibliografia	319
Indeks	323
O autorze	344
Kolofon	345