

Spis treści

Wstęp	9
1. Podstawy przetwarzania danych w SAS Viya	13
1.1. Wprowadzenie	13
1.2. Połączenie z serwerem CAS w SAS Studio	14
1.3. Podstawy programowania w SAS Viya	16
1.4. Praca ze źródłami danych	16
1.4.1. Wczytywanie danych	16
1.4.2. Eksportowanie danych	22
1.5. Przeglądanie danych w SAS Data Explorer	23
1.6. Wyświetlanie atrybutów i przeglądanie danych	24
1.7. Podstawowe przekształcenia danych	30
1.7.1. Filtrowanie danych	30
1.7.2. Sortowanie danych i przetwarzanie w grupach	33
1.7.3. Formatowanie danych	34
1.7.4. Tworzenie podstawowych zapytań SQL	35
1.8. Wprowadzenie do raportowania w plikach zewnętrznych	39
1.8.1. Tworzenie wykresu w kodzie programu	40
1.8.2. Tworzenie raportu w kodzie programu – procedury drukujące	41
1.8.3. Wykresy interaktywne w SAS Visual Analytics	43
2. Wybrane zagadnienia z zakresu metody reprezentacyjnej	61
2.1. Wprowadzenie	61
2.2. Wybrane schematy losowania próby	63
2.2.1. Losowanie proste	63

2.2.2. Losowanie warstwowe	65
2.2.3. Losowanie systematyczne	68
2.2.4. Losowanie dwu- i wielostopniowe	71
2.3. Próbkowanie na potrzeby modelowania predykcyjnego	72
2.3.1. Losowanie z nadreprezentacją klasy mniejszościowej – oversampling	73
2.3.2. Losowanie warstwowe z ograniczeniem udziału klasy większościowej – undersampling	74
2.3.3. Losowanie warstwowe – k -krotny podział próby	75
3. Analiza struktury	77
3.1. Wprowadzenie	77
3.2. Tabele liczebności i częstości	78
3.3. Miary analizy struktury	85
3.3.1. Miary położenia rozkładu	86
3.3.2. Miary zróżnicowania	91
3.3.3. Miary asymetrii	93
3.3.4. Miary koncentracji	94
3.4. Obserwacje odstające i miary odporne	106
4. Rozkłady teoretyczne zmiennych losowych	117
4.1. Wprowadzenie	117
4.2. Graficzna prezentacja rozkładów wybranych zmiennych losowych	118
4.2.1. Rozkład zero-jedynkowy (Bernoulliego)	118
4.2.2. Rozkład dwumianowy	121
4.2.3. Rozkład beta	123
4.2.4. Rozkład normalny	126
4.2.5. Rozkład lognormalny	128
4.2.6. Dobór parametrów rozkładu za pomocą interaktywnego wykresu. Zastosowanie pakietu R Shiny	131
4.3. Nieparametryczny estymator funkcji gęstości prawdopodobieństwa	134
4.4. Badanie zgodności rozkładu empirycznego z rozkładem teoretycznym	141
4.5. Generowanie danych z określonego rozkładu za pomocą procedury IML	152

5. Analiza wariancji i kowariancji	157
5.1. Wprowadzenie	157
5.2. Analiza wariancji z pojedynczą klasyfikacją	160
5.3. Analiza wariancji z podwójną klasyfikacją	184
5.4. Nieparametryczna analiza wariancji – test Kruskala–Wallisa	204
5.5. Analiza kowariancji	211
5.6. Porównanie modeli analizy wariancji i kowariancji	220
6. Analiza korelacji i zależności	225
6.1. Wprowadzenie	225
6.2. Współczynnik korelacji liniowej Pearsona	226
6.3. Nieparametryczne miary zależności	233
6.3.1. Współczynnik korelacji rang Spearmana	233
6.3.2. Współczynnik τ -Kendalla	234
6.3.3. Miara zależności Hoeffdinga	235
6.4. Test niezależności χ^2 i współczynnik V Craméra w ocenie asocjacji	238
6.5. Współczynnik Cronbacha	244
6.6. Ocena związków pomiędzy zmiennymi na potrzeby modelowania	247
7. Regresja liniowa	249
7.1. Wprowadzenie	249
7.2. Model regresji liniowej	249
7.2.1. Klasyczny model regresji liniowej	249
7.2.2. Regresja wieloraka	253
7.3. Diagnostyka modelu	255
7.3.1. Ocena zdolności predykcyjnej	255
7.3.2. Identyfikacja obserwacji odstających i wpływowych	260
7.3.3. Dobór zmiennych do modelu	272
7.3.4. Ocena specyfikacji modelu	284
7.4. Regresja adaptacyjna – dopasowanie funkcjami sklejanymi	287

8. Regresja kwantylowa	295
8.1. Wprowadzenie	295
8.2. Model regresji kwantylowej	296
8.3. Estymacja modelu regresji kwantylowej	299
8.4. Identyfikacja obserwacji odstających za pomocą modelu regresji kwan- tylowej	306
8.5. Zarys budowy aplikacji za pomocą biblioteki R Shiny na przykładzie modeli regresji	311
Aneks	317
Bibliografia	319