

---

# Spis treści

|                                                 |           |
|-------------------------------------------------|-----------|
| <b>Przedmowa .....</b>                          | <b>7</b>  |
| <b>1. Eksploracyjna analiza danych .....</b>    | <b>13</b> |
| Dowody                                          | 13        |
| Badania NSFG                                    | 15        |
| Wczytywanie danych                              | 16        |
| Sprawdzanie poprawności                         | 18        |
| Transformacja                                   | 21        |
| Statystyki podsumowujące                        | 22        |
| Interpretacja                                   | 23        |
| Słowniczek                                      | 24        |
| Ćwiczenia                                       | 25        |
| <b>2. Rozkłady .....</b>                        | <b>26</b> |
| Tabele częstości                                | 26        |
| Rozkłady danych z badań NSFG                    | 28        |
| Wartości odstające                              | 31        |
| Pierwsze dzieci                                 | 33        |
| Wielkość efektu                                 | 34        |
| Prezentacja wyników                             | 36        |
| Słowniczek                                      | 36        |
| Ćwiczenia                                       | 37        |
| <b>3. Funkcje masy prawdopodobieństwa .....</b> | <b>39</b> |
| Funkcje masy prawdopodobieństwa                 | 39        |
| Generowanie podsumowań dotyczących obiektu Pmf  | 42        |
| Paradoks wielkości grup                         | 43        |
| Dane z badań NSFG                               | 46        |
| Inne formy wizualizacji                         | 47        |
| Słowniczek                                      | 48        |
| Ćwiczenia                                       | 49        |

|                                                                                     |            |
|-------------------------------------------------------------------------------------|------------|
| <b>4. Dystrybuanta .....</b>                                                        | <b>51</b>  |
| Percentyle i wyniki centylowe                                                       | 51         |
| Dystrybuanty                                                                        | 54         |
| Porównywanie dystrybuant                                                            | 57         |
| Statystyki oparte na percentylach                                                   | 59         |
| Liczby losowe                                                                       | 62         |
| Słowniczek                                                                          | 64         |
| Ćwiczenia                                                                           | 64         |
| <b>5. Modelowanie rozkładów prawdopodobieństwa .....</b>                            | <b>66</b>  |
| Rozkład dwumianowy                                                                  | 66         |
| Rozkład Poissona                                                                    | 71         |
| Rozkład wykładniczy                                                                 | 74         |
| Rozkład normalny                                                                    | 77         |
| Rozkład logarytmicznie normalny                                                     | 80         |
| Po co tworzyć modele?                                                               | 83         |
| Słowniczek                                                                          | 84         |
| Ćwiczenia                                                                           | 84         |
| <b>6. Funkcje gęstości prawdopodobieństwa .....</b>                                 | <b>86</b>  |
| Porównywanie rozkładów                                                              | 86         |
| Funkcja gęstości prawdopodobieństwa                                                 | 89         |
| Funkcja gęstości prawdopodobieństwa dla rozkładu wykładniczego                      | 92         |
| Porównanie funkcji masy prawdopodobieństwa i funkcji gęstości<br>prawdopodobieństwa | 94         |
| Estymacja jądrowa gęstości                                                          | 95         |
| Model reprezentacji rozkładów                                                       | 99         |
| Słowniczek                                                                          | 104        |
| Ćwiczenia                                                                           | 104        |
| <b>7. Zależności między zmiennymi .....</b>                                         | <b>107</b> |
| Wykresy punktowe                                                                    | 107        |
| Wykresy decylowe                                                                    | 111        |
| Korelacja                                                                           | 113        |
| Siła korelacji                                                                      | 117        |
| Korelacja rangowa                                                                   | 118        |
| Korelacja i przyczynowość                                                           | 122        |
| Słowniczek                                                                          | 122        |
| Ćwiczenia                                                                           | 124        |

|                                                 |            |
|-------------------------------------------------|------------|
| <b>8. Szacowanie .....</b>                      | <b>127</b> |
| Ważenie pingwinów                               | 127        |
| Odporność                                       | 131        |
| Szacowanie wariancji                            | 132        |
| Rozkłady próbkowania                            | 133        |
| Błąd standardowy                                | 136        |
| Przedziały ufności                              | 137        |
| Źródła błędów                                   | 138        |
| Słowniczek                                      | 138        |
| Ćwiczenia                                       | 140        |
| <b>9. Testowanie hipotez .....</b>              | <b>144</b> |
| Rzuty monetą                                    | 144        |
| Testowanie różnicy średnich                     | 146        |
| Inne statystyki testowe                         | 149        |
| Badanie korelacji                               | 150        |
| Testy proporcji                                 | 152        |
| Słowniczek                                      | 156        |
| Ćwiczenia                                       | 156        |
| <b>10. Metoda najmniejszych kwadratów .....</b> | <b>158</b> |
| Metoda najmniejszych kwadratów                  | 158        |
| Współczynnik determinacji                       | 162        |
| Minimalizowanie błędu średniokwadratowego       | 163        |
| Szacowanie                                      | 165        |
| Wizualizowanie niepewności                      | 166        |
| Przekształcenia                                 | 168        |
| Słowniczek                                      | 173        |
| Ćwiczenia                                       | 174        |
| <b>11. Regresja wieloraka .....</b>             | <b>176</b> |
| StatsModels                                     | 176        |
| Regresja wieloraka                              | 179        |
| Zmiennie kontrolne                              | 181        |
| Zależności nieliniowe                           | 185        |
| Regresja logistyczna                            | 187        |
| Słowniczek                                      | 191        |
| Ćwiczenia                                       | 192        |

|                                                      |            |
|------------------------------------------------------|------------|
| <b>12. Analiza szeregów czasowych .....</b>          | <b>194</b> |
| Energia elektryczna                                  | 194        |
| Dekompozycja danych                                  | 195        |
| Predykcje                                            | 201        |
| Model multiplikatywny                                | 205        |
| Autoregresja                                         | 209        |
| Średnia ruchoma                                      | 211        |
| Retrodykcje z wykorzystaniem autoregresji            | 213        |
| ARIMA                                                | 214        |
| Generowanie predykcji z wykorzystaniem funkcji ARIMA | 216        |
| Słowniczek                                           | 217        |
| Ćwiczenia                                            | 218        |
| <b>13. Analiza przeżycia .....</b>                   | <b>222</b> |
| Funkcje przeżycia                                    | 222        |
| Funkcja hazardu                                      | 224        |
| Dane o stanie cywilnym                               | 226        |
| Bootstrapping z wagami                               | 229        |
| Szacowanie funkcji hazardu                           | 230        |
| Szacowanie funkcji przeżycia                         | 233        |
| Pakiet lifelines                                     | 235        |
| Przedziały ufności                                   | 236        |
| Przewidywany pozostały czas do zdarzenia             | 238        |
| Słowniczek                                           | 241        |
| Ćwiczenia                                            | 241        |
| <b>14. Metody analityczne .....</b>                  | <b>244</b> |
| Wykresy prawdopodobieństwa normalnego                | 244        |
| Rozkłady normalne                                    | 248        |
| Rozkład średnich z prób                              | 251        |
| Rozkład różnic                                       | 253        |
| Centralne twierdzenie graniczne                      | 255        |
| Ograniczenia centralnego twierdzenia granicznego     | 257        |
| Zastosowanie centralnego twierdzenia granicznego     | 259        |
| Test korelacji                                       | 262        |
| Test chi-kwadrat                                     | 265        |
| Informatyka i analiza algorytmów                     | 267        |
| Słowniczek                                           | 268        |
| Ćwiczenia                                            | 269        |