Statystyka dla studentów kierunków technicznych i przyrodniczych

Jacek Koronacki, Jan Mielniczuk

Spis treści

Przedmowa

Miniony wiek bywa nazywany wiekiem informacji. Moc obliczeniowa komputerów oraz pojemność ich pamięci rosły w ostatnich dziesięcioleciach nieomal z dnia na dzień. Doświadczyliśmy i nadal doświadczamy niebywałego rozwoju możliwości komunikacji w sieciach komputerowych. Oczywiście wiązał się z tym wszystkim ogromny wzrost możliwości gromadzenia informacji. W przypadku dużych baz danych mamy często do czynienia z megabajtami i nierzadko z terabajtami danych. Wielkiego znaczenia nabrała zatem potrzeba inteligentnego przetwarzania zebranych informacji.

Niejako w cieniu owej rewolucji informatycznej przez cały XX wiek trwał też niezwykły rozwój analizy danych i wnioskowania statystycznego, czyli – statystyki. Obydwa procesy nie były przy tym od siebie niezależne. Z jednej strony, bez statystyki nie ma możliwości pełnego zrozumienia i zinterpretowania wiedzy ukrytej w danych. Z drugiej zaś, techniczny rozwój komputerów umożliwił zalgorytmizowanie procedur statystycznych i rozwiązywanie zadań, z którymi człowiek sam nie mógłby sobie poradzić przez dziesiątki lat wytężonej pracy.

Jednym ze skutków rozwoju informatyki i statystyki jest upowszechnienie badań statystycznych w niemal we wszystkich dziedzinach nauki i praktyki. We wszystkich sferach naszej działalności zbieramy dane, które – nie poddane odpowiedniej analizie – jawią się raczej jako niewiele mówiący chaos niż pewne uporządkowane *uniwersum*. Dzięki statystyce dostrzegamy ów ukryty w danych porządek oraz patrzymy na dane z właściwej perspektywy, tak jak autor obrazu, którego reprodukcję zamieszczamy na okładce, spojrzał na ulicę Madrytu – z bliska pełną południowego zamętu i chaotycznie poruszających się ludzi, a z perspektywy będącą częścią pięknego i zrozumiałego ładu.

Podręcznik ten jest wprowadzeniem w ład oferowany przez statystykę. Jest adresowany przede wszystkim do przyszłych techników i przyrodników, ale

uważamy, że będzie przydatny także dla studentów innych kierunków, zwłaszcza ekonomicznych, rolniczych, społecznych i medycznych. Powinien również zainteresować tych absolwentów wszystkich wymienionych kierunków, którzy uważają, że ich podstawowa wiedza statystyczna jest niedostateczna.

Oddawany do rąk Czytelnika podręcznik odbiega stylem od książek ze statystyki matematycznej już choćby dlatego, że jest adresowany do niematematyków. Swoją konstrukcją nawiązuje do anglosaskiej tradycji uczenia statystyki, którego celem jest danie dogłębnego i szerokiego, ale zarazem możliwie przystępnego wprowadzenia do przedmiotu.

Jesteśmy przekonani, że pojawienie się tego rodzaju podręcznika jest potrzebne, ponieważ przyczyni się do podniesienia poziomu zrozumienia i popularności statystyki wśród studentów. Niezbędne jest ukazanie się książki ujmującej wprowadzenie do statystyki w sposób całościowy, w pełni wykorzystującej nowoczesne techniki obliczeniowe. Jednocześnie Czytelnik musi nauczyć się wykorzystywania owych narzędzi w sposób odpowiedzialny i oparty na dobrym zrozumieniu przedmiotu. Nie trzeba nikogo przekonywać, że obecność na rynku licznych komputerowych pakietów statystycznych jest tyleż błogosławieństwem, co i przekleństwem, bowiem łatwo z nich korzystać bez żadnego zrozumienia oferowanych przez pakiet wyników.

Środkiem do lepszego zrozumienia przedmiotu nie może być nadmierna ścisłość formalna i przytaczanie wielu dowodów, lecz oparcie się na pouczających choć prostych przykładach i szerokiej argumentacji, odwołującej się do zdrowego rozsądku. Tak właśnie pisany jest nasz podręcznik. Jesteśmy przekonani, że w ten sposób można doskonale przekazać istotę rozumowania statystycznego. Za złożonym nawet matematycznym wywodem zawsze kryje się przejrzysta intuicja. To ją przede wszystkim powinen posiąść Czytelnik.

Zakres trzech pierwszych rozdziałów książki (rozdział zawierający wstępną analizę danych, rozdział poświęcony przejściu od modelu probabilistycznego do wnioskowania statystycznego oraz rozdział opisujący podstawy wnioskowania statystycznego) odpowiada typowym uczelnianym kursom ze statystyki; jednak różni się sposobem ujęcia materiału. W sposobie wykładu oraz wyborze tematów szczegółowych kierujemy się potrzebami praktyki oraz swymi doświadczeniami dydaktycznymi z uczelni w Warszawie (w ostatnich latach PJWSTK), The University of Michigan w Ann Arbor, Rice University w Houston i The University of New South Wales w Sydney. Niewątpliwy wpływ wywarły na nas najlepsze podręczniki anglosaskie, zwłaszcza książka Moore'a i McCabe'a „Introduction to the Practice of Statistics", Freeman & Co 1998, którą najczęściej sami wykorzystywaliśmy w nauczaniu.

Trzy pierwsze rozdziały tego podręcznika uzupełnione o omówioną w rozdziale 4 analizę regresji są pomyślane jako podstawa kursu semestralnego,

wprowadzającego słuchacza w zagadnienia statystyki i obejmującego tygodniowo przynajmniej dwugodzinny wykład oraz dwugodzinne laboratorium. W ramach kursu semestralnego udaje się omówić tylko zasadnicze kwestie analizy regresji. W pięciu następnych rozdziałach przedstawiono wybrane, najbardziej istotne dla praktyka zagadnienia statystyki: analizę wariancji i analizę zależności cech jakościowych, metody próbkowania, zagadnienia symulacji komputerowej i metod rangowych. Na podstawie tych rozdziałów oraz zaprezentowanej obszerniej analizy regresji, wykładowca może zaplanować drugi semestr wykładu ze statystyki.

Wśród zagadnień szczegółowych nie znalazło się miejsce dla niezwykle ważnych metod statystyki wielowymiarowej, które – mamy nadzieję – staną się treścią naszego następnego podręcznika. Nie mamy przy tym wątpliwości, że ze względu na występującą obecnie złożoność danych metody wielowymiarowe staną się już wkrótce elementem podstawowych wykładów ze statystyki.

Chcielibyśmy podkreślić fakt, że książka powstała w ramach działalności statutowej Instytutu Podstaw Informatyki Polskiej Akademii Nauk. W trakcie przygotowywania kolejnych wersji manuskryptu, korzystaliśmy z wnikliwych uwag Stanisława Gnota, Andrzeja Dąbrowskiego i Andrzeja Michalskiego oraz Elżbiety Ferenstein, którym serdecznie dziękujemy. Jesteśmy bardzo wdzięczni naszemu najbliższemu współpracownikowi, Janowi Ćwikowi, który sporządził wszystkie rysunki, przygotował ostateczny skład książki, przeliczył i sprawdził wiele przykładów oraz pomagał nam w trakcie kolejnych korekt. Składamy podziękowania Polsko-Japońskiej Wyższej Szkole Technik Komputerowych za finansowe wsparcie wydania naszego podręcznika. Dziękujemy Muzeum Narodowemu w Warszawie za wyrażenie zgody na reprodukcję obrazu Józefa Pankiewicza „Ulica w Madrycie". Z wdzięcznością myślimy o znakomitej pracy redakcyjnej Pani Lilianny Szymańskiej i o opiece Pani Redaktor Zofii Leszczyńskiej nad całością przedsięwzięcia. Bez ich wielkiego poświęcenia i zaangażowania szybkie wydanie tej książki byłoby niemożliwe.

Jacek Koronacki i Jan Mielniczuk

Warszawa, w lipcu 2001