

Ewa Pogorzala
Zamość
ewapogo@interia.pl

„Zastosowania statystyki i data mining w badaniach naukowych”. Seminarium szkoleniowe, Warszawa, 23 października 2008 r.

We wrześniu i październiku StatSoft Polska zorganizowała w Warszawie seminarium pt. „Zastosowania statystyki i data mining”. Spotkania poświęcone były przykładom praktycznego wykorzystania statystyki i data mining w różnych obszarach zastosowań oraz przykładom pracy w programach *STATISTICA*. Na cykl seminariów złożyło się zorganizowane w dn. 30 września spotkanie dotyczące analizy danych w programie *STATISTICA*, którego celem było przedstawienie praktycznych aspektów pracy w programach z rodziny *STATISTICA*. Kolejne spotkanie odbyło się w dn. 23 października, a złożyły się na nie dwa równoległe panele. Jeden z nich poświęcony był praktycznemu wykorzystaniu analizy i data mining, a adresowany był do osób zainteresowanych wykorzystaniem analizy danych do wspomagania codziennej działalności przedsiębiorstw i administracji publicznej. Drugi panel dotyczył zastosowania statystyki i data mining w badaniach naukowych. W trakcie tego spotkania zaprezentowane zostały metody analizy danych i data mining oraz ich zastosowanie w różnych dziedzinach badań naukowych. Warto nadmienić, iż firma StatSoft od 1998 r. organizuje, pod patronatem Polskiego Towarzystwa Statystycznego konkurs na najlepszą pracę doktorską i magisterską przygotowaną z zastosowaniem narzędzi statystyki i analizy danych zawartych w programach z rodziny *STATISTICA*. Celem konkursu jest promocja autorów najlepszych prac oraz popularyzacja nowoczesnych metod analizy danych, dostępnych w oferowanym przez firmę oprogramowaniu¹.

Seminarium w Warszawie rozpoczęło wystąpienie prof. Andrzeja Barczaka z Katedry Ekonometrii Akademii Ekonomicznej w Katowicach. Zaprezentował on referat dotyczący analizy ekonomicznych szeregów czasowych. Szczególnie interesujące, z punktu widzenia badań politologicznych, czy szerzej społecznych, było wy-

¹ Więcej informacji na stronie: <http://www.statsoft.pl/>

stąpienie dr Piotra Jabkowskiego z Instytut Socjologii Uniwersytetu im. Adama Mickiewicza na temat wykorzystania metod statystycznych w badaniach nad poczuciem bezpieczeństwa i poziomem przestępczości w Poznaniu. Wyniki przedstawianych w referacie badań stanowiły efekt kilkuletniej pracy zespołu badawczego zajmującego się socjologiczną diagnozą poziomu jakości życia mieszkańców Poznania. Zasadniczy trzon projektu stanowiły badania survey'owe oparte na technice wywiadów kwestionariuszowych, prowadzonych na losowych i reprezentatywnych próbach badawczych dorosłych mieszkańców Poznania. Drugi zasadniczy elementem projektu stanowiły badania desk research, których zadaniem był opis poziomu życia w Poznaniu w oparciu o dane statystyczne GUS oraz dane innych jednostek zajmujących się zbieraniem danych dla statystyki publicznej. Łącznie przeprowadzono pięć serii badań w latach 2002, 2003, 2004, 2006 oraz 2008.

Przedmiotem wybranego do prezentacji fragmentu badań była jakość życia mieszkańców Poznania w zakresie poziomu poczucia bezpieczeństwa oraz poziomu przestępczości w mieście. W referacie omówione zostały najważniejsze wnioski płynące z przeprowadzonych badań. W pierwszej kolejności przedstawione zostały wyniki reprezentatywnych badań na populacji dorosłych mieszkańców Poznania, a następnie zaprezentowane zostały wyniki analiz kryminologiczno-socjologicznych opartych na analizie danych statystycznych dla wszystkich polskich miast na prawach powiatów grodzkich. Prelegent podkreślił, iż w swoim wystąpieniu skupił się jednak bardziej na korzyściach, jakie dało zastosowanie określonych procedur statystycznych, niż na samych wynikach.

W trakcie realizacji projektu w analizach wyników badań sondażowych wykorzystano log-liniową analizę tabel liczebności do zdiagnozowania wpływu cech społeczno-demograficznych respondentów na deklarowany przez nich poziom poczucia bezpieczeństwa, po drugie regresję wielokrotną z eliminacją krokową wsteczną do określenia powiązań poczucia bezpieczeństwa i innych aspektów jakości życia w mieście oraz następujące metody: analizę skupień, standardowe testy nieparametryczne, jak również analizę rzetelności skal pomiarowych. Przy analizie danych statystycznych GUS-u wykorzystano z kolei następujące procedury: regresji wielokrotnej, analizy korelacji kanonicznych, skalowanie wielowymiarowe, analizę skupień (opartą na metodzie k-średnich oraz grupowanie obiektów i cech w analizie skupień), drzewa klasyfikacyjne oraz szeregi czasowe z prognozowaniem.

Równie interesujące, z punktu widzenia możliwości zastosowania metod statystycznych w badaniach politologicznych, było wystąpienie dr Beaty Krawczyk-Bryłki z Katedry Nauk Społecznych Politechniki Gdańskiej poświęcone kontekstowemu modelowi skuteczności negocjacji. Zdaniem prelegentki znaczenie umiejętności ne-

gocjowania na konkurencyjnym, globalnym rynku potęguje zainteresowanie problematyką oceny skuteczności negocjacji w badaniach z zakresu psychologii społecznej. W tym zakresie istotną jest zarówno metodologia oceniania wyników rokowań, jak i określenie determinantów tego procesu. W przeprowadzonych badaniach referentka przyjęła dwupoziomowy model oceny negocjacji obejmujący parametry przebiegu (płynność i klimat) i skuteczności (wynik i użyteczność) negocjacji. Prelegentka wskazała związek tych parametrów z satysfakcją uczestników rokowań oraz zmiennymi, takimi jak: osobowość, płeć, rola i pozycja negocjacyjna oraz kanał komunikacji wykorzystany w trakcie rokowań (negocjacje bezpośrednie i internetowe). Opracowanie kontekstowego modelu negocjacji było możliwe dzięki wsparciu prowadzonych badań systemem komputerowym, który umożliwił zarówno gromadzenie danych, ich archiwizowanie, prowadzenie negocjacji przy użyciu funkcji czat oraz przekazywanie informacji do systemu STATISTICA, w którym były opracowywane.

Pozostałe wystąpienia seminaryjne dotyczyły wybranych metod analizy danych jakościowych na przykładzie wyników badań kardiologicznych (prof. Jerzy Moczko, Katedra Informatyki i Statystyki, Uniwersytet Medyczny w Poznaniu), zastosowania rentgenowskiej mikroskopii fluorescencyjnej i technik wielowymiarowej analizy wariancji w badaniach nad rolą metali śladowych w procesach neurodegeneracyjnych (dr inż. Joanna Chwiej, Katedra Zastosowań Fizyki Jądrowej, AGH) oraz sieci neuronowych i regresji wielorakiej w kontekście złożoności w badaniach naukowych (dr Maciej Szalaniec, Instytut Katalizy i Fizykochemii Powierzchni PAN w Krakowie).

Zasadniczym celem seminarium było ukazanie użyteczności metod statystycznych w wielu różnych dyscyplinach badawczych. W ramach konferencji przedstawiciele firmy STATISTIC zaprezentowali uczestnikom ofertę szkoleń specjalistycznych, która obejmuje kursy takie jak na przykład *Komputerowe wspomaganie planowania i analizy statystycznej doświadczalnych badań innowacyjnych. Metody wizualizacji danych*, czy też *Analiza badań ankietowych*. Są to kursy organizowane na zasadach komercyjnych, przy czym firma oferuje specjalne zniżki dla pracowników uczelni oraz placówek PAN. Wychodząc z założenia, iż w poznawaniu zjawisk i procesów otaczającej nas rzeczywistości przyrodniczej, społecznej i gospodarczej coraz większą rolę odgrywają badania empiryczne, zaś umiejętność odpowiedniego planowania, przeprowadzania i poprawnej analizy wyników takich badań jest obecnie niezbędnym elementem profesjonalnego przygotowania osób, które prowadzą badania naukowe, firma StatSoft w ramach specjalnej „Oferty akademickiej” proponuje uczelniom specjalną licencję na programy i systemy z rodziny *STATISTICA* zapewniającą nieograniczony dostęp do najnowocześniejszych rozwiązań w zakresie statystycznej analizy danych, data mining oraz wizualizacji i graficznej analizy danych.

Jak zasygnalizowano powyżej znajomość metod i technik badań społecznych, w tym z wykorzystaniem metod statystycznych, stanowi istotny element warsztatu badawczego politologów. Znajomość tych metod jest również istotna dla efektywnego wykorzystywania rezultatów badań prowadzonych w różnych dziedzinach, a przydatnych dla analiz politologicznych. Warto zaznaczyć, iż coraz więcej firm jest zainteresowanych współpracą z instytucjami naukowymi. W tym zakresie wspomnieć należy również o Programie ARIADNA², czyli Programie Współpracy SPSS z Jednostkami Naukowymi. SPSS to międzynarodowa firma o ponad trzydziestopięcioletnim doświadczeniu, specjalizująca się w produkcji oprogramowania analitycznego i data mining. Celem programu ARIADANA jest wspieranie jednostek edukacyjnych prowadzących działalność naukową, badawczą lub dydaktyczną z wykorzystaniem oprogramowania SPSS.

² Więcej informacji na stronie: <http://www.ariadna.edu.pl>.