SPIS TREŚCI

WPROWADZENIE (*Jan Paradysz*) 5

1. ZAGADNIENIA WSTĘPNE (*Jan Paradysz*) 11
1.1. Geneza i historia statystyki 11
1.2. Przedmiot badań statystycznych 14
1.3. Podstawowe pojęcia 15
1.4. Źródła statystyczne 18

2. PREZENTACJA DANYCH STATYSTYCZNYCH (*Marcin Szymkowiak*) . 23
2.1. Tablice statystyczne 23
2.2. Wykresy statystyczne 28
2.3. Wykresy statystyczne w Excelu 35
Zadania 48

3. ANALIZA POPULACJI JEDNOWYMIAROWYCH 50
3.1. Zmienne losowe i ich rozkłady (*Aleksandra Witkowska*) 50
3.2. Charakterystyki rozkładu jednej zmiennej (*Elżbieta Gołata*) 74
3.3. Wnioskowanie statystyczne w populacjach jednowymiarowych (*Aleksandra Witkowska*) 127
3.4. Badania próbkowe (*Jan Paradysz*) 169
Zadania 193

4. ANALIZA WSPÓŁZALEŻNOŚCI (*Marek Witkowski*) 225
4.1. Wstęp 225
4.2. Mierniki korelacji dwóch zmiennych 229
4.3. Wybór postaci analitycznej funkcji regresji 239
4.4. Estymacja parametrów funkcji regresji 245

4.5. Model korelacji i regresji liniowej dwóch zmiennych 246
4.6. Korelacja i regresja wielokrotna 267
4.7. Wnioskowanie statystyczne w analizie korelacji i regresji 277
Zadania 282

5. METODY ANALIZY SZEREGÓW CZASOWYCH 293
5.1. Składniki szeregów czasowych i ich pomiar (*Grażyna Dehnel*) 293
5.2. Metody wyodrębniania trendu (*Elżbieta Gołata, Jan Paradysz*) 298
5.3. Wahania sezonowe (*Tomasz Klimanek*) 341
5.4. Składnik koniunktury (*Grażyna Dehnel*) 366
5.5. Indeksy dynamiki (*Grażyna Dehnel*) 371
5.6. Wnioskowanie statystyczne w analizie dynamiki (*Elżbieta Gołata, Marek Witkowski*) 395
Zadania 407

ANEKS 426

LITERATURA 444

WPROWADZENIE

Obecne wydanie podręcznika zostało zmienione w stosunku do poprzednich edycji z lat 1995–2001. Praca została znacznie poszerzona i uzupełniona o dwa nowe rozdziały i kilka podrozdziałów. Wprowadzono nowe przykłady i zadania. W stosunku do pierwszych wydań zamieściliśmy okna dialogowe z Excela, które umożliwiają prześledzenie wykorzystania wybranych funkcji statystycznych, i dodaliśmy najczęściej wykorzystywane tablice statystyczne dla zajęć przeprowadzonych metodą tradycyjną poza laboratorium komputerowym.

Prezentowany podręcznik nie jest tylko zbiorem zadań z przedmiotu statystyka wykładanego w akademiach ekonomicznych. Został przygotowany z uwzględnieniem ostatnich zmian w programach kształcenia oraz współczesnego rozwoju informatyki. Mikrokomputeryzacja umożliwia redukcję czasu na pracochłonne niegdyś obliczenia do niezbędnego minimum. Wprowadzenie komputerów do obsługi procesu dydaktycznego ma wpływ także na zmiany w zakresie programu wykładu. W dużo większym zakresie niż poprzednio można uwzględnić pewne zagadnienia, na które w okresie przedkomputerowym nie było czasu. W miarę bowiem redukowania liczby godzin na statystykę w wyższych szkołach ekonomicznych kilka lat temu, zmuszeni byliśmy ograniczyć metody graficzne, badanie przyczynowości zjawisk gospodarczych oraz niektóre bardziej pracochłonne procedury w analizie dynamiki. Teraz, dzięki komputeryzacji zajęć ze statystki, możemy poświęcić więcej uwagi metodom graficznym, analizie wariancji czy też pogłębionej analizie szeregów czasowych. Możemy także uczyć studentów posługiwania się pakietami statystycznymi. Od bieżącego roku akademickiego część zajęć na Wydziale Zarządzania odbywa się z wykorzystaniem pakietu statystycznego SAS. W przyszłości trzeba będzie również, chociażby informacyjnie, zapoznać studentów ze specjalnymi narzędziami analizy i prezentacji danych, na przykład z komputerowymi systemami

informacji terenowej, zwanymi też Geographic Information Systems (MapInfo, ArcInfo, GeoInfo). Jest to jednak zadanie na przyszłość, gdyż należy starannie przemyśleć, co usunąć z programu statystyki dla ekonomistów i co na to miejsce wprowadzić.

Wybór arkusza kalkulacyjnego jako narzędzia prezentacji metod statystycznych jest podyktowany wieloma względami. Znajomość arkuszy kalkulacyjnych jest coraz powszechniejsza wśród ekonomistów ze względu na uniwersalność ich zastosowań. Każda nowsza wersja Excela ma coraz więcej coraz bardziej rozbudowanych funkcji statystycznych. Arkusz kalkulacyjny umożliwia także prezentację metod statystycznych w sposób „tradycyjny", za pomocą tablicy roboczej, z ukazaniem etapów obliczeń. Tej właściwości nie mają profesjonalne pakiety statystyczne. Dużym udogodnieniem dla użytkujących arkusz kalkulacyjny do analizy statystycznej jest system funkcji, między innymi matematycznych i statystycznych, oraz opcja *Analiza danych*. Funkcje statystyczne umożliwiają szybkie znajdowanie wielkości najczęściej charakteryzujących rozkłady empiryczne (średnia, wariancja, odchylenie standardowe, częstość, kwartyle, percentyle, kurtoza itd.) Bardzo duże znaczenie mają dla nas wbudowane w arkusze kalkulacyjne najczęściej wykorzystywane rozkłady teoretyczne: dwumianowy, Poissona, normalny, *t*-Studenta, χ^2, *F* Fishera-Snedecora. Niektóre z nich (normalny, *t*-Studenta, *F*, beta, χ^2) występują w dwóch postaciach: a) dla zadanego kwantyla zwracają wartość prawdopodobieństwa, b) dla zadanego prawdopodobieństwa zwracają wartość kwantyla. W przypadku rozkładu *t*-Studenta otrzymuje się wartości pola pod krzywą bądź tylko z jednej strony, bądź też z obu stron razem. Opcja *Analiza danych*, która znajduje się w oknie dialogowym *Narzędzia*, może w wielu przypadkach zastąpić profesjonalny pakiet statystyczny. Dzięki niej można wykonać między innymi analizy wariancji, Fouriera, korelacji i regresji, testy statystyczne oraz statystykę opisową. Chcąc korzystać z *Analizy danych*, należy w momencie instalowania Excela wybrać *wersję użytkownika*. Zapewniamy sobie wówczas wszystkie możliwości analizy danych. Ponieważ pełna wersja Excela spowalnia pracę starszych typów komputerów, można czasowo wyłączyć opcje *Analiza danych*, wchodząc do okna dialogowego *Narzędzia*, wybierając *dodatki* i odznaczając trzy pierwsze opcje. Pod względem wbudowanych funkcji statystycznych Excel znacznie przewyższa pozostałe arkusze kalkulacyjne, ale ma duże wymagania sprzętowe.

Układając zadania, staraliśmy się dobierać przykłady z życia gospodarczego. Najczęściej ich źródłem były publikacje Głównego Urzędu Statystycznego bądź też niepublikowane dane statystyki publicznej, udostępnione

nam dla badań własnych. Tylko w wyjątkowych przypadkach, ze względów dydaktycznych, uciekaliśmy się do danych umownych. Podjęliśmy też próbę rewizji z zakresu przedmiotowego podręcznika. Stopniowo opuszczamy te zagadnienia, które w statystyce dla ekonomistów pełniły rolę pomocniczą dla obliczenia zasadniczych miar struktury, zależności i dynamiki. Staraliśmy się zatem pominąć lub zredukować do niezbędnego minimum takie zagadnienia, jak rachunek momentów, uproszczone metody „z punktem arbitralnym" w analizie struktury czy też pewne przekształcenia w analizie współzależności i dynamik. Przy usuwaniu tych procedur z podręcznika należy zachować ostrożność, gdyż niektóre metody ułatwiające niegdyś obliczenia ciągle posiadają swoją wartość. Na przykład usunięcie z wielu współczesnych podręczników statystyki wzorów na obliczenie miar współzależności z wykorzystaniem tablic korelacyjnych uważamy za nieuzasadnione tylko dlatego, że pełniły one kiedyś rolę procedur ułatwiających rachunki. Należy mieć na uwadze, że ciągle wiele danych statystycznych istnieje, i zapewne będzie jeszcze długo istnieć, w postaci właśnie tablicy korelacyjnej.

Statystyka jest zbiorem metod różnego pochodzenia. Różne nauki wniosły do niej wkład teoretycznie i empirycznie: matematyka, fizyka, ekonomia, nauki o ziemi, historia, socjologia, psychologia i demografia. W związku z tym metodologia badań ilościowych jest bardzo obszerna i żywa, gdyż ciągle się rozwija. Potrzeby opisania badanej rzeczywistości ekonomicznej, stanu ducha osoby lub społeczności, objaśnienia zmian i uwarunkowań tych procesów owocują w liczne propozycje metod analizy. Oto kilka przykładów takich inspiracji metodologicznych. Potrzeba objaśnienia zależności między wzrostem synów i ojców dała F. Galtonowi asumpt do konstrukcji analizy regresji[1] i korelacji. Potrzeba wyodrębnienia wspólnych czynników sprawczych wśród dużej liczby zmiennych doprowadziła psychologa Spearmana do stworzenia analizy czynnikowej. Potrzeba klasyfikacji obiektów według stopnia podobieństwa naprowadziła znakomitego polskiego antropologa, Jana Czekanowskiego, na myśl diagramu jego imienia. Dążenie do obniżenia wielkości próby i kosztów badań statystycznych przywiodła Jerzego Spławę-Neymana do konstrukcji jednego z wariantów losowania zwanego wariantem optymalnym lub „złotą regułą Spławy-Neymana". Potrzeba objaśniania bieżących procesów demograficznych

[1] Przy okazji warto zwrócić uwagę na fakt, że użyta wówczas w bardzo szczególnym przypadku nazwa „regresja" jako „cofanie się wzrostu synów do wzrostu ich ojców" upowszechniła się tak dalece, iż trudno byłoby ją zastąpić innym terminem.

zainspirowała P. K. Whelptona do stworzenia analizy kohortowej. Każda z tych idei była kamieniem milowym w rozwoju statystyki. Oprócz tych, jakże udanych, wynalazków metodologicznych istnieje też wiele pomysłów pośledniejszego znaczenia lub takich, które „sprawdzały się tylko w ściśle określonych warunkach". Jak wspomniano, wiele propozycji metodologicznych miało tylko przejściowo duże znaczenie jako usprawnienie obliczeń numerycznych i wraz z rozwojem techniki obliczeniowej zaczęły tracić na znaczeniu. Z drugiej jednak strony, technika obliczeniowa może sprawić, że wiele interesujących wynalazków z przeszłości należy ponownie ocenić pod względem ich użyteczności. Przykładem takiej nowej oceny i rehabilitacji metody jest analiza ścieżkowa (path analysis). Można zatem powiedzieć, że uprawianie statystyki jest także sztuką wyborów. Do rozwiązania danego problemu, do udzielenia odpowiedzi na określone pytanie, mamy wiele metod statystycznych i wiele źródeł informacji. Badacz musi dokonać wyboru, którą z nich użyje. Wymaga to znacznej wiedzy metod i źródeł statystycznych. Pomocą dla mniej doświadczonych badaczy może być pewien konsensus wśród statystyków co do metod stanowiących kanon w określonej dziedzinie wiedzy. Trudno sobie wyobrazić podręcznik ze statystki dla ekonomistów, który pomijałby analizę dynamiki, co może się zdarzyć na przykład w podręczniku statystyki dla psychologów. Z kolei metodologia badań doświadczalnych, mocno obecna w statystyce dla rolników i przyrodników, nie występuje w naszych podręcznikach. Wśród wybranych metod statystycznych zwracamy uwagę na hierarchię ich ważności i na to żeby, tworzyły w miarę zwarty system poznawczy. Na przykład do oceny asymetrii rozkładu możemy użyć kilku wskaźników, między innymi α^3, opartego na momentach, i wskaźnika skośności, opartego na miarach pozycyjnych. Jeśli analizę struktury i współzależności przeprowadzono za pomocą metody momentów, to użycie wskaźnika skośności zamiast α^3 jest pewnym dysonansem, a nawet błędem w sztuce[2]. Hierarchię ważności metod statystycznych, oprócz tradycji[3] w danej dziedzinie wiedzy, wyznacza także miejsce w systemie analizy, uniwersalność zastosowań i prostota interpretacji. Jak można zauważyć, w naszym podręczniku

[2] W uzasadnionych przypadkach można użyć więcej metod. Jest to uzasadnione wówczas, gdy chcemy upewnić się, czy nie mamy do czynienia ze szczególnym przypadkiem rozkładu.

[3] Abstrahując od dobrej i gorszej tradycji, należy podkreślić jej rolę jako pewną umowę badaczy co do używanych metod. Nawet bardzo dobre metody tracą na znaczeniu, jeśli niewiele osób posługuje się nimi. Do obiektywnie dobrych metod statystycznych należy przekonać środowisko naukowe.

staramy się unikać „ogólnych" zaleceń, które nadmiernie krępowałyby badaczy, czyniąc w krańcowym przypadku niemożliwym zastosowanie danej metody. Jako przykład podamy bardzo rozpowszechniony w podręcznikach pogląd, że „dobry" model regresji powinien charakteryzować się współczynnikiem determinacji powyżej 90% albo 95%. Ci, którzy prowadzą badania w konkretnych dziedzinach, takich jak gospodarka nieruchomości, badania marketingowe czy ekonomika gospodarstw domowych, nigdy nie zbudowaliby żadnego modelu. Osiągnięcie w tych przypadkach R^2 równego 40–50% jest już znacznym osiągnięciem badawczym[4]. Z tego względu w rozwiązaniach klasyfikacyjnych dajemy pierwszeństwo kryteriom merytorycznym nad formalnymi. Przykładem takiego „formalizmu" mogą być eksponowane w wielu podręcznikach wzory na optymalną liczbę przedziałów klasowych w danej populacji, por. R. A. Szmajłowa (1998, s. 76–77) oraz P. H. Waszkiw i inni (2001, s. 40). Wprowadzenie własnych klasyfikacji, jeśli nawet są one optymalne dla konkretnych przypadków, może utrudnić lub uniemożliwić wykorzystanie dodatkowych informacji z innych źródeł. W takich przypadkach zalecamy dostosowanie się raczej do powszechnie przyjętych i uzgodnionych wcześniej rozwiązań w danej dziedzinie wiedzy.

Na zakończenie pozostaje nam miły obowiązek podziękowania za uwagi recenzyjne Profesorowi dr. hab. Michałowi Woźniakowi z Akademii Ekonomicznej w Krakowie. Jego obszerna i wnikliwa recenzja pozwoliła nam poprawić błędy oraz lepiej wyeksponować własne stanowisko. Dziękuję także Panu mgr. Marcinowi Szymkowiakowi za pomoc w przygotowaniu maszynopisu podręcznika.

Jan Paradysz

---

[4] Zwracamy także uwagę na nowe podejście do badań statystycznych w duchu modnej obecnie „statystyki małych obszarów", gdzie dąży się do maksymalnego wykorzystania wszystkich dostępnych informacji.