

Statystyka opisowa

Mieczysław Sobczyk

Wydawnictwo C.H. Beck 

Statystyka opisowa

Statystyka opisowa

Mieczysław Sobczyk



WYDAWNICTWO C.H. BECK
WARSZAWA 2010

Wydawca: Dorota Ostrowska-Furmanek
Redakcja merytoryczna: Beata Kurcz
Recenzent: prof. dr hab. Anna Malina
Projekt okładki i stron tytułowych: Maryna Wiśniewska
Ilustracja na okładce: ©MarkEvans/iStockphoto

Seria: Metody ilościowe

Złożono programem T_EX



© Wydawnictwo C.H. Beck 2010

Wydawnictwo C.H. Beck Sp. z o.o.
ul. Bonifraterska 17, 00-203 Warszawa

Skład i łamanie: Wydawnictwo C.H. Beck
Druk i oprawa: Poznańskie Zakłady Graficzne

ISBN 978-83-255-1607-9

Spis treści

Wstęp	7
Rozdział 1. Podstawowe pojęcia statystyki opisowej, ich definicje i klasyfikacje	11
1.1. Przedmiot statystyki opisowej	11
1.2. Zbiorowość, jednostka statystyczna i jednostka sprawozdawcza	13
1.3. Cecha statystyczna, jej rodzaje i pomiar	15
1.4. Etapy badania statystycznego	18
1.4.1. Przygotowanie (programowanie) badania	20
1.4.2. Obserwacja statystyczna	21
1.4.3. Opracowanie i prezentacja materiału statystycznego	25
1.4.4. Opis statystyczny	38
Pytania kontrolne	39
Zadania	40
Rozdział 2. Statystyczny opis struktury jednowymiarowych rozkładów empirycznych	44
2.1. Rodzaje rozkładów jednowymiarowej zmiennej	44
2.2. Opisowe parametry struktury zjawisk masowych	48
2.2.1. Miary położenia (przeciętne, średnie)	49
2.2.2. Miary zmienności (zróżnicowania, rozproszenia, dyspersji)	60
2.2.3. Miary asymetrii (skośności)	67
2.2.4. Miary koncentracji i spłaszczenia (kurtozy)	71
2.3. Kompleksowa analiza struktury	76
2.4. Podobieństwo struktur	80
Pytania kontrolne	81
Zadania	82
Rozdział 3. Analiza współzależności cech (korelacja i regresja)	97
3.1. Cecha wielowymiarowa i jej rozkład	97
3.2. Funkcyjna, stochastyczna i korelacyjna zależność między zmiennymi	98
3.3. Prezentacja dwuwymiarowego rozkładu empirycznego	100
3.3.1. Szereg dwuwymiarowy	100
3.3.2. Diagram korelacyjny	101
3.3.3. Tablica korelacyjna	102

3.4. Opisowe miary współzależności cech mierzalnych	104
3.4.1. Współczynnik korelacji liniowej Pearsona	104
3.4.2. Wskaźniki (stosunki) korelacyjne Pearsona	109
3.4.3. Korelacja cząstkowa i wieloraka	111
3.5. Opisowe miary współzależności cech jakościowych	117
3.5.1. Współczynnik korelacji rang Spearmana	117
3.5.2. Korelacja rang Kendalla	119
3.5.3. Mierniki korelacji obliczane z tablic czteropolowych	120
3.5.4. Mierniki korelacji wykorzystujące statystykę chi- kwadrat	122
3.6. Liniowy model regresji dwóch zmiennych	128
3.6.1. Funkcja regresji – pojęcie i elementy składowe	129
3.6.2. Wyznaczanie parametrów strukturalnych liniowej funkcji regresji .	131
3.6.3. Ocena dopasowania liniowej funkcji regresji do danych empirycznych	134
3.7. Prognozowanie na podstawie liniowej funkcji regresji	137
Pytania kontrolne	143
Zadania	144
Rozdział 4. Metody analizy dynamiki zjawisk	165
4.1. Ogólna charakterystyka szeregów dynamicznych	165
4.2. Mierniki dynamiki zjawisk jednorodnych	166
4.2.1. Przyrosty absolutne i względne (jednopolstawowe i łańcuchowe) .	167
4.2.2. Indeksy indywidualne i średnie tempo zmian	168
4.2.3. Przekształcenia szeregów indeksów indywidualnych	171
4.3. Badanie dynamiki zjawisk złożonych	173
4.3.1. Indeksy agregatowe dla wielkości absolutnych (wartości, cen, ilości)	174
4.3.2. Indeksy agregatowe dla wielkości stosunkowych	180
4.4. Addytywny i multiplikatywny model wahań w czasie	185
4.4.1. Wyodrębnianie tendencji rozwojowej (trendu)	187
4.4.2. Pomiar wahań sezonowych	194
4.4.3. Eliminacja wahań przypadkowych	203
4.5. Wykorzystanie modelu wahań w czasie do prognozowania	206
Pytania kontrolne	211
Zadania	213
Odpowiedzi do zadań	231
Rozdział 2	231
Rozdział 3	232
Rozdział 4	234
Bibliografia	237

Wstęp

Do podejmowania prawidłowych decyzji i analizowania procesów gospodarczych, demograficznych, społecznych itp. niezbędna jest znajomość metod ilościowych. Z tego też względu w programach nauczania uczelni wyższych przewidziano zajęcia ze statystyki. Programy te obejmują zazwyczaj statystykę opisową (studia licencjackie) oraz statystykę matematyczną (studia magisterskie).

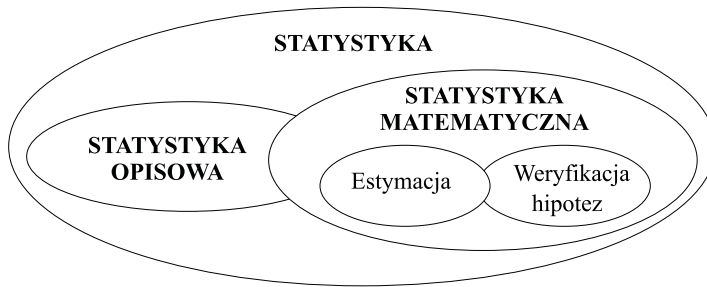
Przedmiotem zainteresowań statystyki opisowej są problemy programowania badań statystycznych, metody obserwacji statystycznej, sposoby opracowywania i prezentacji materiału statystycznego oraz syntetyzująca (sumaryczna) charakterystyka – za pomocą odpowiednich parametrów – właściwości zbioru danych. Metody statystyki opisowej wykorzystywane są wtedy, gdy obserwacją objęta jest cała zbiorowość (populacja generalna). Mówimy wówczas o badaniach pełnych, wyczerpujących.

Statystyka matematyczna (zwana też statystyką indukcyjną) zajmuje się metodami wnioskowania – estymacji i weryfikacji hipotez o całej zbiorowości (populacji generalnej) na podstawie wybranej w sposób losowy pewnej jej części określanej mianem próby losowej. Metody statystyki matematycznej znajdują szerokie zastosowanie w badaniach częściowych (niepełnych, niewyczerpujących). Fundamentalną sprawą w badaniu częściowym jest to, aby próba była reprezentatywna, tzn. jej struktura pod względem badanej właściwości (cechy) była zbliżona do struktury tej cechy w populacji generalnej. Zapewnienie wysokiej reprezentatywności próby jest zagadnieniem dość skomplikowanym. Wynika to z faktu, że populacja generalna jest nieznaną, a dopiero badanie częściowe ma dostarczyć informacji o niej. Wykorzystanie metod wnioskowania statystycznego (statystyki matematycznej) wymaga znajomości rachunku prawdopodobieństwa.

Zakres przedmiotowy statystyki przedstawiono na rysunku 1.

Metody statystyczne dotyczące zarówno opisu, jak i wnioskowania mogą być użyteczne w liczbowym rozpoznaniu struktury, współzależności oraz dynamiki zjawisk masowych. O ile jednak opis statystyczny odpowiada na pytanie: „jak jest?”, o tyle wnioskowanie statystyczne związane jest z pytaniem: „jak prawdopodobnie może być?”. Tak więc, wnioskowanie statystyczne w obydwu zasadniczych procedurach, tj. estymacji i weryfikacji hipotez, odbywa się zawsze w warunkach niepewności. Mówimy, że jest ono wnioskowaniem indukcyjnym o charakterze stochastycznym (opartym na rachunku prawdopodobieństwa). Wzajemne relacje pomiędzy metodami analizy, opisem i wnioskowaniem statystycznym przedstawiono w tabeli 1.

Całość problematyki statystyki jest ujęta w dwóch odrębnych, wzajemnie uzupełniających się pozycjach noszących tytuły: *Statystyka opisowa* i *Statystyka matematyczna*. Prezentowane w nich zagadnienia podawane są od podstaw i nie są zbyt skomplikowane



Rysunek 1. Zakres statystyki

Źródło: [Cieciura, Zacharski, 2007, s. 239].

Tabela 1. Podstawowe metody analizy statystycznej

Metody analizy	Opis statystyczny	Wnioskowanie statystyczne	
Struktury zjawisk	Miary średnie Miary zmienności Miary asymetrii Miary koncentracji i spłaszczenia (kurtozy)	Estymacja parametrów (punktowa i przedziałowa)	Weryfikacja hipotez (parametrycznych i nieparametrycznych)
Współzależności zjawisk	Rachunek korelacji Rachunek regresji		
Dynamiki zjawisk	Szeregi czasowe bez sezonowości Szeregi czasowe z sezonowością Indeksy statystyczne		

Źródło: opracowanie własne na podstawie: [Luszniewicz, Słaby, 2008, s. 14].

pod względem matematycznym. Cenną zaletą tych pozycji jest zamieszczenie wielu rozwiązanych przykładów, które ilustrują istotę omawianych procedur statystycznych oraz pokazują możliwości interpretacyjne otrzymanych wyników. Ponadto zawierają one bogaty zestaw zadań do samodzielnego rozwiązania. Kontrolę stopnia opanowania wiedzy ułatwiają zamieszczone na końcu podręczników odpowiedzi do wszystkich zadań. Służą temu również pytania kontrolne.

W pracach zamieszczono obszerną bibliografię, zawierającą publikacje książkowe z zakresu statystyki. Pozwoli ona na pogłębienie znajomości prezentowanych zagadnień, jak również na poszerzenie wiedzy w zakresie pominiętych z konieczności treści.

We współczesnym świecie nabycie umiejętności praktycznego stosowania metod statystycznych jest nieodzowne. Serwisy informacyjne, gazety codzienne, pisma popularne zawierają „gąszcz liczb”. Należy umieć je interpretować. W przeciwnym razie nie jesteśmy w stanie trafnie ocenić tych doniesień, jak też wyciągać z nich odpowiednich wniosków. Autor daleki jest od popierania stwierdzeń typu: „statystyka jest prosta i oczywista” czy „jest intuicyjnie zrozumiała”. Tak jak każda dziedzina wiedzy, wymaga ona znacznego wysiłku i żmudnej pracy w celu zdobycia określonych umiejętności.

Autor żywi nadzieję, że zawarty w pracach zakres przedmiotowy i zrozumiały sposób ujęcia rozpatrywanych zagadnień, uczynią je przyjaznymi dla studentów wydziałów ekonomicznych i humanistycznych (socjologów, psychologów, pedagogów, historyków

itp.), jak również dla szerokiego grona praktyków prowadzących analizy statystyczne. Tym samym wierzy, że przekonanie o związku „nudy” i „statystyki” okaże się nieuzasadnione, a wątpliwości w zakresie możliwości zrozumienia świata opisanego za pomocą statystyki – błędne¹.

Podręcznik *Statystyka opisowa* zawiera cztery merytoryczne rozdziały. Pierwszy z nich wprowadza Czytelnika w problematykę badań statystycznych. W szczególności omówiono tu przedmiot zainteresowań statystyki opisowej, podstawowe pojęcia statystyczne (zbiorowość, jednostka statystyczna i sprawozdawcza, cecha statystyczna), rodzaje, sposoby pozyskiwania i formy prezentacji danych statystycznych oraz etapy badania statystycznego.

Rozdział drugi poświęcono prezentacji metod statystycznego opisu struktury jednowymiarowych rozkładów empirycznych. Obejmują one miary średnie (przeciętne, położenia), zmienności (zróżnicowania, rozproszenia, dyspersji), asymetrii (skośności) oraz koncentracji i spłaszczenia (kurtozy). Miary te pozwalają na wszechstronną, kompleksową analizę struktury zjawisk masowych.

W rozdziale trzecim zaprezentowano metody analizy rozkładów wielowymiarowych, z położeniem nacisku na rozkłady dwuwymiarowe. Obejmują one problematykę badania siły i kierunku związków przyczynowo-skutkowych (korelację) oraz analizę mechanizmu powiązań między zmiennymi (regresję). W analizie korelacji przedstawiono różne miary określone rodzajem zmiennych (ilościowe, jakościowe), sposobem prezentacji danych wyjściowych (szeregi i tablice korelacyjne) oraz kształtem związku (prostoliniowy, krzywoliniowy). W badaniu mechanizmu powiązań między zmiennymi skupiono uwagę na liniowych funkcjach regresji z jedną zmienną objaśniającą. Rozdział ten zakończono prezentacją wykorzystania funkcji regresji dla celów prognozowania (predykcji).

W ostatnim, czwartym rozdziale, omówiono metody analizy dynamiki zjawisk masowych. Podstawą tego rodzaju analizy są szeregi czasowe (dynamiczne, chronologiczne). W zakresie analizy dynamiki przedstawiono dekompozycję szeregów czasowych (trend, wahania sezonowe i przypadkowe), jak też zaprezentowano rachunek indeksów statystycznych (indywidualnych i agregatowych wskaźników dynamiki). W treści tego rozdziału zwrócono również uwagę na proste metody badania zmian w szeregach chronologicznych (przyrosty absolutne, względne, indeksy indywidualne, średnie tempo zmian).

Na ostateczny kształt tej pracy składa się nie tylko trud autora. Słowa wdzięczności kieruję do Recenzentki Pani Profesor Anny Maliny z Uniwersytetu Ekonomicznego w Krakowie za cenne i szczegółowe uwagi, które przyczyniły się do udoskonalenia treści i formy opracowania.

Pragnę podziękować także córce Ewelinie oraz Monice Sobolewskiej za pomoc i ciężką pracę przy technicznym redagowaniu tekstu.

Autor

¹ Jak mawiał jeden ze studentów: „jeśli pozostałby mi tylko jeden dzień życia, spędziłbym go na statystyce – dzięki temu wydałby mi się on znacznie dłuższy”. Por.: [Szwed, 2009, s. 11].

Rozdział 1. Podstawowe pojęcia statystyki opisowej, ich definicje i klasyfikacje

1.1. Przedmiot statystyki opisowej

Termin **statystyka** wywodzi się od łacińskiego słowa **status** czyli państwo. W piśmiennictwie słowo **statystyka** po raz pierwszy zostało użyte w połowie XVIII w. przez Gottfrieda Achenwalla (1719–1772), profesora uniwersytetu w Marburgu, a następnie w Getyndze. Do połowy XIX w. określenie **statystyka** oznaczało zbiór danych liczbowych dotyczących stanu państwa, a następnie wszelkie gromadzone informacje opisujące świat zewnętrzny. Duże zbiory danych liczbowych wymagały odpowiedniej analizy. Ujmowano je w postaci zestawień tabelarycznych. Autorów tego rodzaju tablic nazywano **tabelarystami**. Za twórcę kierunku tabelarystycznego uważa się Duńczyka J.P. Anchersena, który swej opublikowanej w 1741 r. pracy nadał znamieny tytuł: *Descriptio statuum cultiorum in tabulis*. Tak więc termin statystyka ma swoje źródło merytoryczne w badaniach państwowznawczych, a metodologiczne – w pracach tabelarystów.

Dynamiczny rozwój statystyki jako narzędzia wykorzystywanego do opisu zjawisk gospodarczych (tablice bilansowe, warunki bytu ludności itp.) datuje się od czasów A. Quetleta (1796–1874). W okresie tym statystykę zaczęto traktować jako naukę, czego wyrazem było powołanie w 1834 r. w Anglii Królewskiego Towarzystwa Statystycznego oraz zwołanie w Brukseli I Międzynarodowego Kongresu Statystycznego (1854 r.).

Współcześnie termin **statystyka** ma wiele znaczeń. W mowie potocznej używa się go do określenia zarówno danych, ich opracowań, charakterystyk opisowych, a także całej dyscypliny naukowej.

1. Statystyka rozumiana jest jako zbiór danych liczbowych, przedstawiających kształtowanie się określonych zjawisk i procesów. W tym znaczeniu mówimy np. o statystyce ludności, transportu, rolnictwa itp.
2. Mianem statystyki określa się wszelkie prace związane z gromadzeniem i opracowywaniem danych liczbowych. Zasady gromadzenia danych i prowadzenia badań statystycznych reguluje Ustawa o statystyce publicznej z 29 czerwca 1995 r. (Dz.U. nr 88, poz. 439). Całokształt działań związanych ze zbieraniem informacji statystycznych, ich przechowywaniem, opracowywaniem,

udostępnianiem i rozpowszechnianiem wyników badań statystycznych jako oficjalnych danych tworzy **System Informacyjny Statystyki Publicznej (SISP)**. Zadania wynikające z funkcjonowania SISP koordynuje utworzony w 1918 r. **Główny Urząd Statystyczny (GUS)**.

3. Pewne charakterystyki opisowe obliczane ze zbiorowości próbnych również nazywa się statystykami. W tym sensie statystyką jest np. średnia arytmetyczna z próby, odchylenie standardowe z próby, współczynnik korelacji liniowej Pearsona z próby.
4. Wreszcie statystykę definiuje się jako dyscyplinę naukową mającą własne, specyficzne metody badawcze. W tym znaczeniu **statystykę** rozumie się jako naukę o „ilościowych metodach badania prawidłowości występujących w zjawiskach masowych scharakteryzowanych za pomocą liczb” [Kurkiewicz, Stonawski, 2005, s. 12].

Metoda – to sposób postępowania prowadzący do osiągnięcia wyznaczonego celu. Istnieją określone kryteria pozwalające na ocenę, czy dane postępowanie można określić mianem metody naukowej. Wchodzą one w zakres filozofii, a ściślej **ontologii** (filozofii bytu) oraz **epistemologii** (filozofii poznania).

Celem analiz statystycznych jest wykrycie **prawidłowości** rządzących badanymi zjawiskami masowymi. Obserwowane w zjawiskach masowych prawidłowości są efektem oddziaływania dwóch grup przyczyn: **głównych** (systematycznych) i **ubocznych** (przypadkowych). Przyczyny **główne** oddziałują na badane zjawisko w sposób jednokierunkowy, kształtując tzw. składnik systematyczny zjawiska masowego. Dzięki nim ujawnia się charakterystyczna dla danego zjawiska masowego prawidłowość.

Przyczyny **uboczne** występują w dużej ilości i działają różnokierunkowo. W obserwowanej zbiorowości wywołują one odchylenia od prawidłowości, które ujawniają się w indywidualnych przypadkach. Rezultatem oddziaływania przyczyn ubocznych jest **przypadkowy składnik zjawiska masowego**.

Skutki działania przyczyn głównych uwidaczniają się tym wyraźniej, im liczniejsza jest poddana badaniu zbiorowość. Mówimy wówczas o działaniu **prawa wielkich liczb**. Treść tego prawa można ująć następująco: przy dostatecznie dużej liczbie obserwacji dotyczących określonego zjawiska skutki przyczyn ubocznych redukują się, natomiast ujawniają się prawidłowości.

Sama masowość nie decyduje o możliwości stosowania w badaniach metod statystycznych. Metody te mogą być wykorzystywane jedynie do badania takich zjawisk masowych, które dotyczą zbiorowości składających się z elementów o podobnych – ale nie identycznych – właściwościach. Przykładowo, zbiór jednakowych krzeseł w sali kinowej nie może być przedmiotem badania statystycznego. Rezultaty badania tylko jednego krzesła można tutaj odnieść do całego ich zbioru.

Statystyczne metody analizy mają uniwersalny charakter i są niemal powszechnie wykorzystywane we wszystkich dyscyplinach naukowych (psychologii, socjologii, demografii, politologii, medycynie itp.). Wśród dyscyplin związanych

ze sferą biznesu, metody statystyczne znajdują zastosowanie w marketingu, ubezpieczeniach, rachunkowości, zarządzaniu itp. Metody te zwielokrotniają bowiem siłę poznawczą danej dyscypliny naukowej, pozwalają na wygodne, ściśle i jednoznaczne oddawanie myśli za pomocą liczb. Językiem statystyki można opisać różnorodne zjawiska masowe nieporównanie dokładniej niż werbalnie.

Za podstawowe zadanie statystyki uznaje się dostarczanie wiarygodnych informacji niezbędnych do podejmowania różnego rodzaju decyzji. Jeśli decyzje te mają być skuteczne i efektywne, to powinny bazować na bogatym, poprawnie zgromadzonym i wszechstronnie przeanalizowanym zbiorze informacji. W szczególności, na wyróżnienie zasługują trzy podstawowe funkcje spełniane przez statystykę:

- **informacyjna** – dająca pełny i obiektywny obraz badanych zjawisk;
- **analityczna** – dzięki której możliwe jest określenie czynników kształtujących konkretne procesy i zjawiska;
- **prognostyczna** – pozwalająca na przewidywanie kierunku rozwoju analizowanych zjawisk.

1.2. Zbiorowość, jednostka statystyczna i jednostka sprawozdawcza

Badanie statystyczne prowadzone jest zawsze na pewnej zbiorowości (osób, rzeczy, zjawisk, faktów itp.). Zbiór dowolnych elementów podobnych – ale nie identycznych – pod względem określonych właściwości i poddanych badaniu statystycznemu nazywamy **zbiorowością (populacją generalną)**. Elementy składowe zbiorowości poddane bezpośredniej obserwacji lub pomiarowi noszą nazwę **jednostek statystycznych, jednostek badania** lub **jednostek obserwacji**. Zbiorowości powinny być ściśle określone pod względem **rzeczowym** (kto lub co jest przedmiotem badania), **przestrzennym** (gdzie znajduje się zbiorowość) oraz **czasowym** (jakiego momentu lub okresu dotyczy badanie). Podkreślić należy, że w tej samej zbiorowości można – w zależności od celu badania – wyróżnić różne jednostki. Przykładowo w zbiorowości osób jednostkami statystycznymi mogą być poszczególne osoby, rodziny lub gospodarstwa domowe.

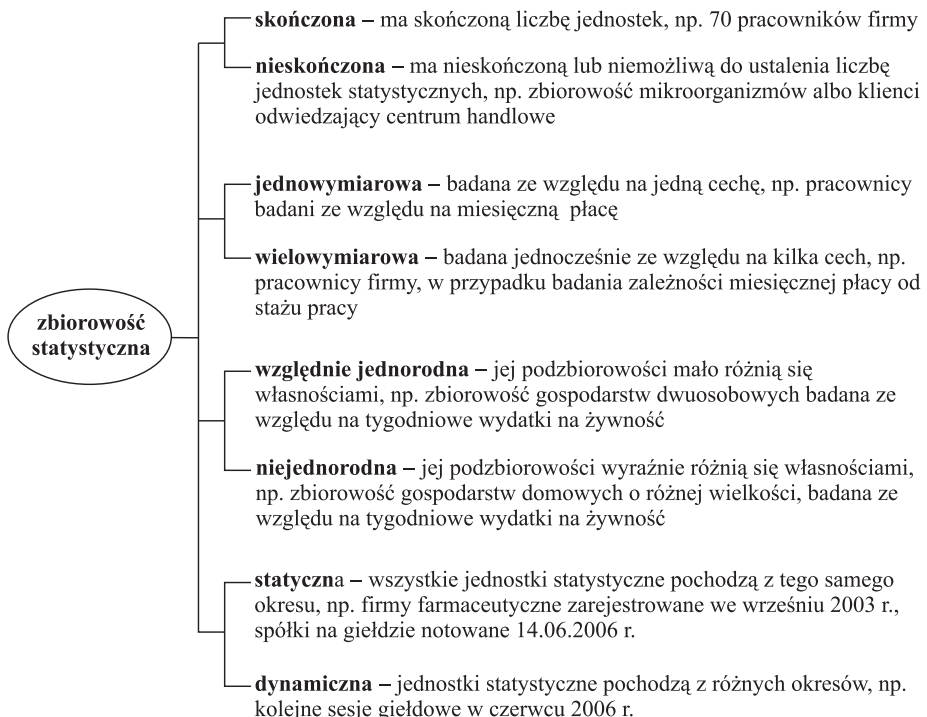
Zbiorowości można wyodrębniać według różnych kryteriów. I tak, ze względu na kryterium czasu wyróżnia się zbiorowości **statyczne** (np. ludność Polski według stanu na 31 grudnia ubiegłego roku) i **dynamiczne** (np. samochody zarejestrowane w Polsce w ubiegłym roku). Zbiorowości **statyczne** tworzą zatem jednostki, które istniały, istnieją lub będą istniały w ściśle określonym momencie. Zbiorowości **dynamiczne** składają się z jednostek obserwowanych w pewnym przedziale czasu.

Według kryterium liczebności, wyróżniamy zbiorowości **skończenie** i **nie-skończenie liczne**. Pierwsze z nich składają się z przeliczalnej liczby jednostek (np. studenci UMCS według stanu na 31 grudnia danego roku), drugie natomiast – tworzą elementy o nieprzeliczalnej liczbie (np. zbiorowość organizmów żywych na kuli ziemskiej).

Wyróżnia się również zbiorowości **jednorodne** i **niejednorodne**. Zbiorowości **jednorodne** tworzą jednostki o identycznym gatunku, typie, rodzaju. Ewentualne różnice występujące między jednostkami mają tutaj wyłącznie charakter ilościowy, a nie jakościowy (zbiorowość jednorodną tworzą np. wszystkie dzieci płci męskiej urodzone w danym dniu). Zbiorowości **niejednorodne** składają się z jakościowo odmiennych jednostek (np. wszystkie urodzone w danym dniu dzieci). Zwrócić należy uwagę na fakt, że jednorodność bądź niejednorodność zbiorowości ma względny charakter. Uzależnione jest to od celu i szczegółowości badania statystycznego. Na przykład w zbiorowości: pracownicy produkcyjni przedsiębiorstwa 15 stycznia danego roku, grupa mężczyzn jest jednorodna z punktu widzenia płci, niejednorodna natomiast ze względu na staż pracy.

Pojęcie jednostki statystycznej należy odróżniać od określenia **jednostka sprawozdawcza**. Jednostką **sprawozdawczą** jest obiekt udzielający informacji o właściwościach jednostek statystycznych tworzących zbiorowość (populację generalną). Jednostką sprawozdawczą jest np. osoba przeprowadzająca wywiad dla potrzeb Powszechnego Spisu Rolnego (w tym przypadku jednostką statystyczną jest gospodarstwo rolne).

Różne klasyfikacje zbiorowości statystycznych przedstawiono na rysunku 1.1.



Rysunek 1.1. Różne klasyfikacje zbiorowości statystycznych

Źródło: [Maksimowicz-Ajchel, 2007, s. 15].

1.3. Cecha statystyczna, jej rodzaje i pomiar

Jednostki statystyczne tworzące określoną populację generalną odznaczają się określonymi własnościami. Własności te nazywamy **cechami statystycznymi**. Cechy statystyczne można podzielić na **stałe** (wspólne dla wszystkich jednostek badanej zbiorowości) oraz **zmienne**, dzięki którym poszczególne jednostki różnią się między sobą. Cechy stałe (rzeczowe, przestrzenne i czasowe) nie podlegają badaniu statystycznemu, a jedynie służą do precyzyjnego zdefiniowania zbiorowości statystycznej. Analizie statystycznej poddawane są tylko cechy zmienne.

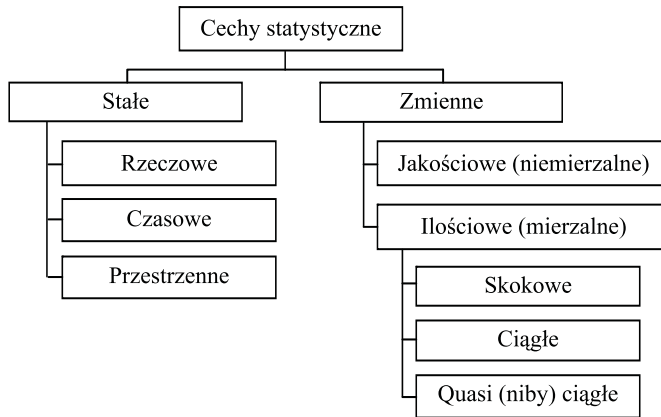
Każda z cech zmiennych może występować w różnych odmianach zwanych **wariantami**. Warianty mogą być określane słownie (np. płeć, zawód, wykształcenie) lub też przyjmować różne wartości liczbowe będące wynikiem zliczania (liczba osób w rodzinie, liczba uczestników kursu, wydatki dziennie na żywność itp.), bądź też rezultatem dokonanego pomiaru (np. wzrost w cm, masa w kg). Cechy, których warianty podawane są w sposób opisowy nazywamy cechami **niemierzalnymi (jakościowymi)**. Cechy, których warianty są wyrażane za pomocą liczb, określa się mianem cech **mierzalnych (ilościowych)**.

Cechy jakościowe, których natężenie określone jest przez przymiotniki (np. ocena wiadomości ucznia: celująca, bardzo dobra, dobra, dostateczna, dopuszczająca, niedostateczna), nazywane są cechami **porządkowymi**. Następującym po sobie wariantom takich cech można przyporządkować kolejne liczby (rangi), odpowiadające zmieniającemu się natężeniu cechy. Cechy jakościowe, których warianty są określone w sposób opisowy i nie jest możliwe ustalenie dla nich hierarchii ważności (np. zawody: ekonomiczne, techniczne, humanistyczne) nazywane są cechami **nominalnymi**.

Wśród cech ilościowych (mierzalnych) wyróżnia się cechy **skokowe (dyskretne)** i **ciągłe**. Pierwsze z nich mogą przyjmować określone wartości ze skończonych i przeliczalnych przedziałów liczbowych, ale bez wartości pośrednich (np. liczba osób w gospodarstwie domowym, liczba pracowników firmy itp.). Warianty liczbowe tego rodzaju cech przyjmują zazwyczaj wartości liczb całkowitych. Cechy ilościowe **ciągłe** mogą natomiast przyjmować każdą wartość z określonego skończonego przedziału liczbowego (np. wiek, wzrost, waga). Cechy ciągłe są wyrażane za pomocą liczb rzeczywistych.

Trzecim rodzajem cechy ilościowej jest cecha **quasi (niby) ciągła**. W rzeczywistości jest to cecha skokowa, ale ze względu na wielką liczbę wariantów, jaką może przyjmować, w badaniach statystycznych traktuje się ją jako cechę ciągłą. Przykładami takich cech są: płace pracowników, ceny towarów, miesięczny dochód gospodarstw domowych, zysk przedsiębiorstwa itp. W obliczeniach statystycznych cechy mierzalne quasi-ciągłe zaliczane są do ciągłych. Podział cech statystycznych przedstawiono na rysunku 1.2.

Warto zauważyć, że podstawą podziału cech statystycznych na mierzalne i niemierzalne jest sposób ich **pomiaru**. Pomiarom nazywamy przyporządkowanie określonych symboli (liczb, nazw, znaków) własnościom mierzonych obiektów.



Rysunek 1.2. Podział cech statystycznych

Źródło: [Pułaska-Turyna, 2005, s. 13].

Niejednokrotnie tę samą własność można zakwalifikować – w zależności od sposobu jej pomiaru – do cech mierzalnych lub niemierzalnych. Na przykład wiek osób mierzony w latach jest cechą mierzalną. Jeśli natomiast wiek badanych osób będziemy określać jako młody, średni lub podeszły – to będziemy mieli do czynienia z cechą niemierzalną.

Typ skali pomiarowej cech może być podstawą ich podziału na cztery grupy: **nominalne**, **porządkowe (rangowe)**, **przedziałowe (interwałowe)**, oraz **stosunkowe (ilorazowe)**. Stosownie do wymienionych wyżej rodzajów cech, wyróżnia się cztery **skale pomiarowe**: nominalną, porządkową, przedziałową i ilorazową.

Skala **nominalna** jest skalą najniższego poziomu. Liczby pełnią tutaj jedynie rolę umownych symboli służących do identyfikacji jednostek statystycznych i ich zaklasyfikowania do wyróżnionych kategorii. Rezultatem pomiaru w tej skali jest więc jedynie stwierdzenie, że n_1 jednostek należy do kategorii pierwszej, n_2 do kategorii drugiej, ..., n_k jednostek do k -tej kategorii. Na wartościach cech mierzonych w skali nominalnej można wykonywać tylko niektóre operacje matematyczne, takie jak: zliczanie, obliczanie częstości występowania, ustalenie klasy najliczniejszej. Przykładami liczb typu nominalnego są: numery na koszulkach zawodników drużyny piłki nożnej, numery tras autobusów, pokoje hotelowych, grupy krwi itp.

Skala **porządkowa (rangowa)** jest wyższym poziomem pomiaru niż skala nominalna. Oprócz klasyfikacji umożliwia ona porządkowanie kategorii ze względu na natężenie danej cechy (np. wykształcenie: podstawowe, średnie, licencjackie, wyższe). Skala ta może ustalać porządek **słaby** lub **mocny**. W słabym uporządkowaniu dopuszczalne jest stwierdzenie „**mniejszy lub równy**” lub „**większy lub równy**”. Nie można natomiast określić różnic między elementami ani stwierdzić, ile razy jeden element jest większy lub mniejszy od innego. W przypadku porządku mocnego: „**większy niż**” bądź „**mniejszy niż**” pojawia się **asymetria**. Przykładowo, jeśli X jest wyższy od Y , to nie może być na odwrót. Na liczbach w skali porządkowej można wykonywać takie same operacje jak dla skali no-

minalnej oraz dodatkowo – wyznaczać charakterystyki porządkowe i niektóre miary korelacji. Skala porządkowa posiada wszystkie własności skali nominalnej oraz dodatkowo umożliwia porządkowanie. Od skali porządkowej można przejść do nominalnej, ale nie odwrotnie. W skali porządkowej określa się np. gatunki owoców i warzyw (I, II, III), stopnie wojskowe, preferencje konsumentów.

Skala **przedziałowa (interwałowa)** nie posiada prawdziwego punktu zerowego, tzw. „zera absolutnego”. Punkt ten jest ustalany arbitralnie (umownie). Przykładowo, w skali temperatury Celsjusza zerem jest punkt, w którym zamraża woda. Niższe temperatury zapisywane są ze znakiem minus, a wyższe – ze znakiem plus. W skali przedziałowej wyrażane są np. lata kalendarzowe, indeksy cen, skala temperatury Fahrenheita. Z uwagi na brak absolutnego punktu zerowego na liczbach w skali przedziałowej nie można wykonywać dzielenia. Wszystkie operacje dopuszczalne na skalach nominalnej i porządkowej są dozwolone na skali przedziałowej, a ponadto można tu obliczać średnie i miary różnicowania. Skala przedziałowa pozwala na określenie różnicy między poszczególnymi jednostkami. Oznacza to, że można tu nie tylko stwierdzić, że jedna wartość cechy jest mniejsza lub większa od drugiej, ale również określić czy różnica (dystans) między nimi jest taka sama (większa lub mniejsza) od różnicy między dwiema innymi wartościami. Niemożliwe jest natomiast stwierdzenie, ile razy jedna wartość jest większa lub mniejsza od drugiej.

Skala **stosunkowa (ilorazowa)** ma własności trzech poprzednich skal, a ponadto posiada naturalny punkt zerowy. Punkt ten oznacza brak danej cechy (fizyczny poziom zerowy). Umożliwia to dokonywanie na liczbach w tej skali wszystkich – łącznie z dzieleniem – operacji matematycznych. W skali stosunkowej mierzy się odległość, dochody, wielkość sprzedaży, ceny towarów, koszty itp. Charakterystykę skal pomiarowych zawarto w tabeli 1.1.

Podział cech na nominalne, porządkowe, przedziałowe i ilorazowe nie jest konkurencyjny w stosunku do klasyfikacji na cechy mierzalne i niemierzalne. Pomiar w skali nominalnej i porządkowej odnosi się bowiem do cech niemierzalnych, natomiast pomiar w skali przedziałowej i ilorazowej – do cech mierzalnych. Skale: nominalna i porządkowa są **skalami słabymi**, natomiast interwałowa i ilorazowa – **skalami mocnymi**. Zgodnie z teorią pomiaru, zmienne ze skali mocniejszej mogą być transformowane na słabsze. Wiąże się to jednak z utratą pewnego zasobu informacji. Wzmacnianie skal, tj. przechodzenie od skal słabszych do mocniejszych, nie jest możliwe. Nie można bowiem z mniejszej ilości informacji uzyskać większego zasobu.

Wyróżnione cztery skale pomiarowe: nominalna, porządkowa, przedziałowa i ilorazowa mają **kumulatywny** charakter. Oznacza to, że charakteryzują się one narastającym stopniem dokładności pomiaru. Każda następna skala jest precyzyjniejsza od poprzedniej: zawiera wszystkie własności poprzedniej oraz dodatkowo nowe.

W wielowymiarowej analizie statystycznej użyteczną rolę pełni podział cech na **stymulanty**, **destymulanty** oraz **nominanty** [Hydzik, Sobolewski, 2007,

Tabela 1.1. Charakterystyka skal pomiarowych

Skala pomiaru	Własności	Przykłady
Nominalna	Pozwala jedynie podzielić elementy zbiorowości na rozłączne podzbiorowości (rozróżnić jednostki).	Płeć – kobiety, mężczyźni Rodzaj wykształcenia – ekonomiczne, techniczne itd.
Porządkowa	Dodatkowo pozwala uporządkować te podzbiorowości (lub jednostki) według intensywności występowania cechy, nie ustalając odległości między nimi.	Ocena produktu – bardzo zła, zła, średnia, dobra, bardzo dobra Wykształcenie – podstawowe, zawodowe, średnie, wyższe
Przedziałowa (interwałowa)	Dodatkowo pozwala ustalić różnice (odległości) intensywności występowania cechy między podzbiorowościami (lub jednostkami). Początek skali to tzw. zero względne (umowne).	Ocena produktu – 1, 2, 3, 4, 5 Temperatura w °C Rok urodzenia Godzina rozpoczęcia egzaminu
Ilorazowa (stosunkowa)	Dodatkowo pozwala ustalić iloraz intensywności występowania cechy między podzbiorowościami (lub jednostkami). Posiada zero bezwzględne.	Dochód, cena, zysk, wiek, liczba zleceń, czas dojazdu, wydatki

Źródło: [Maksimowicz-Ajchel, 2007, s. 16].

s. 26]. Zmienna jest stymulantą, jeśli wyższe jej wartości kwalifikują daną jednostkę statystyczną jako lepszą z punktu widzenia prowadzonego badania (np. dochody ludności w badaniach poziomu życia). W przypadku, gdy niższe wartości zmiennej stanowią o lepszej pozycji jednostki z punktu widzenia prowadzonych badań – to cecha ta jest destymulantą. Nominanta – to taka cecha, od której każde odchylenie (zarówno *in plus*, jak i *in minus*) jest zjawiskiem negatywnym z punktu widzenia prowadzonego badania (np. poziom substancji czynnej leku w organizmie człowieka w przypadku badań farmakologicznych). Wyróżnienie w zbiorze danych stymulant, destymulant i nominant stanowi punkt wyjścia przy liniowym porządkowaniu obiektów (np. województw, firm), czy wyodrębnianiu jednorodnych rejonów.

1.4. Etapy badania statystycznego

Badanie statystyczne to „zespół czynności zmierzających do uzyskania za pomocą metod statystycznych informacji charakteryzujących badaną zbiorowość” [Makać, 1999, s. 20]. Przedmiotem statystyki opisowej są badania pełne (wyczerpujące, całkowite), w których biorą udział wszystkie jednostki danej zbiorowości statystycznej. Badaniami pełnymi są: **spis statystyczny**, **inwentaryzacja**, **rejestracja bieżąca** oraz **sprawozdawczość statystyczna**.

Spisem statystycznym nazywamy okresowe lub doraźne, specjalne badanie, obejmujące wszystkie jednostki danej zbiorowości w ściśle określonym czasie

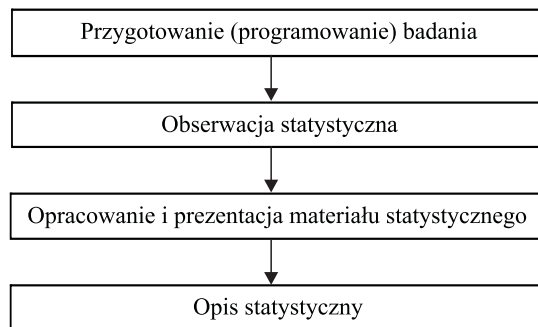
i przestrzeni. Jest on przeprowadzany za pomocą specjalnie przygotowanych formularzy. Formularz statystyczny jest odpowiednio przygotowanym zbiorem pytań i rubryk dotyczących badanych cech danej zbiorowości. Do najważniejszych **spisów okresowych** przeprowadzanych w Polsce należy zaliczyć powszechne spisy ludności, spisy rolne, spisy pogłowia zwierząt gospodarskich itp. **Spisy doraźne** są przeprowadzane – w zależności od potrzeb – w różnych okresach (np. spisy środków trwałych).

Inwentaryzacja jest spisem ujmującym faktyczny stan ilościowy i wartościowy majątku określonej jednostki gospodarczej lub administracyjnej. **Rejestracja bieżąca** polega na systematycznym notowaniu różnych faktów w momencie ich zaistnienia (np. rejestracja urodzeń, zgonów, małżeństw). **Sprawozdawczość statystyczna** obejmuje informacje statystyczne przekazywane obligatoryjnie na jednolitych formularzach przez osoby fizyczne, podmioty gospodarcze i instytucje do władz nadrzędnych lub organów statystycznych (np. zeznania podatkowe) [Wawrzynek, 2007, s. 17].

W przypadku braku możliwości przeprowadzania badania całkowitego (jak również częściowego), stosuje się postępowanie zwane **szacunkiem statystycznym** [Sobczyk, 1982, s. 25]. Polega ono na ustaleniu nieznanych wielkości lub właściwości zbiorowości na podstawie innych zbiorowości pozostających w określonym związku przyczynowo-skutkowym z interesującą nas nieznaną populacją. Z koniecznością szacunku statystycznego spotykamy się często w statystyce historycznej (np. szacowanie liczby ludności na podstawie tzw. dymów, tj. domów podlegających opodatkowaniu).

Szczególną formą szacunku statystycznego jest **rachunek ekstrapolacyjny** lub **interpolacyjny**. **Ekstrapolacja** pozwala na ustalenie nieznanych wartości znajdujących się poza przedziałem wartości znanych. **Interpolacja** polega na szacowaniu nieznanych wartości badanej cechy na podstawie jej wcześniejszych lub późniejszych realizacji.

Każde badanie statystyczne wymaga pewnych koniecznych czynności, których kolejność jest uwarunkowana logicznymi przesłankami (rys. 1.3).



Rysunek 1.3. Etapy badania statystycznego

Źródło: opracowanie własne.