

Zarządzanie

# Wybrane metody ilościowe w finansach

Dorota Witkowska



# **Wybrane metody ilościowe w finansach**



WYDAWNICTWO  
UNIWERSYTETU  
ŁÓDZKIEGO

Zarządzanie

# Wybrane metody ilościowe w finansach

Dorota Witkowska

Dorota Witkowska (ORCID: 0000-0001-9538-9589)  
– Uniwersytet Łódzki, Wydział Zarządzania  
Katedra Zarządzania Finansami Przedsiębiorstwa, 90-237 Łódź, ul. Matejki 22/26

RECENZENT

*Małgorzata Tarczyńska-Luniewska*

REDAKTOR INICJUJĄCY

*Monika Borowczyk*

REDAKCJA

*Aleksandra Urzędowska*

KOREKTA

*Małgorzata Mazur*

SKŁAD I ŁAMANIE

*Munda – Maciej Torz*

KOREKTA TECHNICZNA

*Anna Jakubczyk*

PROJEKT OKŁADKI

*efectoro.pl*

*agencja komunikacji marketingowej*

Zdjęcie wykorzystane na okładce: © Depositphotos.com/pressmaster

© Copyright by Dorota Witkowska, Łódź 2023

© Copyright for this edition by Uniwersytet Łódzki, Łódź 2023

<https://doi.org/10.18778/8142-304-5>

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego

Wydanie I. W.10213.20.0.M

Ark. wyd. 17; ark. druk. 23,375

ISBN 978-83-8142-304-5

e-ISBN 978-83-8142-305-2

Wydawnictwo Uniwersytetu Łódzkiego

90-237 Łódź, ul. Jana Matejki 34A

[www.wydawnictwo.uni.lodz.pl](http://www.wydawnictwo.uni.lodz.pl)

e-mail: [ksiegarnia@uni.lodz.pl](mailto:ksiegarnia@uni.lodz.pl)

tel. 42 635 55 77

*Pamięci moich Mistrzów, w podzięce moim Bliskim*

# Spis treści

Wstęp	11
Rozdział 1	
<b>Wprowadzenie do analizy danych</b>	<b>15</b>
1.1. Podstawowe pojęcia	15
1.2. Pomiar zjawisk, skale pomiarowe	16
1.3. Metody pozyskiwania danych do badania, źródła danych	20
1.4. Grupowanie i prezentacja zebranych danych	24
1.5. Elementy rachunku prawdopodobieństwa	39
Rozdział 2	
<b>Analiza struktury danych</b>	<b>51</b>
2.1. Wskaźniki struktury	52
2.2. Miary położenia	54
2.3. Miary zróżnicowania	59
2.4. Miary asymetrii i koncentracji	67
Rozdział 3	
<b>Metody wnioskowania statystycznego</b>	<b>81</b>
3.1. Podstawowe pojęcia związane z estymacją nieznanych parametrów	81
3.2. Estymacja podstawowych parametrów opisowych zbiorowości	85
3.3. Wprowadzenie do weryfikacji hipotez statystycznych	92
3.4. Weryfikacja hipotez dotyczących parametrów opisowych populacji	97
Rozdział 4	
<b>Analiza współzależności zjawisk</b>	<b>109</b>
4.1. Rozkład dwuwymiarowy	109
4.2. Występowanie współzależności	111
4.3. Podstawowe miary korelacji dwóch cech mierzalnych i porządkowych	116

## 8 Spis treści

4.4. Podstawowe miary współzależności dwóch cech niemierzalnych	126
4.5. Analiza regresji	136

### Rozdział 5

<b>Analiza dynamiki zjawisk</b>	<b>147</b>
5.1. Mierniki dynamiki	148
5.2. Dekompozycja szeregu czasowego	153
5.3. Metody wygładzania szeregów czasowych	170

### Rozdział 6

<b>Modelowanie ekonometryczne</b>	<b>179</b>
6.1. Sformułowanie modelu ekonometrycznego	179
6.2. Klasyfikacja modeli ekonometrycznych	183
6.3. Budowa modeli ekonometrycznych	186
6.4. Rodzaje zmiennych w modelach ekonometrycznych	192
6.5. Modele zmiennych jakościowych	201
6.6. Modele szeregów czasowych	211

### Rozdział 7

<b>Prognozowanie na podstawie szeregów czasowych</b>	<b>219</b>
7.1. Podstawowe pojęcia	220
7.2. Sformalizowane metody prognozowania	223
7.3. Błędy prognoz	238

### Rozdział 8

<b>Elementy wielowymiarowej analizy statystycznej</b>	<b>245</b>
8.1. Podstawowe pojęcia	246
8.2. Mierniki taksonomiczne	248
8.3. Metody klasyfikacji	266
8.4. Metody grupowania	273
8.5. Metody doboru cech diagnostycznych	275

### Rozdział 9

<b>Metody badania finansowych szeregów czasowych</b>	<b>279</b>
9.1. Stopy zwrotu	279
9.2. Metody analizy własności szeregów stóp zwrotu	282
9.3. Modele opisujące kształtowanie się stóp zwrotu	288
9.4. Wykorzystanie metod statystycznych w analizach finansowych szeregów czasowych	294
9.4.1. Analiza notowań wybranej spółki	294
9.4.2. Analiza szeregu stóp zwrotu	298
9.4.3. Modele wyceny aktywów kapitałowych	307



## Rozdział 10

<b>Analiza obiektów wielowymiarowych</b>	<b>311</b>
10.1. Ocena zdolności kredytowej	311
10.1.1. Analiza dyskryminacyjna	315
10.1.2. Klasyfikacja bezwzorcowo	322
10.1.3. Porównanie regresji logistycznej i modeli dyskryminacyjnych	325
10.1.4. Porównanie regresji binarnej, bayesowskiej analizy dyskryminacyjnej	329
10.1.5. Klasyfikacja indywidualnych kredytobiorców za pomocą drzew klasyfikacyjnych	331
10.2. Ocena sytuacji finansowej spółek za pomocą mierników syntetycznych	334
10.2.1. Opis danych	336
10.2.2. Budowa mierników syntetycznych	338

## Rozdział 11

<b>Badanie zmian cen na rynku dóbr heterogenicznych</b>	<b>341</b>
11.1. Dzieła sztuki jako przedmiot inwestycji alternatywnych	341
11.2. Indeksy hedoniczne cen	343
11.3. Hedoniczne indeksy cen obrazów najbardziej popularnych polskich malarzy	346
11.3.1. Baza danych i wybór zmiennych hedonicznych	347
11.3.2. Modele regresji hedonicznej	353
11.3.3. Hedoniczne indeksy cen	362
Zakończenie	365
Literatura	367

# Wstęp

Metody ilościowe umożliwiają prowadzenie analiz zjawisk społeczno-gospodarczych, tym samym wspomagają podejmowanie decyzji i obiektywną ocenę istniejącej sytuacji. Oprócz tego pozwalają rozpoznać mechanizmy funkcjonowania układów gospodarczych, zidentyfikować istniejące determinanty zjawisk ekonomicznych, a także przewidywać skutki podjętych działań oraz zmiany zachodzące w badanych procesach. Tworzą zatem przesłanki do określania przyszłego rozwoju społeczno-gospodarczego. Metody ilościowe są również wykorzystywane do weryfikacji teorii ekonomicznych, a nawet krążących poglądów i opinii.

Przez ponad 40 lat pracy zawodowej autorka stara się aktywnie wykorzystywać bogatą paletę metod statystycznych, ekonometrycznych i optymalizacyjnych, zarówno w badaniach naukowych, jak i tych realizowanych dla praktyki gospodarczej. Celem tej monografii jest przekazanie jej doświadczeń badawczych z zakresu aplikacji wybranych metod statystycznych.

Kiedy autorka zaczynała pracę jako nauczyciel akademicki, wszyscy studenci kierunków ekonomicznych musieli realizować znaczną liczbę godzin dydaktycznych przeznaczonych na przedmioty ilościowe, takie jak matematyka ekonomiczna, statystyka, ekonometria rozumiana w jej szerokim pojęciu, tzn. obejmująca nie tylko modele opisowe, ale również tzw. modele przepływów międzygałęziowych (stworzone przez Wassily'ego Leontiefa, a dzisiaj zwane analizami *input-output*) i metody programowania liniowego (tzn. modele optymalizacyjne będące elementem badań operacyjnych). Wówczas każdy absolwent posiadał podstawową wiedzę z tego zakresu i rozumiał przynajmniej elementarne pojęcia, a przeszkodę do aplikacji metod ilościowych w praktyce stanowiły możliwości obliczeniowe, ponieważ nie każdy wykształcony ekonomista miał nawet pośrednio dostęp do komputerów. Wynikało to przynajmniej z dwóch powodów. Po pierwsze, komputery były w posiadaniu jedynie dużych i „strategicznym” jednostek, takich jak np. uczelnie, duże przedsiębiorstwa i niektóre agencje rządowe. Wprawdzie istniały tzw. Zakłady Elektronicznej Techniki Obliczeniowej (ZETO), będące komercyjnymi ośrodkami przetwarzania danych, z których usług można było odpłatnie korzystać, ale nie były one powszechnie dostępne, chociażby z powodu

ograniczonych (sprzętowych i pracowniczych) możliwości wykonywania usług. Po drugie, obsługa komputerów, nawet jeśli się miało do nich dostęp, wymagała wiedzy informatycznej<sup>1</sup> daleko bardziej zaawansowanej niż obsługa współczesnych przyjaznych użytkownikom urządzeń.

W okresie gospodarki socjalistycznej nie do końca też wierzone, że obiektywne i racjonalnie prowadzone metodami ilościowymi analizy mogą być przydatne w praktyce gospodarczej, czego wyrazem były wątpliwości pracujących studentów: „co tu optymalizować, jak mam plan do wykonania”. Z podobnym podejściem autorka spotkała się również wiele lat później, tj. już w okresie gospodarki rynkowej, kiedy to jeden z dyrektorów dużego banku działającego w Polsce stwierdził, że nie są mu potrzebne nowoczesne metody oceny zdolności kredytowej, bo jego rozliczają z procedur, a straty wynikające z błędnych decyzji kredytowych są wliczone w koszty obsługi klientów.

Ostatnie 30 lat to dla autorki nie tylko przekształcenie polskiej gospodarki z centralnie planowanej w rynkową, ale przede wszystkim przemiany technologiczne, których nieustannie doświadczamy. Jednakże prawdziwą rewolucją tego pokolenia okazał się wręcz niewyobrażalny rozwój technologii, który sprawił, że koncepcje rodem z powieści *science fiction* Lema stały się naszą codziennością. Pojawiły się bowiem komputery osobiste, oprogramowanie rozwiązujące skomplikowane problemy numeryczne poprzez jedno kliknięciem, Internet będący nieograniczoną wręcz skarbnicą wiedzy wszelakiej i komunikatory będące imitacją bliskości. Konieczne stało się zatem nabywanie innych umiejętności – w tym obsługi komputerów, smartfonów i portali społecznościowych. Jednocześnie (niestety) uznano, że technologia i nowe kompetencje zastąpią logiczne rozumowanie.

Szkolnictwo wyższe również ulegało stopniowej zmianie – komercjalizacji, bowiem środki budżetowe dla uczelni „szły za studentem” zatem w „walce konkurencyjnej” pozbywano się z programów nauczania „nieprzyjemnych” i nie lubianych przez studentów przedmiotów ilościowych. W konsekwencji od ponad 10 lat absolwenci wielu kierunków ekonomicznych mają znikomy dostęp lub w ogóle nie poznają metod ilościowych, a co najwyżej przechodzą kurs „klikania” bez rozumienia istoty wykorzystywanych metod, a zatem nie wiedzą, jakie informacje otrzymują w wyniku zastosowania profesjonalnego oprogramowania, i nie umieją ani zinterpretować uzyskanych wyników, ani przeprowadzić ich krytycznej analizy. Co więcej, w wielu przypadkach nawet nie potrafią odpowiednio przygotować danych empirycznych, będących „wsadem” do programów komputerowych. Znana reguła

1 Przykładowo, aby przygotować pracę magisterską, musiałam sobie sama napisać program w języku ALGOL, a dostęp do komputera w Ośrodku Obliczeniowym Uniwersytetu Łódzkiego miałam ograniczony do kilku półgodzinnych sesji około północy. Natomiast do doktoratu program (liczący około 2 tys. kart perforowanych) napisał dla mnie programista z Politechniki Warszawskiej, a obliczenia robiłam w Instytucie Badań Jądrowych w Świerku pod Warszawą, bo na Uniwersytecie Łódzkim nie było dostępnych terminów na wielogodzinne przetwarzanie danych.

*garbage in, garbage out* wydaje się ich nie dotyczyć, a brak znajomości metod i zasad ich stosowania, pozostawia pseudoanalitykom nieograniczone pole do aplikacji najbardziej wyszukanych i skomplikowanych metod, które wymagają jedynie kliknięcia w odpowiedni klawisz na klawiaturze lub w ikonę na ekranie monitora.

Autorka jest zdecydowanie przeciwna takim praktykom i choć popiera stosowanie metod ilościowych w analizach społeczno-ekonomicznych, woli, aby używano nieskomplikowanych, ale zrozumiałych (przez analityków) metod, zamiast bełkotania przez ignorantów o „wyklikanych” wynikach. Wszelkiego rodzaju dane pozwalają na wyciągnięcie wielu interesujących wniosków, które uzyskać można za pomocą relatywnie prostych narzędzi, a problemem jest jedynie wielkość i gwałtowny rozrost zbiorów danych (tzw. *big data*). Wykorzystywanie bardziej zaawansowanych metod często wiąże się z wieloma rozczarowaniami i koniecznością stosowania wielu podejść w celu uzyskania zadowalających wyników<sup>2</sup>, czego też jakby się nie zauważa, pokazując bezkrytycznie (m.in. w artykułach „ni-by-naukowych”) bezsensowne wyniki, które mają świadczyć o „kompetencjach” autora lub braku przydatności metod. Na drugim biegunie niekompetencji zdarza się, że do rangi osiągnięcia naukowego podnosi się obliczenie procentów, prostego indeksu lub estymację funkcji trendu. Innym przykładem może być próba aplikacji dla polskich jednostek gospodarczych modelu predykcji bankructwa, oszacowanego przez Edwarda I. Altmana w 1968 roku dla gospodarki amerykańskiej. Przy czym nie dostrzega się, że minęło już ponad pół wieku od estymacji parametrów tego modelu, a przedsiębiorstwa działające w USA podlegają innemu prawu i regułom funkcjonowania niż polskie firmy. Jednocześnie w pseudonaukowych publikacjach bezmyślnie poszukuje się przyczyn braku skuteczności tego modelu, np. w zaokrągleniach lub przekłamaniach wartości parametru<sup>3</sup>, podczas gdy przyczyna jest prosta i dawno odkryta – trzeba ten model dyskryminacyjny oszacować na nowo, uwzględniając istniejące (tu i teraz) realia gospodarcze, o czym zresztą pisał sam Altman.

Rozwój technologii gromadzenia i przetwarzania danych oraz globalizacja rynków finansowych sprawiają, że podejmowanie właściwych decyzji wymaga posługiwania się metodami ilościowymi. Dlatego niniejsze opracowanie zawiera omówienie wybranych metod statystycznych i ekonometrycznych wykorzystywanych w finansach. Zadaniem, które postawiła przed sobą autorka, było pokazanie,

2 Niejednokrotnie zdarzało się, że np. podczas budowania modelu ekonometrycznego akceptowalny okazywał się jego dwudziesty lub pięćdziesiąty wariant, co pokazuje, jak wiele wysiłku wymaga tak – zdawałoby się – proste zadanie.

3 Celowo nie zacytowano tutaj tego typu prac z dwóch powodów: po pierwsze, aby nie wytykać nikogo palcem i nie zawstydząć, zwłaszcza jeśli autor jest na początku kariery, bo istnieje nadzieja, że się doksztalci; po drugie, w dzisiejszych czasach ocena „jakości” naukowca dokonywana jest na podstawie liczby cytowań i to niezależnie od tego, w jakim kontekście omawia się daną pracę. Nie warto zatem przysparzać tym autorom dodatkowych punktów w ocenie ich dorobku naukowego, przytaczając ich błędne lub bezwartościowe prace.

że stosowanie nawet relatywnie prostych metod istotnie wspomaga podejmowanie decyzji.

Gromadzenie danych rzeczywistych i prowadzenie zaawansowanych analiz wymaga oczywiście korzystania z komputerów i (najlepiej) profesjonalnego oprogramowania, ale to analityk musi zdecydować, które z metod są adekwatne do charakteru zgromadzonych obserwacji, i dokonać ich świadomego wyboru z katalogu dostępnych w pakiecie. Niezbędna jest również umiejętność wyciągania poprawnych wniosków na podstawie uzyskanych wyników oraz krytyczna ocena ich akceptowalności. Innymi słowy, trzeba wiedzieć, co otrzymuje się w wyniku zastosowania konkretnego narzędzia oraz w jaki sposób można przeprowadzić badanie poprawności uzyskanych wyników.

Monografia składa się z jedenastu rozdziałów, w których omówiono wybrane zagadnienia z zakresu statystyki opisowej i matematycznej, modelowania ekonometrycznego i wielowymiarowej analizy porównawczej. Większość przedstawionych metod została zilustrowana przykładami z zakresu finansów<sup>4</sup>. Prezentowane wyniki badań empirycznych wskazują na pojawiające się problemy i ograniczenia w stosowalności omawianych metod. Kolejność rozdziałów wynika z chęci systematyzacji rozpatrywanych zagadnień.

I tak rozdział pierwszy zawiera omówienie zagadnień związanych z: pomiarem zjawisk społeczno-gospodarczych, metodami pozyskiwania danych oraz ich właściwą prezentacją. Rozdział drugi poświęcono metodom opisu struktury danych, wskazując na ich ograniczenia. W kolejnej części pokazano, jak uogólniać wyniki uzyskane na podstawie badania niepełnego. Czwarty rozdział dotyczy analizy współzależności. Następny zawiera omówienie metod analiz dynamiki zjawisk. Kolejny dedykowany jest modelom ekonometrycznym i to zarówno tym klasycznym, jak i modelom zmiennych binarnych oraz modelom szeregów czasowych. Rozdział siódmy stanowi wprowadzenie do prognozowania, a ósmy do wielowymiarowej analizy porównawczej<sup>5</sup>. Natomiast ostatnie trzy rozdziały stanowią pogłębioną ilustrację aplikacji w finansach większości metod omawianych wcześniej. Tak więc rozdział dziewiąty to opis przeprowadzonych analiz finansowych szeregów czasowych na przykładzie szeregu notowań cen akcji wybranej spółki giełdowej. Natomiast dziesiąty dotyczy dwóch problemów – pierwszy związany jest z zastosowaniem wybranych metod klasyfikacji do oceny wiarygodności kredytowej klientów banku, drugi polega na ocenie kondycji ekonomiczno-financej spółek z uwzględnieniem różnych aspektów ich funkcjonowania, przy zastosowaniu mierników taksonomicznych. Ostatni rozdział poświęcono budowie hedonicznych indeksów cen dzieł sztuki, które uwzględniają ich charakterystyki jakościowe.

---

4 Wszędzie tam, gdzie było to możliwe bez nadmiernej komplikacji obliczeniowej, w rozdziałach 1–8 pokazano rzeczywiste dane.

5 Ilustrację metod omawianych w rozdziale 8 i częściowo 6 przedstawiono w rozdziale 10.

## Rozdział 1

# Wprowadzenie do analizy danych

Niniejszy rozdział ma charakter wprowadzający, zatem poświęcony będzie podstawowym zagadnieniom związanym z obserwacją, pomiarem i grupowaniem danych wykorzystywanych w finansach. Omówione zostaną podstawowe pojęcia, gdyż każda dziedzina wiedzy posługuje się charakterystycznym dla niej językiem. Rozważania rozpocznie omówienie zagadnień pomiaru, a dalej wskazane zostaną źródła danych i sposób ich prezentacji, a w ostatnim podrozdziale podstawowe pojęcia z rachunku prawdopodobieństwa.

### 1.1. Podstawowe pojęcia

Statystyka jest nauką o metodach badania prawidłowości występujących w zjawiskach masowych, czyli takich, które mogą zachodzić nieograniczoną liczbę razy i w dużej masie zdarzeń wykazują pewne prawidłowości, jakich nie można zaobserwować w pojedynczym przypadku. Przedmiotem badań statystycznych są zbiorowości statystyczne (populacje), które stanowią zbiór elementów (jednostek) powiązanych ze sobą logicznie i jednocześnie nieidentycznych. Zbiorowość statystyczna to zbiór jednostek (osób, rzeczy lub zjawisk) objętych badaniem statystycznym. Prawidłowe prowadzenie badania wymaga, aby zbiorowość statystyczna była jednoznacznie określona i wyodrębniona. Dokonuje się tego, ustalając cel badania i precyzując, kto lub co należy do badanej zbiorowości, czyli definiuje się jej elementy. Poszczególne jednostki (elementy), które wchodzi w skład badanej zbiorowości statystycznej, posiadają wspólną cechę (właściwość, charakterystykę, atrybut), a jednocześnie różnią się między sobą innymi, sobie właściwymi cechami. Pojedynczą cechę oznacza się przez  $X$ , a warianty, jakie może przyjmować, przez  $x_i$  (dla  $i = 1, 2, \dots, k$ ), czyli:  $x_1, x_2, \dots, x_k$ . W analizach uwzględnia się tylko te cechy, które są istotne z punktu widzenia celu realizowanego badania.

Analizy zjawisk i obiektów można prowadzić na podstawie:

- 1) badania pełnego (wyczerpującego lub całkowitego), obejmującego wszystkie jednostki danej zbiorowości statystycznej,
- 2) badania niepełnego (częściowego), obejmującego jedynie wybrane jednostki badanej zbiorowości statystycznej.

Wybór sposobu badania wynika z celu i zakresu badania, a także wiąże się z zastosowaniem odpowiednich narzędzi badawczych. W pierwszym przypadku będą to metody umożliwiające statystyczny opis badanej zbiorowości. Natomiast w drugim istotny jest sposób doboru elementów do badania, aby możliwe było zastosowanie metod statystyki matematycznej w celu wnioskowania statystycznego o całej zbiorowości na podstawie danych pochodzących z próby<sup>1</sup>.

## 1.2. Pomiar zjawisk, skale pomiarowe

W analizach finansowych wykorzystuje się różnego rodzaju informacje dotyczące zarówno pojedynczych charakterystyk, jak i złożonych zjawisk, obiektów lub procesów społeczno-gospodarczych. Badania prowadzone w naukach społecznych, w tym w szeroko rozumianych finansach, charakteryzują się:

- 1) złożonością przedmiotów badania (tj. obiektów, procesów i zjawisk),
- 2) trudnością prowadzenia obserwacji i dokonywania pomiaru,
- 3) zmiennością w czasie.

Złożoność badań społecznych wynika z konieczności jednoczesnego uwzględnienia w analizach wielu różnych charakterystyk. Innymi słowy, rozpatrywane zjawiska i obiekty są wielowymiarowe (wielocechowe) i do ich opisu wykorzystuje się wiele cech (zmiennych). Przykładem obiektu badania jest przedsiębiorstwo lub kredytobiorca, a zjawiska – notowania cen akcji wybranej spółki. Atrybutami przedsiębiorstwa mogą być: liczba pracowników, osiągnięty przychód, forma własności, udział w rynku etc. Wiarygodność kredytowa klienta indywidualnego oceniana jest m.in. z punktu widzenia uzyskiwanych dochodów, wykonywanego zawodu, miejsca pracy, płci itp. W przypadku analiz cen akcji bierze się pod uwagę wiele czynników, zarówno związanych ze spółką (np. wskaźniki ekonomiczno-finansowe, charakter działalności, sposób zarządzania), jak i dotyczących jej szeroko rozumianego otoczenia (np. parametry makroekonomiczne, takie jak poziom stóp procentowych, faza wzrostu gospodarczego, sytuacja na rynkach kapitałowych, branża itp.).

1 Szersze omówienie tych problemów znaleźć można w bogatej literaturze przedmiotu, np. w pracach: [Aczel 2000; Luszniwicz, Staby 2003; Witkowska 2004].



Trudność obserwacji i pomiaru związana jest z faktem, że większość zjawisk nie jest bezpośrednio obserwowalna i występują problemy z ich pomiarem. Przykładowo popyt na dane dobro oznacza chęć nabycia go, ale nie każdy chętny może dokonać zakupu, zatem popytu się nie obserwuje, a jedynie tzw. popyt zrealizowany, który mierzony jest wartością lub wielkością zakupu danego dobra. Innym przykładem jest inflacja, która oznacza wzrost kosztów utrzymania, ale z uwagi na to, że ceny na różne towary i usługi nie rosną w taki sam sposób, to wzrost kosztów utrzymania zależy od koszyka dóbr i usług, będących przedmiotem konsumpcji poszczególnych gospodarstw domowych. Inflacja jest zatem różnie odczuwalna, a więc nie można jej bezpośrednio zaobserwować w skali gospodarki, a jej pomiar zależy od przyjętych założeń, np. odnośnie do koszyka dóbr i usług oraz formuły obliczeniowej wskaźnika zmiany cen. Innym przykładem jest wiarygodność kredytowa, która nie może być bezpośrednio obserwowana ze względu na swoją złożoność, jest natomiast określana za pomocą różnych metod, odmiennych dla klientów indywidualnych i przedsiębiorstw.

Rozwój społeczno-gospodarczy sprawia, że warunki, w jakich funkcjonują jednostki gospodarcze oraz gospodarstwa domowe, ulegają ciągłym zmianom. W konsekwencji badane obiekty i zjawiska również podlegają przeobrażeniom w czasie. Ta zmienność sprawia, że konieczna jest ciągła aktualizacja danych i prowadzenie analiz na bieżąco.

Obserwacje zjawisk i obiektów badania prowadzone są w różnorodny sposób, w różnych miejscach i momentach. W związku z tym uzyskane w wyniku obserwacji statystycznej informacje mogą przyjmować postać danych przekrojowych, czasowych lub panelowych. Dane przekrojowe powstają, gdy w określonym czasie obserwuje się różne obiekty badania, np. w konkretnym okresie analizuje się ofertę różnych banków w celu założenia lokaty lub zaciągnięcia kredytu. Innym przykładem może być analiza wyników finansowych spółek na koniec danego roku obliczeniowego. Dla tego typu danych moment dokonania pomiaru nie jest istotny, a numer każdej obserwacji zazwyczaj oznacza się przez  $i$ , ( $i = 1, 2, \dots, n$ ). O danych czasowych mówi się, kiedy analiza prowadzona jest dla jednego obiektu w różnych momentach lub okresach czasu. Przy czym przez moment rozumie się punkt na osi czasu, a przez okres przedział na osi czasu. Przykładowo badanie dotyczy rozwoju konkretnego przedsiębiorstwa w kilkuletnim okresie, wówczas liczba obserwacji  $T$  równa jest liczbie okresów (lub momentów) objętych badaniem, np. kolejnych lat. Szereg czasowy tworzą notowania kursów akcji rejestrowane w trakcie kolejnych sesji giełdowych. W szeregach czasowych numer każdej obserwacji oznacza się przez  $t$ ,  $t = 1, 2, \dots, T$ , a kolejne obserwacje są porządkowane chronologicznie. W przypadku danych panelowych obserwacja prowadzona jest jednocześnie w dwóch wymiarach. Przy czym w praktyce najczęściej wykorzystuje się dane przekrojowo-czasowe oznaczające obserwacje poszczególnych jednostek statystycznych w różnym czasie. Przykładem tego typu danych jest ocena rozwoju wielu spółek w kilkuletnim horyzoncie czasowym w celu budowy portfela



inwestycyjnego, wówczas liczba obserwacji  $nT$  równa jest iloczynowi liczby spółek i liczby okresów (lub momentów) objętych badaniem.

Kluczowym pojęciem używanym w statystyce są cechy (statystyczne), czyli właściwości charakteryzujące jednostki wchodzące w skład badanej zbiorowości. Jak wcześniej wspomniano, cechę statystyczną oznacza się przez  $X$ , a jej warianty podlegające obserwacji zapisuje się jako  $x_1, x_2, \dots, x_k$ , gdzie  $k$ , to liczba wyróżnionych wariantów badanej charakterystyki. Wśród cech statystycznych wyróżnia się:

- cechy mierzalne (ilościowe, wymierne), które można wyrazić za pomocą liczb z podaniem odpowiednich uniwersalnych jednostek miary (np. wartość majątku w mln PLN, dochód w tys. PLN, wartość środków trwałych w tys. EUR, kurs euro w złotychkach, liczba rat, wiek w latach, masa w kilogramach, długość w metrach, czas w godzinach, wielkość firmy mierzona liczbą zatrudnionych);
- cechy niemierzalne (jakościowe, niewymierne), których nie można zmierzyć za pomocą uniwersalnych jednostek miary, a jedynie stwierdza się występowanie lub nie określonego wariantu danej cechy (np. forma własności, branża, w jakiej działa przedsiębiorstwo, waluta kredytu, płeć, zawód, wykształcenie, wielkość firmy według podziału na: mikroprzedsiębiorstwa, firmy małe, średnie i duże).

Z terminem „cecha” ściśle związane jest pojęcie zmiennej, którą często określa się jako pomiar cechy. Wśród cech ilościowych wyróżnia się zmienne ciągłe lub skokowe. W pierwszym przypadku zmienna może przyjmować każdą wartość z określonego, skończonego przedziału liczbowego. W drugim przypadku może przyjmować ona jedynie określone wartości z tego przedziału, często należące do zbioru liczb całkowitych lub naturalnych (np. liczba pracowników to cecha skokowa, liczba etatów może być cechą ciągłą). Sposób pomiaru przedmiotu badania pozwala wyróżnić cztery podstawowe skale pomiarowe: nominalną, porządkową, przedziałową i ilorazową (por. [Aczel 2000, s. 36–37; Stanimir 2006, s. 21–22; Sobczyk 2010, s. 15–18]).

Skala nominalna dotyczy cech jakościowych, między którymi można wyróżnić tylko dwie relacje, tj. równości (identyczności) i różności (tj. braku identyczności). Innymi słowy o cechach mierzonych na tej skali w dwóch i więcej jednostkach statystycznych (lub obiektach) można jedynie powiedzieć, czy posiadają taki sam wariant tego atrybutu, lub czy różnią się pod względem tej charakterystyki. W przypadku cech mierzonych na skali nominalnej stosuje się kodowanie binarne za pomocą zmiennych przyjmujących wartość 1, oznaczającą występowanie danego wariantu cechy, lub 0 w sytuacji przeciwnej. Szczególnym przypadkiem skali nominalnej jest skala dychotomiczna (dwudzielna), z której korzysta się w przypadku cechy dwuwariantowej, której przykładem jest płeć lub wyróżnienie wśród spółek tych z udziałem Skarbu Państwa. Kiedy wariantów danej cechy jest więcej niż dwa (np. branża, zawód, forma własności, sektor gospodarki według klasyfikacji PKD, numery tras autobusowych), to mówi się o cechach wie-

lodzielnych (politomicznych). W przypadku cech wielodzielných do kodowania danych używa się  $k$  zmienných binarných (gdzie  $k > 1$ ).

Skala porządkowa (rangowa) pozwala określić relacje większości i mniejszości między wyróżnionymi wariantami cech oraz umożliwia porządkowanie obserwacji według wyników pomiaru. Jednakże pomiar i porządkowanie są dokonywane według subiektywnie ustalonych jednostek pomiarowych, zatem działania matematyczne inne niż wymienione są niedopuszczalne. Przykładami pomiarów na tej skali mogą być: grupa ryzyka kredytowego, poziom wykształcenia (podstawowe, średnie, wyższe), ocena produktu (dobry, średni, zły) lub wielkość przedsiębiorstwa (małe, średnie i duże).

Zmienne mierzone na skali przedziałowej (interwałowej) przyjmują wartości liczbowe, które można uporządkować na osi liczbowej z podaniem stałej i obiektywnie istniejącej jednostki pomiarowej. Skala ta nie posiada jednak naturalnego punktu zerowego, który wyznaczany jest arbitralnie na potrzeby konkretnego badania, zazwyczaj według przyjętej konwencji, np. skala temperatury, mierzona stopniami Celsjusza, ma punkt zero w temperaturze zamarzania wody. Umowność punktu zerowego implikuje również nierówność różnic między wartościami na przyjętej skali (np. przy temperaturze  $40^{\circ}\text{C}$  nie jest dwa razy cieplej niż przy  $20^{\circ}\text{C}$ ). Skala przedziałowa umożliwia wykonywanie operacji dodawania i odejmowania. Niemożliwe jest natomiast mnożenie i dzielenie dwóch wielkości interwałowych, ich potęgowanie oraz używanie średniej kwadratowej i geometrycznej (por. [Zeliaś 2000, s. 34]). Przykładem pomiaru na skali interwałowej może być wynik finansowy przedsiębiorstwa (zysk lub strata), rok rozpoczęcia działalności firmy. Innym przykładem jest opinia na zadany temat podana w punktach na skali Likerta, która podporządkowana jest zasadom skalowania zrównoważonego, posiada bowiem tyle samo wariantów odpowiedzi pozytywnych, co negatywnych i jest skalą niewymuszoną z neutralną odpowiedzią, kiedy nie ma się zdania na zadany temat. Przyjmuje się, że odległości między kolejnymi wariantami odpowiedzi są identyczne, co pozwala traktować skalę Likerta jako interwałową. Dodatkowo można do poszczególnych odpowiedzi przyporządkować wartości liczbowe, co dopuszcza późniejsze wykonywanie operacji matematycznych (por. [Jajuga, Walesiak 2015, s. 8–10; Zeliaś 2000, s. 34]).

Zmienna jest mierzona na skali ilorazowej (stosunkowej), jeśli zbiór jej wartości można jednoznacznie uporządkować na osi liczbowej z podaniem stałej jednostki pomiarowej i jednocześnie występuje w niej naturalny punkt zerowy. Innymi słowy zmienna jest na skali ilorazowej, jeśli stosunki między dwoma pomiarami na tej skali mają interpretację w świecie rzeczywistym. Skala stosunkowa nie nakłada żadnych ograniczeń dotyczących stosowania operacji matematycznych i metod statystycznych. Zatem skala ilorazowa dodatkowo (tj. w stosunku do skali przedziałowej) pozwala na wykonywanie operacji mnożenia i dzielenia. Przykładem cech mierzonych na tej skali są: temperatura w stopniach Kelvina, liczba zatrudnionych, inflacja, cena złota, kurs wymiany walut, wskaźnik cena/zysk

w stosunku do akcji<sup>2</sup>, wiek (w latach), wzrost (w centymetrach), długość (w metrach), dochód kredytobiorcy lub zysk firmy.

Omówione skale pomiarowe przedstawiono w kolejności od najsłabszej do najmocniejszej. Warto odnotować, że mają one właściwości kumulatywne, czyli charakteryzują się narastającym stopniem dokładności pomiaru, a każda silniejsza skala zawiera wszystkie przymioty skali słabszej. Dwie pierwsze skale (tj. nominalną i przedziałową) określa się jako słabe (niemetryczne) i stosuje się je w przypadku cech jakościowych. Natomiast skale przedziałowa i ilorazowa określa się jako mocne (metryczne) i wykorzystuje w przypadku cech ilościowych. W zastosowaniach praktycznych określenie skali pomiaru zmiennych jest bardzo ważne, ponieważ determinuje możliwość zastosowania konkretnych metod analiz. Problem z wykorzystywaniem niektórych metod może się pojawić, gdy w analizowanym zbiorze danych występują zmienne mierzone na skalach różnych typów lub tylko na niemetrycznych. Zazwyczaj przyjmuje się, że metody, które stosuje się do wyników pomiaru w skali słabszej, można też zastosować w przypadku zmiennych mierzonych na skali mocniejszej, ale nie odwrotnie. Oznacza to, że teoretycznie nie jest możliwe wzmacnianie skal. Jednak, jak twierdzą Gatnar i Walesiak [2004, s. 22], w praktycznych badaniach ekonomiczno-społecznych na ogół sztucznie wzmacnia się skale pomiaru, traktując zmienne mierzone na skalach niemetrycznych jak zmienne przedziałowe i ilorazowe.

### 1.3. Metody pozyskiwania danych do badania, źródła danych

Jak wcześniej wspomniano, badanie może obejmować wszystkie obiekty będące jednostkami danej zbiorowości lub tylko ich część. Zazwyczaj to prowadzący badanie musi dokonać wyboru między badaniem pełnym i częściowym, chociaż zdarzają się przypadki, kiedy badanie całkowite nie jest wykonalne, np. obserwacje niektórych danych ekonomicznych możliwe są jedynie w wybranych punktach pomiarowych, np. zysk przedsiębiorstwa dostępny jest na koniec roku, notowanie kursów akcji podaje się na otwarciu i zamknięciu sesji giełdowej. Przykładem badania pełnego są spisy statystyczne, rejestracje statystyczne i sprawozdawczość statystyczna, np. lista obecności, zapis księgowy majątku przedsiębiorstwa, rejestracja operacji wykonanych na koncie bankowym.

<sup>2</sup> Wskaźnik cena/zysk (oznaczany również jako C/Z lub P/E) jest powszechnie używaną przez inwestorów giełdowych miarą opłacalności zakupu akcji danej spółki, którą oblicza się, dzieląc cenę rynkową jednej akcji przez zysk netto przypadający na jedną akcję.