

OPIS STATYSTYCZNY W ZARZĄDZANIU

ZBIÓR ZADAŃ

Anna Małgorzata Olszewska
Beata Madras-Kobus

OPIŚ STATYSTYCZNY W ZARZĄDZANIU

ZBIÓR ZADAŃ

Anna Małgorzata Olszewska
Beata Madras-Kobus



OFICyna WYDAWNICZA POLITECHNIKI BIAŁOSTOCKIEJ
BIAŁYSTOK 2025

Recenzenci:
dr Marta Kusterka-Jefmańska
dr Elżbieta I. Misiewicz

Redaktor naukowy dyscypliny nauki o zarządzaniu i jakości:
dr hab. Ewa Glińska, prof. PB

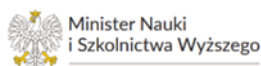
Redakcja i korekta językowa:
Katarzyna Duniewska

Skład, grafika i okładka:
Marcin Dominów

© Copyright by Politechnika Białostocka, Białystok 2025

ISBN 978-83-68077-97-1 (e-Book)

DOI: 10.24427/978-83-68077-97-1



Zadanie finansowane w ramach programu Ministerstwa Nauki i Szkolnictwa Wyższego
pod nazwą „Politechniczna Sieć VIA CARPATIA im. Prezydenta RP Lecha Kaczyńskiego”
w latach 2022-2025, kwota finansowania 10 990 546 zł



POLITECHNIKA
LUBELSKA



Publikacja jest udostępniona na licencji
Creative Commons Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 4.0
(CC BY-NC-ND 4.0).

Pełną treść licencji udostępniono na stronie
creativecommons.org/licenses/by-nc-nd/4.0/legalcode.pl.

Publikacja jest dostępna w Internecie na stronie Oficyny Wydawniczej PB.
Oficina Wydawnicza Politechniki Białostockiej
ul. Wiejska 45C, 15-351 Białystok
www.pb.edu.pl

Spis treści

Oznaczenia.....	5
Wstęp.....	7
1. Prezentacja danych.....	9
1.1. Zadania z zakresu prezentacji i klasyfikacji danych	25
1.2. Pytania testowe z zakresu klasyfikacji danych.....	32
1.3. Odpowiedzi do pytań testowych z zakresu klasyfikacji danych	35
2. Miary przeciętne opisu zbiorowości	36
2.1. Przeciętne miary klasyczne	36
2.2. Przeciętne miary pozycyjne	44
2.3. Zadania z zakresu miar przeciętnych	55
2.4. Odpowiedzi do zadań z zakresu miar przeciętnych	65
2.5. Pytania testowe z zakresu miar przeciętnych	67
2.6. Odpowiedzi do pytań testowych z zakresu miar przeciętnych.....	73
3. Miary dyspersji (zróznicowania, rozproszenia)	74
3.1. Klasyczne miary dyspersji.....	74
3.2. Pozycyjne miary dyspersji.....	85
3.3. Zadania z zakresu miar dyspersji	90
3.4. Odpowiedzi do zadań z zakresu miar zmienności	100
3.5. Pytania testowe z zakresu miar dyspersji.....	102
3.6. Odpowiedzi do pytań testowych z zakresu miar dyspersji.....	107

4. Miary asymetrii i koncentracji.....	108
4.1. Miary asymetrii (skośności).....	108
4.2. Miary koncentracji.....	115
4.3. Zadania z zakresu miar asymetrii i koncentracji	125
4.4. Odpowiedzi do zadań z zakresu miar asymetrii i koncentracji	132
4.5. Pytania testowe z zakresu miar asymetrii i koncentracji.....	133
4.6. Odpowiedzi do pytań testowych z zakresu miar asymetrii i koncentracji	138
Spis wykresów.....	139
Spis tabel	140
Literatura.....	143

Oznaczenia

x_i	– wartość cechy w i -tej klasie
n_i	– liczebność w i -tej klasie
f_i	– wskaźnik struktury (częstość, liczebność względna, odsetek) cechy w i -tej klasie
N	– liczebność badanej zbiorowości
$n_{i\ sk}$	– liczebność skumulowana w i -tej klasie
$f_{i\ sk}$	– częstość skumulowana w i -tej klasie
x_{0i}	– początek i -tego przedziału
x_{1i}	– koniec i -tego przedziału
h_i	– rozpiętość (długość) i -tego przedziału
$\overset{\circ}{x}_i$	– środek i -tego przedziału
x_{max}	– wartość maksymalna cechy
x_{min}	– wartość minimalna cechy
k	– liczba klas/przedziałów klasowych
$\bar{x}$	– średnia arytmetyczna
$\bar{x}_H$	– średnia harmoniczna
$\bar{x}_G$	– średnia geometryczna
D, Mo	– dominanta (moda, modalna)
Q_1	– kwartył pierwszy
Q_2, Me	– kwartył drugi (mediana, wartość środkowa)
Q_3	– kwartył trzeci
s^2	– wariancja
s	– odchylenie standardowe
Q	– odchylenie ćwiartkowe
R	– rozstęp
V_s	– współczynnik zmienności klasyczny (oparty na odchyleniu standardowym)
V_Q	– współczynnik zmienności pozycyjny (oparty na odchyleniu ćwiartkowym)

m_r	– moment zwykły r -ego rzędu
μ_r	– moment centralny r -ego rzędu
A_k	– współczynnik asymetrii (skośności) klasyczny
A_p	– współczynnik asymetrii (skośności) pozycyjny
A_s	– współczynnik asymetrii (skośności) mieszany
K	– współczynnik skupienia (kurtoza)
K'	– współczynnik ekscesu (eksces)
z_i	– procenty funduszu cechy
$z_{i\ sk}$	– skumulowane procenty funduszu cechy
P_K	– pole między krzywą Lorenza a przekątną (pole koncentracji)
K_L	– miara koncentracji Lorenza.

Wstęp

Opis statystyczny w zarządzaniu. Zbiór zadań, będący uzupełnieniem podręcznika *Elementy statystyki. Opis statystyczny*¹, został przygotowany z myślą o studentach – zarówno tych, którzy kształcą się na kierunkach inżynierskich, jak i licencjackich, gdzie statystyka opisowa jest jednym z przedmiotów obowiązkowych. To właśnie z myślą o tej szerokiej grupie odbiorców powstał niniejszy materiał, którego głównym celem jest ułatwienie opanowania podstawowych narzędzi służących do analizy i interpretacji danych liczbowych.

Zbiór zadań został opracowany w oparciu o wieloletnie doświadczenie Autorów w nauczaniu statystyki studentów uczelni technicznych i uniwersytetu. Z biegiem lat miałyśmy okazję obserwować, z jakimi trudnościami najczęściej mierzą się osoby rozpoczynające naukę tego przedmiotu – zwłaszcza, gdy ich wcześniejszy kontakt z analizą danych był niewielki. Wiemy, że największą przeszkodą nie jest rachunkowa strona statystyki, ale sposób, w jaki trzeba połączyć obliczenia z rozumieniem danych, ich interpretacją i właściwym doбором metod. Niniejszy zbiór został opracowany właśnie po to, by wspierać studentów w tym procesie poprzez praktykę, konkretne przykłady i stopniowe zwiększanie poziomu trudności.

Każdy rozdział rozpoczyna się krótkim wprowadzeniem zawierającym podstawowe definicje oraz wzory potrzebne do rozwiązywania zadań. Główną część stanowią przykłady rachunkowe obejmujące zarówno zadania rozwiązywane krok po kroku, jak i zadania do samodzielnego wykonania, opatrzone odpowiedziami. Uzupełnieniem są pytania testowe, które pomagają uporządkować i sprawdzić wiedzę z danego zakresu.

Materiał podzielony został tematycznie: od klasyfikacji i prezentacji danych, przez miary przeciętne i miary zróżnicowania, aż po miary asymetrii i koncentracji. Taki układ pozwala stopniowo rozwijać umiejętności niezbędne do opisu zbiorowości statystycznych i rozumienia ich struktury.

Mamy nadzieję, że zbiór ten okaże się dla studentów pomocny – nie tylko w przygotowaniach do zajęć, kolokwίων czy egzaminów, ale również jako

¹ D. Tarka, A.M. Olszewska, *Elementy statystyki. Opis statystyczny*, Oficyna Wydawnicza Politechniki Białostockiej, Białystok 2018.

punkt wyjścia do samodzielnego i świadomego korzystania z metod statystycznych w pracy naukowej i zawodowej.

Anna Małgorzata Olszewska
Beata Madras-Kobus

1. Prezentacja danych

W badaniach statystycznych konieczne jest precyzyjne określenie podstawowych pojęć, które stanowią fundament dalszej analizy danych. Zrozumienie takich terminów, jak populacja, próba czy jednostka statystyczna pozwala właściwie uporządkować i interpretować wyniki badań.

Populacja (zbiorowość całkowita, populacja generalna) – zbiór wszystkich obiektów podlegających badaniu statystycznemu. Jeżeli badaniu poddaje się wszystkie elementy populacji, mówimy o badaniu całościowym.

Próba (zbiorowość częściowa, podpopulacja) – część populacji poddana badaniu, gdy z różnych względów nie można (lub nie ma potrzeby) obejmować badaniem całej zbiorowości. Próby mogą być losowe lub nielosowe.

Jednostka statystyczna – pojedynczy element populacji, który podlega badaniu i posiada badaną cechę.

Cecha statystyczna – własność jednostek statystycznych, która różnicuje je między sobą.

Jednostka pomiaru – miara, w której wyrażana jest badana cecha statystyczna.

W tabeli 1.1 podano przykłady badań, określono dla nich populację, jednostkę statystyczną, cechę i jednostkę pomiaru.

Tabela 1.1. Przykłady badań statystycznych ze zdefiniowaną dla nich: populacją, jednostką statystyczną, cechą i jednostką pomiaru

Badanie statystyczne	Populacja	Jednostka statystyczna	Cecha	Jednostka pomiaru
Analizie poddano zarobki uzyskane w grudniu przez pracowników działu logistyki w przedsiębiorstwie A. Zebrano dane o miesięcznym wynagrodzeniu brutto wszystkich osób zatrudnionych w tym dziale	wszyscy pracownicy działu logistyki w tym przedsiębiorstwie, którzy pobierali wynagrodzenie w grudniu danego roku	każdy pojedynczy pracownik tego działu, który pobierał wynagrodzenie w grudniu danego roku	miesięczne wynagrodzenie brutto w grudniu	złoty (PLN)

Badanie statystyczne	Populacja	Jednostka statystyczna	Cecha	Jednostka pomiaru
Badanie statystyczne dotyczy ocen studentów z egzaminu ze statystyki na Wydziale Inżynierii Zarządzania Politechniki Białostockiej w danym roku akademickim	wszyscy studenci WIZ PB, którzy zdawali egzamin ze statystyki w danym roku akademickim	pojedynczy student WIZ PB, który pisał egzamin ze statystyki w danym roku akademickim	ocena uzyskana z egzaminu ze statystyki	skala ocen (np. 2–5)
Badanie statystyczne dotyczy ciśnienia tętniczego pacjentów przychodni X, którzy zgłosili się na wizytę w maju bieżącego roku	wszyscy pacjenci przychodni X przyjeżdżący w maju bieżącego roku	pojedynczy pacjent przychodni X przyjeżdżący w maju	wartość ciśnienia tętniczego zarejestrowana podczas wizyty	milimetr słupka rtęci (mmHg)
Badanie statystyczne dotyczy wydatków gospodarstw domowych na żywność w Białymstoku w styczniu bieżącego roku	wszystkie gospodarstwa domowe w Białymstoku w styczniu bieżącego roku	pojedyncze gospodarstwo domowe w Białymstoku	wysokość miesięcznych wydatków na żywność	złoty (PLN)
Poddano analizie wyniki (czas biegu) biegaczy uczestniczących w półmaratonie, który odbył się w Warszawie w kwietniu bieżącego roku	wszyscy zawodnicy, którzy ukończyli półmaraton w Warszawie w kwietniu bieżącego roku	pojedynczy biegacz uczestniczący w półmaratonie	czas biegu	minuta
Badanie statystyczne dotyczy plonów pszenicy zebranych w gospodarstwach rolnych w województwie podlaskim w sezonie żniwnym w 2025 roku	wszystkie gospodarstwa rolne w województwie podlaskim uprawiające pszenicę w sezonie 2025	pojedyncze gospodarstwo rolne uprawiające pszenicę	wielkość plonu pszenicy zebranej w gospodarstwie	tona
Badanie statystyczne dotyczy liczby noclegów zarezerwowanych przez turystów w hotelach w Krakowie w okresie wakacyjnym w 2025 roku	wszystkie rezerwacje noclegów w hotelach w Krakowie w lipcu i sierpniu 2025 roku	pojedyncza rezerwacja noclegu w hotelach w Krakowie	liczba noclegów w danej rezerwacji	doba hotelowa

Badanie statystyczne	Populacja	Jednostka statystyczna	Cecha	Jednostka pomiaru
Analizie poddano uczestnictwo mieszkańców miasta Y w wyborach samorządowych w kwietniu 2024 roku	wszyscy mieszkańcy miasta Y uprawnieni do głosowania w wyborach samorządowych w 2024 roku	pojedynczy wyborca miasta Y uprawniony do głosowania w wyborach samorządowych w 2024 roku	udział w wyborach (tak/nie)	odpowiedź dychotomiczna (np. 0 – nie, 1 – tak)
Analizie poddano średnią liczbę godzin w tygodniu spędzanych na zajęciach pozalekcyjnych przez uczniów klas 4–6 szkoły podstawowej w gminie X w semestrze letnim w 2025 roku	wszyscy uczniowie klas 4–6 szkół podstawowych w gminie X w semestrze letnim w 2025 roku	pojedynczy uczeń klas 4–6 szkoły podstawowej w gminie X	średnia liczba godzin spędzonych na zajęciach pozalekcyjnych w tygodniu	godzina (godz., h)
Badanie statystyczne dotyczy częstości korzystania z narzędzi opartych na sztucznej inteligencji przez pracowników firm IT w Polsce w 2025 roku	wszyscy pracownicy firm IT w Polsce w roku 2025	pojedynczy pracownik firmy IT	częstotliwość korzystania z narzędzi AI (np. codziennie, kilka razy w tygodniu, raz w miesiącu, wcale)	kategoria porządkowa (częstotliwość użycia)

Źródło: opracowanie własne.

Mając zdefiniowane pojęcia populacji, próby, jednostki statystycznej, cechy statystycznej oraz jednostki pomiaru, możemy uporządkować zebrane dane w formie szeregu statystycznego, który obrazuje, jak badana cecha rozkłada się w zbiorowości. W badaniach statystycznych najczęściej dysponujemy dużą liczbą danych, dlatego aby móc je analizować, formułować wnioski i obliczać odpowiednie miary, najwygodniej jest je zagregować w postaci szeregów rozdzielczych. Wyróżnia się trzy typy szeregów:

- proste,
- punktowe,
- przedziałowe.

Szeregi proste wykorzystuje się, gdy liczba obserwacji jest niewielka i można zaprezentować wszystkie wartości cechy w postaci nieskomplikowanego zestawienia, bez potrzeby grupowania ich w przedziały.

Szeregi rozdzielcze punktowe budujemy, gdy:

- dane są niemierzalne (jakościowe) albo mierzalne skokowe (dyskretne, przyjmujące konkretne wartości);
- liczba możliwych wartości jest niewielka lub łatwo rozróżnialna;
- poszczególne wartości nie tworzą płynnego ciągu (np. liczba aut, liczba dostaw dziennie, liczba kontenerów).

Przykłady zmiennych jakościowych, dla których można tworzyć szeregi rozdzielcze punktowe:

- cel podróży (np. wypoczynek, biznesowy, edukacyjny, pielgrzymkowy),
- forma zatrudnienia (np. umowa o pracę, działalność gospodarcza, umowa zlecenie),
- język używany w domu,
- kategoria opakowania (np. jednorazowe, wielokrotnego użytku, paletowe),
- najczęściej odwiedzany typ instytucji kultury (np. kino, teatr, muzeum, galeria),
- narodowość lub pochodzenie etniczne respondentów,
- poziom wykształcenia (np. podstawowe, średnie, wyższe),
- preferowany typ filmu (np. komedia, dramat, horror, science-fiction),
- religia/wyznanie lub brak przynależności (np. katolicyzm, islam, ateizm, protestantyzm),
- rodzaj materiału używanego w produkcji (np. stal, aluminium, plastik, drewno),
- rodzaj stosowanej diety (np. wegetariańska, wegańska, tradycyjna, bezglutenowa),
- środek transportu towarów (np. samochód ciężarowy, kolej, statek, samolot),
- wykonywany zawód (np. nauczyciel, lekarz, kierowca, informatyk),
- rodzaj zmiany produkcyjnej (np. poranna, popołudniowa, nocna),
- sposób motywowania pracowników (np. finansowy, pozafinansowy, mieszany),
- stan cywilny osób (np. kawaler/panna, żonaty/zamężna, rozwiedziony/rozwiedziona, wdowiec/wdowa),
- styl zarządzania (np. autokratyczny, demokratyczny, liberalny),
- środek transportu używany przez turystów (np. samolot, samochód, pociąg, autokar),
- typ ładunku (np. sypki, ciekły, drobnicowy, kontenerowy),
- typ maszyny produkcyjnej (np. tokarka, frezarka, prasa, robot przemysłowy),
- typ miejsca zamieszkania (miasto, wieś, przedmieścia),

- typ mieszkania (np. własnościowe, wynajmowane, komunalne),
- typ struktury organizacyjnej (np. liniowa, macierzowa, funkcjonalna),
- typ szkoły (np. publiczna, prywatna, społeczna),
- typ zakwaterowania (np. hotel, pensjonat, hostel, agroturystyka),
- ulubiony gatunek muzyczny (np. rock, pop, muzyka klasyczna, hip-hop),
- źródło dochodu (np. praca, emerytura, zasiłek, inwestycje).

Przykłady zmiennych mierzalnych, dla których można tworzyć szeregi rozdzielcze punktowe:

- liczba dzieci w rodzinie;
- liczba dostaw zrealizowanych przez kierowców w tygodniu;
- liczba goli strzelonych przez drużynę w sezonie w poszczególnych meczach;
- liczba godzin spędzanych tygodniowo przez respondenta na wolontariacie;
- liczba książek przeczytanych przez studenta w semestrze;
- liczba klientów obsłużonych przez kasjera w ciągu godziny;
- liczba osób biorących udział w wydarzeniu kulturalnym;
- liczba osób mieszkających w jednym gospodarstwie domowym;
- liczba pacjentów obsłużonych przez lekarza w ciągu dnia;
- liczba palet załadowanych na jeden samochód;
- liczba poprawnie rozwiązanych zadań przez ucznia na teście;
- liczba produktów kupionych przez jednego klienta w sklepie spożywczym;
- liczba punktów zdobytych przez zawodnika w turnieju;
- liczba reklamacji złożonych w firmie w ciągu miesiąca;
- liczba stron przeczytanych przez użytkownika e-booka w ciągu dnia;
- liczba transakcji gotówkowych przeprowadzanych dziennie w małym sklepie;
- liczba turystów odwiedzających dane miasto każdego dnia w sezonie;
- liczba tirów przyjeżdżających do magazynu dziennie;
- liczba wizyt w muzeach, galeriach czy na spektaklach w ciągu roku;
- liczba wyprodukowanych elementów przez jednego pracownika w ciągu zmiany;
- liczba wypożyczeń książek z biblioteki w tygodniu;
- liczba zajęć sportowych odbytych w miesiącu przez ucznia;
- liczba pracowników zatrudnionych w dziale firmy;
- liczba źródeł dochodu w gospodarstwie domowym.

Szeregi rozdzielcze przedziałowe budujemy, gdy:

- dane są mierzalne ciągle (mogą przyjmować dowolne wartości w pewnym zakresie);

- możliwych wartości jest bardzo dużo (lub wręcz nieskończenie wiele);
- chcemy pogrupować dane w przedziały klasowe, by je czytelnie analizować.

Przykłady zmiennych mierzalnych, dla których można tworzyć szeregi rozdzielcze przedziałowe:

- czas dojazdu do pracy (przykładowe przedziały: 0–15 min., 15–30 min., ...);
- czas dostawy (przykładowe przedziały: 0–12 godz., 12–24 godz., 24–36 godz., ...);
- czas oczekiwania klienta na usługę (przykładowe przedziały: 0–5 min., 5–10 min., 10–15 min., ...);
- czas realizacji zamówienia (przykładowe przedziały: 0–1 dzień, 1–2 dni, 2–3 dni, ...);
- długość trasy przewozowej (przykładowe przedziały: 0–100 km, 100–200 km, ...);
- dzienne zużycie energii w kWh (przykładowe przedziały: 0–50 kWh, 50–100 kWh, 100–150 kWh, 150–200 kWh, ...);
- liczba błędów produkcyjnych na milion wytworzonych sztuk (przykładowe przedziały: 0–5 ppm, 6–10 ppm, 11–15 ppm, ...);
- liczba godzin pracy kierowcy dziennie (przykładowe przedziały: 0–2 godz., 2–4 godz., 4–6 godz., ...);
- liczba godzin spędzanych online dziennie (przykładowe przedziały: 0–2 godz., 2–4 godz., 4–6 godz., ...);
- liczba klientów obsługiwanych dziennie (przykładowe przedziały: 0–10 os., 10–20 os., 20–30 os., ...);
- liczba kroków wykonywanych dziennie (przykładowe przedziały: 0–3000 kroków, 3000–6000 kroków, 6000–9000 kroków, ...);
- liczba połączeń telefonicznych wykonywanych dziennie (przykładowe przedziały: 0–5 poł., 5–10 poł., 10–15 poł., ...);
- liczba stron czytanych przez ucznia tygodniowo (przykładowe przedziały: 0–20 s., 21–40 s., 41–60 s., ...);
- liczba wyprodukowanych jednostek w ciągu zmiany (przykładowe przedziały: 0–100 szt., 100–200 szt., ...);
- liczba zgłoszeń wpływających do działu obsługi klienta tygodniowo (przykładowe przedziały: 0–50 zgł., 50–100 zgł., 100–150 zgł., ...);
- miesięczne wydatki na jedzenie (przykładowe przedziały: 0–500 zł, 500–1000 zł, ...);
- tygodniowy czas oglądania telewizji (przykładowe przedziały: 0–5 godz., 5–10 godz., 10–15 godz., ...);
- waga ładunku przewożonego w tonach (przykładowe przedziały: 0–10 t, 10–20 t, ...);

- wartość zamówienia w złotych (przykładowe przedziały: 0–500 zł, 500–1000 zł, 1000–1500 zł, ...);
- miesięczna wysokość czynszu (przykładowe przedziały: 0–1000 zł, 1001–2000 zł, 2001–3000 zł, ...).

Agregacja danych w postaci szeregów rozdzielczych stanowi istotny etap analizy statystycznej, jednak równie często dane te przedstawia się w sposób graficzny – najczęściej przy użyciu wykresów, które ułatwiają ich interpretację i porównanie. W praktyce najczęściej stosuje się wykresy:

- kolumnowe (słupkowe),
- histogramy,
- kołowe,
- liniowe.

Graficzne prezentacje danych odgrywają istotną rolę w analizie statystycznej, ponieważ:

- umożliwiają szybkie dostrzeżenie ogólnych tendencji, zależności oraz zmian zachodzących w danych;
- pozwalają na łatwiejszą identyfikację wartości odstających;
- wspierają analizę porównawczą poprzez czytelne zestawienie danych.

Wykresy kolumnowe, zwane także słupkowymi, są stosowane w przypadku danych jakościowych oraz skokowych (dyskretnych). Służą przede wszystkim do porównywania liczebności lub udziałów poszczególnych kategorii. Dzięki swojej przejrzystości umożliwiają szybką ocenę różnic między grupami lub kategoriami.

Przykłady zastosowań wykresów kolumnowych:

- liczba dostaw według przewoźników,
- liczba reklamacji według typów przesyłek,
- liczba zamówień według regionów,
- liczba klientów według przedziału wiekowego,
- liczba zatrudnionych pracowników według działów,
- liczba kursów autobusów według linii.

Histogram jest wykresem służącym do prezentowania danych mierzalnych ciągłych, pogrupowanych w przedziały klasowe. Umożliwia analizę rozkładu częstości oraz identyfikację takich cech, jak koncentracja, symetria czy obecność wartości odstających. Różni się od wykresu słupkowego tym, że słupki są ze sobą połączone, co odzwierciedla ciągłość zmiennej.

Przykłady zastosowań histogramów:

- rozkład czasów dostawy,
- rozkład masy ładunków,

- rozkład długości tras przewozowych,
- rozkład wysokości faktur sprzedażowych,
- rozkład temperatur w magazynie chłodniczym,
- rozkład zużycia paliwa w pojazdach.

Wykresy kołowe stosujemy w przypadku danych jakościowych, gdy celem jest zaprezentowanie udziałów procentowych poszczególnych kategorii w całości. Najlepiej sprawdzają się przy niewielkiej liczbie kategorii, ponieważ zbyt duża liczba elementów obniża czytelność wykresu.

Przykłady zastosowań wykresów kołowych:

- udział poszczególnych rodzajów w całości przewożonych towarów,
- struktura przychodów według klientów,
- udział form płatności w całości sprzedaży,
- struktura kosztów według kategorii (np. paliwo, serwis, kadry),
- udział kanałów sprzedaży (np. online, telefonicznie, stacjonarnie),
- podział zamówień według typów klientów (indywidualni, firmy, instytucje).

Wykresy liniowe służą do prezentacji danych w czasie, czyli w przypadku szeregów czasowych. Pokazują zmiany wartości analizowanej zmiennej w kolejnych okresach (np. dni, miesiące, lata). Pozwalają dostrzec trendy, sezonowość i momenty przełomowe w obserwowanym zjawisku.

Przykłady zastosowań wykresów liniowych:

- zmiany liczby zrealizowanych zleceń w kolejnych miesiącach,
- dynamika wzrostu przewozów w kolejnych latach,
- rozwój sprzedaży w kanałach e-commerce w czasie,
- liczba awarii pojazdów w poszczególnych kwartałach,
- liczba klientów odwiedzających stronę internetową w kolejnych tygodniach,
- zmiany kosztów logistycznych miesiąc do miesiąca.

Przykład 1.1

Zapytano studentów różnych wydziałów Politechniki Białostockiej o liczbę osób tworzących zespoły projektowe, w których brali udział w ostatnim semestrze w ramach różnych przedmiotów. Każdy student wskazał liczbę osób w zespole, w którym pracował najczęściej. Wyniki były następujące: 3, 5, 6, 8, 4, 3, 7, 5, 2, 4, 6, 9, 3, 5, 4, 5, 2, 6, 7, 2, 5, 6, 5, 5, 4, 6, 3, 5, 8, 4, 5, 6, 7, 4, 4, 5, 6, 5, 5, 6. Chcąc dokonać analizy otrzymanych w ten sposób danych, w pierwszym kroku należy określić zbiorowość statystyczną, jednostkę statystyczną, cechę oraz jednostkę pomiaru i dokonać jej klasyfikacji.

Badana populacja obejmuje studentów Politechniki Białostockiej, natomiast jednostką statystyczną jest pojedynczy student Politechniki Białostockiej

poddany badaniu statystycznemu. Cechą różnicującą zbiorowość jest liczba osób w zespole projektowym, w którym student najczęściej pracował. Cecha ta jest wyrażona w osobach, ma charakter skokowy i przyjmuje wartości na skali ilorazowej.

Następnie należy uporządkować dane w kolejności rosnącej. W ten sposób uzyskujemy następujący ciąg: 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9.

Dzięki uporządkowaniu danych można utworzyć szereg rozdzielczy cechy skokowej. Szereg ten został przedstawiony w tabeli 1.2.

Tabela 1.2. Rozkład liczby osób w zespołach projektowych, w których pracowali studenci

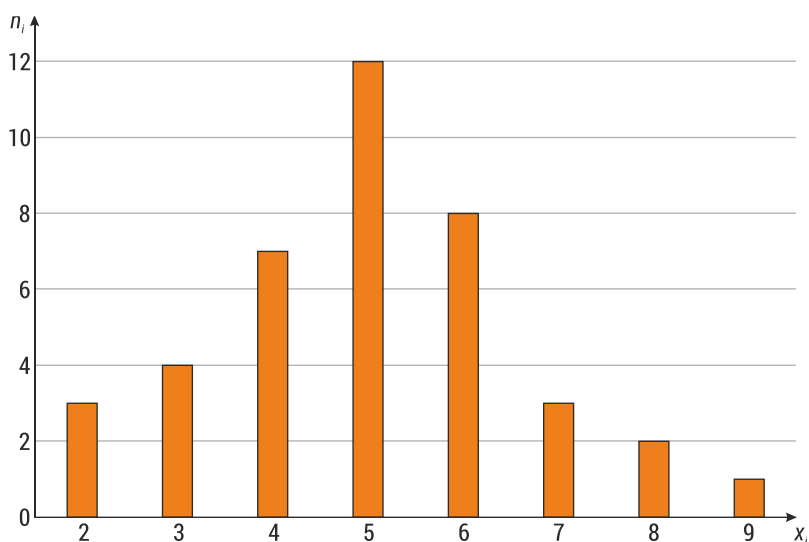
Liczba osób w zespole projektowym, w którym studenci najczęściej pracowali x_i	Liczba studentów n_i	Częstość, $f_i = \frac{n_i}{N}$
2	3	0,050
3	4	0,125
4	7	0,200
5	12	0,300
6	8	0,175
7	3	0,075
8	2	0,050
9	1	0,025
Suma	40	1,000

Źródło: dane umowne.

W analizowanym przypadku wyróżniono osiem wariantów cechy: od 2 do 9 osób w zespołach projektowych. Najwięcej studentów (30%) zadeklarowało udział w zespołach pięcioosobowych, natomiast tylko jeden student (2,5%) wskazał zespół liczący aż 9 osób.

Dla utworzonego w ten sposób szeregu można przygotować wykres prezentujący rozkład liczby osób w zespołach projektowych, w których najczęściej pracowali studenci (wykres 1.1). Na osi poziomej umieszcza się uzyskane liczby osób w zespole, natomiast na osi pionowej – liczebność studentów deklaruujących udział w zespołach o danej wielkości. Taki wykres pozwala zidentyfikować najczęściej występujące warianty cechy oraz ocenić kształt rozkładu.

Na podstawie wykresu można zauważyć, że najwięcej studentów wskazało zespoły pięcioosobowe, co oznacza, że była to najczęściej występująca wielkość zespołu. Stosunkowo dużo studentów pracowało również w zespołach cztero- i sześcioosobowych. Mniej popularne były zespoły dwu-, trzy- i siedmioosobowe, natomiast najrzadziej wskazywano zespoły ośmio- i dziewięcioosobowe. Rozkład ma charakter asymetryczny, co sugeruje, że największe zespoły zdarzają się rzadko, jak też najwięcej studentów funkcjonuje w grupach średniej wielkości.



Wykres 1.1. Wykres prezentujący rozkład liczby osób w zespołach projektowych, w których pracowali studenci

Źródło: opracowanie własne na podstawie danych umownych.

Przykład 1.2

Dane dotyczą stawki godzinowej brutto (w złotych) pracownika działu logistyki w 100 przedsiębiorstwach: 12,4; 10,1; 14,7; 11,3; 13,5; 16,8; 15,2; 18,1; 17,4; 13,0; 14,2; 11,6; 13,3; 19,0; 16,3; 20,5; 14,6; 12,3; 18,4; 16,2; 15,9; 14,1; 13,7; 17,5; 19,2; 12,6; 15,4; 11,8; 16,1; 14,8; 13,6; 12,5; 15,1; 17,8; 16,9; 13,2; 18,9; 19,4; 20,7; 17,1; 15,3; 13,4; 14,5; 11,5; 16,6; 12,8; 17,0; 14,4; 14,2; 15,5; 11,7; 16,4; 13,9; 18,3; 13,1; 12,7; 15,7; 19,3; 17,2; 14,3; 15,6; 13,0; 11,4; 15,6; 11,9; 17,7; 13,8; 14,6; 19,6; 19,6; 19,9; 14,9; 13,5; 15,8; 16,5; 14,7; 13,3; 12,2; 17,6; 15,5; 10,2; 10,8; 11,6; 14,0; 12,9; 14,5; 20,9; 21,1; 15,9; 13,7; 19,1; 16,8; 18,4; 15,7; 11,5; 10,8; 17,7; 15,0; 11,9; 12,6. Przystępując do analizy danych, należy najpierw określić zbiorowość statystyczną, jednostkę statystyczną, cechę i jednostkę pomiaru. Zbiorowość statystyczną stanowią tu wszystkie badane przedsiębiorstwa posiadające dział logistyki. Jednostkami statystycznymi są pojedyncze firmy poddane badaniu statystycznemu, zaś mierzoną cechą statystyczną – stawka godzinowa brutto pracownika działu logistycznego. Jest to cecha mierzalna (ilościowa) i ciągła. Jednostką pomiaru są złotówki.

Zanim przystąpimy do budowy szeregu, należy najpierw uszeregować wartości w kolejności od najmniejszej do największej: 10,1; 10,2; 10,8; 10,8; 11,3; 11,4; 11,5; 11,5; 11,6; 11,6; 11,7; 11,8; 11,9; 11,9; 12,2; 12,3; 12,4; 12,5; 12,6; 12,6; 12,7; 12,8; 12,9; 13,0; 13,0; 13,1; 13,2; 13,3; 13,3; 13,4; 13,5; 13,5; 13,6; 13,7; 13,7; 13,8; 13,9; 14,0; 14,1; 14,2; 14,2; 14,3; 14,4; 14,5; 14,5; 14,6; 14,6; 14,7; 14,7; 14,8; 14,9; 15,0; 15,1; 15,2; 15,3; 15,4; 15,5; 15,5; 15,6; 15,6; 15,7;

15,7; 15,8; 15,9; 15,9; 16,1; 16,2; 16,3; 16,4; 16,5; 16,6; 16,8; 16,8; 16,9; 17,0; 17,1; 17,2; 17,4; 17,5; 17,6; 17,7; 17,7; 17,8; 18,1; 18,3; 18,4; 18,4; 18,9; 19,0; 19,1; 19,2; 19,3; 19,4; 19,6; 19,6; 19,9; 20,5; 20,7; 20,9; 21,1.

W kolejnym etapie należy przystąpić do budowy szeregu rozdzielczego. Proces ten rozpoczyna się od określenia liczby klas (przedziałów). Granice przedziałów można wyznaczyć na podstawie jednego z kilku wzorów dostępnych w literaturze. W praktyce najczęściej stosuje się dwa spośród nich, przedstawione poniżej:

$$k = \sqrt{N} \quad (1.1)$$

$$k = 1 + 3,322 \cdot \log N \quad (1.2)$$

Wzór 1.2 nazywamy **wzorem Sturgesa**. Należy podkreślić, że podane wzory stanowią jedynie wskazówkę, od jakiej liczby klas można rozpocząć budowę szeregu rozdzielczego. W praktyce zaleca się, aby liczba klas nie była mniejsza niż 5 i nie przekraczała 25, co pozwala zachować równowagę między szczegółowością a czytelnością prezentowanych danych.

W związku z tym, że analizie poddano dane dotyczące stawki godzinowej brutto pracowników działu logistyki w $N = 100$ przedsiębiorstwach, zalecana liczba klas wynosi ok. 10, zgodnie ze wzorem (1.1):

$$k = \sqrt{100} = 10$$

W kolejnym kroku należy wyznaczyć szerokość (rozpiętość) przedziałów, korzystając ze wzoru:

$$h = \frac{x_{max} - x_{min}}{k} \quad (1.3)$$

W analizowanym przykładzie obliczenia mają postać:

$$h = \frac{21,1 - 10,1}{10} = 1,1 \approx 1$$

Granice przedziałów można wyznaczyć, rozpoczynając od najmniejszej wartości w zbiorze danych lub od jej przybliżenia zaokrąglonego w dół. W tym przypadku, ponieważ najmniejsza wartość wynosi 10,1, jako początek przedziału przyjmujemy wartość 10.

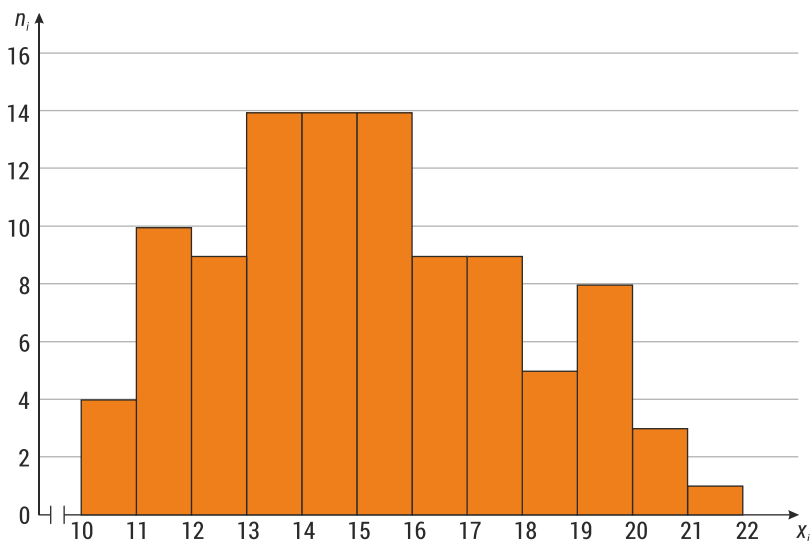
Tak skonstruowany szereg rozdzielczy przedstawiono w tabeli 1.3.

Tabela 1.3. Rozkład stawki godzinowej brutto pracowników działu logistyki ($h = 1$)

Stawka godzinowa brutto [w zł] $[x_{0i} - x_{1i})$	Liczba pracowników n_i
10–11	4
11–12	10
12–13	9
13–14	14
14–15	14
15–16	14
16–17	9
17–18	9
18–19	5
19–20	8
20–21	3
21–22	1
Suma	100

Źródło: opracowanie własne na podstawie danych umownych.

Ze względu na zastosowanie zaokrąglenia w dół przy wyznaczaniu szerokości przedziałów, w szeregu utworzono 12 klas. Histogram odpowiadający temu podziałowi przedstawiono na wykresie 1.2.



Wykres 1.2. Histogram prezentujący rozkład stawki godzinowej brutto pracowników działu logistyki w zł ($h = 1$)

Źródło: opracowanie własne na podstawie danych umownych.

Analiza wykresu wskazuje, że nie tworzy on zwartej, regularnej struktury – ma charakter nieregularny i poszarpany. Może to świadczyć o niepoprawnie dobranych parametrach grupowania. W celu oceny poprawności zastosowanego podziału na klasy przedziałowe warto przeprowadzić dodatkową weryfikację. Potrzeba takiego działania wynika z faktu, że zarówno liczba klas, jak i szerokość przedziałów są zazwyczaj ustalane na podstawie intuicji oraz doświadczenia osoby analizującej dane.

Grupowanie uznaje się za prawidłowe, jeżeli różnice między środkami poszczególnych przedziałów a średnimi arytmetycznymi wartości cechy przyporządkowanymi do tych przedziałów nie przekraczają połowy szerokości przedziału. Obliczenia dotyczące tej weryfikacji zostały przedstawione w tabeli 1.4.

Tabela 1.4. Wyliczenia związane ze sprawdzeniem poprawności grupowania danych dotyczących stawki godzinowej brutto pracowników działu logistyki ($h = 1$)

Przedziały $[x_{0i} - x_{1i})$	Pomiary należące do przedziałów	Liczebność n_i	Średnia z wartości z danego przedziału $\bar{x}_i$	Środek przed- ziału $\bar{x}_i$	$ \bar{x}_i - \bar{x}_i $
10–11	10,1; 10,2; 10,8; 10,8	4	10,48	10,5	0,03
11–12	11,3; 11,4; 11,5; 11,5; 11,6; 11,6; 11,7; 11,8; 11,9; 11,9	10	11,62	11,5	0,12
12–13	12,2; 12,3; 12,4; 12,5; 12,6; 12,6; 12,7; 12,8; 12,9	9	12,56	12,5	0,06
13–14	13,0; 13,0; 13,1; 13,2; 13,3; 13,3; 13,4; 13,5; 13,5; 13,6; 13,7; 13,7; 13,8; 13,9	14	13,43	13,5	0,07
14–15	14,0; 14,1; 14,2; 14,2; 14,3; 14,4; 14,5; 14,5; 14,6; 14,6; 14,7; 14,7; 14,8; 14,9	14	14,46	14,5	0,04
15–16	15,0; 15,1; 15,2; 15,3; 15,4; 15,5; 15,5; 15,6; 15,6; 15,7; 15,7; 15,8; 15,9; 15,9	14	15,51	15,5	0,01
16–17	16,1; 16,2; 16,3; 16,4; 16,5; 16,6; 16,8; 16,8; 16,9	9	16,51	16,5	0,01
17–18	17,0; 17,1; 17,2; 17,4; 17,5; 17,6; 17,7; 17,7; 17,8	9	17,44	17,5	0,06
18–19	18,1; 18,3; 18,4; 18,4; 18,9	5	18,42	18,5	0,08

Przedziały $[x_{0i} - x_{1i})$	Pomiary należące do przedziałów	Liczebność n_i	Średnia z wartości z danego przedziału $\bar{x}_i$	Środek prze- działu $\bar{x}_i$	$ \bar{x}_i - \bar{x} $
19-20	19,0; 19,1; 19,2; 19,3; 19,4; 19,6; 19,6; 19,9	8	19,39	19,5	0,11
20-21	20,5; 20,7; 20,9	3	20,70	20,5	0,20
21-22	21,1	1	21,10	21,5	0,40
Suma					1,18

Źródło: opracowanie własne na podstawie danych umownych.

W związku z tym, że w tabeli 1.4 suma modułów różnic między średnimi arytmetycznymi wartości cechy w poszczególnych przedziałach a środkami tych przedziałów przekracza połowę szerokości przedziału ($\frac{h}{2} = \frac{1}{2} = 0,5 < 1,18$), należy uznać, że zastosowane grupowanie jest nieadekwatne. W związku z tym konieczna jest zmiana liczby klas oraz szerokości przedziałów.

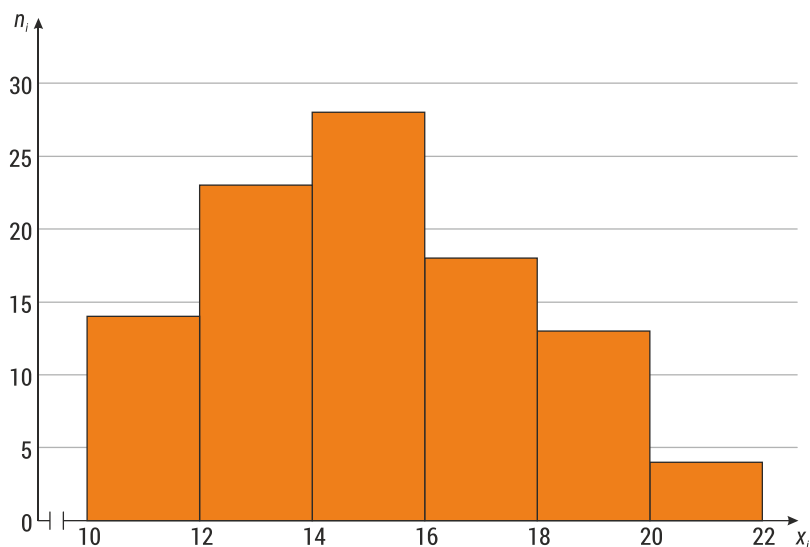
Na podstawie przeprowadzonej analizy przygotowano nowy szereg, w którym szerokość przedziałów ustalono na 2. Zestawienie danych przedstawiono w tabeli 1.5, natomiast odpowiadający mu histogram zaprezentowano na wykresie 1.3.

Tabela 1.5. Rozkład stawki godzinowej brutto pracowników działu logistyki ($h = 2$)

Stawka godzinowa brutto [w zł] $[x_{0i} - x_{1i})$	Liczba pracowników n_i
10-12	14
12-14	23
14-16	28
16-18	18
18-20	13
20-22	4
Suma	100

Źródło: opracowanie własne na podstawie danych umownych.

Wykres przedstawiony na rysunku 1.3 jest zwarty i regularny, co potwierdza prawidłowy dobór parametrów grupowania. Histogram ten ma charakter jednomodalny, co oznacza, że rozkład danych posiada jedno wyraźne maksimum. Najwięcej obserwacji odnotowano w przedziale stawki godzinowej brutto wynoszącej od 12 do 14 zł, natomiast najmniej – w przedziale od 20 do 22 zł. Rozkład cechuje się asymetrią prawostronną, co wskazuje na mniejsze zagęszczenie wyższych wartości oraz ich rozciągnięcie w prawą stronę wykresu.



Wykres 1.3. Histogram prezentujący rozkład stawki godzinowej brutto wśród pracowników działu logistyki w zł ($h = 2$)

Źródło: opracowanie własne na podstawie danych umownych.

Poprawność tego grupowania została dodatkowo zweryfikowana rachunkowo – stosowne wyliczenia zamieszczono w tabeli 1.6.

Tabela 1.6. Wyliczenia związane ze sprawdzeniem poprawności grupowania danych dotyczących stawki godzinowej brutto pracowników działu logistyki ($h = 2$)

Przedziały $[x_{0i} - x_{1i})$	Pomiary należące do przedziałów	Liczebność n_i	Średnia wartości z danego przedziału $\bar{x}_i$	Środek przedziału $\dot{x}_i$	$ \bar{x}_i - \dot{x}_i $
10–12	10,1; 10,2; 10,8; 10,8; 11,3; 11,4; 11,5; 11,5; 11,6; 11,6; 11,7; 11,8; 11,9; 11,9	14	11,29	11	0,29
12–14	12,2; 12,3; 12,4; 12,5; 12,6; 12,6; 12,7; 12,8; 12,9; 13,0; 13,0; 13,1; 13,2; 13,3; 13,3; 13,4; 13,5; 13,5; 13,6; 13,7; 13,7; 13,8; 13,9	23	13,09	13	0,09

Przedziały $[x_{0i} - x_{1i})$	Pomiary należące do przedziałów	Liczebność n_i	Średnia wartości z danego przedziału $\bar{x}_i$	Środek przedziału x_i	$ \bar{x}_i - x_i $
14–16	14,0; 14,1; 14,2; 14,2; 14,3; 14,4; 14,5; 14,5; 14,6; 14,6; 14,7; 14,7; 14,8; 14,9; 15,0; 15,1; 15,2; 15,3; 15,4; 15,5; 15,5; 15,6; 15,6; 15,7; 15,7; 15,8; 15,9; 15,9	28	14,99	15	0,01
16–18	16,1; 16,2; 16,3; 16,4; 16,5; 16,6; 16,8; 16,8; 16,9; 17,0; 17,1; 17,2; 17,4; 17,5; 17,6; 17,7; 17,7; 17,8	18	16,98	17	0,02
18–20	18,1; 18,3; 18,4; 18,4; 18,9; 19,0; 19,1; 19,2; 19,3; 19,4; 19,6; 19,6; 19,9	13	19,02	19	0,02
20–22	20,5; 20,7; 20,9; 21,1	4	20,80	21	0,20
Suma					0,63

Źródło: opracowanie własne na podstawie danych umownych.

W nowym szeregu suma modułów różnic między średnimi arytmetycznymi wartości cechy w przedziałach a środkami tych przedziałów wyniosła 0,63, co jest wartością mniejszą od połowy szerokości przedziału ($\frac{h}{2} = \frac{2}{2} = 1 > 0,63$). Oznacza to, że zastosowane grupowanie należy uznać za poprawne.

1.1. Zadania z zakresu prezentacji i klasyfikacji danych

- 1.1. Ze względu na planowane zmiany w ofercie leasingowej zarząd firmy AUTO-LEAS zdecydował o przeprowadzeniu badania statystycznego wśród 50 przedsiębiorstw transportowych w celu oceny ich zapotrzebowania na nowe pojazdy. Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj.
- 1.2. W związku z rozwojem usług zdrowotnych w regionie samorząd województwa postanowił przeprowadzić badanie statystyczne mające na celu ocenę poziomu satysfakcji pacjentów z opieki ambulatoryjnej. Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj.
- 1.3. W celu dostosowania programu lojalnościowego sieć supermarketów MAXI zleciła badanie statystyczne mające na celu poznanie preferencji zakupowych swoich klientów. Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj.
- 1.4. W związku z planowanym rozszerzeniem oferty kont oszczędnościowych bank FINANSER postanowił przeprowadzić badanie statystyczne wśród 1000 obecnych klientów, aby określić ich skłonność do regularnego oszczędzania. Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj.
- 1.5. Zarząd banku KAPITAŁ+ zlecił wykonanie badania statystycznego wśród osób ubiegających się o kredyt hipoteczny, aby określić średni poziom dochodów gospodarstw domowych starających się o finansowanie zakupu nieruchomości. Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj.
- 1.6. Departament ryzyka banku OMEGA postanowił przeprowadzić badanie statystyczne wśród firm posiadających kredyty inwestycyjne w celu określenia częstotliwości występowania opóźnień w spłacie rat (liczby spóźnionych spłat). Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj.
- 1.7. W banku DOMFIN przeprowadzono badanie wśród 25 klientów, pytając ich o liczbę posiadanych rachunków oszczędnościowych. Otrzymano następujące odpowiedzi:
1, 2, 1, 3, 2, 2, 1, 3, 2, 1, 2, 3, 2, 1, 3, 2, 2, 1, 3, 2, 1, 3, 2, 2, 1.
Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

- 1.8.** W dziale kredytów hipotecznych banku zebrano dane o liczbie kredytów hipotecznych przypadających na jednego doradcę (20 osób):

5, 6, 7, 5, 6, 7, 5, 6, 7, 5, 6, 7, 5, 6, 7, 5, 6, 7, 5, 6.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

- 1.9.** W oddziale banku zbadano miesięczną liczbę transakcji gotówkowych obsługiwanych przez 15 kasjerów:

20, 18, 22, 19, 21, 20, 18, 23, 21, 20, 19, 22, 21, 20, 19.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

- 1.10.** W firmie transportowej EUROTRANS zebrano dane o liczbie przewiezionych ładunków (w setkach ton) przez 12 kierowców w ostatnim miesiącu:

8, 7, 9, 8, 7, 9, 8, 7, 9, 8, 7, 9.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

- 1.11.** W miejskim przedsiębiorstwie komunikacyjnym sprawdzono, ile kursów w ciągu dnia wykonują poszczególne autobusy (badano 20 autobusów):
10, 12, 11, 10, 12, 11, 10, 12, 11, 10, 12, 11, 10, 12, 11, 10, 12, 11, 10, 12.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

- 1.12.** W porcie lotniczym policzono, ile samolotów przyjęto każdego dnia przez ostatnie 15 dni:

40, 42, 43, 40, 41, 42, 40, 41, 42, 40, 41, 42, 43, 40, 41.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

- 1.13.** W zakładzie produkującym elementy stalowe zbadano dzienną liczbę wyprodukowanych części na jednej linii produkcyjnej (dane z 20 dni):

100, 102, 101, 100, 102, 101, 100, 102, 101, 100, 102, 101, 100, 102,
101, 100, 102, 101, 100, 102.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.14. W fabryce mebli przez 30 dni zapisywano liczbę wyprodukowanych dziennie krzeseł:

50, 52, 51, 53, 52, 50, 52, 51, 53, 52, 50, 52, 51, 53, 52,
50, 52, 51, 53, 52, 50, 52, 51, 53, 52, 50, 52, 51, 53, 52.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.15. W zakładzie produkującym żarówki sprawdzono liczbę wadliwych sztuk w 25 losowych partiach produkcyjnych:

2, 1, 3, 2, 1, 3, 2, 1, 3, 2, 1, 3, 2, 1, 3, 2, 1, 3, 2, 1, 3, 2, 1, 3, 2.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.16. W zakładzie produkującym napoje sprawdzono liczbę wyprodukowanych dziennie skrzynek w ciągu 15 dni:

120, 130, 125, 120, 130, 125, 120, 130, 125, 120, 130, 125, 120, 130, 125.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.17. W banku KAPITAŁ PLUS zmierzono czas oczekiwania (w sekundach) 30 klientów na obsługę w okienku:

55, 63, 47, 51, 49, 60, 58, 45, 50, 53, 61, 56, 49, 52, 54,
57, 48, 50, 59, 62, 53, 51, 56, 54, 48, 50, 52, 53, 49, 60.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.18. W pewnym banku zmierzono kwoty wpłat gotówkowych (w tys. zł) dla 20 klientów:

15,2; 13,5; 14,7; 15,0; 13,8; 14,9; 15,3; 14,0; 14,8; 15,1;
13,7; 14,5; 15,0; 13,9; 14,6; 15,2; 14,4; 13,6; 14,3; 15,0.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.19. W dziale obsługi kart kredytowych zmierzono średni czas (w min.) wydania nowej karty dla 25 klientów:

20, 18, 22, 21, 19, 23, 22, 20, 21, 19, 22, 21,
20, 18, 23, 22, 20, 21, 19, 22, 21, 20, 19, 22, 21.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.20. W firmie kurierskiej zbadano czas dostawy (w godzinach) 20 przesyłek na trasie międzymiastowej:

3,5; 3,2; 3,8; 3,7; 3,4; 3,6; 3,5; 3,3; 3,7; 3,8;
3,6; 3,4; 3,5; 3,3; 3,7; 3,8; 3,6; 3,4; 3,5; 3,3.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.21. W porcie morskim zmierzono czas rozładunku kontenerowca (w godzinach) dla 15 statków:

48, 52, 50, 47, 49, 51, 53, 48, 50, 52, 49, 51, 50, 52, 53.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.22. W przedsiębiorstwie kolejowym zmierzono długość przewiezionych składów towarowych (w metrach) dla 20 pociągów:

750, 800, 780, 790, 770, 760, 800, 780, 790, 770, 750, 800, 780, 790,
770, 760, 800, 780, 790, 770.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.23. W zakładzie produkcji rur stalowych zmierzono średnicę rur (w mm) dla 25 losowo wybranych rur:

100,2; 100,5; 100,1; 100,3; 100,4; 100,2; 100,3; 100,5; 100,4;
100,2; 100,3; 100,4; 100,5; 100,2; 100,3; 100,5; 100,4; 100,3;
100,5; 100,2; 100,3; 100,5; 100,4; 100,3; 100,5.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.24. W zakładzie farmaceutycznym zbadano masę tabletek (w mg) w partii 30 sztuk:

499, 502, 501, 500, 498, 503, 501, 500, 499, 501, 502, 500,
499, 500, 501, 503, 498, 502, 501, 499, 500, 503, 498, 502,
501, 500, 499, 501, 503, 500.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.25. W fabryce sprzętu AGD zmierzono czas montażu jednego urządzenia (w min.) dla 20 stanowisk:

32, 34, 33, 35, 32, 34, 33, 35, 32, 34, 33, 35, 32, 34, 33, 35, 32, 34, 33, 35.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.26. W zakładzie produkującym kable zmierzono długość gotowych odcinków kabla (w metrach) dla 15 rolek:

100,5; 100,7; 100,6; 100,4; 100,8; 100,5; 100,7; 100,6;

100,4; 100,8; 100,5; 100,7; 100,6; 100,4; 100,8.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj. Zbuduj szereg rozdzielczy i zaprezentuj go graficznie.

1.27. W ramach analizy efektywności procesu pakowania zarejestrowano czasy pakowania 50 jednostek towaru wyrażone w sekundach. Pomiaru uzyskane dla poszczególnych jednostek przedstawiono poniżej:

38, 72, 59, 45, 83, 91, 56, 62, 78, 55, 67, 48, 81, 74, 69, 53, 85,

61, 49, 77, 80, 54, 70, 65, 60, 66, 75, 58, 64, 52, 47, 76, 82, 68,

71, 57, 63, 51, 79, 50, 84, 73, 46, 44, 43, 42, 40, 41, 39, 37.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj oraz jednostkę pomiaru. Na podstawie otrzymanych wyników zbuduj szereg rozdzielczy, a następnie dokonaj graficznej prezentacji i interpretacji otrzymanych szeregów, przyjmując:

a) przedział o długości 5,

b) przedział o długości 10,

c) przedział o długości 20,

d) przedział o długości wyznaczonej na podstawie wzoru.

1.28. W celu oceny zgodności dostaw z wymaganiami jakościowymi dokonano ważenia 70 jednostek pewnego towaru. Pomiaru masy (w kg) przedstawiono poniżej:

2,3; 3,1; 2,9; 3,7; 4,2; 3,5; 2,7; 4,0; 3,3; 3,8; 4,1; 2,8; 3,6; 4,3; 3,0;

3,9; 2,5; 3,4; 4,5; 2,6; 4,4; 2,4; 3,2; 4,6; 3,7; 2,9; 3,1; 4,0; 3,5; 2,8;

3,6; 4,2; 3,3; 4,1; 2,7; 3,9; 4,3; 2,5; 4,4; 3,0; 3,8; 2,6; 3,2; 4,5; 2,4;

3,4; 4,6; 2,9; 3,7; 3,1; 3,5; 3,8; 4,0; 2,5; 2,6; 3,3; 4,1; 2,4; 3,0; 4,2;

2,7; 3,9; 4,3; 2,8; 3,6; 2,3; 3,2; 4,4; 2,9; 4,6.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj oraz jednostkę pomiaru. Na podstawie otrzymanych wyników zbuduj szereg rozdzielczy, a następnie dokonaj graficznej prezentacji i interpretacji otrzymanych szeregów, przyjmując:

a) przedział o długości 0,2;

b) przedział o długości 0,5;

- c) przedział o długości 1,0;
- d) przedział o długości wyznaczonej na podstawie wzoru.

1.29. W ramach analizy tempa pracy uczniów zmierzono czas (w min.) potrzebny 60 osobom na rozwiązanie testu matematycznego. Pomiary przedstawiono poniżej:

5,4; 6,7; 5,9; 7,1; 8,2; 6,5; 5,1; 7,3; 6,8; 7,0; 8,0; 5,3; 6,9; 7,2; 5,7;
 6,4; 8,1; 5,5; 7,4; 6,2; 7,6; 5,2; 7,7; 6,3; 5,8; 7,5; 6,0; 8,3; 5,6; 7,8;
 6,1; 8,4; 5,0; 7,9; 6,6; 5,4; 7,0; 8,5; 6,7; 5,9; 7,1; 6,8; 5,3; 7,2; 6,5;
 8,0; 5,7; 7,4; 6,3; 8,2; 5,6; 7,3; 6,9; 7,6; 5,8; 6,4; 8,1; 5,2; 7,5; 6,6.

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj oraz jednostkę pomiaru. Na podstawie otrzymanych wyników zbuduj szereg rozdzielczy, a następnie dokonaj graficznej prezentacji i interpretacji otrzymanych szeregów, przyjmując:

- a) przedział o długości 0,5;
- b) przedział o długości 1,0;
- c) przedział o długości 2,0;
- d) przedział o długości wyznaczonej na podstawie wzoru.

1.30. Poniższa tabela przedstawia zestawienie firm transportowych na Podlasiu według rodzaju obsługiwanych tras.

Lp.	Trasy	
	krajowe	międzynarodowe
1.		
2.	-	+
3.	-	+
4.	+	+
5.	-	+
6.	-	+
7.	+	+
8.	-	-
9.	+	+
10.	+	+
11.	+	+
12.	+	+
13.	-	+
14.	-	+
15.	+	+
16.	+	-
17.	-	+
18.	-	+
19.	+	-
20.	-	-
21.	+	+

Lp.	Trasy	
22.	–	+
23.	+	–
24.	+	–
25.	+	+

Określ: zbiorowość statystyczną, jednostkę statystyczną, cechę badanej zbiorowości statystycznej i jej rodzaj, a następnie:

- pogrupuj firmy według typu obsługiwanych tras,
- zbuduj szereg rozdzielczy dla tras krajowych,
- zbuduj szereg rozdzielczy dla tras międzynarodowych,
- zbuduj tablicę przedstawiającą firmy według obu rodzajów obsługiwanych tras,
- dokonaj graficznej prezentacji otrzymanego szeregu.

1.2. Pytania testowe z zakresu klasyfikacji danych

1. Wielkość transportu morskiego (w tonach) to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna.
 - d) Żadna z powyższych.
2. Liczba statków w porcie w danym dniu to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna.
 - d) Żadna z powyższych.
3. Kolor kadłuba statku to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna (jakościowa).
 - d) Żadna z powyższych.
4. Czas realizacji zamówienia (w dniach) to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna.
 - d) Żadna z powyższych.
5. Liczba palet w magazynie to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna (jakościowa).
 - d) Żadna z powyższych.
6. Stan jakościowy dostawy (zgodna/niezgodna) to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna (jakościowa).
 - d) Żadna z powyższych.
7. Temperatura przechowywania produktów (w stopniach Celsjusza) to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,
 - c) niemierzalna (jakościowa).
 - d) Żadna z powyższych.
8. Liczba reklamacji na 1000 dostaw to cecha:
 - a) mierzalna typu ciągłego,
 - b) mierzalna typu skokowego,

- c) niemierzalna (jakościowa).
 - d) Żadna z powyższych.
9. Dane jakościowe to:
- a) wartości wyrażone liczbowo;
 - b) dane mierzalne na skali interwałowej;
 - c) dane opisowe, np. kolory, nazwy;
 - d) wyniki pomiaru w kilogramach.
10. Zmienna ilościowa to:
- a) kolor samochodu,
 - b) marka samochodu,
 - c) liczba przejechanych kilometrów,
 - d) typ samochodu.
11. Szereg rozdzielczy przedziałowy jest:
- a) wykresem liniowym zmian w czasie,
 - b) grupowaniem danych w przedziały klasowe,
 - c) analizą korelacji między dwiema zmiennymi,
 - d) sposobem liczenia średniej.
12. Liczebność w szeregu rozdzielczym oznacza:
- a) liczbę przedziałów klasowych,
 - b) szerokość klasy,
 - c) liczbę obserwacji w danym przedziale,
 - d) średnią wartość zmiennej.
13. Wzór Sturgesa służy do:
- a) wyznaczania średniej,
 - b) wyznaczania mediany,
 - c) określania liczby przedziałów klasowych,
 - d) obliczania odchylenia standardowego.
14. Histogram przedstawia:
- a) zmiany wartości w czasie,
 - b) rozkład częstości w przedziałach klasowych,
 - c) powiązania między dwiema zmiennymi,
 - d) strukturę jakościową próby.
15. Pojęcie „częstość” oznacza:
- a) liczbę przedziałów klasowych,
 - b) procentowy udział obserwacji w ogólnej liczbie danych,
 - c) różnicę między największą a najmniejszą wartością,
 - d) liczbę pomiarów powyżej średniej.
16. Dane jakościowe najlepiej prezentować w formie:
- a) histogramu,
 - b) wykresu kołowego,

- c) diagramu rozrzutu,
 - d) wykresu słupkowego dla szeregu czasowego.
17. Klasyfikację według cech ilościowych stosujemy, gdy:
- a) dane opisują kategorie,
 - b) nie znamy jednostek miary,
 - c) dane dotyczą płci,
 - d) dane są liczbowe i mierzalne.
18. Liczba klas obliczona wg wzoru Sturgesa dla liczebności $N = 50$ wynosi (zaokrąglając):
- a) 5,
 - b) 7,
 - c) 10,
 - d) 12.
19. Do przedstawienia udziału procentowego poszczególnych kategorii najlepiej nadaje się:
- a) histogram,
 - b) wykres liniowy,
 - c) wykres kołowy,
 - d) wykres rozrzutu.
20. W tabeli szeregu rozdzielczego liczebność skumulowana to:
- a) suma wszystkich wartości w szeregu,
 - b) suma wartości częstości względnych,
 - c) różnica między największą a najmniejszą liczebnością,
 - d) suma kolejnych liczebności od pierwszego przedziału do bieżącego.
21. Za pomocą szeregów czasowych można przedstawić:
- a) liczbę rozwodów w Polsce w latach 2000–2022,
 - b) rozkład wieku pracowników korporacji w 2024 roku,
 - c) preferencje zakupowe klientów dwóch przedsiębiorstw – dane uzyskane z badania ankietowego w lipcu 2024 roku,
 - d) liczbę punktów uzyskanych z kolokwium przez studentów w semestrze letnim.

1.3. Odpowiedzi do pytań testowych z zakresu klasyfikacji danych

1. a)
2. b)
3. c)
4. a)
5. b)
6. c)
7. a)
8. b)
9. c)
10. c)
11. b)
12. c)
13. c)
14. b)
15. b)
16. b)
17. d)
18. b)
19. c)
20. d)
21. a)

2. Miary przeciętne opisu zbiorowości

2.1. Przeciętne miary klasyczne

W analizie statystycznej jednym z podstawowych celów jest określenie tendencji centralnej, czyli wartości reprezentatywnej dla całego zbioru danych. W tym celu stosuje się różne miary przeciętne, spośród których najczęściej wykorzystywane są miary przeciętne klasyczne. Do miar przeciętnych klasycznych zaliczamy:

- średnią arytmetyczną,
- średnią harmoniczną,
- średnią geometryczną.

Średnia arytmetyczna ($\bar{x}$) jest jedną z najczęściej stosowanych miar tendencji centralnej. Jej użycie jest uzasadnione, gdy spełnione są następujące warunki:

1. Dane są mierzalne (liczbowe).
2. Wszystkie wartości mają jednakowe znaczenie (wagę).
3. Rozkład danych jest dość symetryczny, czyli nie występują skrajne wartości odstające.
4. Celem analizy jest określenie przeciętnego poziomu zjawiska.

Średnia arytmetyczna pozwala na syntetyczne ujęcie zbioru danych i jest punktem odniesienia w wielu analizach statystycznych. Wzory do obliczania średniej arytmetycznej to:

- dla szeregu prostego:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.1)$$

- dla szeregu rozdzielczego punktowego:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i \cdot n_i = \sum_{i=1}^k x_i \cdot f_i \quad (2.2)$$

- dla szeregu rozdzielczego przedziałowego:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k \overset{\circ}{x}_i \cdot n_i = \sum_{i=1}^k \overset{\circ}{x}_i \cdot f_i \quad (2.3)$$

Średnia harmoniczna ($\bar{x}_H$) jest stosowana, gdy wartości zmiennej podane są w jednostkach względnych, np. km/h (prędkość) czy kg/os. (np. spożycie owoców), natomiast wagi w jednostkach liczników tychże jednostek względnych, czyli odpowiednio: km i kg. Wzory do obliczania średniej harmonicznej to:

- dla szeregu prostego:

$$\bar{x}_H = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}} \quad (2.4)$$

- dla szeregu punktowego:

$$\bar{x}_H = \frac{N}{\sum_{i=1}^k \frac{n_i}{x_i}} \quad (2.5)$$

- dla szeregu przedziałowego:

$$\bar{x}_H = \frac{N}{\sum_{i=1}^k \frac{n_i}{x_i}} \quad (2.6)$$

Schemat interpretacyjny średniej arytmetycznej lub harmonicznej może przyjąć następującą postać:

Przeciętna wartość cechy [nazwa cechy]
w zbiorowości [nazwa zbiorowości] wynosi ... [wartość $\bar{x}$]
[jednostek pomiaru].

Średnia geometryczna ($\bar{x}_G$) jest stosowana w analizie zjawisk dynamicznych, w których kluczowe znaczenie mają tempo zmian i względne przyrosty lub spadki. Jej stosowanie jest uzasadnione w sytuacjach, gdy:

- analizujemy procesy zmian w czasie (dynamika);
- interesuje nas tempo wzrostu lub spadku zjawiska;
- dane mają postać wskaźników względnych, indeksów lub procentowych zmian.

Wzory do obliczania średniej geometrycznej to:

- dla szeregu prostego:

$$\bar{x}_G = \sqrt[N]{\prod_{i=1}^N x_i} \quad (2.7)$$

- dla szeregu rozdzielczego punktowego:

$$\bar{x}_G = \sqrt[N]{\prod_{i=1}^k x_i^{n_i}} \quad (2.8)$$

Schemat interpretacyjny średniej geometrycznej może przyjąć następującą postać:

W analizowanym okresie [od ... do...] wartość cechy
[nazwa cechy] z okresu na okres
(np. z roku na rok, średniorocznie lub średniookresowo)
[rosła dla $\bar{x}_G > 0$ / malała dla $\bar{x}_G < 0$] o ...%
[wartość $|(\bar{x}_G - 1) \cdot 100\%|$].

Miary przeciętne klasyczne nie powinny być stosowane:

- w przypadku danych silnie asymetrycznych, gdzie bardziej adekwatną miarą tendencji centralnej jest np. mediana;
- gdy w zbiorze danych występują wartości odstające (ekstremalne), które mogą istotnie zniekształcić wynik;
- w przypadku danych jakościowych, dla których nie stosuje się miar przeciętnych liczbowych, ponieważ nie mają one sensu interpretacyjnego.

Ponadto w szeregach przedziałowych zawierających klasy otwarte (np. „poniżej 1000 zł”, „powyżej 200 cm”) nie można wyznaczyć średnich klasycznych, ponieważ nie da się obliczyć środka takiego przedziału. Wyjątkiem są sytuacje, gdy liczebność danych w otwartej klasie jest niewielka (mniejsza niż 10% całej populacji) lub gdy dysponujemy dodatkowymi informacjami pozwalającymi ustalić najmniejszą wartość dla pierwszego otwartego przedziału bądź największą dla ostatniego. W sytuacji, gdy liczba obserwacji w otwartej klasie nie przekracza 10% liczebności całkowitej, dopuszcza się sztuczne domknięcie takiego przedziału zgodnie z logiką szeregu rozdzielczego.

Przykład 2.1 (kontynuacja przykładu 1.1)

Wróćmy do szeregu rozdzielczego punktowego dotyczącego liczby osób w zespołach, w których najczęściej pracowali analizowani studenci Politechniki Białostockiej i obliczmy średnią liczbę osób w zespołach projektowych. Wyliczenia przedstawiono w tabeli 2.1.

Tabela 2.1. Wyliczenia średniej dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności

x_i	n_i	$x_i \cdot n_i$
2	3	6
3	4	12
4	7	28
5	12	60
6	8	48
7	3	21
8	2	16
9	1	9
Suma	40	200

Źródło: wyliczenia na podstawie danych umownych.

Korzystając ze wzoru na średnią arytmetyczną dla szeregu rozdzielczego cechy skokowej, otrzymujemy:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^8 x_i \cdot n_i = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_i \cdot n_i + \dots + x_8 \cdot n_8}{N} = \frac{200}{40} = 5 \text{ os.}$$

Te same wyniki można uzyskać, wykorzystując częstość zamiast liczebności. Wyliczenia podano w tabeli 2.2.

Tabela 2.2. Wyliczenia średniej dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości

x_i	f_i	$x_i \cdot f_i$
2	0,050	0,150
3	0,125	0,300
4	0,200	0,700
5	0,300	1,500
6	0,175	1,200
7	0,075	0,525
8	0,050	0,400
9	0,025	0,225
Suma	1,000	5,000

Źródło: wyliczenia na podstawie danych umownych.

Korzystając ze wzoru na średnią arytmetyczną dla szeregu rozdzielczego cechy skokowej, otrzymujemy:

$$\bar{x} = \sum_{i=1}^8 x_i \cdot f_i = x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_i \cdot f_i + \dots + x_8 \cdot f_8 = 5 \text{ os.}$$

Analizowani studenci Politechniki Białostockiej pracowali w zespołach składających się średnio z 5 osób.

Przykład 2.2 (kontynuacja przykładu 1.2)

Wróćmy do szeregu rozdzielczego przedziałowego dotyczącego stawki godzinowej brutto pracowników działu logistyki i obliczmy średnią liczbę osób w zespołach projektowych. Wyliczenia przedstawiono w tabeli 2.3.

Tabela 2.3. Wyliczenia średniej dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności

$x_{0i} - x_{1i}$	n_i	$\overset{\circ}{x}_i$	$\overset{\circ}{x}_i \cdot n_i$
10–12	14	11	154
12–14	23	13	299
14–16	28	15	420
16–18	18	17	306
18–20	13	19	247
20–22	4	21	84
Suma	100		1510

Źródło: wyliczenia na podstawie danych umownych.

Korzystając ze wzoru na średnią arytmetyczną dla szeregu rozdzielczego cechy ciągłej, otrzymujemy:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^6 \overset{\circ}{x}_i \cdot n_i = \frac{1510}{100} = 15,1 \text{ zł}$$

Te same wyniki można uzyskać, wykorzystując częstość zamiast liczebności – wyliczenia przedstawiono w tabeli 2.4.

Tabela 2.4. Wyliczenia średniej dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości

$x_{0i} - x_{1i}$	n_i	$\overset{\circ}{x}_i$	f_i	$\overset{\circ}{x}_i \cdot f_i$
10–12	14	11	0,14	1,54
12–14	23	13	0,23	2,99
14–16	28	15	0,28	4,20
16–18	18	17	0,18	3,06
18–20	13	19	0,13	2,47
20–22	4	21	0,04	0,84
Suma	100			15,10

Źródło: wyliczenia na podstawie danych umownych.

Korzystając ze wzoru na średnią arytmetyczną dla szeregu rozdzielczego przedziałowego cechy ciągłej, otrzymujemy:

$$\bar{x} = \sum_{i=1}^6 \overset{\circ}{x}_i \cdot f_i = 15,10 \text{ zł}$$

Średnia stawka godzinowa brutto analizowanych pracowników działu logistyki wynosi 15,10 zł.

Przykład 2.3

Firma transportowa analizuje średnią prędkość swoich ciężarówek na trasie Warszawa–Gdańsk w różnych warunkach drogowych:

1. Na pierwszym odcinku trasy (240 km) średnia prędkość wynosiła 60 km/h (korki i remonty).
2. Na drugim odcinku (150 km) średnia prędkość wynosiła 100 km/h (autostrada).

Oblicz średnią prędkość ciężarówek na całej trasie.

W tym zadaniu analizowana zmienna, czyli prędkość, wyrażona jest w jednostkach względnych: km/h. Wagi, czyli przejechane odcinki trasy, wyrażono natomiast w jednostkach z liczników tychże jednostek względnych, czyli km. W takiej sytuacji do obliczenia średniej prędkości należy użyć średniej harmonicznnej. Wyliczenia przedstawiono w tabeli 2.5.

Tabela 2.5. Wyliczenia średniej harmonicznnej prędkości ciężarówek

x_i [w km/h]	n_i [w km]	$\frac{n_i}{x_i}$ [w h]
60	240	4
100	150	1,5
Suma	390	5,5

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając otrzymane wartości do wzoru na średnią harmoniczną, otrzymujemy:

$$\bar{x}_H = \frac{N}{\sum_{i=1}^2 \frac{n_i}{x_i}} = \frac{390}{5,5} = 70,91 \text{ km/h}$$

Średnia prędkość ciężarówek na trasie Warszawa–Gdańsk w opisanych warunkach wynosi ok. 71 km/h.

Przykład 2.4

Firma transportowa analizuje średnią prędkość swoich ciężarówek na trasie Warszawa–Gdańsk w różnych warunkach drogowych:

1. Na pierwszym odcinku trasy kierowcy jechali z prędkością 60 km/h przez 4 godziny (korki i remonty).
2. Na drugim odcinku prędkość wynosiła 100 km/h przez 1,5 godziny (autostrada).

Oblicz średnią prędkość ciężarówek na całej trasie.

W przedstawionej sytuacji należy obliczyć średnią prędkość jako średnią arytmetyczną ważoną, ponieważ znane są czasy przejazdu na poszczególnych odcinkach oraz odpowiadające im średnie prędkości. W takim przypadku wagami w obliczeniach są czasy trwania przejazdu. Wyliczenia przedstawiono w tabeli 2.6

Tabela 2.6. Wyliczenia średniej arytmetycznej prędkości ciężarówek

x_i [w km/h]	n_i [w h]	$n_i \cdot x_i$ [w km]
60	4	240
100	1,5	150
Suma	5,5	390

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając otrzymane wartości do wzoru (2.2), otrzymujemy:

$$\bar{x} = \frac{\sum_{i=1}^2 x_i \cdot n_i}{N} = \frac{390}{5,5} = 70,91 \text{ km/h}$$

Średnia prędkość ciężarówek na trasie Warszawa–Gdańsk w opisanych warunkach wynosi ok. 71 km/h.

Przykład 2.5

W ciągu kolejnych lat firma logistyczna zwiększała liczbę obsługiwanych zleceń transportowych:

- w pierwszym roku o 10%,
- w drugim roku o 20%,
- w trzecim roku o 5%.

Oblicz średnie roczne tempo zmian (średnie tempo wzrostu) w analizowanych latach.

Do analizy średniego tempa zmian badanego zjawiska w czasie wykorzystujemy średnią geometryczną. W celu określenia jej wartości najpierw należy obliczyć wskaźniki zmian (tabela 2.7).

Tabela 2.7. Wyliczenia średniej geometrycznej dla rozkładu liczby obsłużonych zleceń transportowych

t	Wzrost [w %]	Wskaźnik zmian [w %]	Wskaźnik zmian x_i
1	10	110	1,10
2	20	120	1,20
3	5	105	1,05

Źródło: wyliczenia na podstawie danych umownych.

Następnie liczymy średnią geometryczną ze wzoru na taką średnią:

$$\bar{x}_G = \sqrt[3]{\prod_{i=1}^3 x_i} = \sqrt[3]{1,1 \cdot 1,2 \cdot 1,05} = \sqrt[3]{1,254} = 1,078 = 107,8\% \approx 108\%$$

W analizowanym okresie liczba obsłużonych zleceń transportowych rosła z roku na rok średnio o 7,8%.

2.2. Przeciętne miary pozycyjne

W analizie statystycznej oprócz miar przeciętnych klasycznych, takich jak średnie, stosuje się również miary przeciętne pozycyjne. Ich wartość zależy od położenia (pozycji) danej obserwacji w uporządkowanym zbiorze, a nie od wszystkich wartości liczbowych. Są one szczególnie przydatne w przypadku danych asymetrycznych lub zawierających wartości odstające. Do miar przeciętnych pozycyjnych zaliczamy:

- dominantę (modę),
- kwartyle.

Miary przeciętne pozycyjne stosujemy, gdy:

1. Dane są asymetryczne lub zawierają wartości odstające – miary pozycyjne, w przeciwieństwie do średnich klasycznych, są odporne na wpływ ekstremalnych obserwacji.
2. Dane mają charakter mierzalny lub porządkowy – medianę i kwartyle można obliczać zarówno dla danych liczbowych, jak i dla danych porządkowych, takich jak przykładowo oceny jakości dostaw wyrażone w skali od 1 (bardzo niska) do 5 (bardzo wysoka).
3. Analiza dotyczy struktury lub rozkładu danych w zbiorowości – interesuje nas to, jaki odsetek obserwacji znajduje się poniżej lub powyżej określonego poziomu, przykładowo, jaki procent dostawców realizuje dostawy w czasie krótszym niż 24 godziny.

Dominanta (moda, modalna) to wartość, która występuje w zbiorze danych z największą częstością. W szeregu punktowym jest równa tej wartości zmiennej, której przypisana jest największa liczba (częstość) wystąpień. Jeżeli kilka różnych wartości występuje z jednakową, najwyższą liczebnością (częstością), wówczas zbiór danych może mieć więcej niż jedną dominantę – mówimy wtedy o rozkładzie wielomodalnym.

W celu wyznaczenia dominanty (mody) z szeregu rozdzielczego przedziałowego należy w pierwszej kolejności zidentyfikować przedział modalny, czyli przedział o największej liczebności (częstości). W tym właśnie przedziale znajduje się dominanta. Jej wartość oblicza się następnie za pomocą odpowiedniego wzoru:

- dla szeregu rozdzielczego przedziałowego z podanymi liczebnościami:

$$D = x_{0D} + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D \quad (2.9)$$

- dla szeregu rozdzielczego przedziałowego z częstościami:

$$D = x_{0D} + \frac{f_D - f_{D-1}}{(f_D - f_{D-1}) + (f_D - f_{D+1})} \cdot h_D \quad (2.10)$$

gdzie

x_{0D} – początek przedziału, w którym znajduje się dominanta;

n_D – liczebność przedziału, w którym znajduje się dominanta;

n_{D-1} – liczebność przedziału poprzedzającego przedział, w którym znajduje się dominanta;

n_{D-1} – liczebność przedziału następnego po przedziale, w którym znajduje się dominanta;

h_D – rozpiętość przedziału dominanty;

f_D – częstość przedziału, w którym znajduje się dominanta;

f_{D-1} – częstość przedziału poprzedzającego przedział, w którym znajduje się dominanta;

f_{D-1} – częstość przedziału następnego po przedziale, w którym znajduje się dominanta.

Schemat interpretacyjny dominanty może przyjąć następującą postać:

Najczęściej występującą wartością cechy [nazwa cechy]
w zbiorowości [nazwa zbiorowości] jest ... [wartość D]
[jednostek pomiaru].

Kwartyle to miary pozycyjne, które dzielą uporządkowany zbiór danych na cztery równe części, wyznaczając wartości graniczne odpowiadające określonym proporcjom obserwacji. W celu wyznaczenia ich w szeregu punktowym należy najpierw uporządkować dane rosnąco, a następnie określić następujące pozycje:

- **kwartyl pierwszy** (Q_1), nazywany również dolnym, czyli wartość, poniżej której znajduje się 25% obserwacji;
- **kwartyl drugi** (Q_2), czyli mediana (wartość środkowa), która dzieli zbiór na połowę – 50% danych znajduje się poniżej tej wartości;
- **kwartyl trzeci** (Q_3), zwany także górnym, czyli wartość, poniżej której znajduje się 75% obserwacji.

W sytuacji, gdy pozycja kwartylowa przypada pomiędzy dwiema obserwacjami, wartość kwartyłu wyznacza się za pomocą interpolacji liniowej, przy uwzględnieniu odpowiednich wartości sąsiednich.

W przypadku zmiennych ciągłych, których wartości zostały pogrupowane w szereg rozdzielczy przedziałowy, kwartyle wyznacza się szacunkowo, z wykorzystaniem metod interpolacyjnych. I tak **kwartyl pierwszy** (Q_1) znajduje się w przedziale, w którym liczebność skumulowana jest równa lub po raz

pierwszy większa od $\frac{N}{4}$. Następnie wartość kwartyła wyznaczana jest ze wzorów:

- dla szeregu rozdzielczego przedziałowego z podanymi liczebnościami:

$$Q_1 = x_{0Q_1} + \frac{h_{Q_1}}{n_{Q_1}} \cdot \left(\frac{N}{4} - \sum_{i=1}^{k_1-1} n_i \right) \quad (2.11)$$

- dla szeregu rozdzielczego przedziałowego z częstościami:

$$Q_1 = x_{0Q_1} + \frac{h_{Q_1}}{f_{Q_1}} \cdot (0,25 - \sum_{i=1}^{k_1-1} f_i) \quad (2.12)$$

gdzie:

- x_{0Q_1} – początek przedziału, w którym znajduje się kwartył pierwszy;
- h_{Q_1} – rozpiętość przedziału, w którym znajduje się kwartył pierwszy;
- n_{Q_1} – liczebność przedziału, w którym znajduje się kwartył pierwszy;
- $\sum_{i=1}^{k_1-1} n_i$ – suma liczebności z przedziałów poprzedzających przedział, w którym znajduje się kwartył pierwszy (lub liczebność skumulowana z przedziału poprzedzającego przedział, w którym znajduje się kwartył pierwszy);
- f_{Q_1} – częstość przedziału, w którym znajduje się kwartył pierwszy;
- $\sum_{i=1}^{k_1-1} f_i$ – suma częstości z przedziałów poprzedzających przedział, w którym znajduje się kwartył pierwszy (lub częstość skumulowana z przedziału poprzedzającego przedział, w którym znajduje się kwartył pierwszy).

Kwartył drugi (Q_2) znajduje się w przedziale, w którym liczebność skumulowana jest równa lub po raz pierwszy większa od $\frac{N}{2}$. Wówczas należy zastosować jeden z poniższych wzorów:

- dla szeregu rozdzielczego przedziałowego z podanymi liczebnościami:

$$Q_2 = Me = x_{0Q_2} + \frac{h_{Q_2}}{n_{Q_2}} \cdot \left(\frac{N}{2} - \sum_{i=1}^{k_2-1} n_i \right) \quad (2.13)$$

- dla szeregu rozdzielczego przedziałowego z częstościami:

$$Q_2 = Me = x_{0Q_2} + \frac{h_{Q_2}}{f_{Q_2}} \cdot (0,5 - \sum_{i=1}^{k_2-1} f_i) \quad (2.14)$$

gdzie:

- x_{0Q_2} – początek przedziału, w którym znajduje się kwartył drugi;
- h_{Q_2} – rozpiętość przedziału, w którym znajduje się kwartył drugi;
- n_{Q_2} – liczebność przedziału, w którym znajduje się kwartył drugi;

$\sum_{i=1}^{k_2-1} n_i$ – suma liczebności z przedziałów poprzedzających przedział, w którym znajduje się kwartył drugi (lub liczebność skumulowana z przedziału poprzedzającego przedział, w którym znajduje się kwartył drugi);

f_{Q_2} – częstość przedziału, w którym znajduje się kwartył drugi;

$\sum_{i=1}^{k_2-1} f_i$ – suma częstości z przedziałów poprzedzających przedział, w którym znajduje się kwartył drugi (lub częstość skumulowana z przedziału poprzedzającego przedział, w którym znajduje się kwartył drugi).

Kwartył trzeci (Q_3) znajduje się w przedziale, w którym liczebność skumulowana jest równa lub po raz pierwszy większa od $\frac{3N}{4}$. Wówczas stosujemy jeden z poniższych wzorów:

- dla szeregu rozdzielczego przedziałowego z podanymi liczebnościami:

$$Q_3 = x_{0Q_3} + \frac{h_{Q_3}}{n_{Q_3}} \cdot \left(\frac{3N}{4} - \sum_{i=1}^{k_3-1} n_i \right) \quad (2.15)$$

- dla szeregu rozdzielczego przedziałowego z częstościami:

$$Q_3 = x_{0Q_3} + \frac{h_{Q_3}}{f_{Q_3}} \cdot (0,75 - \sum_{i=1}^{k_3-1} f_i) \quad (2.16)$$

gdzie:

x_{0Q_3} – początek przedziału, w którym znajduje się kwartył trzeci;

h_{Q_3} – rozpiętość przedziału, w którym znajduje się kwartył trzeci;

n_{Q_3} – liczebność przedziału, w którym znajduje się kwartył trzeci;

$\sum_{i=1}^{k_3-1} n_i$ – suma liczebności z przedziałów poprzedzających przedział, w którym znajduje się kwartył trzeci (lub liczebność skumulowana z przedziału poprzedzającego przedział, w którym znajduje się kwartył trzeci);

f_{Q_3} – częstość przedziału, w którym znajduje się kwartył trzeci;

$\sum_{i=1}^{k_3-1} f_i$ – suma częstości z przedziałów poprzedzających przedział, w którym znajduje się kwartył trzeci (lub częstość skumulowana z przedziału poprzedzającego przedział, w którym znajduje się kwartył trzeci).

Schemat interpretacyjny kwartyli może przyjąć postać:

Przynajmniej [25% / 50% / 75%] zbiorowości [nazwa zbiorowości] ma wartość cechy [nazwa cechy] mniejszą lub równą ... [wartość Q_1 / Me / Q_3] [jednostek pomiaru] oraz przynajmniej [75% / 50% / 25%] zbiorowości [nazwa zbiorowości] ma wartość cechy [nazwa cechy] większą lub równą ... [wartość Q_1 / Me / Q_3] [jednostek pomiaru].

Przykład 2.6 (kontynuacja przykładu 1.1)

Wracając do szeregu rozdzielczego dotyczącego liczby osób w zespołach, w których najczęściej pracowali analizowani studenci Politechniki Białostockiej, wyznaczmy pozycyjne miary tendencji centralnej.

Tabela 2.7. Wyliczenia liczebności i częstości skumulowanej dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci

x_i	n_i	$n_{i\ sk}$	f_i	$f_{i\ sk}$	
2	3	3	0,050	0,050	
3	4	7	0,125	0,175	
4	7	14	0,200	0,375	$\Leftarrow Q_1$
5	12	26	0,300	0,675	$\Leftarrow D, Me$
6	8	34	0,175	0,850	$\Leftarrow Q_3$
7	3	37	0,075	0,925	
8	2	39	0,050	0,975	
9	1	40	0,025	1,000	
Suma	40		1,000		

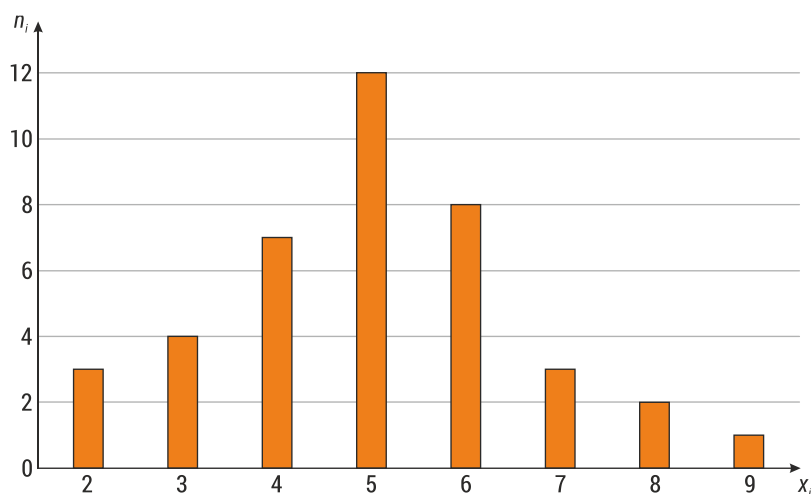
Źródło: wyliczenia na podstawie danych umownych.

Analizując dane zawarte w tabeli 2.8, można zauważyć, że najczęściej – bo aż 12 razy – studenci wskazywali, że brali udział w realizacji projektów w zespołach pięcioosobowych. Jest to łatwo dostrzegalne również na wykresie przedstawiającym rozkład liczby osób w zespołach (wykres 2.1). Kategoria ta stanowi dominantę rozkładu, co oznacza, że wartością cechy najczęściej występującą wśród badanych jest 5 ($D = 5$).

Analizowany rozkład charakteryzuje się jednym wyraźnym szczytem, dlatego nazywany jest rozkładem jednomodalnym. Wyłącznie w przypadku takich rozkładów można wyznaczyć dominantę.

Przejdźmy teraz do wyznaczenia kwartyli. Łączna liczba studentów wynosi 40, co oznacza, że do wyznaczenia mediany – czyli wartości dzielącej

uporządkowany zbiór danych na połowy – należy określić środkowe pozycje: 20 i 21 jednostkę. Wartości skumulowane pokazują, że do klasy o wartości cechy równej 4 włącznie przypada 14 studentów, natomiast po dodaniu kolejnej klasy ($x = 5$) skumulowana liczebność osiąga 26. Oznacza to, że zarówno 20, jak i 21 student znajdują się w klasie odpowiadającej wartości 5. Zatem mediana wynosi 5 ($Me = \frac{x_{20} + x_{21}}{2} = \frac{5 + 5}{2} = 5$), co oznacza, że przynajmniej połowa badanych studentów pracowała najczęściej w zespołach nie większych niż pięcioosobowe i przynajmniej połowa w zespołach nie mniejszych niż pięcioosobowe.



Wykres 2.1. Wykres prezentujący rozkład liczby osób w zespołach projektowych, w których pracowali studenci – wyznaczenie dominanty

Źródło: opracowanie własne na podstawie danych umownych.

Podobnie wyznacza się pozostałe kwartyle – kwartył pierwszy (dolny) i trzeci (górny). Pierwszy kwartył (Q_1) to wartość odpowiadająca 25 percentylowi, czyli 10 studentowi w szeregu uporządkowanym ($\frac{n}{4} = \frac{40}{4} = 10$). Ponieważ do klasy $x = 3$ włącznie przypada 7 osób, a po dodaniu klasy $x = 4$ liczebność rośnie do 14, to 10 student należy do klasy o wartości 4. Stąd $Q_1 = 4$. Zatem przynajmniej 25% studentów pracowało, przygotowując projekty na zajęciach, najczęściej w grupach co najwyżej czteroosobowych i przynajmniej 75% studentów najczęściej w grupach przynajmniej czteroosobowych.

Trzeci kwartył (Q_3), odpowiadający 75 percentylowi, przypada na 30 osobę ($\frac{3n}{4} = \frac{3 \cdot 40}{4} = 30$). Do klasy $x = 5$ włącznie mamy 26 osób, a po uwzględnieniu klasy $x = 6$ już 34, w związku z czym 30 student znajduje się w klasie

o wartości 6, a więc $Q_3 = 6$. Zatem przynajmniej 75% studentów realizowało na zajęciach projekty najczęściej w grupach co najwyżej sześćoosobowych i przynajmniej 25% studentów w grupach nie mniejszych niż sześćoosobowe.

Przykład 2.7 (kontynuacja przykładu 1.2)

Wracając do szeregu rozdzielczego przedziałowego dotyczącego stawek godzinowych brutto pracowników działu logistyki, wyznaczmy teraz miary średnie pozycyjne. Rozpocznijmy od dominanty. W pierwszym kroku należy sprawdzić, czy szereg spełnia warunki umożliwiające jej wyznaczenie, tj.: równa szerokość przedziałów, jednomodalność rozkładu oraz brak dominanty w skrajnym przedziale.

W analizowanym przypadku wszystkie warunki są spełnione: przedziały mają jednakową szerokość, rozkład jest jednomodalny, a przedział dominanty nie znajduje się na skraju rozkładu. W oparciu o dane zawarte w tabeli 2.9 ustalono, że dominanta występuje w trzecim przedziale, obejmującym stawki od 14 do 16 zł. Można zatem zastosować wzór interpolacyjny (2.9) do wyznaczania dominanty:

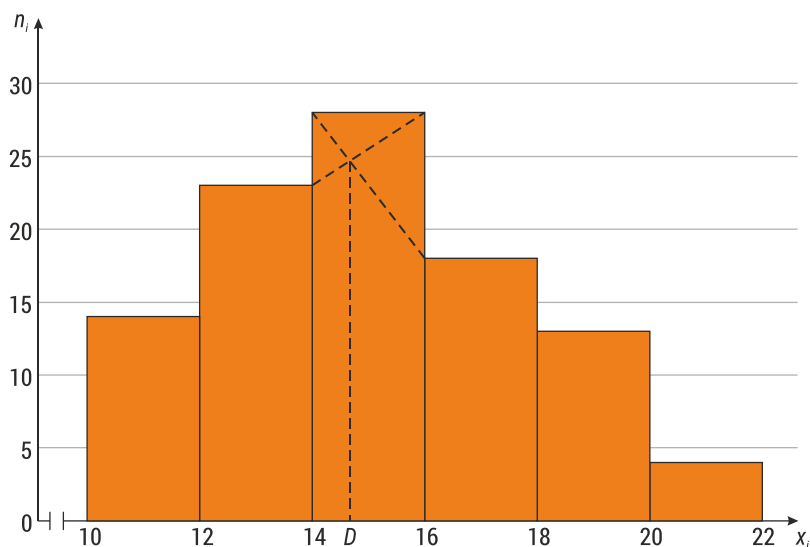
$$\begin{aligned} D &= x_{0D} + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D = \\ &= 14 + \frac{28 - 23}{(28 - 23) + (28 - 18)} \cdot 2 = \\ &= 14 + \frac{5}{15} \cdot 2 = 14 + 0,67 = 14,67 \text{ zł} \end{aligned}$$

Podany wynik można również uzyskać, korzystając ze wzoru (2.10) wykorzystującego częstość:

$$\begin{aligned} D &= x_{0D} + \frac{f_D - f_{D-1}}{(f_D - f_{D-1}) + (f_D - f_{D+1})} \cdot h_D = \\ &= 14 + \frac{0,28 - 0,23}{(0,28 - 0,23) + (0,28 - 0,18)} \cdot 2 = \\ &= 14 + \frac{0,05}{0,15} \cdot 2 = 14 + 0,67 = 14,67 \text{ zł} \end{aligned}$$

Uzyskany wynik oznacza, że pracownicy działu logistyki najczęściej mają godzinową stawkę wynoszącą ok. 14,67 zł.

Dominantę można również wyznaczyć graficznie, korzystając z histogramu. Graficzne wyznaczenie dominanty dla analizowanych danych dotyczących godzinowej stawki pracowników działu logistyki przedstawiono na wykresie 2.2.



Wykres 2.2. Graficzne wyznaczenie dominanty dla rozkładu stawki godzinowej brutto pracowników działu logistyki

Źródło: opracowanie własne na podstawie danych umownych.

Przystąpmy teraz do wyznaczenia kwartyli. W celu ich obliczenia niezbędne jest wykorzystanie liczebności (lub częstości) skumulowanej. Odpowiednie wyliczenia przedstawiono w tabeli 2.9.

Tabela 2.8. Wyliczenia liczebności i częstości skumulowanej dla rozkładu stawki godzinowej brutto pracowników działu logistyki

$x_{0i} - x_{1i}$	n_i	n_{isk}	f_i	f_{isk}	
10-12	14	14	0,14	0,14	
12-14	23	37	0,23	0,37	$\Leftarrow Q_1$
14-16	28	65	0,28	0,65	$\Leftarrow D, Me$
16-18	18	83	0,18	0,83	$\Leftarrow Q_3$
18-20	13	96	0,13	0,96	
20-22	4	100	0,04	1,00	
Suma	100		1,00		

Źródło: wyliczenia na podstawie danych umownych.

Przed wyznaczeniem wartości kwartyli należy określić ich położenie w szeregu. W analizowanym przypadku liczebność zbiorowości wynosi 100, a dane dotyczą cechy ciągłej. Ze względu na brak dokładnych wartości wewnątrz przedziałów nie stosuje się uśredniania dwóch środkowych jednostek, jak ma to

miejsce w przypadku cechy skokowej. Pozycje kwartyli wyznacza się zatem według zależności:

$$N_{Q_1} = \frac{N}{4} = \frac{100}{4} = 25, \quad \text{czyli} \quad Q_1 = X_{25}$$

$$N_{Q_2} = \frac{N}{2} = \frac{100}{2} = 50, \quad \text{czyli} \quad Q_2 = X_{50}$$

$$N_{Q_3} = \frac{3N}{4} = \frac{3 \cdot 100}{4} = 75, \text{czyli} \quad Q_3 = X_{75}$$

Na podstawie danych zawartych w tabeli 2.9 można stwierdzić, że wartość pierwszego kwartyla mieści się w drugim przedziale. W pierwszym przedziale znajduje się łącznie 14 jednostek, natomiast drugi przedział obejmuje jednostki od 15 do 37 włącznie. Ponieważ 25. pozycja przypada właśnie na ten przedział, kwartył pierwszy zostanie wyznaczony w jego obrębie.

Po ustaleniu położenia kwartyła kolejnym krokiem jest jego obliczenie i interpretacja. W przypadku pierwszego kwartyła zastosowano wzór (2.11):

$$\begin{aligned} Q_1 &= x_{0Q_1} + \frac{h_{Q_1}}{n_{Q_1}} \cdot \left(\frac{N}{4} - \sum_{i=1}^{k_1-1} n_i \right) = 12 + \frac{2}{23} \cdot (25 - 14) = \\ &= 12 + \frac{2}{23} \cdot 11 = 12 + 0,96 = 12,96 \text{ zł} \end{aligned}$$

lub wzór (2.12):

$$\begin{aligned} Q_1 &= x_{0Q_1} + \frac{h_{Q_1}}{f_{Q_1}} \cdot \left(0,25 - \sum_{i=1}^{k_1-1} f_i \right) = 12 + \frac{2}{0,23} \cdot (0,25 - 0,14) = \\ &= 12 + \frac{2}{0,23} \cdot 0,11 = 12 + 0,96 = 12,96 \text{ zł} \end{aligned}$$

Uzyskany wynik oznacza, że co najmniej 25% badanych pracowników działu logistyki otrzymuje stawkę godzinową nieprzekraczającą 12,96 zł, a jednocześnie przynajmniej 75% – stawkę wynoszącą co najmniej 12,96 zł.

Analogicznie wyznaczono kwartył drugi, czyli medianę. Przy liczebności zbiorowości wynoszącej 100 jednostek, mediana przypada na 50 pozycję. Z danych przedstawionych w tabeli 2.9 wynika, że znajduje się ona w trzecim przedziale, obejmującym jednostki od 38 do 65 włącznie. Wartość mediany oblicza się na podstawie wzoru (2.13):

$$Q_2 = Me = x_{0Q_2} + \frac{h_{Q_2}}{n_{Q_2}} \cdot \left(\frac{N}{2} - \sum_{i=1}^{k_2-1} n_i \right) =$$

$$= 14 + \frac{2}{28} \cdot (50 - 37) = 14 + \frac{2}{28} \cdot 13 = 14 + 0,93 = 14,93 \text{ zł}$$

lub wzoru (2.14):

$$\begin{aligned} Q_2 = Me &= x_{0Q_2} + \frac{h_{Q_2}}{f_{Q_2}} \cdot \left(0,5 - \sum_{i=1}^{k_2-1} f_i \right) = 14 + \frac{2}{0,28} \cdot (0,5 - 0,37) = \\ &= 14 + \frac{2}{0,28} \cdot 0,13 = 14 + 0,93 = 14,93 \text{ zł} \end{aligned}$$

Uzyskany wynik oznacza, że co najmniej 50% badanych pracowników działu logistyki otrzymuje stawkę godzinową nie wyższą niż 14,93 zł, natomiast przynajmniej 50% – nie niższą niż 14,93 zł.

Pozycja trzeciego kwartyła odpowiada 75 jednostce, która – zgodnie z danymi zawartymi w tabeli 2.9 – mieści się w czwartym przedziale. Przedział ten obejmuje obserwacje od 66 do 83 jednostki włącznie. W celu wyznaczenia wartości trzeciego kwartyła zastosowano wzór (2.15):

$$\begin{aligned} Q_3 &= x_{0Q_3} + \frac{h_{Q_3}}{n_{Q_3}} \cdot \left(\frac{3N}{4} - \sum_{i=1}^{k_3-1} n_i \right) = 16 + \frac{2}{18} \cdot (75 - 65) = \\ &= 16 + \frac{2}{19} \cdot 10 = 16 + 1,05 = 17,05 \text{ zł} \end{aligned}$$

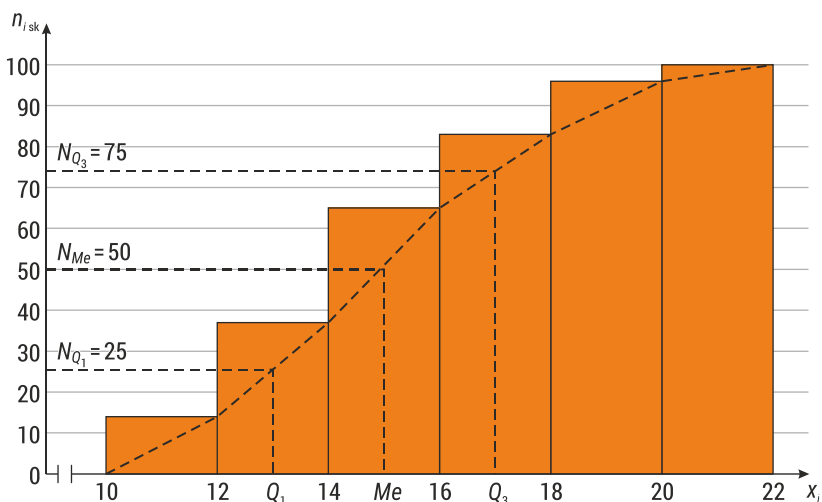
lub wzór (2.16):

$$\begin{aligned} Q_3 &= x_{0Q_3} + \frac{h_{Q_3}}{f_{Q_3}} \cdot \left(0,75 - \sum_{i=1}^{k_3-1} f_i \right) = 16 + \frac{2}{0,18} \cdot (0,75 - 0,65) = \\ &= 16 + \frac{2}{0,19} \cdot 0,10 = 16 + 1,05 = 17,05 \text{ zł} \end{aligned}$$

Uzyskana wartość oznacza, że co najmniej 75% pracowników działu logistyki otrzymuje stawkę godzinową nieprzekraczającą 17,05 zł, zaś przynajmniej 25% – wynoszącą co najmniej 17,05 zł.

Wartości kwartyli można wyznaczać nie tylko metodami analitycznymi, lecz również graficznie – z wykorzystaniem wykresu liczebności lub częstości skumulowanej.

W przypadku analizowanego przykładu odpowiedni wykres przedstawiono na rysunku 2.3.



Wykres 2.3. Graficzne wyznaczenie kwartyli dla rozkładu stawki godzinowej brutto pracowników działu logistyki

Źródło: opracowanie własne na podstawie danych umownych.

Na rysunku 2.3 przedstawiono skumulowany histogram wag, na podstawie którego wyznaczono krzywą skumulowaną utworzoną poprzez połączenie końców kolejnych przedziałów klasowych. Na osi rzędnych (OY) zaznaczono wartości odpowiadające $N/4$ (25%), $N/2$ (50%) oraz $3N/4$ (75%). Z każdego z tych punktów poprowadzono proste równoległe do osi odciętych (OX), a następnie wyznaczono punkty przecięcia tych prostych z krzywą skumulowaną. Ostatecznie, rzutując te punkty na oś poziomą, odczytano wartości odpowiadające kwartyłom. Analiza wykresu wskazuje, że odczytane graficznie wartości kwartyli są zgodne z wynikami uzyskanymi metodą analityczną.

2.3. Zadania z zakresu miar przeciętnych

- 2.1.** W ramach badania przeprowadzonego wśród małych i średnich przedsiębiorstw z województwa podlaskiego zebrano informacje na temat liczby kanałów sprzedaży wykorzystywanych przez każde z przedsiębiorstw do dystrybucji swoich produktów lub usług. Pod uwagę brano różne kanały, takie jak sprzedaż stacjonarna, platformy aukcyjne, e-commerce, sprzedaż telefoniczna, B2B, sprzedaż mobilna itp. Zebrane dane (liczba kanałów wykorzystywanych przez każde przedsiębiorstwo) są następujące:

5, 4, 5, 3, 2, 3, 5, 5, 5, 6, 1, 2, 4, 4, 1, 1, 2, 5, 3, 3, 4, 3, 1, 2, 6, 4, 4,
3, 4, 3, 3, 3, 4, 1, 5, 3, 2, 2, 2, 3, 1, 4, 1, 6, 4, 5, 4, 3, 3, 2.

Wyznacz średnią, dominantę i kwartyle. Otrzymane wyniki zinterpretuj.

- 2.2.** W ramach badania przeprowadzonego wśród właścicieli mikroprzedsiębiorstw zapytano, ilu stałych pracowników zatrudniają w swojej firmie (nie licząc właściciela). Poniżej przedstawiono dane z dwóch niezależnych grup badanych:

1. W pierwszym badaniu udział wzięło 30 przedsiębiorców. Liczba zatrudnianych stałych pracowników w poszczególnych firmach wynosiła:

1, 3, 4, 4, 4, 6, 7, 3, 2, 2, 3, 4, 4, 4, 5,
8, 1, 2, 2, 4, 4, 4, 5, 3, 2, 3, 1, 2, 5, 4.

2. W drugim badaniu udział wzięło 29 przedsiębiorców. Oto zebrane dane:

1, 3, 4, 4, 4, 6, 7, 3, 2, 2, 3, 4, 4, 4, 5,
8, 1, 2, 2, 4, 4, 4, 5, 3, 2, 3, 1, 2, 5.

Podaj kwartyle tego rozkładu. Wyznacz dominantę i średnią.

- 2.3.** W ramach analizy operacyjnej firm logistycznych działających na terenie Polski zebrano dane dotyczące liczby tras dostaw realizowanych codziennie przez jedno z przedsiębiorstw. Pod uwagę wzięto tylko trasy stałe, wykonywane regularnie każdego dnia roboczego. Dane zebrano w formie szeregu rozdzielczego w poniższej tabeli.

Liczba codziennych tras dostaw	Liczba przedsiębiorstw
2	4
3	11
4	17
5	6
6	2

Wyznacz średnią, dominantę i kwartyle liczby codziennych tras dostaw. Otrzymane wyniki zinterpretuj.

- 2.4. Poniższa tabela prezentuje liczbę punktów sprzedaży przedsiębiorstwa handlowego.

Liczba punktów sprzedaży	Odsetek miast [w %]
1	40,5
2	32,6
3	17,3
4	9,6

Wyznacz wartość średnią, dominującą oraz środkową liczby punktów sprzedaży prowadzonych przez analizowane przedsiębiorstwo w poszczególnych miastach. Otrzymane wyniki zinterpretuj.

- 2.5. W zakładzie produkcyjnym przeanalizowano wyniki kontroli jakości w kolejnych partiach produkcyjnych. Oceniano liczbę braków wykrytych w każdej partii. Wyniki zestawiono w poniższej tabeli.

Liczba braków	Liczba partii
0	10
1	15
2	8
3	7
4	4
5	3

Wyznacz dominującą, środkową i średnią liczbę braków w partii. Otrzymane wyniki zinterpretuj.

- 2.6. W poniższej tabeli zebrano dane dotyczące wieku grupy pracowników zakładu produkcyjnego.

Wiek [w latach]	Liczba pracowników
20-25	10
25-30	20
30-35	110
35-40	35
40-45	20
45-50	5

W jakim wieku jest najwięcej pracowników? Jak kształtują się średnia i środkowa wartość wieku analizowanych osób oraz jakie wartości przyjmują kwartył pierwszy i trzeci? Wyznacz odpowiednie miary również graficznie, a wyniki zinterpretuj.

- 2.7. W pewnym regionie przeprowadzono badanie dotyczące liczby pracowników zatrudnionych w przedsiębiorstwach działających w sektorze przetwórstwa spożywczego. Zebrane dane zestawiono w poniższej tabeli.

Liczba pracowników	Odsetek firm [w %]
50–150	0,01
150–250	0,12
250–350	0,19
350–450	0,33
450–550	0,25
550–650	0,10

Wyznacz średnią oraz dominantę i kwartyle liczby pracowników zatrudnionych w przedsiębiorstwach działających w sektorze przetwórstwa spożywczego. Dominantę i kwartyle wyznacz również graficznie. Otrzymane wyniki zinterpretuj.

- 2.8. Mając dane o liczbie szkół w miastach pewnego regionu, wyznacz wszystkie średnie klasyczne i pozycyjne (dominanta i kwartyle).

Liczba szkół	Odsetek miast [w %]
0–5	0,11
5–10	0,32
10–15	0,46
15–20	0,09
20–25	0,02

Wyznacz podane miary pozycyjne także graficznie. Otrzymane wyniki zinterpretuj.

- 2.9. Poniższe dane dotyczą średniej wielkości produkcji w poszczególnych zakładach. Wyznacz średnią oraz dominantę i kwartyle (także graficznie) wielkości produkcji analizowanych zakładów. Otrzymane wyniki zinterpretuj.

Wielkość produkcji [w tys. szt.]	Liczba zakładów
20–40	9
40–60	24
60–80	18
80–100	16
100–120	13

- 2.11.** Dane zawarte w poniższej tabeli przedstawiają koszty projektów (w mln złotych) zrealizowanych przez różne przedsiębiorstwa w danym roku kalendarzowym. Wyznacz średnią klasyczną i średnie pozycyjne (także graficznie). Otrzymane wyniki zinterpretuj.

Koszt projektu [w mln zł]	Liczba przedsiębiorstw
0,05–0,15	2
0,15–0,25	10
0,25–0,35	24
0,35–0,45	26
0,45–0,55	18

- 2.12.** Pobrano próbę losową złożoną ze 100 studentów Politechniki Białostockiej. Zbadano, ile czasu (w min.) zajmuje im dojazd z miejsca zamieszkania do uczelni. Otrzymane dane pogrupowano w klasy i przedstawiono w formie tabeli poniżej.

Czas dojazdu [w min.]	Liczba studentów
5–15	6
15–25	34
25–35	30
35–45	20
45–55	10

Wyznacz średnią oraz kwartale i dominantę (także graficznie), a otrzymane wyniki zinterpretuj.

- 2.13.** Mediana liczby osób zatrudnionych w przedsiębiorstwach pewnego regionu wynosi 58 osób. W regionie jest 35 firm zatrudniających od 50 do 60 osób, zaś przedsiębiorstw zatrudniających mniej niż 50 osób jest 184. Ile przedsiębiorstw jest umiejscowionych w analizowanym regionie?
- 2.14.** Dominanta zarobków osób zatrudnionych w pewnym przedsiębiorstwie to 4540 zł. Wiadomo, że 24 osoby zarabiają od 4500 do 4600 zł, zaś 18 osób zarabia od 4600 do 4700 zł. Ilu pracowników zarabia poniżej 4500 zł, ale nie mniej niż 4400 zł?
- 2.15.** Na zawodach sportowych co czwarty zawodnik osiągnął co najwyżej 78 punktów. Było 25 osób, które otrzymały od 70 do 80 punktów. Ile osób osiągnęło mniej niż 70 punktów, jeśli startowało 120 zawodników?
- 2.16.** W pewnym przedsiębiorstwie przeanalizowano czasy realizacji zleceń produkcyjnych (w godzinach). Ustalono, że połowa zleceń była realizowana w mniej niż 94 godziny, zaś 20% zleceń było realizowanych w czasie od 80 do 100 godzin. Jaki procent zleceń był realizowany dłużej niż 100 godzin?

2.17. W grupie 10 studentów piszących pracę kontrolną zbadano rozkład otrzymanych ocen. Uzyskano następujące wyniki: 2, 3, 3, 3, 4, 4, 5, 3, 5, 4. Oblicz średnią ocen uzyskanych przez studentów.

2.18. Informacje o liczbie pracowników zatrudnianych przez 30 mikroprzedsiębiorstw usługowych podane są w postaci następującego szeregu rozdzielczego:

Liczba pracowników	Liczba mikroprzedsiębiorstw
0	6
1	13
2	8
3	2
4	1
Ogółem	30

Oblicz średnią liczbę pracowników zatrudnionych w tych firmach.

2.19. Liczbę punktów sprzedaży firmy Sukces w badanych miastach przedstawia tabela:

Liczba punktów sprzedaży	Odsetek miast [w %]
1	37,5
2	35,0
3	17,5
4	10,0
Ogółem	100,0

Jaka jest przeciętna liczba punktów sprzedaży firmy Sukces w badanych miastach?

2.20. Dane dotyczą płacy pracowników zakładu Y:

Płaca [w zł]	Odsetek pracowników [w %]
5000–6000	4
6000–7000	10
7000–8000	46
8000–9000	20
9000–10 000	15
10 000–11 000	5
Ogółem	100

Jaki jest przeciętna płaca pracowników tego zakładu?

2.21. Rozkład przedsiębiorstw produkcyjnych pod względem liczby pracowników przedstawia się następująco:

Liczba pracowników	0-100	100-200	200-300	300-400	400-500	500-600
Odsetek przedsiębiorstw [w %]	0,02	0,08	0,10	0,40	0,30	0,10

Wyznacz przeciętną liczbę pracowników w badanych przedsiębiorstwach.

2.22. Do użytku oddano dwa nowe bloki mieszkalne. W bloku pierwszym 10% mieszkańców stanowią osoby samotne, 40% małżeństwa bezdzietne, 25% małżeństwa z jednym dzieckiem, 20% małżeństwa z dwojgiem dzieci i 5% małżeństwa z trojgiem dzieci. W bloku drugim mieszka razem 80 rodzin i ich struktura według liczby osób w rodzinie przedstawia się następująco:

Liczba osób w rodzinie	1	2	3	4	5	6
Liczba rodzin	14	20	30	10	4	2

Oblicz odpowiednie średnie arytmetyczne i przy ich pomocy scharakteryzuj strukturę badanych rodzin według liczby osób w rodzinie.

2.23. Mamy dwa szeregi rozdzielcze obrazujące rozkłady miesięcznych wydatków na transport (mężczyzn i kobiet). Ustal i porównaj przeciętny poziom wydatków na transport w obu zbiorowościach osobno i łącznie.

Wydatki [w zł]	Liczba kobiet
190-230	8
230-270	18
270-310	12
310-350	8
350-390	4
Ogółem	50

Wydatki [w zł]	Liczba mężczyzn
190-230	5
230-270	20
270-310	30
310-350	25
350-390	15
390-430	5
Ogółem	100

2.24. Średnia miesięczna płaca pierwszej brygady stażystów wynosiła 5800 zł, a średnia płaca drugiej brygady 7000 zł. Jaka była przeciętna płaca stażystów liczona dla dwóch brygad łącznie, jeżeli pierwsza brygada była 1,5 raza liczniejsza?

2.25. Zbadano grupę studentów studiów zaocznych pod względem wieku i zestawiono zebrane informacje w następujący szereg rozdzielczy:

Wiek [w latach]	do 20	20-22	22-24	24 i więcej	Ogółem
Liczba studentów	2	3	5	10	20

Dodatkowo wiadomo, że średni wiek 10 najstarszych studentów wynosi 26 lat. Zbadaj przeciętny wiek w całej grupie za pomocą miar klasycznych.

2.26. W pewnym roku akademickim na poszczególnych kierunkach Wydziału Inżynierii Zarządzania Politechniki Białostockiej do egzaminu ze statystyki przystąpiło łącznie 300 studentów I roku. Średnia ocen 85 studentów na kierunku zarządzanie wyniosła 3,8; 30 studentów na kierunku ZiIP – 3,7; 70 studentów na kierunku ZFiR – 4,0. Średnia ocena studentów na kierunku logistyka była równa 4,2. Jaka była średnia ocena ze statystyki ogółu studentów I roku na badanym wydziale?

2.27. W oparciu o poniższe dane ustal przeciętny czas eksploatacji maszyn.

Czas eksploatacji maszyn [w latach]	Liczba maszyn
do 2	3
do 4	10
do 6	16
do 8	20

2.28. Wiadomo, że samochód jechał:

- 5 km z prędkością 20 km/h i 10 km z prędkością 30 km/h;
 - 0,5 h z prędkością 20 km/h, 0,75 h z prędkością 30 km/h.
- Oblicz średnią prędkość samochodu. Jakiej użyjesz średniej i dlaczego?

2.29. W pewnej hurtowni:

- zakupiono dwa rodzaje towaru: 50 szt. towaru po 3 zł za sztukę i 40 szt. towaru po 5 zł za sztukę;
 - wydano 150 zł na zakup towaru, którego cena za jedną sztukę wynosi 3 zł i 200 zł na zakup towaru, którego jedna sztuka kosztuje 5 zł.
- Oblicz średnią cenę zakupionych towarów w podpunktach a) i b).

2.30. Załóżmy, że gęstość zaludnienia w dwóch 50-tysięcznych miastach wynosi odpowiednio: 400 osób/km² i 600 osób/km². Jaka jest przeciętna gęstość zaludnienia obu tych miast?

2.31. Okręg składa się z trzech regionów:

- I zamieszkuje 5000 osób z gęstością 20 os./km²;
- II zamieszkuje 4000 osób z gęstością 40 os./km²;
- III zamieszkuje 3000 osób z gęstością 10 os./km².

Jaka jest średnia gęstość zaludnienia w całym okręgu?

2.32. Oblicz średnią wydajność pracy czterech robotników na podstawie następujących danych:

Wydajność [w szt./h]	10	11	12	113
Liczba wyprodukowanych sztuk	110	220	330	144

2.34. Ceny za 1 kg truskawek w zależności od ich jakości kształtowały się następująco: 13 zł, 15 zł i 6 zł. Plantator uzyskał ze sprzedaży truskawek odpowiednio: 2 tys. zł, 3 tys. zł i 6 tys. zł. Ile wynosiła średnia cena 2 kg truskawek?

2.35. Zbadaj średnie tempo przyrostu produkcji pewnego przedsiębiorstwa w latach 2021–2025.

Lata	Produkcja przemysłowa [w tys. szt.]
2021	20
2022	24
2023	30
2024	42
2025	63

2.36. Z danych o ludności pewnej miejscowości wynika, że w kolejnych trzech okresach liczba ludności wynosiła odpowiednio 5 tys., 7 tys. i 8 tys. osób. Oblicz średni przyrost względny ludności.

2.37. Niżej podana jest tablica przedstawiająca liczbę strzelonych goli podczas jednej kolejki jesiennej rundy rozgrywek ligi piłki nożnej.

Liczba goli	0	1	2	3	4	5	Ogółem
Liczba meczów	15	16	4	5	2	1	43

Wyznacz dominującą i środkową liczbę strzelonych goli.

2.38. Poniższy szereg przedstawia strukturę według wieku w pewnym przedsiębiorstwie:

Wiek [w latach]	do 25	do 30	do 35	do 40	do 45	do 50	do 55
Liczba pracowników	8	32	47	59	84	95	100

Wyznacz dominujący wiek pracowników.

2.39. W spółdzielni mieszkaniowej Słoneczny Stok dominują mieszkania o powierzchni w przedziale 45–50 m² (30%). Na podstawie badań statystycznych ustalono, że najwięcej mieszkań ma powierzchnię 48,125 m², zaś 20% mieszkań ma powierzchnię pomiędzy 40 a 45 m². Jaki odsetek mieszkań ma powierzchnię w przedziale od 50 do 55 m²?

2.40. Szereg rozdzielczy jest następujący:

Szacunkowa wartość środków obrotowych [w mln zł]	do 20	20–30	30–40	40–50	50–60	60–70	Ogółem
Liczba zakładów rzemieślniczych	130	100	25	15	8	2	280

Na podstawie powyższych informacji:

1. Oblicz medianę i Q₃, a wyniki zinterpretuj.

2. Odpowiedz, czy za pomocą innych średnich obliczenie przeciętnej wartości byłoby poprawne i możliwe.
3. Odpowiedz, czy dodanie informacji o tym, że łączna wartość środków obrotowych wynosi 7 mld zł, zmieni odpowiedź na poprzednie pytanie. Jeśli tak, to oblicz tę średnią.

2.41. Mediana wzrostu 150 dzieci w wieku 14 lat znajdowała się w przedziale 155–160 cm, do którego zaliczało się 40 dzieci, i wynosiła 158 cm. Ile dzieci miało wzrost poniżej 155 cm?

2.42. Firma przewozowa MaxBus obsługuje trzy typy tras pasażerskich: T1 – trasy lokalne, T2 – trasy regionalne oraz T3 – trasy międzymiastowe. W ciągu tygodnia wykonano następującą liczbę kursów i przewieziono średnio tylu pasażerów podczas jednego kursu:

Typ trasy	Średnia liczba pasażerów na kurs	Liczba kursów w tygodniu
T1	25	70
T2	38	50
T3	60	30

W kolejnym tygodniu:

- liczba kursów tras T1 spadła o 20% (np. z powodu mniejszego zapotrzebowania),
- liczba kursów tras T2 wzrosła o 10%,
- liczba kursów tras T3 wzrosła o 50%.

Średnia liczba pasażerów poszczególnych kursów na każdej trasie pozostała bez zmian.

Oblicz i zinterpretuj średnią liczbę pasażerów po wystąpieniu opisanych zmian.

2.43. Firma logistyczna EcoTrans obsługuje trzy typy pojazdów dostawczych: małe (typ S), średnie (typ M) i duże (typ L). Każdy typ pojazdu ma inną ładowność i odbywa różną liczbę kursów w ciągu tygodnia. Dane za miniony tydzień przedstawiono w tabeli poniżej.

Typy pojazdów	Średnia masa ładunku na kurs [w t]	Liczba kursów w tygodniu
S (małe)	1,2 t	60
M (średnie)	2,5 t	45
L (duże)	4,0 t	25

W kolejnym tygodniu firma planuje zwiększyć liczbę kursów pojazdów typu M o 20%, a liczbę kursów pojazdów typu S zmniejszyć o 15% (przy zachowaniu tych samych średnich mas ładunku na kurs). Liczba

kursów pojazdów typu L się nie zmienia. Oblicz i zinterpretuj średnią masę ładunku po zmianach.

2.44. Kurier pokonał trasę z miasta A do miasta B (o długości 60 km) z prędkością 70 km/h. W drodze powrotnej z miasta B do A jechał z prędkością 90 km/h. Oblicz średnią prędkość kuriera na całej trasie (tam i z powrotem).

2.45. Samochód pokonał trasę złożoną z trzech odcinków:

- odcinek 1: 120 km – zużycie paliwa 7 l/100 km,
- odcinek 2: 80 km – zużycie paliwa 10 l/100 km,
- odcinek 3: 100 km – zużycie paliwa 8 l/100 km.

Oblicz średnie zużycie paliwa na 100 km na całej trasie.

2.46. W firmie A, w której pracuje 25 osób, wynagrodzenie każdego pracownika jest wyższe od 4000 zł. Czworo z nich zarabia miesięcznie nie więcej niż 4400zł, osiem zarabia nie więcej niż 4800 zł, piętnaście otrzymuje nie więcej niż 5200 zł, zaś dwadzieścia jeden nie więcej niż 5600 zł. Spośród pozostałych osób żadna nie zarabia więcej niż 6000 zł. Jaka jest przeciętna miesięczna płaca w firmie A?

2.47. W pewnym banku zbadano miesięczne wydatki 100 klientów korzystających z kart kredytowych. Wyniki przedstawiono w poniższej tabeli (wydatki w tys. zł).

Wydatki [w tys. zł]	Liczba klientów
0–1	10
1–2	25
2–3	30
3–5	20
5–8	10
8–10	5

- Oblicz średnie miesięczne wydatki klienta.
- Oszacuj dominantę i medianę dla tego rozkładu.
- Oblicz przeciętny wydatek przypadający na klienta w grupie wydającej powyżej 3 tys. zł.

2.4. Odpowiedzi do zadań z zakresu miar przeciętnych

- 2.1. $D = 3$; $Q_1 = 2$; $Me = 3$; $Q_3 = 4$; $\bar{x} = 3,28$ kanału
- 2.2. a) $D = 4$, $Q_1 = 2$, $Me = 4$, $Q_3 = 4$, $\bar{x} = 3,53$ zatrudnionych;
b) $D = 4$, $Q_1 = 2$, $Me = 4$, $Q_3 = 4$, $\bar{x} = 3,52$ zatrudnionych
- 2.3. $D = 4$; $Q_1 = 3$; $Me = 4$; $Q_3 = 4$; $\bar{x} = 3,775$ trasy
- 2.4. $D = 1$; $Q_1 = 1$; $Me = 2$; $Q_3 = 3$; $\bar{x} = 1,96$ punktów sprzedaży
- 2.5. $D = 1$; $Q_1 = 1$; $Me = 1$; $Q_3 = 3$; $\bar{x} = 1,766$ braków
- 2.6. $D = 32,73$; $Q_1 = 30,9$; $Me = 33,18$; $Q_3 = 36,43$; $\bar{x} = 33,75$ roku
- 2.7. $D = 413,64$; $Q_1 = 313,16$; $Me = 404,55$; $Q_3 = 490,00$; $\bar{x} = 399$ pracowników
- 2.8. $D = 11,373$; $Q_1 = 7,188$; $Me = 10,761$; $Q_3 = 13,478$; $\bar{x} = 10,45$ szkoły
- 2.9. $D = 54,29$; $Q_1 = 49,17$; $Me = 67,78$; $Q_3 = 91,25$; $\bar{x} = 70$ tys. szt.
- 2.10. $D = 0,370$; $Q_1 = 0,283$; $Me = 0,365$; $Q_3 = 0,442$; $\bar{x} = 0,36$ mln zł
- 2.11. $D = 23,750$; $Q_1 = 20,588$; $Me = 28,333$; $Q_3 = 37,500$; $\bar{x} = 29,4$ min.
- 2.12. 424 przedsiębiorstwa
- 2.13. 20 pracowników
- 2.14. 10 zawodników
- 2.15. 44% zleceń
- 2.16. $\bar{x} = 3,6$
- 2.17. $\bar{x} = 1,3$ pracownika
- 2.18. $\bar{x} = 2$ punkty sprzedaży
- 2.19. $\bar{x} = 7970$ zł
- 2.20. $\bar{x} = 368$ pracowników
- 2.21. $\bar{x} = 2,7$ osoby w obu blokach
- 2.22. $\bar{x}_K = 275,6$; $\bar{x}_M = 306$ zł
- 2.23. $\bar{x} = 6280$ zł
- 2.24. $\bar{x} = 23,8$ roku
- 2.25. $\bar{x} = 3,99$
- 2.26. $\bar{x} = 4,1$ lat
- 2.27. a) średnia harmoniczna $\bar{x}_H = 25,7$ km/h; b) średnia arytmetyczna $\bar{x} = 26$ km/h
- 2.28. a) średnia arytmetyczna $\bar{x} = 3,89$ zł/szt.; b) średnia harmoniczna $\bar{x}_H = 3,89$ zł/szt.
- 2.29. $\bar{x}_H = 480$ osób/km²
- 2.30. $\bar{x}_H = 18,5$ osoby/km²
- 2.31. $\bar{x}_H = 13,45$ szt./h
- 2.32. $\bar{x}_H = 8,1$ zł
- 2.33. wzrost średnio o 33% z roku na rok
- 2.34. wzrost średnio o 26% z okresu na okres
- 2.35. $D = 1$; $Me = 1$ gol

2.36. $D = 42,4$ roku

2.37. 24%

2.38. a) $Me = 21$ mln zł, $Q_3 = 28$ mln zł; b) Nie można domknąć pierwszego przedziału, gdyż jego liczebność przekracza 10% zbiorowości.

c) $\bar{x} = 25$ mln zł

2.39. 51 osób

2.40. $\bar{x} = 40$ osób

2.41. $\bar{x} = 2,38$ t

2.42. $\bar{x}_H = 78,75$ km/h

2.43. $\bar{x} = 8,13$ l/100 km

2.44. $\bar{x} = 5032$ zł

2.45. a) $\bar{x} = 3,075$; b) $Me = 2,5$, $D = 2,33$; c) $\bar{x} = 5,43$ tys. zł

2.5. Pytania testowe z zakresu miar przeciętnych

1. Dlaczego nie należy wyznaczać średniej arytmetycznej dla danych silnie asymetrycznych?
 - a) Ponieważ nie da się jej obliczyć.
 - b) Ponieważ zawsze jest równa medianie.
 - c) Ponieważ wymaga danych jakościowych.
 - d) Ponieważ może nie oddawać centralnego położenia danych.
2. Dominanta w analizie statystycznej oznacza:
 - a) najczęściej występującą wartość cechy,
 - b) wartość najbardziej skrajną,
 - c) wartość najmniejszą,
 - d) wartość przeciętną.
3. Czym różni się mediana od średniej arytmetycznej?
 - a) Średnia jest wartością środkową.
 - b) Mediana nie uwzględnia wartości skrajnych.
 - c) Mediana nie może być stosowana do danych mierzalnych.
 - d) Mediana zawsze jest większa od średniej.
4. Co można wnioskować, gdy rozkład liczby napraw komputerów ma dwa wierzchołki w histogramie?
 - a) Zmienna jest losowa.
 - b) Dane są niedokładne.
 - c) Histogram nie jest skonstruowany.
 - d) Występują dwie różne grupy urządzeń.
5. Co mierzy średnia geometryczna?
 - a) Zmienność wartości cechy.
 - b) Średnie tempo zmian zjawiska.
 - c) Średnią wartość cechy szeregu przekrojowego.
 - d) Środkową wartość cechy szeregu przekrojowego.
6. Średnia tygodniowa płaca pierwszej brygady stażystów wynosiła 900 zł, a średnia płaca drugiej brygady 1000 zł. Jeżeli pierwsza brygada była trzykrotnie liczniejsza niż druga, to średnia płaca stażystów liczona dla dwóch brygad wynosiła łącznie:
 - a) 975 zł,
 - b) 925 zł,
 - c) 950 zł,
 - d) 900 zł.
7. Dominanta miesięcznej liczby sprzedanych produktów wynosi 20. Co to oznacza?
 - a) Sprzedaż wynosiła średnio 20 sztuk.
 - b) Nie sprzedawano mniej niż 20 sztuk.

- c) Najczęściej sprzedawano 20 sztuk miesięcznie.
 - d) Każdego miesiąca sprzedawano 20 sztuk.
8. Dominanta liczby książek przeczytanych w roku wynosi 0. Co to oznacza?
- a) Każdy przeczytał 0 książek.
 - b) Najwięcej osób nie przeczytało żadnej książki.
 - c) Większość przeczytała jedną książkę.
 - d) Nie wiadomo, ile książek przeczytano.
9. Średnia długość snu wśród badanych studentów wynosi 6,2 godziny. Co to oznacza?
- a) Minimum snu to 6 godzin.
 - b) Każdy student śpi dokładnie 6,2 godziny.
 - c) Przeciętnie student śpi 6,2 godziny na dobę.
 - d) Wszyscy spali poniżej średniej.
10. Mediana wynagrodzeń w firmie wynosi 4800 zł. Co z tego wynika?
- a) Jest to wysokość maksymalnej pensji.
 - b) Połowa pracowników zarabia 4800 zł lub mniej.
 - c) Tylko jedna osoba zarabia 4800 zł.
 - d) Wszyscy pracownicy zarabiają 4800 zł.
11. Pierwszy kwartył poziomu hałasu w biurze to 45 dB. Co to znaczy?
- a) Połowa wyników to 45 dB.
 - b) Jest to poziom maksymalny.
 - c) 75% pomiarów wykazało dokładnie 45 dB.
 - d) 25% pomiarów hałasu dało wynik niższy niż 45 dB.
12. Trzeci kwartył miesięcznego zużycia wody wynosi 9 m^3 . Jak to interpretować?
- a) Większość zużywała ponad 9 m^3 .
 - b) Średnie zużycie to 9 m^3 .
 - c) Minimum to 9 m^3 .
 - d) 75% gospodarstw zużywało nie więcej niż 9 m^3 .
13. Mediana poziomu stresu w populacji wynosi 3 (w skali 1–5). Co to oznacza?
- a) Wszyscy są zestresowani.
 - b) Połowa osób oceniła stres na 3 lub mniej.
 - c) Stres jest maksymalny.
 - d) Średnia ocena to 3.
14. Miarami przeciętnymi są:
- a) średnia arytmetyczna, odchylenie standardowe, dominanta;
 - b) średnia harmoniczna, kwartył drugi, dominanta;
 - c) średnia arytmetyczna, mediana, kurtoza;
 - d) średnia arytmetyczna, mediana, rozstęp.
15. Analizując czas załadunku standardowej palety z towarem do ciężarówki w centrum dystrybucyjnym, wyznaczono $Q_1 = 3,2 \text{ min.}$, co oznacza, że:

- a) 25% palet jest ładowanych w czasie powyżej 3,2 min.;
 - b) 75% palet jest ładowanych w czasie poniżej 3,2 min.;
 - c) 50% palet jest ładowanych w czasie poniżej 3,2 min.;
 - d) 25% palet jest ładowanych w czasie poniżej 3,2 min.
16. Dominanta rozkładu stypendiów pobieranych na uczelni wynosi 1200 zł. Oznacza to, że:
- a) 25% studentów otrzymuje co najwyżej 1200 zł stypendium;
 - b) 50% studentów otrzymuje 1200 zł stypendium;
 - c) wokół 1200 zł jest skupiona największa liczba stypendiów;
 - d) gdyby fundusz stypendialny podzielić po równo między studentów pobierających stypendia, to każdy otrzymałby po 1200 zł.
17. Jeżeli wartości zmiennej podane są w jednostkach względnych (np. km/h, kg/os.), natomiast wagi w jednostkach liczników tych jednostek względnych (czyli odpowiednio: km, kg), to do obliczenia średniego poziomu badanej zmiennej należy użyć:
- a) średniej harmonicznej,
 - b) średniej geometrycznej,
 - c) średniej arytmetycznej,
 - d) dominanty.
18. Badanie oszczędności 300 gospodarstw domowych wykazało, że 225 gospodarstw ma oszczędności mniejsze niż 1000 zł. Mediana w tym rozkładzie:
- a) jest większa niż 1000 zł,
 - b) jest większa niż 225 gospodarstw,
 - c) jest mniejsza niż 1000 zł,
 - d) jest mniejsza niż 225 gospodarstw.
19. Które ze stwierdzeń jest fałszywe?
- a) Średnia arytmetyczna jest miarą pozycyjną.
 - b) Średnia arytmetyczna jest sumą wartości zmiennej wszystkich jednostek badanej zbiorowości podzieloną przez liczbę tych jednostek.
 - c) Suma odchyleń poszczególnych wartości zmiennej od średniej arytmetycznej jest równa zero.
 - d) Średnia arytmetyczna jest nie mniejsza niż minimalna wartość badanej cechy i nie większa niż wartość maksymalna.
20. Które ze stwierdzeń jest prawdziwe w odniesieniu do kwartyła pierwszego?
- a) Dzieli zbiorowość tak, że przynajmniej 25% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 75% nie większe od kwartyła pierwszego.
 - b) Dzieli zbiorowość tak, że przynajmniej 25% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 25% nie większe od kwartyła pierwszego.

- c) Dzieli zbiorowość tak, że przynajmniej 75% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 25% nie większe od kwartyła pierwszego.
 - d) Dzieli zbiorowość tak, że przynajmniej 75% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 75% nie większe od kwartyła pierwszego.
21. Które ze stwierdzeń jest prawdziwe w odniesieniu do kwartyła trzeciego?
- a) Dzieli zbiorowość tak, że przynajmniej 25% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 75% nie większe od kwartyła pierwszego.
 - b) Dzieli zbiorowość tak, że przynajmniej 25% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 25% nie większe od kwartyła pierwszego.
 - c) Dzieli zbiorowość tak, że przynajmniej 75% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 25% nie większe od kwartyła pierwszego.
 - d) Dzieli zbiorowość tak, że przynajmniej 75% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 75% nie większe od kwartyła pierwszego.
22. Które ze stwierdzeń jest prawdziwe w odniesieniu do kwartyła drugiego?
- a) Dzieli zbiorowość tak, że przynajmniej 25% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 75% nie większe od kwartyła pierwszego.
 - b) Dzieli zbiorowość tak, że przynajmniej 25% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 25% nie większe od kwartyła pierwszego.
 - c) Dzieli zbiorowość tak, że przynajmniej 75% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 25% nie większe od kwartyła pierwszego.
 - d) Dzieli zbiorowość tak, że przynajmniej 50% jednostek ma wartości nie mniejsze od kwartyła pierwszego, a przynajmniej 50% nie większe od kwartyła pierwszego.
23. Które ze stwierdzeń jest fałszywe?
- a) Kwartył dzieli zbiorowość na dwie części o określonych proporcjach.
 - b) Kwartył jest najczęściej występującą wartością.
 - c) Kwartył jest miarą pozycyjną.
 - d) Kwartył może być wartością środkową.
24. Które ze stwierdzeń jest prawdziwe?
- a) Dominanta dzieli zbiorowość na dwie części o określonych proporcjach.
 - b) Dominanta jest najczęściej występującą wartością.
 - c) Dominanta jest miarą klasyczną.
 - d) Dominanta jest wartością środkową.

25. W trzyosobowej rodzinie (ojciec, matka i syn) wyznaczono średnią płacę w maju. Ojciec miał płacę wyższą od średniej o 250 zł, zaś matka niższą o 200 zł. Płaca syna była:
- a) wyższa od średniej o 50 zł,
 - b) niższa od średniej o 50 zł,
 - c) wyższa od średniej o 225 zł,
 - d) niższa od średniej o 225 zł.
26. Średnia arytmetyczna nie jest miarą:
- a) tendencji centralnej,
 - b) klasyczną,
 - c) pozycyjną,
 - d) znajdującą się w przedziale od najmniejszej wartości badanej cechy.
27. Średnia płaca brutto w maju w przedsiębiorstwie spedycyjnym wynosiła 6250 zł. W czerwcu wszyscy pracownicy otrzymali premię w wysokości 500 zł. Średnia płaca w czerwcu wynosiła:
- a) 6250 zł,
 - b) 6500 zł,
 - c) 6750 zł,
 - d) 7000 zł.
28. Średnia płaca brutto w maju w przedsiębiorstwie spedycyjnym wynosiła 6250 zł. W czerwcu z powodu zawarcia nowego kontraktu wynagrodzenie pracowników wzrosło dwukrotnie. Średnia płaca w czerwcu wynosiła:
- a) 6250 zł,
 - b) 9375 zł,
 - c) 12 500 zł,
 - d) 18 750 zł.
29. Odwrotność średniej arytmetycznej z odwrotności wartości zmiennych to:
- a) średnia arytmetyczna,
 - b) średnia harmoniczna,
 - c) średnia geometryczna,
 - d) średnia pozycyjna.
30. Badano liczbę studentów mieszkających w akademiku i uzyskano następujące wyniki: w 10 pokojach mieszka po 2 studentów, w 20 pokojach po 3 studentów, a w 5 pokojach po 4 studentów. Do obliczenia średniej liczby studentów przypadających na pokój należy użyć:
- a) średniej harmonicznej,
 - b) średniej geometrycznej,
 - c) średniej arytmetycznej,
 - d) dominanty.
31. Badano liczbę studentów mieszkających w akademiku i uzyskano następujące wyniki: 20 studentów mieszka w pokojach dwuosobowych, 60 studentów w pokojach trzyosobowych, a 20 studentów w pokojach

czteroosobowych. Do obliczenia średniej liczby studentów przypadających na pokój należy użyć:

- a) średniej harmonicznej,
- b) średniej geometrycznej,
- c) średniej arytmetycznej,
- d) dominanty.

32. Badano liczbę studentów mieszkających w akademiku i uzyskano następujące wyniki: 20 studentów mieszka w pokojach dwuosobowych, 60 studentów w pokojach trzyosobowych, a 20 studentów w pokojach czteroosobowych. Do obliczenia liczby studentów najczęściej przypadających na pokój należy użyć:

- a) średniej harmonicznej,
- b) średniej geometrycznej,
- c) średniej arytmetycznej,
- d) dominanty.

33. Badano liczbę studentów przypadających średnio na jeden pokój w kolejnych semestrach i uzyskano następujące wyniki: w pierwszym semestrze były to 2 osoby, w drugim semestrze 2,5 osoby, a w trzecim semestrze 3 osoby. Do obliczenia średniej zmiany liczby studentów przypadających na pokój w badanym okresie należy użyć:

- a) średniej harmonicznej,
- b) średniej geometrycznej,
- c) średniej arytmetycznej,
- d) dominanty.

2.6. Odpowiedzi do pytań testowych z zakresu miar przeciętnych

1. d)
2. a)
3. b)
4. d)
5. b)
6. b)
7. c)
8. b)
9. c)
10. b)
11. d)
12. d)
13. b)
14. b)
15. d)
16. c)
17. a)
18. c)
19. a)
20. c)
21. b)
22. d)
23. b)
24. b)
25. b)
26. c)
27. c)
28. c)
29. b)
30. c)
31. a)
32. d)
33. b)

3. Miary dyspersji (zróźnicowania, rozproszenia)

3.1. Klasyczne miary dyspersji

W analizie statystycznej oprócz miar tendencji centralnej istotne znaczenie mają również miary dyspersji (zróźnicowania), które informują o tym, jak bardzo dane rozpraszają się wokół wartości przeciętnej. Miary te pozwalają ocenić stopień zmienności cechy w badanej zbiorowości. Do klasycznych miar dyspersji zaliczamy:

- wariancję,
- odchylenie standardowe,
- współczynnik zmienności oparty na odchyleniu standardowym,
- typowy obszar zmienności oparty na odchyleniu standardowym.

Wariancja (s^2) to jedna z podstawowych klasycznych miar dyspersji, która informuje o tym, jak bardzo poszczególne wartości cechy w analizowanej zbiorowości różnią się od średniej arytmetycznej. Mówiąc inaczej, wariancja mierzy przeciętną kwadratową odległość obserwacji od średniej. Im większa jest wariancja, tym większe jest zróźnicowanie danych. Obliczamy ją, wykorzystując następujące wzory:

- dla szeregu prostego:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3.1)$$

- dla szeregu rozdzielczego punktowego:

$$s^2 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i = \sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i \quad (3.2)$$

- dla szeregu rozdzielczego przedziałowego:

$$s^2 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^2 \cdot n_i = \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^2 \cdot f_i \quad (3.3)$$

Alternatywnie wariancję można obliczyć w sposób uproszczony, bez konieczności wcześniejszego wyznaczania odchyleń od średniej, stosując tzw. wzór skrócony, który opiera się na różnicy między średnią arytmetyczną kwadratów wartości a kwadratem średniej arytmetycznej:

- dla szeregu prostego:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i)^2 - (\bar{x})^2 \quad (3.4)$$

- dla szeregu rozdzielczego punktowego:

$$s^2 = \frac{1}{N} \sum_{i=1}^k (x_i)^2 \cdot n_i - (\bar{x})^2 = \sum_{i=1}^k x_i^2 \cdot f_i - (\bar{x})^2 \quad (3.5)$$

- dla szeregu rozdzielczego przedziałowego:

$$s^2 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i)^2 \cdot n_i - (\bar{x})^2 = \sum_{i=1}^k (\overset{\circ}{x}_i)^2 \cdot f_i - (\bar{x})^2 \quad (3.6)$$

Wariancję stosuje się przede wszystkim w sytuacjach, gdy:

- potrzebna jest ilościowa miara rozrzutu danych względem średniej arytmetycznej pozwalająca ocenić, jak bardzo poszczególne obserwacje są oddalone od wartości przeciętnej;
- wynik ma służyć jako podstawa do dalszych analiz statystycznych, takich jak: testy istotności (np. test t-Studenta, analiza wariancji), estymacja parametrów populacji czy obliczanie współczynnika korelacji liniowej, regresji oraz innych metod wnioskowania statystycznego.

Wadą wariancji jest jej ograniczona interpretowalność w zastosowaniach praktycznych – wynik wyrażany jest w jednostkach kwadratowych (np. cm^2 , zł^2), które nie mają bezpośredniego odniesienia do skali badanej cechy.

Odchylenie standardowe (s) to jedna z ważniejszych klasycznych miar zmienności, która określa, jak bardzo poszczególne wartości w zbiorze danych odchylają się przeciętnie od średniej arytmetycznej. Obliczamy ją jako pierwiastek kwadratowy z wariancji, a mianowicie:

$$s = \sqrt{s^2} \quad (3.7)$$

Odchylenie standardowe posiada istotną zaletę: jest wyrażone w tej samej jednostce co badana cecha (np. min., kg, zł), co czyni je bardziej intuicyjnym i łatwiejszym do interpretacji miernikiem niż wariancja.

Stosujemy je, gdy:

- chcemy posłużyć się miarą dyspersji, która zachowuje jednostkę badanej cechy, co umożliwia bezpośrednie porównanie z wartościami średnimi;
- analizujemy, jak bardzo dane są zróżnicowane wokół średniej, czyli oceniamy stopień rozproszenia wyników w zbiorze danych.

Schemat interpretacyjny odchylenia standardowego może przyjąć postać:

Wartości cechy [nazwa cechy] w zbiorowości [nazwa zbiorowości] przeciętnie odchylają się (różnią się) od średniej o ok. ... [wartość s] [jednostek pomiaru].

Współczynnik zmienności (V_s) to względna miara dyspersji, która informuje, jaki procent średniej arytmetycznej stanowi odchylenie standardowe. Dzięki temu pozwala porównywać zróżnicowanie różnych cech – nawet wtedy, gdy są one wyrażone w różnych jednostkach lub mają różne poziomy przeciętne.

Wzór na współczynnik zmienności wyraża się następująco:

$$V_s = \frac{s}{\bar{x}} \quad (3.8)$$

Dla uproszczenia interpretacji można przyjąć następującą umowną klasyfikację zmienności:

- do 5% – nieistotna (bardzo słaba) statystycznie zmienność (niektórzy autorzy podają do 10%),
- od 5% do 25% – niewielka (słaba) zmienność,
- od 25% do 50% – umiarkowana zmienność,
- od 50% do 100% – duża (silna) zmienność,
- powyżej 100% – bardzo duża (bardzo silna) zmienność.

Współczynnik zmienności stosujemy, gdy:

- chcemy porównać zmienność różnych cech, które różnią się jednostką miary i poziomem średnim;
- zależy nam na ocenie stabilności procesów lub zjawisk – im mniejsza wartość współczynnika zmienności, tym mniejsze względne zróżnicowanie, a więc większa stabilność analizowanego procesu;
- analizujemy dane w ujęciu porównawczym – przykładowo chcemy porównać zmienność sprzedaży w różnych krajach, branżach lub okresach.

Schemat interpretacyjny klasycznego współczynnika zmienności może przyjąć postać:

Rozkład zbiorowości [nazwa zbiorowości] ze względu na cechę [nazwa cechy] cechuje [bardzo słabe/słabe/umiarkowane/silne/bardzo silne] zróżnicowanie.

Typowy obszar zmienności klasyczny (obszar typowy, klasyczny obszar typowy) to przedział, który obejmuje wartości znajdujące się w określonej odległości od średniej arytmetycznej mierzonej w jednostkach odchylenia standardowego. Określa on, w jakim zakresie można oczekiwać wystąpienia większości (ponad 50%) zbiorowości, jeśli dane są stosunkowo regularnie rozłożone wokół średniej. Typowy obszar zmienności wyrażany jest wzorem:

$$x_{typ} \in (\bar{x} - s; \bar{x} + s) \quad (3.9)$$

Stosujemy go, gdy:

- chcemy sprawdzić, czy wartości są rozproszone wokół średniej w oczekiwanym zakresie;
- interesuje nas, jaki zakres obejmuje większość danych – bez konieczności analizowania całego rozkładu.

Schemat interpretacyjny klasycznego obszaru typowego może przyjąć postać:

Większość (ponad 50%) zbiorowości [nazwa zbiorowości] ma wartości cechy [nazwa cechy] w przedziale od ... [wartość $\bar{x} - s$] [jednostek pomiaru] do ... [wartość $\bar{x} + s$] [jednostek pomiaru].

Jeżeli rozkład jest zgodny z rozkładem normalnym, to:

- ok. 68% obserwacji mieści się w przedziale: $(\bar{x} - s; \bar{x} + s)$,
- ok. 95% obserwacji mieści się w przedziale: $(\bar{x} - 2s; \bar{x} + 2s)$,
- ok. 99,7% obserwacji mieści się w przedziale: $(\bar{x} - 3s; \bar{x} + 3s)$.

Przykład 3.1 (kontynuacja przykładu 1.1)

Wracając do szeregu rozdzielczego dotyczącego liczby osób w zespołach, w których najczęściej pracowali analizowani studenci Politechniki Białostockiej, przejdźmy do wyznaczenia klasycznych miar zmienności. W poprzednim rozdziale przedstawiono obliczenia dotyczące średniej liczby członków zespołu, która wyniosła 5 osób.

Obliczenia związane z wyznaczeniem wariancji przedstawiono w tabeli 3.1.

Tabela 3.1. Wyliczenia wariancji (wzór podstawowy) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności

x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2 \cdot n_i$
2	3	-3	27
3	4	-2	16
4	7	-1	7
5	12	0	0
6	8	1	8
7	3	2	12
8	2	3	18
9	1	4	16
Suma	40		104

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę z tabeli 3.1 do wzoru (3.2), obliczamy wariancję liczby osób w zespołach:

$$s^2 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i = \frac{104}{40} = 2,6 \text{ os.}^2$$

Wyliczenia można również przeprowadzić, korzystając z przekształconego wzoru wariancyjnego, który pozwala uniknąć obliczania odchyłeń poszczególnych wartości od średniej. Pomocnicze wyliczenia przedstawiono w tabeli 3.2.

Tabela 3.2. Wyliczenia wariancji (wzór przekształcony) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności

x_i	n_i	$(x_i)^2$	$(x_i)^2 \cdot n_i$
2	3	4	12
3	4	9	36
4	7	16	112
5	12	25	300
6	8	36	288
7	3	49	147
8	2	64	128
9	1	81	81
Suma	40		1104

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając wyliczenia z tabeli 3.2 do wzoru (3.5), obliczamy:

$$s^2 = \frac{1}{N} \sum_{i=1}^k x_i^2 \cdot n_i - (\bar{x})^2 = \frac{1104}{40} - 5^2 = 27,6 - 25 = 2,6 \text{ os.}^2$$

W obu przypadkach uzyskano ten sam wynik.

Inną metodą wyznaczania wariancji jest wykorzystanie częstości. W tym podejściu każda wartość zmiennej jest ważona przez swoją częstość, analogicznie jak w przypadku liczebności. Wyliczenia pomocnicze dla podstawowego wzoru przedstawiono w tabeli 3.3.

Tabela 3.3. Wyliczenia wariancji (wzór podstawowy) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości

x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2 \cdot f_i$
2	0,075	-3	0,675
3	0,100	-2	0,400
4	0,175	-1	0,175
5	0,300	0	0,000
6	0,200	1	0,200
7	0,075	2	0,300
8	0,050	3	0,450
9	0,025	4	0,400
Suma	1,000		2,600

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę uzyskaną w tabeli 4.3 do wzoru (4.2), otrzymujemy:

$$s^2 = \sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i = 2,6 \text{ os.}^2$$

Analogicznie można przeprowadzić obliczenia wariancji przy użyciu przekształconego wzoru, uwzględniając częstości poszczególnych wartości zmiennej. Szczegółowe wyliczenia przedstawiono w tabeli 3.4.

Tabela 3.4. Wyliczenia wariancji (wzór przekształcony) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości

x_i	f_i	$(x_i)^2$	$(x_i)^2 \cdot f_i$
2	0,075	4	0,300
3	0,100	9	0,900
4	0,175	16	2,800
5	0,300	25	7,500
6	0,200	36	7,200
7	0,075	49	3,675
8	0,050	64	3,200
9	0,025	81	2,025
Suma	40		27,600

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając wyliczenia z tabeli 3.4 do wzoru (3.5), obliczamy:

$$s^2 = \sum_{i=1}^k x_i^2 \cdot f_i - (\bar{x})^2 = 27,6 - 5^2 = 27,6 - 25 = 2,6 \text{ os.}^2$$

Niezależnie od zastosowanej metody obliczeń – klasycznej czy przekształconej – uzyskano ten sam wynik wariancji, co potwierdza poprawność przeprowadzonych wyliczeń oraz równoważność obu podejść rachunkowych.

Otrzymany wynik wyrażony jest w jednostkach kwadratowych, czyli w osobach². W celu uzyskania wartości możliwej do bezpośredniej interpretacji należy obliczyć odchylenie standardowe, czyli pierwiastek z wariancji. Wówczas wynik wyrażony jest już w jednostkach pomiaru, czyli w tym przypadku w osobach:

$$s = \sqrt{s^2} = \sqrt{2,6} = 1,61 \text{ os.}$$

Na tej podstawie możemy stwierdzić, że liczba osób w zespołach, w których pracowali analizowani studenci, różni się średnio od przeciętnej liczby 5 osób o 1,61 osoby.

Kontynuując analizę liczby osób w zespołach projektowych, w których pracowali studenci, przejdźmy do wyznaczenia typowego obszaru zmienności oraz współczynnika zmienności tej cechy.

Korzystając ze wzoru na typowy obszar zmienności (3.9), otrzymujemy przedział:

$$x_{typ} \in (5 - 1,61; 5 + 1,61)$$

$$x_{typ} \in (3,39; 6,61) \text{ os.}$$

Możemy zatem przyjąć, że większość zespołów liczy od 4 do 6 osób, co oznacza, że typowe rozmiary grup projektowych mieszczą się w podanym zakresie. Jeśli spojrzymy na dane empiryczne, możemy zauważyć, że znaczna część wartości (67,5%) rzeczywiście znajduje się w wyznaczonym przedziale, co potwierdza jego reprezentatywność.

Obliczmy teraz współczynnik zmienności, korzystając ze wzoru (3.8):

$$V_s = \frac{s}{\bar{x}} = \frac{1,61}{5} = 0,322 = 32,2\%$$

Otrzymany wynik oznacza, że zróżnicowanie liczby osób w zespołach jest umiarkowane.

Dla porównania, gdyby inna grupa studentów tworzyła zespoły, w których średnia liczba osób również wynosiłaby 5, ale odchylenie standardowe byłoby

mniejsze (przykładowo wynosiłoby 1 osobę), to współczynnik zmienności wyniósłby tylko 20%. W takim przypadku można byłoby stwierdzić, że hipotetyczna druga grupa pracuje w bardziej jednorodnych zespołach, gdyż liczba ich członków jest mniej zróżnicowana. Z kolei, gdyby inna grupa miała współczynnik zmienności równy przykładowo 60%, oznaczałoby to znacznie większe zróżnicowanie liczebności zespołów, co mogłoby wskazywać na dużą swobodę w ich tworzeniu ze względu na liczebność.

Przykład 3.2 (kontynuacja przykładu 1.2)

Wracając do szeregu rozdzielczego przedziałowego dotyczącego stawek godzinowych brutto pracowników działu logistyki, wyznaczmy klasyczne miary zmienności. Zaczniemy od wyliczenia wariancji na podstawie klasycznego wzoru, w którym obliczane są odchylenia kwadratowe poszczególnych wariantów cechy od średniej arytmetycznej. Wyliczenia pomocnicze przedstawiono w tabeli 3.5.

Tabela 3.5. Wyliczenia wariancji (wzór podstawowy) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności

$x_{0i} - x_{1i}$	n_i	$\overset{\circ}{x}_i$	$\overset{\circ}{x}_i - \bar{x}$	$(\overset{\circ}{x}_i - \bar{x})^2 \cdot n_i$
10–12	14	11	–4,1	235,34
12–14	23	13	–2,1	101,43
14–16	28	15	–0,1	0,28
16–18	18	17	1,9	64,98
18–20	13	19	3,9	197,73
20–22	4	21	5,9	139,24
Suma	100			739,00

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę z tabeli 3.5 do wzoru (3.3), obliczamy wariancję stawek godzinowych brutto pracowników działu logistyki:

$$s^2 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^2 \cdot n_i = \frac{739}{100} = 7,39 \text{ zł}^2$$

Alternatywnie obliczenia można przeprowadzić z wykorzystaniem przekształconego wzoru wariancyjnego, który pozwala uniknąć uprzedniego obliczania odchyleń. Wyliczenia pomocnicze wykonane tą metodą przedstawiono w tabeli 3.6.

Tabela 3.6. Wyliczenia wariancji (wzór przekształcony) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności

$x_{0i} - x_{1i}$	n_i	$\overset{\circ}{x}_i$	$\left(\overset{\circ}{x}_i\right)^2$	$\left(\overset{\circ}{x}_i\right)^2 \cdot n_i$
10-12	14	11	121	1694
12-14	23	13	169	3887
14-16	28	15	225	6300
16-18	18	17	289	5202
18-20	13	19	361	4693
20-22	4	21	441	1764
Suma	100			23 540

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę z tabeli 3.6 do wzoru (3.6), obliczamy wariancję stawek godzinowych brutto pracowników działu logistyki:

$$s^2 = \frac{1}{N} \sum_{i=1}^k \left(\overset{\circ}{x}_i\right)^2 \cdot n_i - (\bar{x})^2 = \frac{23\,540}{100} - (15,1)^2 = 235,40 - 228,01 = 7,39 \text{ zł}^2$$

Kolejna tabela (3.7) prezentuje wyliczenia związane z wariancją dla szeregu rozdzielczego przedziałowego z częstością.

Tabela 3.7. Wyliczenia wariancji (wzór podstawowy) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości

$x_{0i} - x_{1i}$	f_i	$\overset{\circ}{x}_i$	$\overset{\circ}{x}_i - \bar{x}$	$\left(\overset{\circ}{x}_i - \bar{x}\right)^2 \cdot f_i$
10-12	0,14	11	-4,1	2,3534
12-14	0,23	13	-2,1	1,0143
14-16	0,28	15	-0,1	0,0028
16-18	0,18	17	1,9	0,6498
18-20	0,13	19	3,9	1,9773
20-22	0,04	21	5,9	1,3924
Suma	1,00			7,3900

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę z tabeli 3.7 do wzoru (3.3), obliczamy wariancję stawek godzinowych brutto pracowników działu logistyki:

$$s^2 = \sum_{i=1}^k \left(\overset{\circ}{x}_i - \bar{x}\right)^2 \cdot f_i = 7,39 \text{ zł}^2$$

Tabela 3.8 prezentuje wyliczenia wariancji ze wzoru przekształconego dla szeregu rozdzielczego przedziałowego, w którym przedziały ważone są częstością.

Tabela 3.8. Wyliczenia wariancji (wzór przekształcony) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości

$x_{0i} - x_{1i}$	f_i	$\overset{\circ}{x}_i$	$(\overset{\circ}{x}_i)^2$	$(\overset{\circ}{x}_i)^2 \cdot f_i$
10-12	0,14	11	121	16,94
12-14	0,23	13	169	38,87
14-16	0,28	15	225	63,00
16-18	0,18	17	289	52,02
18-20	0,13	19	361	46,93
20-22	0,04	21	441	17,64
Suma	1,00			235,40

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę z tabeli 3.8 do wzoru (3.6), obliczamy wariancję stawek godzinowych brutto pracowników działu logistyki:

$$s^2 = \sum_{i=1}^k (\overset{\circ}{x}_i)^2 \cdot f_i - (\bar{x})^2 = 235,40 - (15,1)^2 = 235,40 - 228,01 = 7,39 \text{ zł}^2$$

Uzyskany wynik wyrażony jest w jednostkach kwadratowych, czyli w złotych², dlatego nie nadaje się do bezpośredniej interpretacji. W związku z tym poniżej przedstawiono obliczenia odchylenia standardowego stawek godzinowych, które – jako pierwiastek z wariancji – pozwala przedstawić zróżnicowanie w tej samej jednostce co analizowana cecha, czyli w złotych:

$$s = \sqrt{s^2} = \sqrt{7,39} = 2,72 \text{ zł}$$

Na tej podstawie można wnioskować, że stawki godzinowe pracowników działu logistyki przeciętnie różnią się od średniej wartości 15,10 zł o ok. 2,72 zł.

W dalszej części analizy stawek godzinowych brutto przystąpmy do wyznaczenia typowego obszaru zmienności oraz współczynnika zmienności tej cechy.

Stosując wzór na typowy obszar zmienności (3.9), otrzymujemy następujący przedział:

$$x_{typ} \in (15,10 - 2,72; 15,10 + 2,72)$$

$$x_{typ} \in (12,38; 17,82) \text{ zł}$$

Można zatem przyjąć, że stawki godzinowe większości pracowników mieszczą się w zakresie od ok. 12,38 zł do 17,82 zł. Analiza danych empirycznych wskazuje, że 66% obserwacji znajduje się w tym przedziale.

Następnie wyznaczmy współczynnik zmienności, korzystając ze wzoru (3.8):

$$V_s = \frac{s}{\bar{x}} = \frac{2,72}{15,10} = 0,180 = 18\%$$

Uzyskany wynik oznacza, że zróżnicowanie stawek godzinowych w analizowanym dziale można określić jako niewielkie.

3.2. Pozycyjne miary dyspersji

W przypadku rozkładów niesymetrycznych lub zawierających wartości odstające szczególnie przydatne są pozycyjne miary dyspersji, które opierają się na rozkładzie obserwacji, a nie na średnich odchyleniach. Są one bardziej odporne na wpływ wartości skrajnych i dobrze sprawdzają się w analizie danych porządkowych lub rozkładów asymetrycznych. Do pozycyjnych miar dyspersji zaliczamy:

- rozstęp,
- odchylenie ćwiartkowe,
- współczynnik zmienności oparty na odchyleniu ćwiartkowym,
- typowy obszar zmienności oparty na odchyleniu ćwiartkowym.

Rozstęp (R) to różnica między największą a najmniejszą wartością w zbiorze danych:

$$R = x_{max} - x_{min} \quad (3.10)$$

Stosujemy go, gdy:

- chcemy szybko ocenić zakres zmienności w danych,
- analizujemy dane porządkowe lub ilościowe (najlepiej przy niewielkich zbiorach),
- dane nie zawierają wartości odstających (rozstęp jest na nie bardzo wrażliwy).

Odchylenie ćwiartkowe (Q) to miara dyspersji oparta na kwartylach, która informuje o rozproszeniu środkowych 50% danych, czyli obserwacji znajdujących się pomiędzy pierwszym (Q_1) a trzecim (Q_3) kwartylem. Jest szczególnie przydatne, gdy chcemy wyeliminować wpływ wartości skrajnych. Wzór na odchylenie ćwiartkowe:

$$Q = \frac{Q_3 - Q_1}{2} \quad (3.11)$$

Należy zaznaczyć, że podana w powyższym wzorze różnica pomiędzy kwartylem trzecim i pierwszym ($Q_3 - Q_1$) nazywana jest **rozstępem międzykwartylowym**.

Odchylenie ćwiartkowe jako miarę zmienności stosujemy, gdy:

- chcemy ocenić rozproszenie danych z pominięciem wartości skrajnych;
- rozkłady są skośne lub zawierają obserwacje odstające – w takich przypadkach odchylenie ćwiartkowe lepiej odzwierciedla zróżnicowanie niż klasyczne odchylenie standardowe;

- pracujemy z danymi o rozkładzie nieznanym lub odmiennym od rozkładu normalnego, gdzie zastosowanie miar klasycznych mogłoby prowadzić do błędnych wniosków.

Schemat interpretacyjny odchylenia ćwiartkowego może przyjąć postać:

Wartości cechy [nazwa cechy] w zbiorowości [nazwa zbiorowości] zawężonej do środkowej połowy (zakresu pomiędzy pierwszym a trzecim kwartylem) przeciętnie odchylają się od mediany o ok. ... [wartość Q] [jednostek pomiaru].

Współczynnik zmienności pozycyjny (V_Q) to względna miara zróżnicowania, która opisuje rozproszenie środkowych 50% danych względem mediany. Jest przydatny przy analizie danych, które są asymetryczne lub zawierają wartości odstające, dla których klasyczny współczynnik zmienności (oparty na średniej) może dawać mylące wyniki. Wyznacza się go następująco:

$$V_Q = \frac{Q}{Me} \quad (3.12)$$

Stosujemy go, gdy bardziej adekwatne jest obliczanie zmienności względem mediany, a nie średniej, przykładowo w przypadku rozkładów skośnych lub zawierających obserwacje odstające. Zakresy interpretacyjne dla pozycyjnego współczynnika zmienności można przyjąć analogicznie jak w przypadku klasycznego.

Schemat interpretacyjny pozycyjnego współczynnika zmienności może przyjąć postać:

Rozkład zbiorowości [nazwa zbiorowości] zawężonej do zakresu pomiędzy pierwszym a trzecim kwartylem ze względu na cechę [nazwa cechy] cechuje [bardzo słabe/słabe/umiarkowane/silne/bardzo silne] zróżnicowanie.

Typowy obszar zmienności pozycyjny to obszar, w którym znajduje się typowe 50% obserwacji:

$$x'_{typ} \in (Me - Q; Me + Q) \quad (3.13)$$

Stosujemy go, gdy:

- chcemy wskazać „typowy” zakres, w którym znajduje się większość z środkowych 50% danych;
- potrzebujemy „odpornej” na wartości odstające miary rozrzutu;
- prezentujemy dane graficznie, np. na wykresie pudełkowym, gdzie przedział między kwartylami stanowi podstawę wizualizacji zmienności.

Schemat interpretacyjny typowego obszaru zmienności pozycyjnego może przyjąć postać:

Większość zbiorowości [nazwa zbiorowości] zawężonej do środkowej połowy (zakresu pomiędzy pierwszym a trzecim kwartylem) przyjmuje wartości cechy [nazwa cechy] w przedziale od ... [wartość $Me - Q$] [jednostek pomiaru] do ... [wartość $Me + Q$] [jednostek pomiaru].

Miary zróżnicowania można również podzielić na **absolutne** – wyrażone w jednostkach takich jak badana cecha statystyczna – oraz **względne** – najczęściej interpretowane w procentach.

Do miar zróżnicowania absolutnych należą:

- odchylenie standardowe,
- współczynnik zmienności,
- rozstęp,
- odchylenie ćwiartkowe,
- obszar typowy (pozycyjny i klasyczny).

Do miar zróżnicowania względnych należą współczynniki zmienności, zarówno klasyczny, jak i pozycyjny. Miary względne można stosować do porównywania zróżnicowania cech statystycznych oraz dyspersji cechy w różnych zbiorowościach.

Przykład 3.3 (kontynuacja przykładu 1.1)

Wracając ponownie do rozkładu liczebności zespołów, w których najczęściej pracowali analizowani studenci Politechniki Białostockiej, obliczmy miary zmienności oparte na pozycjach. Z wcześniejszych analiz wynika, że mediana (Me) liczby osób w zespole wyniosła 5, pierwszy kwartył (Q_1) – 4 osoby, a trzeci kwartył (Q_3) – 6 osób. Podstawiając te dane do wzoru (3.11), otrzymujemy:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{6 - 4}{2} = \frac{2}{2} = 1 \text{ os.}$$

Uzyskany wynik oznacza, że typowe odchylenie liczby osób w zespołach, ograniczone do środkowych 50% obserwacji, wynosiło średnio 1 osobę względem mediany.

Wykorzystując tę wartość do wyznaczenia typowego zakresu zmienności, czyli tzw. obszaru typowego, otrzymujemy:

$$x'_{typ} \in (5 - 1; 5 + 1)$$

$$x'_{typ} \in (4; 6) \text{ os.}$$

Oznacza to, że zdecydowana większość zespołów – mieszcząca się pomiędzy pierwszym a trzecim kwartylem, a więc obejmująca ponad 50% badanych – składała się z 5 osób, przy czym typowy rozrzut zawierał się w przedziale od 4 do 6 osób.

W końcowym etapie obliczamy pozycyjny współczynnik zmienności (V_Q), określający względne zróżnicowanie w obrębie środkowych 50% danych, co przebiega następująco:

$$V_Q = \frac{Q}{Me} = \frac{1}{5} = 0,2 = 20\%$$

Otrzymany wynik wskazuje na niewielkie zróżnicowanie liczebności zespołów projektowych w obrębie zakresu między pierwszym a trzecim kwartylem.

Przykład 3.4 (kontynuacja przykładu 1.2)

Biorąc pod uwagę stawki godzinowe brutto pracowników działu logistyki oraz wykorzystując wcześniej obliczone wartości kwartylowe ($Q_1 = 12,96$ zł; $Me = 14,93$ zł; $Q_3 = 17,05$ zł), można wyznaczyć pozycyjne miary zmienności.

Odchylenie ćwiartkowe wynosi:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{17,05 - 12,96}{2} = \frac{4,09}{2} = 2,045 \approx 2,05 \text{ zł}$$

Otrzymany wynik oznacza, że przeciętnie stawka godzinowa pracowników w zakresie między pierwszym a trzecim kwartylem różniła się od mediany o 2,05 zł. Na tej podstawie można wyznaczyć pozycyjny obszar typowy:

$$x'_{typ} \in (14,93 - 2,05; 14,93 + 2,05)$$

$$x'_{typ} \in (12,88; 16,98) \text{ zł}$$

Oznacza to, że większość pracowników, których stawki godzinowe mieściły się w środkowych 50% rozkładu, zarabiała od 12,88 zł do 16,98 zł brutto

za godzinę. Obszar ten jest również wyraźnie węższy niż typowy obszar wyznaczony na podstawie klasycznych miar zmienności, który wyniósł (12,38; 17,82).

Pozycyjny współczynnik zmienności obliczamy w następujący sposób:

$$V_Q = \frac{Q}{Me} = \frac{2,05}{14,93} = 0,137 = 13,7\%$$

Wynik ten wskazuje na niewielkie zróżnicowanie stawek godzinowych wśród pracowników działu logistyki po odrzuceniu skrajnych 25% najniższych i najwyższych wartości. Zmienność w środkowej części rozkładu jest zatem niższa niż ta, którą obserwujemy przy uwzględnieniu całej populacji (18%).

3.3. Zadania z zakresu miar dyspersji

- 3.1.** W pewnym zakładzie odsetek braków w kolejnych dniach kształtował się na poziomie: 0,08; 0,035; 0,22; 0,07; 0,01; 0,15; 0,24; 0,16; 0,30; 0,02. Korzystając z otrzymanych danych, wyznacz:
- rozstęp odsetka braków w badanym okresie,
 - odchylenie standardowe odsetka braków w badanym przedsiębiorstwie,
 - typowy obszar zmienności klasyczny,
 - klasyczny współczynnik zmienności.
- 3.2.** W Biurze Podróży „Kosmita” wprowadzono nową ofertę wycieczkową na Marsa. Chcąc określić, jakim zainteresowaniem cieszy się nowa oferta, od poniedziałku do niedzieli, czyli przez siedem dni, pracownicy skrzętnie notowali liczbę klientów pytających o jej szczegóły. Zebrane dane przedstawiały się następująco: 4, 8, 9, 7, 10, 3, 1. Oblicz:
- przeciętną dzienną liczbę klientów zainteresowanych wycieczką na Marsa;
 - wariancję, odchylenie standardowe i współczynnik zmienności dziennej liczby klientów pytających o nową ofertę wycieczkową;
 - rozstęp.
- 3.3.** W poniższej tabeli przedstawiono dane dotyczące liczby projektów realizowanych przez małe firmy zlokalizowane w jednej ze stref przemysłowych.

Liczba projektów	Liczba przedsiębiorstw
0	20
1	30
2	26
3	15
4	8
5	1

W celu określenia, jaki jest poziom zróżnicowania przedsiębiorstw ze względu na liczbę zrealizowanych projektów, wyznacz: odchylenie standardowe, odchylenie ćwiartkowe, typowy obszar zmienności (klasyczny i pozycyjny) oraz współczynnik zmienności (klasyczny i pozycyjny). Otrzymane wyniki zinterpretuj.

- 3.5. Przedsiębiorstwo posiada 100 maszyn, których czas dotychczasowej eksploatacji przedstawiono w poniższej tabeli.

Czas eksploatacji maszyn [w latach]	Liczba maszyn
0–2	20
2–4	15
4–6	10
6–8	25
8–10	20
10–12	10

- Zbadaj zróżnicowanie rozkładu czasu poprzez wyznaczenie: odchylenia standardowego, klasycznego współczynnika zmienności i typowego obszaru zmienności.
 - Korzystając z otrzymanych w poprzednim podpunkcie danych oraz wiedząc, że w przedsiębiorstwie średnia arytmetyczna wartości badanych maszyn wyniosła 200 mln zł, a wariancja charakteryzująca zróżnicowanie maszyn pod względem wartości była równa 400 (mln zł)², oceń, czy zbiorowość badanych maszyn jest bardziej zróżnicowana pod względem czasu eksploatacji czy też pod względem wartości.
- 3.6. Mając dane dotyczące wieku grupy pracowników pewnego zakładu produkcyjnego przedstawione w poniższej tabeli, wyznacz zróżnicowanie rozkładu wieku pracowników badanego zakładu za pomocą miar klasycznych i pozycyjnych. Otrzymane wyniki zinterpretuj.

Wiek [w latach]	Liczba pracowników
20–25	10
25–30	20
30–35	110
35–40	35
40–45	20
45–50	5

- 3.7. Rozkład liczby pracowników zatrudnionych w przedsiębiorstwach działających w sektorze przetwórstwa spożywczego zaprezentowano w poniższej tabeli.

Liczba pracowników	Odsetek firm [w %]
50–150	0,01
150–250	0,12
250–350	0,19
350–450	0,33
450–550	0,25
550–650	0,10

Wyznacz zróżnicowanie rozkładu liczby pracowników w przedsiębiorstwach z sektora przetwórstwa spożywczego za pomocą miar klasycznych i pozycyjnych. Otrzymane wyniki zinterpretuj.

- 3.8.** Poniżej przedstawione dane dotyczą średniej wielkości produkcji w poszczególnych zakładach.

Wielkość produkcji [w tys. szt.]	Liczba zakładów
20–40	9
40–60	24
60–80	18
80–100	16
100–120	13

Wyznacz zróżnicowanie wielkości produkcji za pomocą miar klasycznych i pozycyjnych. Otrzymane wyniki zinterpretuj.

- 3.9.** Dane zawarte w poniższej tabeli przedstawiają koszty projektów (w mln złotych) zrealizowanych przez różne przedsiębiorstwa w danym roku kalendarzowym.

Koszt projektu [w mln zł]	Liczba firm
0,05–0,15	2
0,15–0,25	10
0,25–0,35	24
0,35–0,45	26
0,45–0,55	18

Wyznacz zróżnicowanie rozkładu kosztów projektów za pomocą miar klasycznych i pozycyjnych. Otrzymane wyniki zinterpretuj.

- 3.10.** Pobrano losową próbę złożoną ze 100 studentów Politechniki Białostockiej. Zbadano, ile czasu (w min.) zajmuje im dojazd z miejsca zamieszkania do uczelni. Otrzymane dane przedstawiono w poniższej tabeli.

Czas dojazdu [w min.]	Liczba studentów
5–15	6
15–25	34
25–35	30
35–45	20
45–55	10

Wyznacz zróżnicowanie rozkładu czasu dojazdu za pomocą miar klasycznych i pozycyjnych. Otrzymane wyniki zinterpretuj.

3.11. Zbadano mieszkańców Białegostoku pod kątem procentowego udziału ich wydatków na żywność w odniesieniu do miesięcznych dochodów.

Procentowy udział wydatków na żywność w dochodzie [w %]	Odsetek mieszkańców [w %]
do 50	2
50-60	15
60-70	18
70-80	22
80-90	33
90-100	10

Wyznacz średni i środkowy udział wydatków na żywność w dochodzie w analizowanej grupie mieszkańców oraz poziom ich zróżnicowania (klasyczny i pozycyjny).

3.12. Wśród grupy składającej się z 20 studentów i 10 studentek przeprowadzono test ze statystyki opisowej. Z testu można było uzyskać maksymalnie 50 punktów. Wyniki w obu podgrupach przedstawiały się następująco:

Studenci	Studentki
$\bar{x}_1 = 33$	$\bar{x}_2 = 32$
$Me_1 = 31$	$Me_2 = 35$
$D_1 = 30$	$D_2 = 38$
$s_1 = 7$	$s_2 = 5$

Wykorzystując powyższe informacje:

- Określ, która podgrupa wypadła lepiej i uzasadnij swoją opinię.
- Wyznacz średnią uzyskanych punktów dla całej grupy.
- Wyznacz wariancję liczby punktów w całej grupie.
- Oceń stopień zróżnicowania wyników w całej grupie.

3.13. W poniższej tabeli przedstawiono wyniki badania dotyczącego długości stażu pracy wykonanego dla trzech grup pracowników.

Grupa	Średnia stażu pracy [w latach]	Odchylenie standardowe stażu pracy [w latach]	Liczebność
I	5	1,5	12
II	7	2,0	28
III	6	1,2	10

Wyznacz odchylenie standardowe, obszar typowy i współczynnik zmienności stażu pracy ogółu pracowników.

- 3.14.** W pewnej przychodni stomatologicznej zapisywano przez miesiąc czas oczekiwania w poczekalni zgłaszających się tam pacjentów. Wyniki zebrano w postaci poniższej tabeli.

Czas oczekiwania [w min.]	Liczba pacjentów przychodni
do 6	24
6–12	80
12–18	32
18–24	24
24 i więcej	10

Zbadaj zróżnicowanie rozkładu czasu oczekiwania na leczenie w badanej przychodni stomatologicznej za pomocą miar klasycznych i pozycyjnych.

- 3.15.** Przeanalizowano klientów pewnego banku ze względu na dwie cechy: czas ich obsługi podczas zawierania umowy kredytowej i kwotę branego kredytu. Wiadomo, że w przypadku wylosowanych 20 klientów banku łączny czas ich obsługi wynosił 54 godziny, zaś suma kwadratów poszczególnych czasów obsługi wyniosła 154 (godz.)². Wiedząc dodatkowo, że zróżnicowanie tychże klientów ze względu na kwotę zaciąganego kredytu wynosi $V_s = 29\%$, określ, która z cech bardziej różnicuje badaną grupę.
- 3.16.** Pewna firma handlowa posiada dwa oddziały. W pierwszym tygodniowa wielkość sprzedaży (liczona w tys. szt.) kształtowała się następująco: 4,7; 5,1; 4,8; 3,6; 4,9; 3,6; 2,3. W drugim średnia ilość sprzedanych sztuk wyniosła 4 tys., zaś średnia z kwadratów wielkości sprzedaży – 18,21. Wyniki sprzedaży którego z oddziałów utrzymywały się na bardziej stabilnym poziomie w badanym okresie?
- 3.17.** W badaniach strat czasu pracy spowodowanych brakiem surowców w pewnym zakładzie produkcyjnym 15% pracowników oceniło te straty na 10%, 40% pracowników na 20%, zaś dwie kolejne grupy stanowiące po 16% załogi stwierdziły, że tracą: jedna 25%, a druga 30% czasu. Pozostała część załogi oceniła straty czasu pracy aż na 40%. Oblicz przeciętny odsetek strat czasu pracy w badanym zakładzie. Wiedząc, że współczynnik zmienności płac wynosi $V_s = 12\%$, porównaj zróżnicowanie płac ze zróżnicowaniem strat czasu pracy.
- 3.18.** W pewnym zakładzie produkcyjnym czas (w godz.) przestoju tokarek na skutek awarii kształtował się na poziomie: 4,0; 2,5; 12,0; 7,0; 0,5; 13,5; 24,0; 18,0; 30,0; 76,0. Oblicz rozstęp czasu postoju spowodowanego awariami. Co można stwierdzić na podstawie otrzymanego wyniku?

3.19. Znajdź rozstęp międzykwartylowy i odchylenie ćwiartkowe dla poniższego zbioru danych: 8, 9, 7, 10, 5, 4, 6.

3.20. Instytut Meteorologii i Gospodarki Wodnej notował (o godz. 12:00 każdego dnia) temperaturę w ciągu kolejnych dni kwietnia 2025 r. w Warszawie (liczby umowne):

Temperatura [w °C]	0	2	6	8	13	15	17	20	23	25	27
Liczba dni	2	3	3	4	5	6	2	2	1	1	1

- Oblicz odchylenie standardowe temperatury w kolejnych dniach w kwietniu.
- Oblicz odchylenie ćwiartkowe temperatury w kolejnych dniach w kwietniu.
- Wyznacz pozycyjny typowy obszar zmienności temperatury dziennej.
- Jaki procent dni w kwietniu charakteryzował się nietypowymi temperaturami?

3.21. W fabryce sznurów przy produkcji sznura typu Z obowiązuje 3% współczynnik zmienności jego wagi. Ustal, czy pracownik X wypełnił normatyw produkcji, jeżeli po zważeniu 17 wykonanych sztuk okazało się, że ważą one: 110, 122, 120, 128, 117, 120, 121, 124, 126, 122, 128, 121, 118, 130, 129, 128, 128 g.

3.22. Maszyna A produkuje detale o średniej wadze 10 g i odchyleniu standardowym 2,4 g. Detale wyprodukowane przez maszynę B mają natomiast średnią wagę 6 g i odchylenie standardowe 1,8 g. Która maszyna pracuje bardziej precyzyjnie?

3.23. Poniższa tabela przedstawia czas trwania budowy budynków o kubaturze 2000–3000m² przedstawia tabela.

Czas trwania budowy [w miesiącach]	Liczba budów
do 7	12
7–13	78
13–19	30
19–25	7
25 i więcej	3
Razem	130

Zbadaj zróżnicowanie danego rozkładu za pomocą miar klasycznych i pozycyjnych.

3.24. Po dokonaniu analizy wyników z egzaminu dla 50 kandydatów na maklerów ustalono, że łączna liczba punktów uzyskanych przez nich na egzaminie wyniosła 6508, a suma kwadratów liczby punktów uzyskanych przez poszczególnych kandydatów była równa 871460. Wiedząc

dodatkowo, że współczynnik zmienności czasu przygotowania kandydatów do egzaminu wynosi 30,7%, ustal, która z badanych cech (czas przygotowania czy wynik) wykazywała większe zróżnicowanie.

3.25. Zakład obuwniczy składa się z trzech wydziałów produkujących różny asortyment obuwia. Element wadliwie wykończony jest kwalifikowany przez brakarzy jako brak. Przeciętne odsetki braków na poszczególnych wydziałach kształtują się odpowiednio: 4, 2 i 4%, zaś ich wariancje wynoszą: 0,64; 0,49; 0,81. Dzienna produkcja wydziału I wynosi 500 par, a wydziału II – 400 par obuwia. Nie posiadamy informacji o dziennej produkcji wydziału III, wiadomo jednak, że średni odsetek braków w całym zakładzie ukształtował się na poziomie 3,5%.

3.26. Uzupełnij informację o wielkości produkcji wydziału III.

- Oblicz wariancję odsetka braków w całym zakładzie obuwniczym.
- Oceń stopień zróżnicowania odsetka produkcji wybrakowanej w całym zakładzie.

3.27. Na podstawie badań budżetów gospodarstw domowych w 2024 roku ustalono, że przeciętne miesięczne wydatki na odzież i obuwie w rodzinach czteroosobowych wyniosły $\bar{x} = 1300$ zł i $s = 200$ zł miesięcznie na osobę. Przeprowadzono badanie wśród 20 rodzin czteroosobowych i otrzymano następujące dane (w zł): 800; 920; 950; 950; 1000; 1000; 1150; 1200; 1200; 1200; 1350; 1400; 1500; 1600; 1600; 1700; 1700; 1800; 2500; 3000. Określ, ile rodzin można nazwać „typowymi” ze względu na badaną cechę. Dla ilu rodzin wydatki te przekraczają statystycznie dopuszczalny obszar zmienności?

3.28. Lekkoatleta A uzyskał w skoku w dal następujące wyniki na zawodach w całym sezonie (w m): 6,82; 6,96; 7,23; 7,05; 7,80; 7,75. Lekkoatleta B, startujący w tych samych zawodach, uzyskał takie wyniki, że ich średnia arytmetyczna wyniosła 7,5 m, a suma ich kwadratów 450,2592 m². Który z tych lekkoatletów osiągał wyniki na bardziej zbliżonym poziomie?

3.29. Mamy następujące dane dotyczące jednodniowych zarobków 10 pracowników w czterech różnych zakładach pracy:

Zarobek jednodniowy [w zł]	Pracownicy w zakładzie			
	A	B	C	D
100	2	5	1	0
150	2	0	2	4
200	2	0	4	2
250	2	0	2	4
400	2	5	1	0

Oceń rozproszenie zarobków w poszczególnych zakładach.

3.30. Rozkład otrzymanych punktów z egzaminu z matematyki 34 studentów scharakteryzowano w sposób następujący: średnia arytmetyczna dla kwadratów wartości poszczególnych ocen wyniosła 25 punktów, a średnia arytmetyczna ważona 4,5 punktu. Maksymalna liczba punktów, jaką można było otrzymać z egzaminu wynosiła 10. Oceń zróżnicowanie studentów ze względu na otrzymaną liczbę punktów.

3.31. W pewnej firmie w czerwcu 2017 r. zbadano dwie grupy pracownicze pod względem długości stażu pracy. Okazało się, że dane charakteryzujące obie grupy były następujące:

Grupa	Średni staż pracy [w latach]	Odchylenie standardowe
I	10	2
II	12	3

Omów, jak można zbadać średni staż pracy i zróżnicowanie ogółu pracowników łącznie, jeśli obie grupy są jednakowo liczne.

3.32. W przedsiębiorstwie A przeciętna wydajność jednego robotnika wynosi 20 szt./h, zaś odchylenie standardowe to 40% średniej arytmetycznej. Dla przedsiębiorstwa B uzyskano natomiast następujący szereg skumulowany:

Wydajność pracy [w szt./h]	Liczba pracowników
do 10	25
do 20	80
do 30	95
do 40	100

Porównaj zróżnicowanie wydajności pracy w obu przedsiębiorstwach.

3.33. Bank przeanalizował czas obsługi klientów w pięciu oddziałach (w min.): 12, 15, 10, 18, 20. Oblicz odchylenie standardowe i współczynnik zmienności czasu obsługi. Oceń stabilność pracy oddziałów.

3.34. W ciągu miesiąca bank udzielił 8 kredytów hipotecznych na następujące kwoty (w mln zł): 0,9; 1,2; 1,0; 0,8; 1,5; 1,3; 0,7; 1,8. Oblicz wariancję, odchylenie standardowe i rozstęp kwot kredytów.

3.35. Bank porównał wyniki sprzedaży produktów inwestycyjnych w dwóch zespołach doradców (kwoty w tys. zł):

- zespół A: 100, 105, 110, 115, 120;
- zespół B: 90, 100, 110, 120, 130.

3.36. Porównaj zróżnicowanie wyników obu zespołów, obliczając odchylenie standardowe i współczynnik zmienności.

- 3.37.** Bank przeanalizował miesięczne wpływy na poszczególne konta klientów VIP (w tys. zł), które wynosiły: 500, 520, 490, 510, 480, 530. Oblicz odchylenie standardowe i wskaż, czy wpływy były stabilne.
- 3.38.** W celu oceny stabilności portfela kredytowego bank przeanalizował wysokość rat kredytów gotówkowych (w zł), które wynosiły: 900, 920, 950, 910, 930, 940, 915, 935. Oblicz odchylenie standardowe wysokości rat i skomentuj wynik.
- 3.39.** Czas dostawy poszczególnych przesyłek kurierskich (w godz.) wynosił: 4, 5, 6, 8, 7, 5, 6, 7, 5, 9. Oblicz odchylenie standardowe i kwartyłowy rozstęp czasu dostaw.
- 3.40.** Koszt transportu 10 kolejnych dostaw (w tys. zł) wynosił: 2,1; 2,3; 2,5; 2,0; 2,2; 2,6; 2,4; 2,7; 2,0; 2,3. Oblicz wariancję i współczynnik zmienności kosztów dostaw.
- 3.41.** Magazyn przeanalizował wahania zapasów surowca A (w t), przy czym waga poszczególnych ładunków wynosiła: 50, 52, 47, 53, 48, 51, 54, 46. Oblicz rozstęp, odchylenie standardowe oraz współczynnik zmienności. Skomentuj stabilność zapasów.
- 3.42.** Firma logistyczna zbadała czas załadunku poszczególnych pojazdów (w min.), który wynosił: 45, 50, 55, 60, 65, 55, 60, 50, 55, 60. Oblicz kwartyłowy rozstęp i odchylenie standardowe.
- 3.43.** Czas przejazdu na trasie A–B w różnych warunkach drogowych (w min.) wynosił: 90, 95, 100, 85, 110, 105, 95, 100. Oblicz współczynnik zmienności i ocen stabilność czasu przejazdu.
- 3.44.** Linia produkcyjna wytwarza detale o wadze (w g): 100, 102, 99, 101, 98, 103, 97, 100, 102, 101. Oblicz wariancję i odchylenie standardowe masy detali.
- 3.45.** Firma produkuje komponenty o długości nominalnej 50 mm. Próbkę pomiarów (w mm) miały masę: 50,1; 49,9; 50,2; 50,0; 49,8; 50,1; 50,3; 49,7. Oblicz odchylenie standardowe i określ, czy proces produkcji jest stabilny (za stabilny uznaj taki z odchyleniem standardowym $< 0,25$ mm).
- 3.46.** W ciągu 10 dni fabryka wyprodukowała (w tys. szt.) następującą ilość towarów: 120, 118, 121, 117, 123, 119, 122, 118, 120, 121. Oblicz rozstęp kwartyłowy i współczynnik zmienności dziennej produkcji.
- 3.47.** Koszty produkcji partii wyrobów (w tys. zł) wyniosły: 300, 310, 320, 340, 330, 350, 360, 340. Oblicz odchylenie standardowe i wskaż, czy koszty są stabilne (przyjmij stabilność przy współczynniku zmienności poniżej 5%).

3.48. Czas realizacji zamówień produkcyjnych (w dniach) wynosił: 10, 11, 12, 9, 13, 14, 10, 12, 11, 10. Oblicz rozstęp kwartyłowy i pozycyjny współczynnik zmienności. Oceń stabilność procesu.

3.4. Odpowiedzi do zadań z zakresu miar zmienności

- 3.1. a) $R = 0,29$; b) $s = 0,096$; c) $x_{typ} \in (0,033; 0,224)$; d) $V_s = 75\%$
- 3.2. a) $\bar{x} = 6$ klientów; b) $s^2 = 9,714$ klienta², $s = 3,117$ klienta, $V_s = 51,9\%$;
c) $R = 9$ klientów
- 3.3. $s^2 = 1,530$ projektu²; $s = 1,237$ projektu; $V_s = 75,4\%$;
 $x_{typ} \in (0,403; 2,877)$ projektu; $Q = 0,5$ projektu; $V_Q = 50\%$;
 $x'_{typ} \in (0,5; 1,5)$ projektu
- 3.4. a) $s^2 = 10,96$ roku², $s = 3,311$ roku, $V_s = 57,1\%$,
 $x_{typ} \in (2,489; 9,111)$ roku; b) Zbiorowość jest bardziej zróżnicowana
pod względem czasu ($V_s = 10\%$).
- 3.5. $s^2 = 25,94$ roku²; $s = 5,09$ roku; $V_s = 15,1\%$;
 $x_{typ} \in (28,66; 38,84)$ roku; $Q = 2,76$ roku; $V_Q = 8,3\%$;
 $x'_{typ} \in (30,42; 35,94)$ roku
- 3.6. $s^2 = 14\,099$ pracowników²; $s = 118,74$ pracownika;
 $V_s = 29,8\%$; $x_{typ} \in (280,26; 517,74)$ pracownika;
 $Q = 88,42$ pracownika; $V_Q = 21,9\%$;
 $x'_{typ} \in (316,12; 492,97)$ pracownika
- 3.7. $s^2 = 640$ (tys. szt.)²; $s = 25,30$ tys. szt.; $V_s = 36,1\%$;
 $x_{typ} \in (44,70; 95,30)$ tys. szt.; $Q = 21,04$ tys. szt.; $V_Q = 31\%$;
 $x'_{typ} \in (46,74; 88,82)$ tys. szt.
- 3.8. $s^2 = 0,0109$ (mln zł)²; $s = 0,104$ mln zł; $V_s = 29\%$;
 $x_{typ} \in (0,256; 0,464)$ mln zł; $Q = 0,079$ mln zł; $V_Q = 21,8\%$;
 $x'_{typ} \in (0,286; 0,445)$ mln zł
- 3.9. $s^2 = 117,64$ min.²; $s = 10,846$ min.; $V_s = 36,9\%$;
 $x_{typ} \in (18,554; 47,954)$ min.; $Q = 8,465$ min.; $V_Q = 29,8\%$;
 $x'_{typ} \in (19,877; 36,789)$ min.
- 3.10. $\bar{x} = 74,9$ pkt. proc.; $s^2 = 168,99$ (pkt. proc.)²;
 $s = 13,00$ pkt. proc.; $V_s = 17,4\%$; $x_{typ} \in (61,9; 87,9)$ pkt. proc.;
 $Me = 76,818$ pkt. proc.; $Q = 10,505$ pkt. proc.; $V_Q = 13,7\%$;
 $x'x'_{typ} \in (66,313; 87,323)$ pkt. proc.
- 3.11. b) $\bar{x} = 32,67$ pkt.; c) $s^2 = 41,22$ (pkt.)²
- 3.12. $s = 1,941$ roku; $V_s = 30,7\%$; $x_{typ} \in (4,379; 8,261)$ roku
- 3.13. $s^2 = 42,034$ min.²; $s = 6,483$ min.; $V_s = 53,9\%$;
 $x_{typ} \in (5,552; 18,519)$ min.; $Q = 4,509$ min.; $V_Q = 42,6\%$;
 $x'_{typ} \in (6,066; 15,084)$ min.
- 3.14. $\bar{x} = 2,7$ godz.; $s^2 = 0,41$ (godz.)²; $V_s = 23,7\%$; Bardziej różnicuje
kwota zaciągniętego kredytu.
- 3.15. Pierwszy, bo: I) $\bar{x} = 4,14$ tys. szt., $s = 0,94$ tys. szt., $V_s = 22,7\%$;
II) $\bar{x} = 4$ tys. szt., $s = 1,5$ tys. szt., $V_s = 37,2\%$.

- 3.16. Większe zróżnicowanie strat czasu pracy, bo $\bar{x} = 23,5\%$; $s = 8,65\%$;
 $V_s = 36,8\%$.
- 3.17. $R = 75,5$ godz.
- 3.18. $Q_3 - Q_1 = 4$; $Q = 2$
- 3.19. a) $s = 7,1$ °C; b) $Q = 4,5$ °C c) $x_{typ} \in (4,9; 19,1)$ °C; d) 67%
- 3.20. Nie, $V_s = 4,2\%$.
- 3.21. maszyna A
- 3.22. $s = 4,8$ miesiąca; $V_s = 40,6\%$; $x_{typ} \in (7,1; 16,7)$ miesiąca;
 $Q = 3$ miesiące; $V_Q = 26,7\%$; $x'_{typ} \in (8,1; 14,0)$ miesiąca
- 3.23. czas przygotowania, $V_{wynik} = 17\%$
- 3.24. a) $n_{III} = 700$ par obuwia; b) $s^2 = 0,75$;
c) zróżnicowanie słabe, $V_s = 25\%$
- 3.25. 6 rodzin; 14 rodzin
- 3.26. lekkoatleta A
- 3.27. $V_A = 47\%$; $V_B = 60\%$; $V_C = 37\%$; $V_D = 22\%$
- 3.28. $V_s = 48,4\%$, zróżnicowanie umiarkowane
- 3.29. $V_I = 20\%$; $V_{II} = 25\%$
- 3.30. Zróżnicowanie wydajności pracy w obu przedsiębiorstwach jest takie samo, $V_s \approx 40\%$.
- 3.31. $s = 3,7$ min.; $V_s = 24,6\%$
- 3.32. $s^2 = 0,123$ (mln zł)²; $s = 0,35$ mln zł; $R = 1,1$ mln zł
- 3.33. $s_A = 7,07$ tys. zł; $V_A = 6,4\%$; $s_B = 14,14$ tys. zł; $V_A = 12,9\%$
- 3.34. $s = 17,08$ tys. zł; $V_s = 3,4\%$; Wpływy były stabilne.
- 3.35. $s = 15,6$ zł
- 3.36. $s = 1,47$ h; $Q_3 - Q_1 = 2$ godz.
- 3.37. $s^2 = 0,053$ (tys. zł)²; $V_s = 10\%$;
- 3.38. $R = 8$ t; $s = 2,71$ t; $V_s = 5,41\%$; Zapasy były stabilne.
- 3.39. $Q_3 - Q_1 = 10$ min.; $s = 5,68$ min.
- 3.40. $V_s = 7,7\%$; Czas przejazdu był stabilny.
- 3.41. $s^2 = 3,2$ g²; $s = 1,79$ g
- 3.42. $s = 0,19$ mm; Proces produkcji jest stabilny.
- 3.43. $Q_3 - Q_1 = 3$ tys. szt; $V_s = 1,5\%$
- 3.44. $s = 19$ tys. zł; Koszty nie są stabilne.
- 3.45. $Q_3 - Q_1 = 2$ dni; $V_Q = 9,1\%$; Czas realizacji zamówień był stabilny.

3.5. Pytania testowe z zakresu miar dyspersji

1. Kiedy warto stosować odchylenie ćwiartkowe?
 - a) Gdy chcemy uniknąć wpływu wartości ekstremalnych.
 - b) Zawsze, gdy dane są skokowe.
 - c) Gdy interesuje nas dominanta.
 - d) Gdy chcemy uwzględnić wpływ wartości ekstremalnych.
2. W jakim celu wyznacza się klasyczny współczynnik zmienności?
 - a) Pozwala on porównać rozrzut różnych cech.
 - b) Wyznacza on poziom asymetrii.
 - c) Określa on zakres zmienności danych.
 - d) Stanowi on miarę położenia.
3. Średnia masa ciała w grupie osób wynosi 70 kg, zaś odchylenie standardowe 0. Jak to interpretujemy?
 - a) Obserwujemy dużą zmienność.
 - b) Każda osoba w grupie waży dokładnie 70 kg.
 - c) Oznacza to silną asymetrię.
 - d) Średnia nie ma w tym przypadku znaczenia.
4. Współczynnik zmienności wynosi 5%. Co to oznacza dla zmiennej?
 - a) Zmienna jest skrajnie losowa.
 - b) Wariancja wynosi 5%.
 - c) Zmienna jest bardzo stabilna.
 - d) Jest to wartość średnia.
5. Współczynnik zmienności dochodów wynosi 90%. Co można powiedzieć?
 - a) Średnia dochodów to 90 zł.
 - b) Dochody są bardzo zróżnicowane.
 - c) Jest to dominanta.
 - d) Dochody są równe.
6. Wskaż zdanie fałszywe:
 - a) Im mniejsza wariancja, tym mniejsze zróżnicowanie zmiennej.
 - b) Jeżeli wszystkie wartości zmiennej są jednakowe, to jej wariancja jest równa 0.
 - c) Jeżeli wszystkie wartości zmiennej pomnożymy przez 3, to wariancja zwiększy się 9 razy.
 - d) Jeżeli do każdej wartości zmiennej dodamy 3, to wariancja zwiększy się o 3.
7. Jeśli średni czas rozładunku w magazynie to 45 min., a współczynnik zmienności wynosi 10%, oznacza to, że:
 - a) Proces jest bardzo niestabilny.
 - b) Czas rozładunku jest równy 10 min.
 - c) Odchylenie standardowe to ok. 4,5 min.
 - d) Wariancja wynosi 100.

8. Odchylenie standardowe płac pracowników przedsiębiorstwa Z wynosi 500 zł. Jeśli wszyscy badani pracownicy otrzymają premię w wysokości 100 zł każdy, to wówczas odchylenie standardowe płac będzie równe:
- 400 zł,
 - 100 zł,
 - 600 zł,
 - 500 zł.
9. Analizując wyniki egzaminu studentów w dwóch grupach, otrzymano: Grupa 1: $\bar{x}_1 = 25$ pkt., $s_1 = 5$ pkt. Grupa 2: $\bar{x}_2 = 40$ pkt., $s_2 = 8$ pkt. W której grupie zmienność wyników studentów była większa?
- W grupie pierwszej.
 - W grupie drugiej.
 - Były jednakowe.
 - Żadne z powyższych.
10. Miarą zróżnicowania jest:
- średnia arytmetyczna,
 - mediana,
 - wariancja,
 - suma.
11. Współczynnik zmienności wyraża:
- bezwzględny poziom rozproszenia,
 - zmienność w odniesieniu do średniej,
 - różnicę między maksymalną i minimalną wartością,
 - sumę kwadratów odchyleń.
12. Jeżeli wariancja wynosi 16, to odchylenie standardowe wynosi:
- 2,
 - 4,
 - 8,
 - 256.
13. W dwóch oddziałach banku odchylenie standardowe czasu obsługi wynosi odpowiednio 2 i 5 min. W którym oddziale obsługa jest bardziej stabilna (zakładając podobną średnią)?
- W pierwszym.
 - W drugim.
 - Jest taka sama.
 - Brak danych.
14. Współczynnik zmienności wyrażamy zazwyczaj:
- bez jednostek,
 - w jednostkach takich jak dane,
 - w procentach,
 - w ułamkach.

15. Rozstęp jest równy 12. Co to oznacza?
- a) Mediana wynosi 12.
 - b) Różnica między minimum a maksimum wynosi 12.
 - c) Średnia wynosi 12.
 - d) Odchylenie standardowe wynosi 12.
16. Która miara jest najbardziej wrażliwa na wartości skrajne?
- a) Odchylenie standardowe.
 - b) Mediana.
 - c) Rozstęp kwartyłowy.
 - d) Średnie odchylenie bezwzględne.
17. Odchylenie standardowe wynosi 0. W związku z tym dane są:
- a) bardzo zróżnicowane,
 - b) idealnie zgodne,
 - c) losowe,
 - d) źle policzone.
18. Stabilność kosztów produkcji przy różnych wielkościach partii najlepiej mierzy:
- a) wariancja,
 - b) średnia,
 - c) rozstęp,
 - d) współczynnik zmienności
19. W firmie kurierskiej czas dostawy waha się od 2 do 6 godzin. Rozstęp wynosi:
- a) 4,
 - b) 8,
 - c) 3,
 - d) 2.
20. Odchylenie standardowe kwoty kredytu hipotecznego wynosi 100 tys. zł, a średnia 500 tys. zł. Współczynnik zmienności wynosi zatem:
- a) 5%,
 - b) 20%,
 - c) 25%,
 - d) 50%.
21. W magazynie surowców rozstęp kwartyłowy wynosi 4 tony. Co oznacza ta wartość?
- a) 50% wartości mieści się w przedziale 4 ton.
 - b) 50% wartości mieści się w przedziale ± 4 tony od mediany.
 - c) Skrajna różnica wynosi 4 tony.
 - d) Średnie odchylenie wynosi 4 tony.
22. W logistyce stabilność czasów dostaw najlepiej ocenić za pomocą:
- a) rozstępu,
 - b) mediany,

- c) odchylenia standardowego,
 - d) średniej.
23. Jeżeli współczynnik zmienności wynosi 90%, to:
- a) dane są mało zróżnicowane,
 - b) dane są umiarkowanie stabilne,
 - c) dane są bardzo zróżnicowane,
 - d) brakuje innych potrzebnych danych.
24. Jeśli w fabryce różnice w długości elementów są bardzo niewielkie, to odchylenie standardowe będzie:
- a) duże,
 - b) małe,
 - c) zerowe,
 - d) niemożliwe do obliczenia.
25. Bank porównał zróżnicowanie wpływów klientów w zł i euro. Rzetelne porównanie umożliwia:
- a) wariancja,
 - b) odchylenie standardowe,
 - c) współczynnik zmienności,
 - d) rozstęp.
26. Dla którego zestawu danych odchylenie standardowe będzie największe?
- a) 100, 100, 100, 100;
 - b) 90, 100, 110, 100;
 - c) 80, 100, 120, 100;
 - d) 95, 100, 105, 100.
27. Która z miar zróżnicowania najlepiej nadaje się dla rozkładu silnie skośnego?
- a) Wariancja.
 - b) Rozstęp.
 - c) Odchylenie kwartylowe.
 - d) Odchylenie standardowe.
28. Które ze stwierdzeń jest prawdziwe?
- a) Wariancja jest miarą pozycyjną.
 - b) Wariancja jest sumą wartości zmiennej wszystkich jednostek badanej zbiorowości podzieloną przez liczbę tych jednostek.
 - c) Suma odchyleń poszczególnych wartości zmiennej od średniej arytmetycznej jest równa wariancji.
 - d) Jeśli wszystkie wartości cechy powiększymy (pomniejszymy) o pewną stałą, to wariancja nie ulegnie zmianie.
29. Odchylenie ćwiartkowe jest miarą zróżnicowania (zmienności, dyspersji):
- a) mieszaną,
 - b) klasyczną,
 - c) pozycyjną,
 - d) Żadna odp. nie jest poprawna.

30. Typowy obszar zmienności jest miarą:
- a) zróżnicowania,
 - b) koncentracji,
 - c) asymetrii,
 - d) przeciętną.
31. Analizując oceny z matematyki w dwóch grupach uczniów w pewnej szkole podstawowej otrzymano następujące wyniki: Grupa 1: średnia = 5, odchylenie standardowe = 0,25; Grupa 2: średnia = 4, odchylenie standardowe = 0,2. W której grupie oceny są bardziej zróżnicowane?
- a) W grupie pierwszej.
 - b) W grupie drugiej.
 - c) Są jednakowe w obu grupach.
 - d) Podano za mało informacji.
32. Odchylenie ćwiartkowe czasu oczekiwania w przychodni wynosi 5 min. Co to sugeruje?
- a) Nie było rozrzutu.
 - b) Czas oczekiwania w środkowej połowie danych różni się przeciętnie o 5 min.
 - c) Wszyscy czekają 5 min.
 - d) Czas oczekiwania wynosi dokładnie 5 min.

3.6. Odpowiedzi do pytań testowych z zakresu miar dyspersji

1. a)
2. a)
3. b)
4. c)
5. b)
6. d)
7. c)
8. d)
9. c)
10. c)
11. b)
12. b)
13. a)
14. c)
15. b)
16. a)
17. b)
18. d)
19. a)
20. b)
21. b)
22. c)
23. c)
24. b)
25. c)
26. c)
27. c)
28. d)
29. c)
30. a)
31. c)
32. b)

4. Miary asymetrii i koncentracji

4.1. Miary asymetrii (skośności)

Miary asymetrii służą do oceny stopnia i kierunku odchylenia rozkładu zmiennej od rozkładu symetrycznego. Wyróżnia się trzy podstawowe grupy miar asymetrii:

1. **Miara klasyczna** – najczęściej stosowaną klasyczną miarą asymetrii jest współczynnik asymetrii, którego wzór ogólny przedstawia się następująco:

$$A_k = \frac{\mu_3}{s^3} \quad (4.1)$$

gdzie μ_3 jest momentem centralnym rzędu trzeciego, który wyznaczany jest następująco w zależności od szeregu statystycznego:

- dla szeregu prostego:

$$\mu_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3 \quad (4.2)$$

- dla szeregu rozdzielczego punktowego:

$$\mu_3 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^3 \cdot n_i = \sum_{i=1}^k (x_i - \bar{x})^3 \cdot f_i \quad (4.3)$$

- dla szeregu rozdzielczego przedziałowego:

$$\mu_3 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^3 \cdot n_i = \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^3 \cdot f_i \quad (4.4)$$

2. **Miara pozycyjna** – opiera się na wartościach kwartyli i jest wyznaczana zgodnie z następującym wzorem:

$$A_p = \frac{Q_3 - 2Me + Q_1}{2Q} \quad (4.5)$$

3. **Miara mieszana** – łączy elementy klasyczne i pozycyjne, wyznacza się ją zgodnie z poniższym wzorem:

$$A_s = \frac{\bar{x} - D}{s} \quad (4.6)$$

Miary asymetrii są miarami względnymi (niemianowanymi). Określają zarówno kierunek, jak i siłę asymetrii, dzięki czemu mogą być wykorzystywane do porównywania asymetrii różnych rozkładów. Znak współczynnika wskazuje kierunek asymetrii, natomiast jego wartość bezwzględna informuje o jej intensywności.

Interpretacja kierunku miar asymetrii:

1. Jeśli $A_k, A_p, A_s > 0$, to rozkład cechy jest prawostronnie (dodatnio) asymetryczny, czyli ogon rozkładu ciągnie się w prawo, a większość wartości znajduje się po lewej stronie średniej.
2. Jeśli $A_k, A_p, A_s < 0$, to rozkład jest lewostronnie (ujemnie) asymetryczny, czyli ogon ciągnie się w lewo, a większość danych leży po prawej stronie średniej.
3. Jeśli $A_k, A_p, A_s = 0$, to rozkład jest symetryczny, w związku z czym dane rozmieszczone są równomiernie względem średniej.

Dla uproszczenia interpretacji można przyjąć następującą umowną klasyfikację siły asymetrii na podstawie wartości bezwzględnej współczynnika asymetrii $|A|$:

- $(0,0; 0,3]$ asymetria słaba,
- $(0,3; 0,7]$ asymetria umiarkowana,
- $(0,7; 1,0]$ asymetria silna.

Uwaga: Przy skrajnej asymetrii rozkładu może wystąpić $|A| > 1$.

Szablon interpretacyjny dla klasycznej lub mieszanej miary asymetrii może przyjąć postać:

Rozkład cechy [nazwa cechy] w zbiorowości [nazwa zbiorowości] charakteryzuje [słaba / umiarkowana / silna] asymetria [prawostronna (dodatnia) dla $A_k > 0$ / lewostronna (ujemna) dla $A_k < 0$].
Oznacza to, że [niewiele / znacznie] ponad połowa jednostek ma wartości cechy [mniejsze dla $A_k > 0$ / większe dla $A_k < 0$] od średniej.

Szablon interpretacyjny dla pozycyjnej miary asymetrii może przyjąć postać:

Rozkład cechy [nazwa cechy] w środkowej połowie zbiorowości [nazwa zbiorowości] charakteryzuje [słaba /umiarkowana/silna] asymetria [*prawostronna (dodatnia) dla $A_p > 0$ /lewostronna (ujemna) dla $A_p < 0$*].

Oznacza to, że odległość między medianą a trzecim kwartylem jest [nieco / wyraźnie] [*większa dla $A_p > 0$ / mniejsza dla $A_p < 0$*] niż między medianą a pierwszym kwartylem, co świadczy o przesunięciu środkowej części rozkładu w stronę [*wyższych dla $A_p > 0$ / niższych dla $A_p < 0$*] wartości.

Przykład 4.1 (kontynuacja przykładu 1.1)

Wracając do szeregu rozdzielczego przedstawiającego liczbę osób w zespołach, w których najczęściej pracowali studenci Politechniki Białostockiej, przejdźmy teraz do wyznaczenia miar asymetrii tego rozkładu. W poprzednich rozdziałach uzyskano już część niezbędnych parametrów: średnia liczba osób w zespole ($\bar{x}$) wyniosła 5, natomiast odchylenie standardowe (s) – 1,61 osoby. Dominanta (D) wyniosła 5 osób, mediana (Me) – 5 osób, pierwszy kwartył (Q_1) – 4 osoby, trzeci kwartył (Q_3) – 6 osób, a odchylenie ćwiartkowe (Q) – 1 osobę.

Obliczenia dotyczące momentu centralnego trzeciego rzędu dla szeregu z liczebnościami przedstawiono w tabeli 4.1, natomiast dla szeregu z częstościami względnymi – w tabeli 4.2.

Tabela 4.1. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności

x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3 \cdot n_i$
2	3	-3	-81
3	4	-2	-32
4	7	-1	-7
5	12	0	0
6	8	1	8
7	3	2	24
8	2	3	54
9	1	4	64
Suma	40		30

Źródło: wyliczenia na podstawie danych umownych.

Tabela 4.2. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości

x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3 \cdot f_i$
2	0,075	-3	-2,025
3	0,100	-2	-0,800
4	0,175	-1	-0,175
5	0,300	0	0,000
6	0,200	1	0,200
7	0,075	2	0,600
8	0,050	3	1,350
9	0,025	4	1,600
Suma	1,000		0,750

Źródło: wyliczenia na podstawie danych umownych.

W związku z tym, że otrzymana w ostatniej kolumnie tabeli 4.2 suma jest dodatnia, rozkład liczby osób w zespołach projektowych, w których najczęściej pracowali studenci Politechniki Białostockiej, charakteryzuje się asymetrią dodatnią, czyli prawostronną. Potwierdza to dodatni znak momentu centralnego trzeciego rzędu – wzór (4.3):

$$\mu_3 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^3 \cdot n_i = \frac{30}{40} = 0,750 \text{ os.}^3$$

Analogiczny jest także wynik momentu centralnego rzędu trzeciego z tabeli 4.2, w której wyliczono go przy wykorzystaniu wzoru z częstością (4.3):

$$\mu_3 = \sum_{i=1}^k (x_i - \bar{x})^3 \cdot f_i = 0,750 \text{ os.}^3$$

W celu określenia siły tej asymetrii należy wartość momentu centralnego rzędu trzeciego podzielić przez sześcian odchylenia standardowego, co pozwala wyznaczyć klasyczny współczynnik asymetrii (4.1):

$$A_k = \frac{\mu_3}{s^3} = \frac{0,750}{(1,61)^3} = 0,180$$

Otrzymana wartość współczynnika asymetrii świadczy o słabej asymetrii rozkładu liczby osób w zespołach. Oznacza to, że większość studentów pracowała w zespołach mniejszych niż przeciętna (średnia), jednak różnice te nie były znaczne, ponieważ siła asymetrii jest niewielka.

Korzystając z wcześniej wyznaczonej dominanty liczby osób w zespole, wynoszącej 5 osób, można również obliczyć mieszany współczynnik asymetrii (4.6):

$$A_s = \frac{\bar{x} - D}{s} = \frac{5 - 5}{1,61} = 0$$

W tym przypadku otrzymany wynik jest zerowy, co wskazuje na rozkład symetryczny.

Pozostaje jeszcze sprawdzić asymetrię w środkowej części rozkładu, czyli pomiędzy pierwszym a trzecim kwartylem – wzór (4.5):

$$A_p = \frac{Q_3 - 2Me + Q_1}{2Q} = \frac{6 - 2 \cdot 5 + 4}{2 \cdot 1} = 0$$

Otrzymana zerowa wartość współczynnika pokazuje, że w środkowej części rozkładu liczba osób w zespołach projektowych cechuje się symetrią.

Przykład 4.2 (kontynuacja przykładu 1.2)

Biorąc pod uwagę stawki godzinowe brutto pracowników działu logistyki oraz wykorzystując wcześniej obliczone miary ($\bar{x} = 15,1$ zł, $s = 2,72$ zł, $D = 14,67$ zł, $Q_1 = 12,96$ zł, $Me = 14,93$ zł, $Q_3 = 17,05$ zł, $Q = 2,05$ zł), wyznaczmy miary asymetrii. Wyliczenia momentu centralnego rzędu trzeciego dla liczebności przedstawiono w tabeli 4.3, a dla częstości 4.4.

Tabela 4.3. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności

$x_{0i} - x_{1i}$	n_i	$\dot{x}_i$	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^3 \cdot n_i$
10-12	14	11	-4,1	-964,89
12-14	23	13	-2,1	-213,00
14-16	28	15	-0,1	-0,03
16-18	18	17	1,9	123,46
18-20	13	19	3,9	771,15
20-22	4	21	5,9	821,52
Suma	100			538,21

Źródło: wyliczenia na podstawie danych umownych.

Tabela 4.4. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości

$x_{0i} - x_{1i}$	f_i	$\overset{\circ}{x}_i$	$\overset{\circ}{x}_i - \bar{x}$	$(\overset{\circ}{x}_i - \bar{x})^3 \cdot f_i$
10-12	0,14	11	-4,1	-9,6489
12-14	0,23	13	-2,1	-2,1300
14-16	0,28	15	-0,1	-0,0003
16-18	0,18	17	1,9	1,2346
18-20	0,13	19	3,9	7,7115
20-22	0,04	21	5,9	8,2152
Suma	1,00			5,3821

Źródło: wyliczenia na podstawie danych umownych.

W związku z tym, że suma wartości w ostatniej kolumnie tabeli 4.3 jest większa od zera, możemy wnioskować, że rozkład stawek godzinowych brutto pracowników działu logistyki wykazuje asymetrię dodatnią, czyli prawostronną. Potwierdza to dodatnia wartość trzeciego momentu centralnego, obliczonego zgodnie ze wzorem (4.4):

$$\mu_3 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^3 \cdot n_i = \frac{538,21}{100} = 5,3821 \text{ zł}^3$$

Ten sam wynik uzyskano przy zastosowaniu wzoru z częstościami:

$$\mu_3 = \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^3 \cdot f_i = 5,3821 \text{ zł}^3$$

W celu oceny natężenia tej asymetrii dzielimy wartość trzeciego momentu centralnego przez sześcian odchylenia standardowego. W ten sposób obliczamy klasyczny współczynnik asymetrii, zgodnie ze wzorem (4.1):

$$A_k = \frac{\mu_3}{s^3} = \frac{5,3821}{(2,72)^3} = 0,267$$

Tak uzyskany wynik świadczy o słabej prawostronnej asymetrii rozkładu – oznacza to, że ponad połowa pracowników otrzymuje stawki niższe niż średnia, jednak różnice te nie są znaczne, ponieważ siła asymetrii pozostaje niewielka.

Na podstawie wcześniej wyznaczonej dominanty stawek godzinowych możemy obliczyć również mieszany współczynnik asymetrii, zgodnie ze wzorem (4.6):

$$A_s = \frac{\bar{x} - D}{s} = \frac{15,1 - 14,67}{2,72} = 0,158$$

Powyższy wynik sugeruje, że rozkład cechuje słaba asymetria prawostronna, czyli niewiele ponad połowa pracowników otrzymuje stawki niższe niż średnia wynosząca 15,10 zł.

Dodatkowo, w celu zbadania asymetrii rozkładu w jego środkowej części (czyli pomiędzy pierwszym a trzecim kwartyłem), wykorzystano wzór (4.5):

Otrzymana wartość potwierdza symetrię w środkowym zakresie rozkładu stawek – oznacza to, że połowa pracowników otrzymuje wynagrodzenie skupione równomiernie wokół mediany:

$$A_p = \frac{Q_3 - 2Me + Q_1}{2Q} = \frac{17,05 - 2 \cdot 14,93 + 12,96}{2 \cdot 2,05} = 0,037$$

Wartość współczynnika pozycyjnego asymetrii jest dodatnia, lecz bardzo niska, co wskazuje na minimalną przewagę wyższych wartości w środkowym zakresie rozkładu. Oznacza to, że mediana znajduje się nieco poniżej środka rozstępu kwartyłowego. W praktyce jednak rozkład można uznać za bliski symetrii w jego środkowej połowie.

4.2. Miary koncentracji

Miary koncentracji służą do oceny stopnia skupienia wartości cechy w obrębie zbiorowości, najczęściej wokół wartości przeciętnej, lub do określania udziału poszczególnych jednostek w ogólnej wartości cechy. W praktyce statystycznej znajdują zastosowanie m.in. w analizie nierówności dochodów, koncentracji majątku, monopolizacji rynku, a także w badaniach rozkładów produkcji czy konsumpcji.

Najpierw przyjrzyjmy się ocenie stopnia skupienia wartości wokół średniej arytmetycznej. Tak rozumianą koncentrację porównuje się zazwyczaj z rozkładem normalnym, który stanowi punkt odniesienia w analizie kształtu rozkładu.

Miary wykorzystywane do opisu tego rodzaju koncentracji to:

1. **Kurtoza** – opisuje stopień skupienia wartości wokół średniej, wyznaczana jest następująco:

$$K = \frac{\mu_4}{s^4} \quad (4.7)$$

gdzie: μ_4 to moment centralny czwartego rzędu, który wyznaczany jest następująco w zależności od szeregu statystycznego:

- dla szeregu prostego:

$$\mu_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4 \quad (4.8)$$

- dla szeregu rozdzielczego punktowego:

$$\mu_4 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^4 \cdot n_i = \sum_{i=1}^k (x_i - \bar{x})^4 \cdot f_i \quad (4.9)$$

- dla szeregu rozdzielczego przedziałowego:

$$\mu_4 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^4 \cdot n_i = \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^4 \cdot f_i \quad (4.10)$$

2. **Eksces** – wartość kurtozy pomniejszona o kurtozę rozkładu normalnego $K = 3$. Wskazuje, czy rozkład jest bardziej, czy mniej skupiony wokół średniej niż rozkład normalny. Wyznaczany jest następująco:

$$K' = K - 3 \quad (4.11)$$

Jeśli:

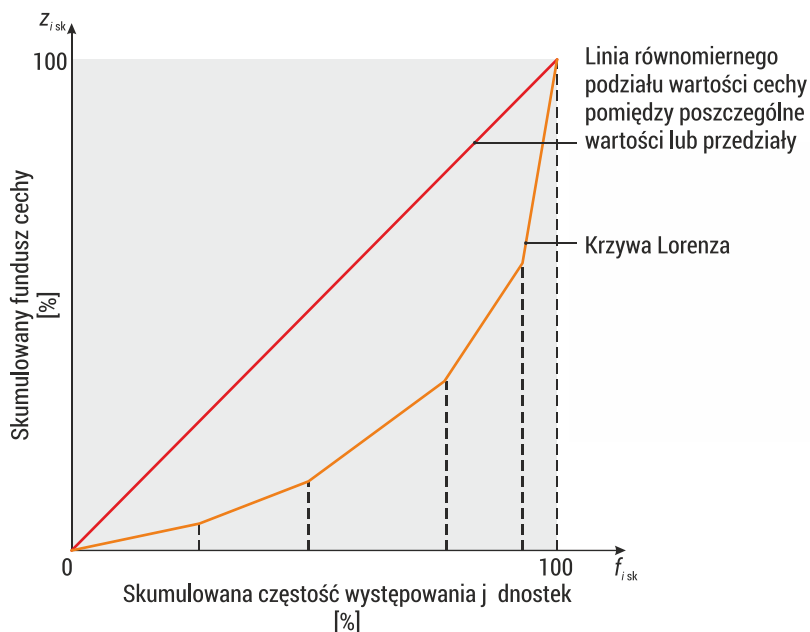
- $K' > 0$ – rozkład, nazywany leptokurtycznym (wysmukłym), jest węższy i wyższy w porównaniu z rozkładem normalnym w środkowej części i jednocześnie wydłużony na brzegach wykresu;
- $K' < 0$ – rozkład, nazywany platokurtycznym (spłaszczonym), jest spłaszczony w porównaniu z rozkładem normalnym w środkowej części (niższa i szersza środkowa część), ale zawężony (krótszy) w „ogonach” wykresu;
- $K' = 0$ – skupienie wartości wokół średniej jest takie jak w rozkładzie normalnym.

Miary takie jak kurtoza i eksces pozwalają ocenić, na ile wartości cechy skupiają się wokół średniej w rozkładzie, odnosząc się do rozkładu normalnego jako punktu odniesienia. W wielu zastosowaniach praktycznych istotniejsze jest jednak zbadanie, jak wartości danej cechy są rozdzielone między jednostki zbiorowości, niezależnie od kształtu rozkładu. Tego rodzaju koncentrację, czyli nierównomierność podziału całkowitej sumy wartości cechy, można mierzyć za pomocą krzywej Lorenza i współczynnika Lorenza.

Krzywa Lorenza służy do graficznej oceny stopnia koncentracji funduszu cechy w zbiorowości, czyli pokazuje, jak dana suma (np. fundusz stypendialny, dochód, produkcja) jest rozdzielona pomiędzy jednostki populacji. Aby ją sporządzić, tworzy się dwa szeregi skumulowane:

- skumulowaną częstość jednostek statystycznych (oznaczony jako $f_{i\ sk}$), odkładaną na osi OX ;
- skumulowane procenty funduszu cechy przypadające tym jednostkom (oznaczony jako $z_{i\ sk}$), odkładane na osi OY .

Jeśli każda jednostka otrzymuje taką samą wartość cechy, krzywa Lorenza pokrywa się z przekątną kwadratu (czyli ilustruje brak koncentracji). Jeśli zaś cały fundusz przypada jednej jednostce, otrzymujemy koncentrację absolutną. W większości przypadków krzywa Lorenza przebiega pomiędzy tymi dwiema skrajnymi sytuacjami – wtedy mówimy o koncentracji pośredniej.



Wykres 4.1. Krzywa Lorenza

Źródło: opracowanie własne.

Miara koncentracji Lorenza K_L opiera się na polu powierzchni między krzywą Lorenza a przekątną kwadratu. Oblicza się ją według wzoru:

$$K_L = \frac{P_k}{5000} \quad (4.12)$$

gdzie P_k to pole między krzywą Lorenza a przekątną (pole koncentracji), zaś 5000 to połowa pola kwadratu koncentracji.

W związku z tym, że w praktyce dysponujemy najczęściej krzywą łamaną, a nie funkcją analityczną, pole pod krzywą Lorenza obliczamy przybliżonym sposobem, np. metodą trapezów, korzystając ze wzoru:

$$K_L = \frac{5000 - P}{5000} \quad (4.13)$$

gdzie P to suma pól trapezów pod krzywą Lorenza.

Interpretacja współczynnika Lorenza:

- Jeśli $K_L = 0$, oznacza to brak koncentracji – każda jednostka otrzymuje dokładnie tyle samo.
- Jeśli $K_L = 1$, oznacza to koncentrację pełną (absolutną) – cały fundusz znajduje się w rękach jednej jednostki.
- Im bardziej krzywa Lorenza odchyła się od przekątnej, tym większa nierówność w podziale wartości cechy między jednostki.

Krzywa Lorenza i współczynnik koncentracji są często używane do badania tego, czy wartości takie jak dochód, majątek, produkcja czy sprzedaż są równo rozdysponowane, czy też skupiają się w rękach nielicznych (np. kilku osób, firm lub regionów).

Przykład 4.3 (kontynuacja przykładu 1.1)

Wracając do szeregu rozdzielczego przedstawiającego liczbę osób w zespołach, w których najczęściej pracowali studenci Politechniki Białostockiej, przejdźmy teraz do wyznaczenia miar koncentracji tego rozkładu. W poprzednich rozdziałach uzyskano już część niezbędnych parametrów, a mianowicie: $\bar{x} = 5$ osób, $s = 1,61$ osoby.

Wyliczenia momentu centralnego rzędu czwartego dla liczebności przedstawiono w tabeli 4.5, a dla częstości 4.6.

Tabela 4.5. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności

x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^4 \cdot n_i$
2	3	-3	243
3	4	-2	64
4	7	-1	7
5	12	0	0
6	8	1	8
7	3	2	48
8	2	3	162
9	1	4	256
Suma	40		788

Źródło: wyliczenia na podstawie danych umownych.

Tabela 4.6. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości

x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^4 \cdot f_i$
2	0,075	-3	6,075
3	0,100	-2	1,600
4	0,175	-1	0,175
5	0,300	0	0,000
6	0,200	1	0,200
7	0,075	2	1,200
8	0,050	3	4,050
9	0,025	4	6,400
Suma	1,000		19,700

Źródło: wyliczenia na podstawie danych umownych.

Podstawiając sumę z ostatniej kolumny tabeli 4.5 do wzoru (4.9), otrzymujemy:

$$\mu_4 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^4 \cdot n_i = \frac{788}{40} = 19,7 \text{ os.}^4$$

Ten sam wynik uzyskano przy zastosowaniu wzoru z częstościami:

$$\mu_4 = \sum_{i=1}^k (x_i - \bar{x})^4 \cdot f_i = 19,7 \text{ os.}^4$$

W celu obliczenia kurtozy należy otrzymany wynik zestandaryzować, czyli podzielić przez odchylenie standardowe w czwartej potęgze, korzystając ze wzoru (4.7):

$$K = \frac{\mu_4}{s^4} = \frac{19,7}{(1,61)^4} = 2,932$$

Otrzymany wynik może być jeszcze ponadto przekształcony na eksces za pomocą wzoru (4.11):

$$K' = K - 3 = 2,932 - 3 = -0,068$$

Wartość ekscesu wynosi $-0,068$ i oznacza, że rozkład liczby osób w zespołach projektowych, w których pracowali studenci, jest nieznacznie bardziej spłaszczony niż rozkład normalny. Oznacza to, że wartości tej cechy są nieco mniej skoncentrowane wokół średniej w porównaniu z rozkładem normalnym.

Obliczmy teraz poziom koncentracji tej cechy w rozumieniu miary Lorenza. Dane dotyczące liczby osób w zespołach oraz odpowiadające im liczebności przedstawiono w postaci szeregu punktowego, co oznacza, że łączny fundusz cechy jest sumą iloczynów liczby osób i odpowiadających im liczebności. Mając te iloczyny, można wyznaczyć procentowe udziały jednostek oraz procentowe udziały funduszu cechy w całej zbiorowości, a także ich wartości skumulowane. Wyliczenia te zostały przedstawione w tabeli 4.7.

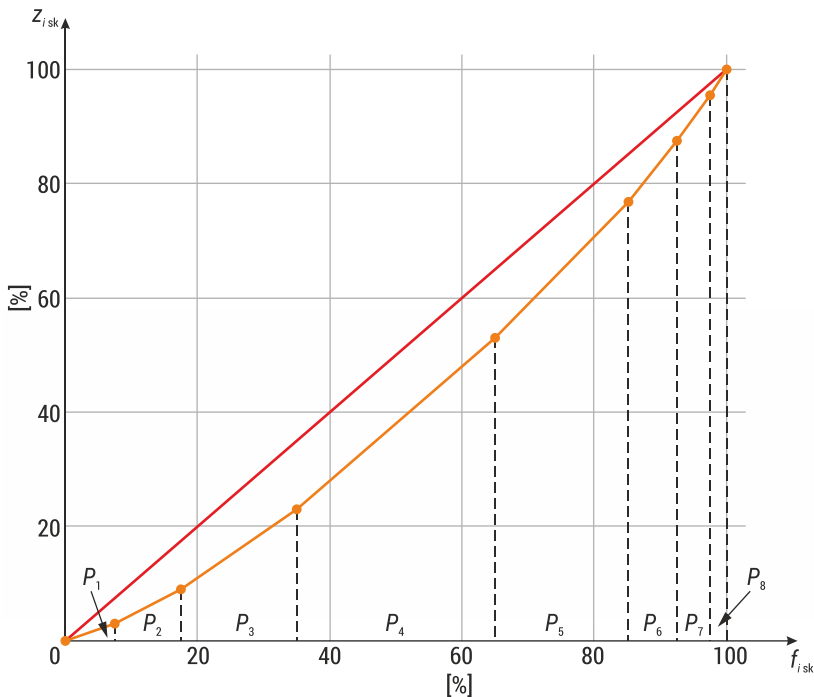
Tabela 4.7. Wyliczenia pomocnicze do miary Lorenza dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci

x_i	n_i	$x_i \cdot n_i$	f_i [w %]	z_i [w %]	f_{isk} [w %]	z_{isk} [w %]
2	3	6	7,5	3,0	7,5	3,0
3	4	12	10,0	6,0	17,5	9,0
4	7	28	17,5	14,0	35,0	23,0
5	12	60	30,0	30,0	65,0	53,0
6	8	48	20,0	24,0	85,0	77,0
7	3	21	7,5	10,5	92,5	87,5
8	2	16	5,0	8,0	97,5	95,5
9	1	9	2,5	4,5	100,0	100,0
Suma	40	200	100,0	100,0		

Źródło: wyliczenia na podstawie danych umownych.

Dwie ostatnie kolumny tabeli (4.7) to współrzędne punktów, które po połączeniu tworzą krzywą Lorenza.

Analizując wykres 4.2, można zauważyć, że krzywa Lorenza odchyła się w tym przypadku zauważalnie od przekątnej, co oznacza, że fundusz cechy (łączna liczba osób w zespołach) jest częściowo skoncentrowany w określonych wartościach liczebności.



Wykres 4.2. Krzywa Lorenza dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci

Źródło: opracowanie własne na podstawie danych umownych.

Obliczmy teraz pole pod tą krzywą (przybliżoną metodą trapezów):

$$P_1 = \frac{1}{2} \cdot 3 \cdot 7,5 = 11,250$$

$$P_2 = \frac{1}{2} \cdot (3 + 9) \cdot 10 = 60,000$$

$$P_3 = \frac{1}{2} \cdot (9 + 23) \cdot 17,5 = 280,000$$

$$P_4 = \frac{1}{2} \cdot (23 + 53) \cdot 30 = 1140,000$$

$$P_5 = \frac{1}{2} \cdot (53 + 77) \cdot 20 = 1300,000$$

$$P_6 = \frac{1}{2} \cdot (77 + 87,5) \cdot 7,5 = 616,875$$

$$P_7 = \frac{1}{2} \cdot (87,5 + 95,5) \cdot 5 = 457,500$$

$$P_8 = \frac{1}{2} \cdot (95,5 + 100,0) \cdot 2,5 = 244,375$$

W związku z powyższym pole utworzonej figury pod krzywą Lorenza wynosi:

$$P = \sum_{i=1}^8 P_i = 4110$$

Miara Lorenza wyznaczona ze wzoru (4.13) wynosi natomiast:

$$K_L = \frac{5000 - 4110}{5000} = 0,178$$

Otrzymany wynik wskazuje na niewielki poziom koncentracji. Oznacza to, że liczba osób w zespołach projektowych, w których pracowali studenci Politechniki Białostockiej, nie jest rozłożona całkowicie równomiernie – część zespołów jest liczniejsza niż pozostałe.

Przykład 4.4 (kontynuacja przykładu 1.2)

Biorąc pod uwagę stawki godzinowe brutto pracowników działu logistyki i obliczone już miary ($\bar{x} = 15,1$ zł, $s = 2,72$ zł), wyznaczmy kurtozę i eksces. Wyliczenia momentu centralnego rzędu czwartego dla liczebności przedstawiono w tabeli 4.8, a dla częstości 4.9.

Tabela 4.8. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności

$x_{0i} - x_{1i}$	n_i	$\overset{\circ}{x}_i$	$\overset{\circ}{x}_i - \bar{x}$	$(\overset{\circ}{x}_i - \bar{x})^4 \cdot n_i$
10-12	14	11	-4,1	3956,07
12-14	23	13	-2,1	447,31
14-16	28	15	-0,1	0,00
16-18	18	17	1,9	234,58
18-20	13	19	3,9	3007,47
20-22	4	21	5,9	4846,94
Suma	100			12 492,38

Źródło: wyliczenia na podstawie danych umownych.

Tabela 4.9. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości

$x_{0i} - x_{1i}$	f_i	$\overset{\circ}{x}_i$	$\overset{\circ}{x}_i - \bar{x}$	$(\overset{\circ}{x}_i - \bar{x})^4 \cdot f_i$
10-12	0,14	11	-4,1	3956,07
12-14	0,23	13	-2,1	4,4731
14-16	0,28	15	-0,1	0,0000
16-18	0,18	17	1,9	2,3458
18-20	0,13	19	3,9	30,0747
20-22	0,04	21	5,9	48,4694
Suma	1,00			124,9238

Źródło: wyliczenia na podstawie danych umownych.

Po podstawieniu sumy z ostatniej kolumny tabeli 4.8 do wzoru (4.10) otrzymujemy wartość czwartego momentu centralnego:

$$\mu_4 = \frac{1}{N} \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^4 \cdot n_i = \frac{12\,492,38}{100} = 124,9238 \text{ zł}^4$$

Taką samą wartość uzyskano również przy obliczeniach opartych na częstościach:

$$\mu_4 = \sum_{i=1}^k (\overset{\circ}{x}_i - \bar{x})^4 \cdot f_i = 124,9238 \text{ zł}^4$$

W celu wyznaczenia kurtozy należy podzielić otrzymany moment czwartego rzędu przez odchylenie standardowe podniesione do czwartej potęgi, zgodnie ze wzorem (4.7):

$$K = \frac{\mu_4}{s^4} = \frac{124,9238}{(2,72)^4} = 2,282$$

Kurtoza ta może zostać przekształcona na eksces, czyli odchylenie względem rozkładu normalnego, przy wykorzystaniu wzoru (4.11):

$$K' = K - 3 = 2,282 - 3 = -0,718$$

Rozkład stawek godzinowych brutto wśród pracowników działu logistyki jest wyraźnie bardziej spłaszczony niż rozkład normalny. Oznacza to, że dane są znacznie mniej skoncentrowane wokół średniej niż w rozkładzie normalnym. W praktyce może to świadczyć o większym zróżnicowaniu stawek i braku wyraźnego skupienia wokół jednej typowej wartości.

Obliczmy teraz poziom koncentracji stawek godzinowych brutto pracowników działu logistyki, posługując się miarą Lorenza.

Wyznaczenie miary Lorenza wymaga wyznaczenia funduszu cechy dla każdego przedziału, czyli łącznej kwoty stawek godzinowych przypadających na pracowników z każdego przedziału. W analizowanym przykładzie był on

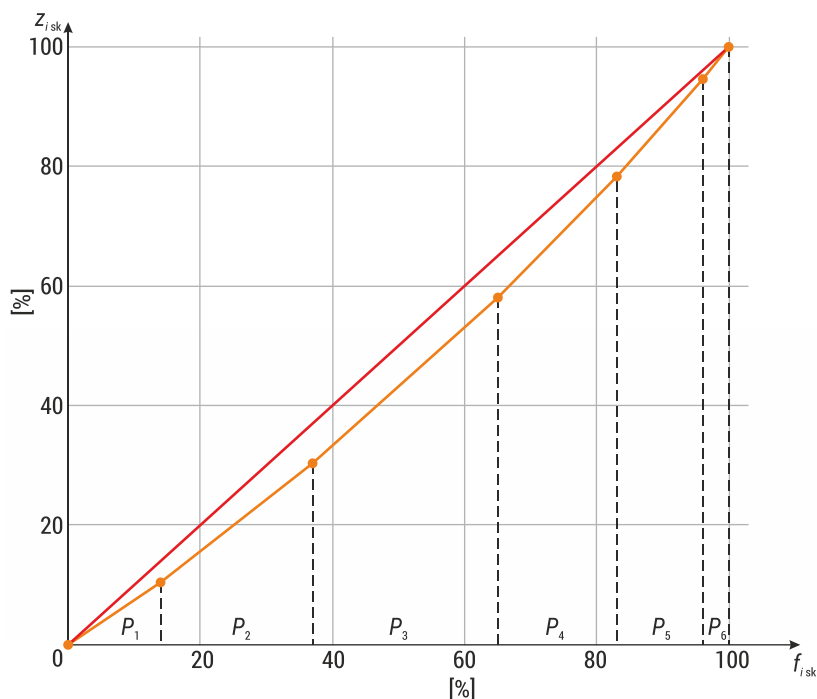
obliczany jako element badania poprawności grupowania (tabela 1.5). Następnie obliczono procentowe udziały liczebności i funduszu cechy w całej zbiorowości. Wyniki przedstawiono w tabeli 4.10.

Tabela 4.10. Wyliczenia pomocnicze do miary Lorenza dla rozkładu stawki godzinowej brutto pracowników działu logistyki

$x_{0i} - x_{1i}$	n_i	$\sum x_i$	f_i [w %]	z_i [w %]	f_{isk} [w %]	z_{isk} [w %]
10–12	14	158,06	14	10,4	14	10,4
12–14	23	301,07	23	19,9	37	30,3
14–16	28	419,72	28	27,7	65	58,0
16–18	18	305,64	18	20,2	83	78,2
18–20	13	247,26	13	16,3	96	94,5
20–22	4	83,20	4	5,5	100	100,0
Suma	100	1 514,95	100	100,0		

Źródło: wyliczenia na podstawie danych umownych.

Na podstawie skumulowanych udziałów liczebności i funduszu cechy wyznaczono krzywą Lorenza (wykres 4.3).



Wykres 4.3. Krzywa Lorenza dla rozkładu stawki godzinowej brutto pracowników działu logistyki

Źródło: opracowanie własne na podstawie danych umownych.

Na podstawie skumulowanych udziałów liczebności i funduszu cechy wyznaczono krzywą Lorenza, a następnie obliczono metodą trapezów pole pod tą krzywą:

$$P_1 = \frac{1}{2} \cdot 10,4 \cdot 14 = 72,80$$

$$P_2 = \frac{1}{2} \cdot (10,4 + 30,3) \cdot 23 = 468,05$$

$$P_3 = \frac{1}{2} \cdot (30,3 + 58,0) \cdot 28 = 1236,20$$

$$P_4 = \frac{1}{2} \cdot (58,0 + 78,2) \cdot 18 = 1225,80$$

$$P_5 = \frac{1}{2} \cdot (78,2 + 94,5) \cdot 13 = 1122,55$$

$$P_6 = \frac{1}{2} \cdot (94,5 + 100) \cdot 4 = 389,00$$

W związku z powyższym pole utworzonej figury pod krzywą Lorenza wynosi:

$$P = \sum_{i=1}^8 P_i = 4514,40$$

Miara Lorenza wyznaczona ze wzoru (4.13) wynosi natomiast:

$$K_L = \frac{5000 - 4514,40}{5000} = 0,097$$

Wynik ten wskazuje, że stawki godzinowe brutto wśród pracowników działu logistyki są rozłożone stosunkowo równomiernie i nie występuje skupienie funduszu cechy w jednym lub kilku przedziałach. Inaczej mówiąc, większość pracowników otrzymuje podobne stawki, a zróżnicowanie poziomu wynagrodzeń jest niewielkie.

4.3. Zadania z zakresu miar asymetrii i koncentracji

- 4.1. W poniższej tabeli przedstawiono dane dotyczące liczby projektów realizowanych przez małe firmy zlokalizowane w jednej ze stref przemysłowych.

Liczba projektów	Liczba przedsiębiorstw
0	20
1	30
2	26
3	15
4	8
5	1

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.2. Mając dane dotyczące wieku grupy pracowników pewnego zakładu produkcyjnego przedstawione w poniższej tabeli, wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

Wiek [w latach]	Liczba pracowników
20–25	10
25–30	20
30–35	110
35–40	35
40–45	20
45–50	5

- 4.3. W tabeli poniżej zaprezentowano rozkład liczby pracowników zatrudnionych w przedsiębiorstwach działających w sektorze przetwórstwa spożywczego.

Liczba pracowników	Odsetek firm [w %]
50–150	0,01
150–250	0,12
250–350	0,19
350–450	0,33
450–550	0,25
550–650	0,10

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.5. Poniższe dane dotyczą średniej wielkości produkcji w poszczególnych zakładach.

Wielkość produkcji [w tys. szt.]	Liczba zakładów
20–40	9
40–60	24
60–80	18
80–100	16
100–120	13

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.6. Dane zawarte w poniższej tabeli przedstawiają koszty projektów (w mln zł) zrealizowanych przez różne przedsiębiorstwa w danym roku kalendarzowym.

Koszt projektu [w mln zł]	Liczba firm
0,05–0,15	2
0,15–0,25	10
0,25–0,35	24
0,35–0,45	26
0,45–0,55	18

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.7. Pobrano próbę losową złożoną ze 100 studentów studiów dziennych na pewnym uniwersytecie. W tabeli poniżej przedstawiono ich rozkład według czasu dojazdu od miejsca zamieszkania do uczelni.

Czas dojazdu [w min.]	Liczba studentów
5–15	6
15–25	34
25–35	30
35–45	20
45–55	10

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.8. Jaka jest asymetria rozkładu czasu poświęcanego na naukę uczniów klasy czwartej, jeżeli dominanta wyniosła 3,5 godz., odchylenie standardowe 1,7 godz., zaś średnia 3,8 godz.?
- 4.9. Analizie poddano zarobki pracowników jednej z firm. Moment centralny rzędu drugiego wyniósł $\mu_2 = 441$, rzędu trzeciego $\mu_3 = -1868$, zaś czwartego $\mu_4 = 689\,345$. Wyznacz i zinterpretuj asymetrię i koncentrację zarobków pracowników analizowanej firmy.
- 4.10. Przeanalizowano wiek klientów pewnego biura turystycznego. Wyznaczono momenty centralne rzędu drugiego ($\mu_2 = 144$), trzeciego

($\mu_3 = 1421$) i czwartego ($\mu_4 = 82\ 901$). Wyznacz i zinterpretuj miarę asymetrii i koncentracji wieku klientów tego biura turystycznego.

- 4.11.** Jaka jest asymetria rozkładu wydatków na komórkę uczniów jednej ze szkół, jeżeli dominanta wyniosła 64 zł, odchylenie standardowe 17 zł, zaś średnia 57 zł?
- 4.12.** Zbadano dwie drużyny koszykówki pod względem wzrostu ich członków. Średnie wzrostu obu drużyn były identyczne. W pierwszej współczynnik skośności wyniósł 0,52, zaś w drugiej 0,45. Którą drużynę koszykówki (biorąc pod uwagę jedynie wzrost) powinien wybrać trener? Odpowiedź uzasadnij.
- 4.13.** Zbadano zawartość oleju palmowego w kremie czekoladowym pewnej marki, wytwarzanym przez dwie różne fabryki. Otrzymany rozkład zawartości tego składnika dla pierwszego wytwórcy cechuje umiarkowana asymetria lewostronna, zaś dla drugiego – prawostronna. Mając do wyboru produkty z obu fabryk, wskaż który należy wybrać i dlaczego.
- 4.14.** W pewnym zakładzie przemysłowym zbadano dzienną produkcję pracowników:

Liczba sztuk	Liczba pracowników
80–100	10
100–120	30
120–140	40
140–160	15
160–180	5

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.15.** Czas dostawy przesyłek kurierskich (w godzinach) kształtował się następująco:

Czas [w h]	Liczba przesyłek
0–2	10
2–4	40
4–6	60
6–8	25
8–10	15

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

- 4.16.** W firmie X płaca środkowa wynosi 6400 zł, przy czym 25% pracowników zarabia poniżej 5900 zł, a 25% pracowników powyżej 7200 zł. Oceń asymetrię płac.

4.18. W poniższej tabeli przedstawiono wiek mieszkańców pewnego osiedla.

Wiek [w latach]	Liczba mieszkańców
0–20	15
21–40	30
41–60	35
61–80	20

Wyznacz miary asymetrii i koncentracji. Wyniki zinterpretuj.

4.19. W tabeli poniżej przedstawiono wydajność pracy pracowników w szt./h.

Wydajność [w szt./h]	Odsetek pracowników [w %]
10–15	15
15–20	25
20–25	35
25–30	20
30–35	5

Oblicz współczynnik skośności. Jaki to rozkład? Zbadaj koncentrację rozkładu.

4.20. Poniższa tabela przedstawia liczbę książek przeczytanych w miesiącu przez pewną grupę czytelników.

Liczba książek	Liczba czytelników
1	40
2	30
3	20
4	7
5	2
6	1

Oblicz współczynnik asymetrii pozycyjny i zinterpretuj asymetrię rozkładu. Zbadaj koncentrację rozkładu.

4.21. Wiadomo, że:

- grupa A: $Me = 40, Q_1 = 35, Q_3 = 55$;
- grupa B: $Me = 35, Q_1 = 30, Q_3 = 42$.

Oblicz współczynnik skośności dla obu grup i porównaj kierunek i siłę asymetrii.

4.22. Analizowano czas oczekiwania kierowców na załadunek (w min.) i otrzymano wyniki: średnia: 55, mediana: 50, dominanta: 45, wariancja: 100. Oblicz współczynnik skośności. Zinterpretuj wynik.

4.23. W grupie studentów badano wzrost i otrzymano wyniki: średnia wzrostu: 175 cm, dominanta: 170 cm, odchylenie standardowe: 7 cm. Oblicz współczynnik asymetrii i odpowiedz, czy większość studentów ma wzrost niższy czy wyższy od średniego.

4.24. Dana jest masa ładunków przewożonych przez ciężarówkę:

Masa [w t]	Liczba kursów
1-3	20
3-5	30
5-7	40
7-9	10

Oblicz i zinterpretuj współczynniki asymetrii. Zbadaj koncentrację rozkładu.

4.25. Zbiory ziemniaków z 1 ha (w t):

Plon [w t/ha]	Liczba gospodarstw
10-15	15
15-20	40
20-25	30
25-30	10
30-35	5

Oblicz i zinterpretuj współczynnik asymetrii. Zbadaj koncentrację rozkładu.

4.26. Wiadomo, że 25% najniższych wynagrodzeń w pewnym banku nie przekracza 5700 zł, natomiast 25% najwyższych przekracza 9600 zł, przy czym $Me = 7200$ zł. Oblicz i zinterpretuj pozycyjny współczynnik asymetrii.

4.27. W pewnej firmie różnica między średnią płacą a dominantą wynosi 300 zł na korzyść średniej, zaś wariancja = 90 000. Oblicz i zinterpretuj współczynnik asymetrii.

4.28. Badając zużycie energii w GWh przez 50 zakładów przemysłowych, otrzymano następujące wyniki:

$$\bar{x} = 4,5 \text{ GWh}$$

$$Me = 4,2 \text{ GWh}$$

$$D = 3,8 \text{ GWh}$$

$$s^2 = 1 \text{ (GWh)}^2$$

Oblicz i zinterpretuj współczynnik skośności.

4.29. W poniższej tabeli przedstawiono miesięczną produkcję mleka (w tys. l).

Wielkość produkcji [w tys. l]	Liczba mleczarni
10-20	5
20-30	15
30-40	25
40-50	35
50-60	20

Zidentyfikuj typ rozkładu i możliwy kierunek asymetrii. Zbadaj koncentrację rozkładu.

4.30. W pewnej szkole uczniowie pisali sprawdzian z matematyki i uzyskali następujące wyniki:

- średnia: 3,5
- mediana: 3,2
- dominanta: 2,8
- odchylenie standardowe: 0,9

Oblicz współczynnik skośności i zinterpretuj otrzymany wynik.

4.31. Poniższa tabela przedstawia czas trwania pobytu turystów w pewnym hotelu (w dniach).

Czas pobytu [w dniach]	Liczba osób
1–3	20
4–6	30
7–9	40
10–12	10

Oceń symetrię rozkładu. Zbadaj koncentrację rozkładu.

4.32. W pewnej spółdzielni rybackiej przeanalizowano połowy ryb (w kg), a wyniki przedstawiono w poniższej tabeli.

Ilość ryb [w kg]	Liczba jednostek
0–100	5
100–200	10
200–300	20
300–400	10
400–500	5

Oblicz i zinterpretuj współczynnik asymetrii. Zbadaj koncentrację rozkładu.

4.33. Poniższa tabela przedstawia dane liczbowe obrazujące strukturę hoteli ze względu na liczbę pokoi.

Liczba pokoi	Odsetek hoteli [w %]	Łączna liczba pokoi
5–20	30	4500
20–40	35	14 000
40–80	20	18 000
80–120	10	10 000
120–200	5	10 500

Zbadaj stopień koncentracji liczby pokoi w badanych hotelach. W tym celu należy wykorzystać:

- kurtozę,
- miarę Lorenza.

4.34. Poniższa tabela przedstawia wydatki losowo wybranej grupy mieszkańców Białegostoku na konsumpcję.

Kwota wydatków [w tys. zł]	Odsetek miesz- kańców [w %]	Łączna kwota wydatków na konsumpcję [w mln zł]
0–0,5	21	3,8
0,5–1,0	35	11,2
1,0–1,5	24	16,4
1,5–2,0	12	10,9
2,0–2,5	8	7,7

Zaprezentuj graficznie oraz wyznacz stopień koncentracji (stopień równomiernego rozmieszczenia, czyli miarę Lorenza) wydatków na konsumpcję w analizowanej zbiorowości.

4.35. Poniższa tabela przedstawia łączną liczbę otrzymanych przez studentów punktów na kolokwium ze statystyki.

Liczba punktów	Liczba studentów	Łączna liczba otrzymanych punktów
0–10	18	122
10–20	36	513
20–30	30	772
30–40	24	672
40–50	12	421

Zaprezentuj graficznie i wyznacz stopień koncentracji (miarę Lorenza) liczby punktów w zbiorowości analizowanych studentów.

4.36. Wiedząc, że 20% wszystkich mieszkańców miasta wydało 70% ogółu wydanej kwoty, graficznie i algebraicznie wyznacz koncentrację wydatków wśród mieszkańców, a wynik zinterpretuj.

4.37. Wiedząc, że 40% wszystkich mieszkańców miasta wydało 80% ogółu wydanej kwoty, graficznie i algebraicznie wyznacz koncentrację wydatków wśród mieszkańców, a wynik zinterpretuj.

4.4. Odpowiedzi do zadań z zakresu miar asymetrii i koncentracji

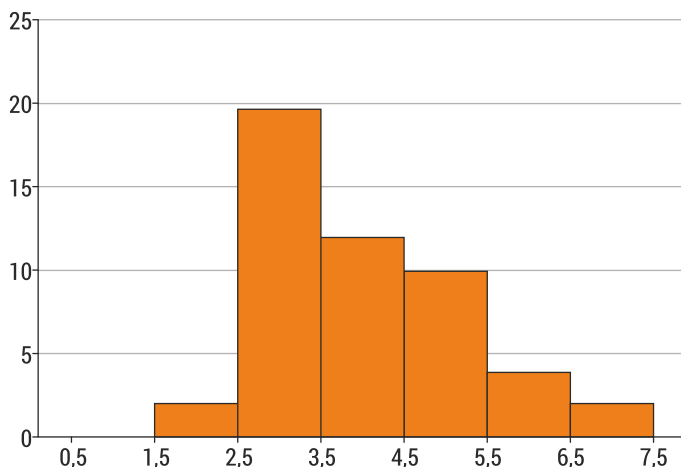
- 4.1. $A_s = 0,517$; $A_k = 0,454$; $A_p = 1$; $K = 2,464$; $K' = -0,536$
4.2. $A_s = 0,201$; $A_k = 0,337$; $A_p = 0,176$; $K = 3,670$; $K' = 0,670$
4.3. $A_s = -0,123$; $A_k = -0,196$; $A_p = -0,034$; $K = 2,393$; $K' = -0,607$
4.4. $A_s = 0,621$; $A_k = 0,148$; $A_p = 0,116$; $K = 1,914$; $K' = -1,086$
4.5. $A_s = -0,096$; $A_k = -0,332$; $A_p = -0,032$; $K = 2,418$; $K' = -0,582$
4.6. $A_s = 0,521$; $A_k = 0,307$; $A_p = 0,084$; $K = 2,290$; $K' = -0,710$
4.7. $A_k = 0,176$
4.8. $A_k = -0,202$; $K = 3,545$; $K' = 0,545$
4.9. $A_k = 0,822$; $K = 3,998$; $K' = 0,998$
4.10. $A_k = -0,412$
4.11. Drużynę drugą, gdyż w niej większość koszykarzy ma wzrost powyżej średniej.
4.12. Z fabryki drugiej, gdyż większość jej produktów ma mniejszą od średniej zawartość oleju palmowego.
4.13. $A_s = -0,036$; $A_k = 0,21$; $A_p = -0,091$; $K = 2,75$; $K' = -0,25$
4.14. $A_s = 0,098$; $A_k = 0,24$; $A_p = -0,032$; $K = 2,59$; $K' = -0,41$
4.15. $A_p = 0,23$
4.16. $A_s = -0,155$; $A_k = -0,119$; $A_p = -0,063$; $K = 2,04$; $K' = -0,96$
4.17. $A_s = -0,138$; $A_k = 0,043$; $A_p = -0,107$; $K = 2,31$; $K' = -0,69$
4.18. $A_p = 0,00$; $K = 3,82$; $K' = 0,82$
4.19. $A_{pA} = 0,50$; $A_{pB} = 0,167$
4.20. $A_s = 1$
4.21. $A_s = 0,714$; Większość studentów ma wzrost niższy od średniego.
4.22. $A_s = -0,382$; $A_k = -0,094$; $A_p = -0,143$; $K = 2,1$; $K' = -0,9$
4.23. $A_s = 0,279$; $A_k = 0,558$; $A_p = 0,118$; $K = 2,96$; $K' = -0,04$
4.24. $A_p = 0,23$
4.25. $A_s = 1$
4.26. $A_s = 0,7$
4.27. $A_s = -0,358$; $A_k = -0,429$; $A_p = -0,138$; $K = 2,41$; $K' = -0,59$
4.28. $A_s = 0,778$
4.29. $A_s = -0,094$; $A_k = -0,2$; $A_p = -0,143$; $K = 2,102$; $K' = -0,898$
4.30. $A_s = 0,0$; $A_k = 0,0$; $A_p = -0,0$; $K = 2,5$; $K' = -0,5$
4.31. $K = 5,21$; $K_L = 0,406$
4.32. $K_L = 0,305$
4.33. $K_L = 0,236$
4.34. $K_L = 0,5$
4.35. $K_L = 0,2$

4.5. Pytania testowe z zakresu miar asymetrii i koncentracji

1. Która miara pozwala na ocenę stopnia asymetrii rozkładu?
 - a) Kurtoza.
 - b) Wariancja.
 - c) Współczynnik skośności.
 - d) Mediana.
2. Co oznacza, że szereg ma asymetrię lewostronną?
 - a) Większość obserwacji ma wartości większe od średniej.
 - b) Dane są skupione wokół mediany.
 - c) Średnia, dominanta i mediana mają zbliżone wartości.
 - d) Większość obserwacji ma wartości mniejsze od średniej.
3. Która z miar wskazuje na poziom koncentracji wokół średniej?
 - a) Asymetria.
 - b) Mediana.
 - c) Kurtoza.
 - d) Kwartyle.
4. Czym jest eksces równy 0?
 - a) Oznacza rozkład o asymetrii lewostronnej.
 - b) Oznacza brak rozrzutu.
 - c) Oznacza ekstremalną wartość dyspersji.
 - d) Oznacza taką samą koncentrację jak w rozkładzie normalnym.
5. Średnia prędkość rowerzystów wynosi 20 km/h, a mediana 25 km/h. Co to sugeruje?
 - a) Prędkość jest stała.
 - b) Rozkład jest prawostronnie asymetryczny.
 - c) Rozkład jest symetryczny.
 - d) Rozkład jest lewostronnie asymetryczny.
6. Eksces dla rozkładu czasu reakcji na bodziec wyniósł 1,2. Co to oznacza?
 - a) Oznacza to silną asymetrię.
 - b) Średnia wynosi 1,2.
 - c) Rozkład ma bardziej smukły szczyt niż rozkład normalny.
 - d) Rozkład jest całkowicie płaski.
7. Współczynnik asymetrii rozkładu wieku pacjentów wynosi $-0,7$. Co to znaczy?
 - a) Oznacza to dominujący młody wiek wśród pacjentów.
 - b) Dane są całkowicie symetryczne.
 - c) Większość pacjentów to osoby starsze.
 - d) Rozkład jest asymetryczny z ogonem z lewej strony.

8. Współczynnik Lorenza dochodów w kraju wynosi 0,15. Co można w związku z tym stwierdzić?
- Dochody są stosunkowo równo rozłożone.
 - Dochód rośnie liniowo.
 - Średni dochód wynosi 1500 zł.
 - Dochody są całkowicie nierówne.
9. Rozkład ilości godzin spędzonych online jest symetryczny. Co można w związku z tym stwierdzić?
- Występuje silna asymetria.
 - Dominanta wynosi 0.
 - Średnia, mediana i dominanta są sobie równe.
 - Nie można nic stwierdzić.
10. Kurtoza czasu dostawy paczek wynosi 2. Jak zinterpretować ten wynik?
- Występuje silna asymetria.
 - Rozkład jest bardziej wysmukły niż normalny.
 - Dostawy są natychmiastowe.
 - Rozkład jest bardziej spłaszczony niż normalny.
11. W rozkładzie asymetrycznym lewostronnym (ujemnym):
- większość obserwacji przyjmuje wartości większe od średniej arytmetycznej,
 - większość obserwacji przyjmuje wartości mniejsze od średniej arytmetycznej,
 - obserwacji większych od średniej arytmetycznej jest tyle samo co obserwacji mniejszych od średniej,
 - większość zbiorowości jest bardziej skupiona wokół średniej arytmetycznej niż w rozkładzie normalnym.
12. Firma logistyczna analizuje czasy dostawy przesyłek. Jeśli mediana wynosi 24 godziny, a średnia 36 godzin, to rozkład czasów dostawy jest prawdopodobnie:
- symetryczny,
 - dwumodalny,
 - prawostronnie (dodatnie) asymetryczny,
 - lewostronnie (ujemnie) asymetryczny.
13. W rozkładzie asymetrycznym prawostronnym (dodatnim):
- większość obserwacji przyjmuje wartości większe od średniej arytmetycznej,
 - większość obserwacji przyjmuje wartości mniejsze od średniej arytmetycznej,
 - obserwacji większych od średniej arytmetycznej jest tyle samo co obserwacji mniejszych od średniej,
 - większość zbiorowości jest bardziej skupiona wokół średniej arytmetycznej niż w rozkładzie normalnym.

14. Na podstawie danych dotyczących wydajności pracy (w szt./h) 50 robotników otrzymano następujący histogram:



Rozkład wydajności pracy charakteryzuje się zatem:

- a) asymetrię ujemną (lewostronną),
 - b) asymetrię dodatnią (prawostronną),
 - c) symetrię.
 - d) Żadna odpowiedź nie jest poprawna.
15. W banku analizowano rozkład miesięcznych wpływów na rachunki klientów. Jeśli współczynnik asymetrii > 0 , oznacza to, że:
- a) większość klientów ma wpływy powyżej średniej,
 - b) rozkład jest symetryczny,
 - c) większość klientów ma wpływy poniżej średniej,
 - d) wszystkie wpływy są równe.
16. Współczynnik Lorenza dla rozkładu dochodów klientów banku wynosi 0,9. Oznacza to:
- a) bardzo wysoką nierówność dochodów,
 - b) pełną równość dochodów,
 - c) średnią koncentrację dochodów,
 - d) Brak danych do oceny.
17. Miarą koncentracji NIE jest:
- a) współczynnik Lorenza,
 - b) kurtoza,
 - c) odchylenie standardowe,
 - d) eksces.
18. W badaniu czasu dojazdu pasażerów do pracy uzyskano współczynnik asymetrii równy 2. Oznacza to, że:
- a) większość dojeżdża bardzo szybko;
 - b) większość dojeżdża w przeciętnym tempie;

- c) większość dojeżdża długo, ale są nieliczne bardzo długie dojazdy;
 - d) rozkład jest symetryczny.
- 19.** Stopień koncentracji przewozów pasażerskich w kilku największych firmach transportowych najlepiej opisze:
- a) mediana liczby pasażerów,
 - b) współczynnik Lorenza,
 - c) średnia liczba kursów,
 - d) rozstęp kwartyłowy.
- 20.** Jeżeli współczynnik asymetrii liczby opóźnień lotów jest bliski 0, oznacza to, że:
- a) opóźnienia są bardzo częste,
 - b) opóźnienia są bardzo rzadkie,
 - c) rozkład opóźnień jest symetryczny,
 - d) wszystkie loty są punktualne.
- 21.** Współczynnik Lorenza dla rynku przewozów kontenerowych wynosi 0,15. Oznacza to:
- a) wysoki stopień koncentracji,
 - b) bardzo niski stopień koncentracji,
 - c) całkowity monopol.
 - d) Nie można ocenić koncentracji.
- 22.** W badaniu czasu dostawy towarów w transporcie drogowym wykazano asymetrię dodatnią. Oznacza to, że:
- a) większość dostaw realizowana jest szybciej niż średnia,
 - b) większość dostaw jest realizowana wolniej niż średnia,
 - c) wszystkie dostawy realizowane są punktualnie,
 - d) odchylenie standardowe jest zerowe.
- 23.** W fabryce zmierzono wymiary produkowanych elementów. Współczynnik asymetrii okazał się ujemny, co oznacza, że:
- a) większość elementów ma wymiar powyżej średniej,
 - b) większość elementów ma wymiar poniżej średniej,
 - c) wszystkie elementy są identyczne,
 - d) rozkład jest równomierny.
- 24.** Współczynnik Lorenza bliski 0 dla liczby wyprodukowanych jednostek przez poszczególne linie produkcyjne oznacza:
- a) równomierne obciążenie linii,
 - b) skrajną nierówność produkcji,
 - c) koncentrację produkcji w jednej linii,
 - d) losowy rozkład produkcji.
- 25.** W firmie produkcyjnej współczynnik Lorenza wzrósł z 0,2 do 0,5. Oznacza to:
- a) większe zróżnicowanie udziałów produkcyjnych,
 - b) mniejszą koncentrację produkcji,
 - c) wzrost koncentracji na korzyść kilku działów,

- d) pełną równowagę produkcji.
- 26.** Gdy wartości odchylają się w prawo od średniej, mamy do czynienia z:
 - a) asymetrią dodatnią,
 - b) asymetrią ujemną,
 - c) brakiem asymetrii,
 - d) koncentracją 100%.
- 27.** Współczynnik asymetrii mieszany opiera się na:
 - a) medianie i średniej,
 - b) tylko na medianie,
 - c) odchyleniu standardowym,
 - d) kwartylach.
- 28.** Całkowitą koncentrację oznacza następująca wartość współczynnika Lorenza:
 - a) 1;
 - b) 0;
 - c) 0,5.
 - d) Nie ma takiej wartości.
- 29.** Jeśli rozkład cechy jest symetryczny, to:
 - a) średnia > mediana,
 - b) średnia = mediana = dominanta,
 - c) dominanta > średnia,
 - d) mediana > średnia.
- 30.** Dlaczego nie należy wyznaczać średniej arytmetycznej dla danych silnie asymetrycznych?
 - a) Ponieważ nie da się jej obliczyć.
 - b) Ponieważ zawsze jest równa medianie.
 - c) Ponieważ wymaga danych jakościowych.
 - d) Ponieważ może nie oddawać centralnego położenia danych.

4.6. Odpowiedzi do pytań testowych z zakresu miar asymetrii i koncentracji

1. c)
2. a)
3. c)
4. d)
5. d)
6. c)
7. d)
8. a)
9. c)
10. d)
11. a)
12. c)
13. b)
14. b)
15. c)
16. a)
17. c)
18. a)
19. b)
20. c)
21. b)
22. a)
23. a)
24. a)
25. b)
26. b)
27. c)
28. a)
29. b)
30. d)

Spis wykresów

Wykres 1.1. Wykres prezentujący rozkład liczby osób w zespołach projektowych, w których pracowali studenci.....	18
Wykres 1.2. Histogram prezentujący rozkład stawki godzinowej brutto pracowników działu logistyki w zł ($h = 1$)	20
Wykres 1.3. Histogram prezentujący rozkład stawki godzinowej brutto wśród pracowników działu logistyki w zł ($h = 2$)	23
Wykres 2.1. Wykres prezentujący rozkład liczby osób w zespołach projektowych, w których pracowali studenci – wyznaczenie dominanty	49
Wykres 2.2. Graficzne wyznaczenie dominanty dla rozkładu stawki godzinowej brutto pracowników działu logistyki.....	51
Wykres 2.3. Graficzne wyznaczenie kwartyli dla rozkładu stawki godzinowej brutto pracowników działu logistyki	54
Wykres 4.1. Krzywa Lorenza	116
Wykres 4.2. Krzywa Lorenza dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci.....	120
Wykres 4.3. Krzywa Lorenza dla rozkładu stawki godzinowej brutto pracowników działu logistyki	123

Spis tabel

Tabela 1.1. Przykłady badań statystycznych ze zdefiniowaną dla nich: populacją, jednostką statystyczną, cechą i jednostką pomiaru.....	9
Tabela 1.2. Rozkład liczby osób w zespołach projektowych, w których pracowali studenci.....	17
Tabela 1.3. Rozkład stawki godzinowej brutto pracowników działu logistyki ($h = 1$).....	20
Tabela 1.4. Wyliczenia związane ze sprawdzeniem poprawności grupowania danych dotyczących stawki godzinowej brutto pracowników działu logistyki ($h = 1$)	21
Tabela 1.5. Rozkład stawki godzinowej brutto pracowników działu logistyki ($h = 2$)	22
Tabela 1.6. Wyliczenia związane ze sprawdzeniem poprawności grupowania danych dotyczących stawki godzinowej brutto pracowników działu logistyki ($h = 2$)	23
Tabela 2.1. Wyliczenia średniej dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności	39
Tabela 2.2. Wyliczenia średniej dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości.....	39
Tabela 2.3. Wyliczenia średniej dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności.....	40
Tabela 2.4. Wyliczenia średniej dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości.....	40
Tabela 2.5. Wyliczenia średniej harmonicznej prędkości ciężarówek.....	41
Tabela 2.6. Wyliczenia średniej arytmetycznej prędkości ciężarówek.....	42
Tabela 2.7. Wyliczenia liczebności i częstości skumulowanej dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci.....	48

Tabela 2.8. Wyliczenia liczebności i częstości skumulowanej dla rozkładu stawki godzinowej brutto pracowników działu logistyki.....	51
Tabela 3.1. Wyliczenia wariancji (wzór podstawowy) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności	78
Tabela 3.2. Wyliczenia wariancji (wzór przekształcony) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności	78
Tabela 3.3. Wyliczenia wariancji (wzór podstawowy) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości.....	79
Tabela 3.4. Wyliczenia wariancji (wzór przekształcony) dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości.....	79
Tabela 3.5. Wyliczenia wariancji (wzór podstawowy) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności.....	81
Tabela 3.6. Wyliczenia wariancji (wzór przekształcony) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności.....	82
Tabela 3.7. Wyliczenia wariancji (wzór podstawowy) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości	82
Tabela 3.8. Wyliczenia wariancji (wzór przekształcony) dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości	83
Tabela 4.1. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności	110
Tabela 4.2. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości.....	111
Tabela 4.3. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności.....	112

Tabela 4.4. Wyliczenia momentu centralnego rzędu trzeciego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości	113
Tabela 4.5. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem liczebności.....	118
Tabela 4.6. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci, z wykorzystaniem częstości.....	118
Tabela 4.7. Wyliczenia pomocnicze do miary Lorenza dla rozkładu liczby osób w zespołach projektowych, w których pracowali studenci.....	119
Tabela 4.8. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem liczebności.....	121
Tabela 4.9. Wyliczenia momentu centralnego rzędu czwartego dla rozkładu stawki godzinowej brutto pracowników działu logistyki z wykorzystaniem częstości	122
Tabela 4.10. Wyliczenia pomocnicze do miary Lorenza dla rozkładu stawki godzinowej brutto pracowników działu logistyki	123

Literatura

- [1] Aczel A.D., Sounderpandian J., *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa 2017.
- [2] Bąk I., Markowicz I., Mojsiewicz M., *Statystyka w zadaniach, cz. I, Statystyka opisowa*, Wydawnictwo Naukowo-Techniczne, Warszawa 2001.
- [3] Józefacka N.M., Kołek M.F., Arciszewska-Leszczuk A., *Metodologia i statystyka, t. I, Przewodnik naukowego turysty*, Wydawnictwo Naukowe PWN, Warszawa 2023.
- [4] Józwiak J., Podgórski J., *Statystyka od podstaw*, PWE, Warszawa 2012.
- [5] Kassyk-Rokicka H. (red.), *Statystyka. Zbiór zadań*, PWE, Warszawa 1994 i nast. wyd.
- [6] Kassyk-Rokicka H., *Mierniki statystyczne*, PWE, Warszawa 2011.
- [7] Kowalski L., *Statystyka*, WAT, Warszawa 2021.
- [8] Lissowski G., Haman J., Jasiński M., *Podstawy statystyki dla socjologów, t. 1, Opis statystyczny*, Wydawnictwo Naukowe SCHOLAR, Warszawa 2011.
- [9] Luszniewicz A. (red.), *Statystyka w zarządzaniu*, Wyższa Szkoła Finansów i Zarządzania w Białymstoku, Białystok 2006.
- [10] Olbrych B., Rudecka-Gutowska S., *Statystyka opisowa w praktyce*, Akademia Handlowa Nauk Stosowanych w Radomiu, Radom 2024.
- [11] Ostasiewicz S., Rusnak Z., Siedlecka U., *Statystyka. Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej Oskara Langego, Wrocław 2006.
- [12] Podgórski J., *Statystyka dla studiów licencjackich*, PWE, Warszawa 2005.
- [13] Sobczyk M., *Statystyka*, Wydawnictwo Naukowe PWN, Warszawa 2010.
- [14] Tarka D., Olszewska A.M., *Elementy statystyki. Opis statystyczny*, Oficyna Wydawnicza Politechniki Białostockiej, Białystok 2018.
- [15] Wagner W., Mantaj A., *Zasady statystyki jedno- i dwuwymiarowej. Metodyka, teoria i zastosowania, t. I, Statystyka opisowa*, Wydawnictwo Oświatowe FOSZE, Rzeszów 2014.

- [16] Wasilewska E., *Statystyka opisowa nie tylko dla socjologów*, Wydawnictwo SGGW, Warszawa 2008.
- [17] Wierzbński J., *Statystyka opisowa*, Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, Warszawa 2006.
- [18] Zeliaś A., Pawełek B., Wanat S., *Metody statystyczne. Zadania i sprawdziany*, PWE, Warszawa 2002.
- [19] Zeliaś A., *Metody statystyczne*, Warszawa, PWE 2000.

