Państwowa Wyższa Szkoła Zawodowa w Koninie

Materiały dydaktyczne 17

ARTUR ZIMNY

STATYSTYKA OPISOWA

Materiały pomocnicze do ćwiczeń

wydanie drugie zmienione

Konin 2010

Tytuł
Statystyka opisowa
Materiały pomocnicze do ćwiczeń
wydanie drugie zmienione

Autor
Artur Zimny

Recenzja naukowa
dr Kazimierz Kruszka

Opracowanie redakcyjne i korekta
Maria Sierakowska

Projekt okładki
Agnieszka Jankowska

Łamanie i skład
Krzysztof Przybylak

Druk i oprawa

ISBN 978-83-88335-56-3

Wydawnictwo Państwowej Wyższej Szkoły Zawodowej w Koninie
ul. kard. S. Wyszyńskiego 3C, 62-510 Konin
tel. (063) 249-72-09
e-mail:wydawnictwo@konin.edu.pl

SPIS TREŚCI

OD AUTORA

Opracowanie, które znalazło się w Państwa rękach, jest efektem mojego kilkuletniego doświadczenia w prowadzeniu ćwiczeń ze statystyki opisowej na specjalnościach ekonomicznych w Państwowej Wyższej Szkole Zawodowej w Koninie. Jest przeznaczone przede wszystkim dla studentów specjalności ekonomicznych PWSZ, choć oczywiście mogą z niego korzystać również poszukujący wiedzy o podstawowych narzędziach statystycznych potrzebnych np. w pracy zawodowej.

Przygotowując ten podręcznik chciałem przybliżyć podstawowe metody analizy statystycznej i ich praktyczne wykorzystanie. Mimo że Czytelnik zetknie się z wieloma wzorami i formułami obliczeniowymi, to ich zastosowanie nie wymaga pogłębionej wiedzy z matematyki. Wystarczy znajomość podstawowych działań arytmetycznych i kolejności ich wykonywania. Obok niezbędnej wiedzy teoretycznej, którą przedstawiłem w możliwie zwięzły i przystępny sposób, zawarłem w opracowaniu również liczne przykłady i zadania przeznaczone zarówno do rozwiązywania podczas ćwiczeń, jak i do samodzielnej pracy. Przy ich konstruowaniu korzystałem, gdy tylko było to możliwe z dydaktycznego punktu widzenia, z danych empirycznych obrazujących aktualne problemy społeczno-ekonomiczne kraju i regionu. Źródłem tych danych były przede wszystkim publikacje Głównego Urzędu Statystycznego oraz Bank Danych Regionalnych GUS.

Opracowanie składa się z pięciu rozdziałów, a jego układ i treść są ściśle związane z tematyką ćwiczeń ze statystyki opisowej na specjalnościach ekonomicznych w PWSZ w Koninie. W rozdziale pierwszym przedstawione zostały podstawowe pojęcia i zagadnienia, których znajomość jest konieczna dla zrozumienia materiału prezentowanego w dalszej części opracowania. Rozdział drugi poświęcono metodom analizy struktury zbiorowości, omawiając przy tym podstawowe grupy miar charakteryzujących rozkład jednej zmiennej. Rozdział trzeci wprowadza Czytelnika w problematykę analizy współzależności zjawisk, czyli sposobów określania kierunku, siły oraz kształtu zależności między badanymi zjawiskami. W rozdziale czwartym natomiast głównym tematem są metody analizy dynamiki zjawisk. W podsumowaniu zawarto zestaw zadań, które pozwolą usystematyzować i utrwalić wiedzę oraz umiejętności zdobyte przez studentów w trakcie cyklu ćwiczeń ze statystyki opisowej. Z kolei dołączona bibliografia przedmiotu powinna ułatwić zainteresowanym pogłębienie wiedzy poprzez dotarcie do innych podręczników ze statystyki.

Mam nadzieję, że to opracowanie spełni swoje zadanie. Jednocześnie mając świadomość, że nie jest pozbawione błędów i nieścisłości, będę bardzo wdzięczny za ich wskazanie, co pozwoli na wprowadzenie odpowiednich udoskonaleń i modyfikacji.

Artur Zimny

1. WPROWADZENIE

1.1. Informacje ogólne

Statystyka to nauka zajmująca się metodami badania prawidłowości zachodzących w procesach masowych oraz ich ilościową lub jakościową analizą z punktu widzenia dyscypliny naukowej, w której skład procesy te wchodzą. Zadaniem statystyki jest dostarczanie wiarygodnych informacji niezbędnych do podejmowania decyzji w różnych dziedzinach.

Statystyka jako nauka dzieli się na:
- **statystykę opisową** (opis statystyczny), która zajmuje się metodami gromadzenia, opracowania i prezentacji danych wraz z ich sumarycznym opisem,
- **statystykę matematyczną** (wnioskowanie statystyczne), która zajmuje się metodami wnioskowania o całej zbiorowości na podstawie zbadania pewnej jej części, czyli próby.

1.2. Zbiorowość, jednostka i cecha statystyczna

Zbiorowość statystyczna (populacja) to zbiór jednostek (osób, przedmiotów, zdarzeń) objętych badaniem statystycznym, które mają jedną lub kilka cech wspólnych oraz wiele cech różnicujących (zmiennych). Zbiorowość statystyczna musi być jednoznacznie określona pod względem rzeczowym, przestrzennym oraz czasowym.

Jednostka statystyczna to najmniejszy element zbiorowości statystycznej objętej badaniem.

Cechy statystyczne (zmienne) to właściwości jednostek statystycznych tworzących badaną zbiorowość. Dzieli się je na cechy:
- jakościowe (niemierzalne),
- ilościowe (mierzalne)
  - skokowe (dyskretne),
  - ciągłe,
  - quasi ciągłe.

1.3. Istota i etapy badania statystycznego

Badanie statystyczne to zespół czynności zmierzających do uzyskania, za pomocą metod statystycznych, informacji charakteryzujących zbiorowość statystyczną objętą badaniem. Badanie statystyczne umożliwia wykrycie lub potwierdzenie istniejących prawidłowości statystycznych.

Etapy badania statystycznego:
- przygotowanie badania,
- obserwacja statystyczna,
- opracowanie i prezentacja materiału statystycznego,
- analiza statystyczna.

1.4. Materiał statystyczny i sposoby jego prezentacji

Materiał statystyczny to zbiór danych uzyskanych w wyniku obserwacji statystycznej. Dzieli się on na materiał:
- pierwotny, który stanowią dane specjalnie gromadzone dla celów określonego badania,
- wtórny, który stanowią dane gromadzone dla innych celów, a które podmiot badający wykorzystuje w swoim badaniu.

Spośród licznych źródeł danych statystycznych na szczególną uwagę zasługują publikacje Głównego Urzędu Statystycznego oraz urzędów statystycznych (m.in. Rocznik Statystyczny RP, Mały Rocznik Statystyczny, roczniki statystyczne województw, branżowe roczniki statystyczne i wiele innych), które rozpowszechniane są zarówno w formie tradycyjnej (książkowej), jak i elektronicznej (na dyskach CD-ROM oraz na stronie internetowej Urzędu <www.stat.gov.pl>).

Szereg statystyczny to zbiór wyników badania jednostek statystycznych przedstawiony w formie uporządkowanej lub uporządkowanej i pogrupowanej według wariantów badanej cechy zmiennej. Szeregi statystyczne dzielą się:
- ze względu na formę na:
  - szeregi proste (wyliczające),
  - szeregi rozdzielcze (strukturalne), a te na:
    - jednostopniowe (punktowe),
    - wielostopniowe (przedziałowe);
- ze względu na treść na:
  - szeregi czasowe (dynamiczne),
  - szeregi przestrzenne (geograficzne).

Tablica statystyczna służy do prezentacji zebranego materiału statystycznego za pomocą liczb. Poprawnie zbudowana tablica powinna składać się z trzech elementów:

- tytułu,
- tablicy właściwej,
- źródła danych statystycznych i ewentualnych uwag wyjaśniających.

Jeżeli niemożliwe jest wypełnienie danego miejsca w tablicy wartością liczbową, to stosuje się następujące **znaki umowne**:

- kreska (–) – zjawisko nie wystąpiło,
- zero (0) lub (0,0) – zjawisko istniało w wielkości mniejszej od 0,5 (0,05),
- kropka (.) – zupełny brak informacji albo brak informacji wiarygodnych,
- znak x – wypełnienie pozycji jest niemożliwe lub niecelowe,
- znak # – dane nie mogą być opublikowane ze względu na konieczność zachowania tajemnicy statystycznej w rozumieniu ustawy o statystyce publicznej,
- „w tym" – oznacza, że nie podaje się wszystkich składników sumy.

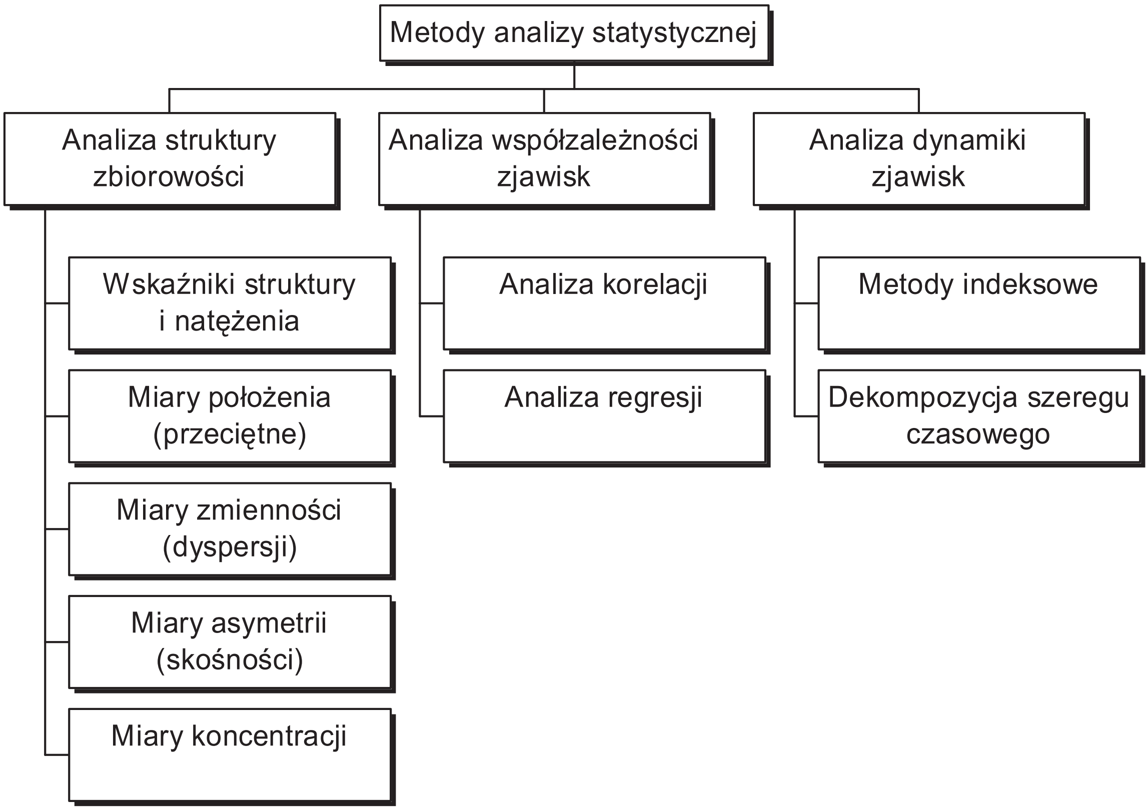
Wykres statystyczny służy do prezentacji zebranego materiału statystycznego za pomocą obrazu graficznego, tj. wielkości, kształtu lub barwy. Poprawnie wykonany wykres powinien składać się z następujących elementów:

- tytułu,
- pola wykresu,
- skali,
- legendy,
- źródła danych statystycznych i ewentualnych uwag wyjaśniających.

1.5. Metody analizy statystycznej

W ramach statystyki opisowej można wyróżnić trzy podstawowe działy analizy:

- **analizę struktury zbiorowości**, która pozwala ustalić, jak są rozłożone poszczególne warianty cechy zmiennej wśród jednostek badanej zbiorowości statystycznej,
- **analizę współzależności zjawisk**, która zajmuje się badaniem powiązań między różnymi cechami zmiennymi charakteryzującymi zbiorowość statystyczną,
- **analizę dynamiki zjawisk**, której zadaniem jest określenie zmian zachodzących w kształtowaniu się cechy zmiennej w czasie.



Schemat 1.1. Podział metod analizy statystycznej

Źródło: Opracowanie własne.

1.6. Zastosowanie programów komputerowych w statystyce

Prowadzenie badań statystycznych, a zwłaszcza opracowanie i analiza dużych zbiorów danych, wymaga wykorzystania komputerów. Zastosowanie programów komputerowych pozwala zredukować czas potrzebny na pracochłonne obliczenia do niezbędnego minimum. Upowszechnianie się metod statystycznych sprawia, że obecnie nawet podstawowe oprogramowanie zawiera elementarne procedury statystyczne. Na szczególną uwagę zasługuje arkusz kalkulacyjny Microsoft Excel, który dzięki wbudowanym funkcjom statystycznym oraz opcji „Analiza danych" może znacznie ułatwić analizę statystyczną. Wykorzystanie Excela sygnalizuje możliwości programów komputerowych do obliczeń statystycznych i powinno stanowić pierwszy krok do samodzielnego stosowania profesjonalnych pakietów statystycznych (Statgraphics, Statistica, SPSS, SAS i innych). Pakiety te, są stosunkowo proste w obsłudze, jednak wymagają pewnej wiedzy ze statystyki, aby można było poprawnie używać zawarte w nich procedury oraz interpretować uzyskane wyniki.

1.7. Służby statystyki publicznej w Polsce i Unii Europejskiej

Centralnym organem administracji rządowej w Polsce, właściwym w sprawach statystyki, jest Prezes Głównego Urzędu Statystycznego, który wykonuje swoje zadania przy pomocy służb statystyki publicznej. Zgodnie z ustawą z 29 czerwca 1995 r. o statystyce publicznej, służby statystyki publicznej stanowi Prezes Głównego Urzędu Statystycznego, podlegli mu dyrektorzy szesnastu urzędów statystycznych oraz inne jednostki statystyki[1]. Do zadań służb statystyki publicznej należy:

- rozpoznawanie zapotrzebowania na informacje i analizy statystyczne oraz przygotowywanie na tej podstawie projektów programów badań statystycznych statystyki publicznej,
- organizowanie i prowadzenie badań statystycznych oraz ustalanie ich metodologii,
- zbieranie, gromadzenie i opracowywanie danych statystycznych oraz ich analizowanie,
- przeprowadzanie spisów powszechnych,
- przechowywanie danych statystycznych,
- opracowywanie standardowych klasyfikacji, nomenklatur i definicji podstawowych kategorii, ustalanie wzajemnych relacji między nimi oraz ich interpretacja,
- udostępnianie i rozpowszechnianie wynikowych informacji statystycznych, w tym podstawowych wielkości i wskaźników,
- opracowywanie i ogłaszanie prognoz demograficznych oraz statystycznych prognoz gospodarczych i społecznych,
- przedstawianie Prezydentowi, Sejmowi i Senatowi, organom administracji rządowej, Najwyższej Izbie Kontroli, Narodowemu Bankowi Polskiemu, organom jednostek samorządu terytorialnego oraz innym instytucjom rządowym wynikowych informacji statystycznych w zakresie, terminach i formach określonych w programie badań statystycznych,
- prowadzenie krajowych rejestrów urzędowych: podmiotów gospodarki narodowej i podziału terytorialnego kraju,
- prowadzenie badań i analiz statystycznych wynikających z przyjętych przez Rzeczpospolitą Polską zobowiązań międzynarodowych,

[1] Ustawa z 29 czerwca 1995 r. o statystyce publicznej, Dz.U. z 1995 r., nr 88, poz. 439, art. 22-29.

- dokonywanie statystycznych porównań międzynarodowych i ogłaszanie ich wyników,
- wykonywanie przyjętych przez Rzeczpospolitą Polską zobowiązań przekazywania danych statystycznych organizacjom międzynarodowym,
- współpraca z wyspecjalizowanymi w dziedzinie statystyki organizacjami międzynarodowymi, regionalnymi oraz organami i urzędami innych krajów,
- prowadzenie prac naukowych i badawczo-rozwojowych w zakresie metodologii badań statystycznych i standardów klasyfikacyjnych oraz zastosowań metod matematycznych i informatyki w statystyce,
- prowadzenie szkolenia, dokształcania i doskonalenia w dziedzinie statystyki,
- popularyzacja wiedzy o statystyce.

Instytucją, która zajmuje się sprawami statystyki w Unii Europejskiej jest Eurostat (The Statistical Office of the European Communities)[2]. Urząd sporządza analizy i prognozy istotne dla podejmowania decyzji przez organy wspólnotowe oraz koordynuje i monitoruje prace narodowych urzędów statystycznych w celu unifikacji stosowanych przez nie metod badań, a także konsoliduje statystyki krajowe, państw członkowskich. Ponadto, do kompetencji Eurostatu należy analizowanie i prognozowanie tendencji rozwoju Unii Europejskiej.

1.8. Przykłady

Przykład 1.8.1.

Obserwacji poddano studentów I roku PWSZ w Koninie w dniu 15 lutego 2010 r.

<u>Zbiorowość statystyczna</u> – studenci I roku PWSZ w Koninie w dniu 15 lutego 2010 r.
<u>Jednostka statystyczna</u> – student I roku PWSZ w Koninie w dniu 15 lutego 2010 r.

[2] <www.ec.europa.eu/eurostat>

Przykłady cech statystycznych (zmiennych)

Zbiorowość statystyczna	Cecha statystyczna	Warianty cechy	Określenie cechy
studenci I roku PWSZ w Koninie w dniu 15 lutego 2010 r.	wiek	19, 20, 22, 21, 19, 23, 19 itd.	ilościowa (mierzalna), ciągła
	płeć	kobieta, mężczyzna	jakościowa (niemierzalna)
	wzrost	173, 181, 185, 179, 176, 169 itd.	ilościowa (mierzalna), ciągła
	kierunek studiów	filologia, fizjoterapia, informacja naukowa i bibliotekoznawstwo, inżynieria środowiska, mechanika i budowa maszyn, pedagogika, pielęgniarstwo, politologia, praca socjalna, turystyka i rekreacja, wychowanie fizyczne, zarządzanie	jakościowa (niemierzalna)
	liczba rodzeństwa	0, 2, 3, 1, 2, 0 itd.	ilościowa (mierzalna), skokowa

Źródło: Dane umowne.

Przykład 1.8.2.

Obserwacji poddano 30 pracowników konińskiej firmy „Zet” biorąc pod uwagę liczbę posiadanych dzieci. Wyniki obserwacji przedstawiają się następująco (stan z 1 lutego 2010 r.)

3, 1, 0, 2, 3, 0, 5, 1, 2, 5, 4, 5, 1, 0, 3, 2, 4, 6, 4, 2, 1, 0, 0, 1, 1, 2, 3, 2, 1, 3

Zbiorowość statystyczna – pracownicy konińskiej firmy „Zet” w dniu 1 lutego 2010 r.
Jednostka statystyczna – pracownik konińskiej firmy „Zet” w dniu w dniu 1 lutego 2010 r.
Cecha statystyczna (zmienna) – liczba posiadanych dzieci

Porządkowanie materiału statystycznego → **szereg prosty**

- rosnąco – 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6
- malejąco – 6, 5, 5, 5, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0

Grupowanie materiału statystycznego → **szereg rozdzielczy jednostopniowy (punktowy)**

Liczba dzieci (x_i)	Liczba pracowników (n_i)
0	5
1	7
2	6
3	5
4	3
5	3
6	1
Razem	30

Źródło: Dane umowne.

Przykład 1.8.3.

Obserwacji poddano 30 pracowników konińskiej firmy „Zet" ze względu wysokość płacy netto w styczniu 2010 r. (w zł). Uzyskane wyniki uporządkowano wzrastająco, otrzymując poniższy ciąg informacji → **szereg prosty**:

1220, 1249, 1258, 1280, 1290, 1310, 1310, 1315, 1318, 1320, 1320, 1320, 1320, 1320, 1328, 1380, 1385, 1385, 1390, 1395, 1395, 1398, 1410, 1420, 1420, 1430, 1430, 1450, 1480, 1499

Zbiorowość statystyczna – pracownicy konińskiej firmy „Zet" w styczniu 2010 r.
Jednostka statystyczna – pracownik konińskiej firmy „Zet" w styczniu 2010 r.
Cecha statystyczna (zmienna) – wysokość płacy netto (w zł)

Określenie rozstępu

$R = x_{max} - x_{min} = 1499 - 1220 = 279$

Ustalenie liczby przedziałów klasowych: przyjęcie $k = 6$

Ustalenie rozpiętości przedziałów klasowych:

$$C_x = \frac{x_{max} - x_{min}}{k} = 279/6 = 46{,}5$$

po zaokrągleniu 47

Grupowanie materiału statystycznego → **szereg rozdzielczy wielostopniowy (przedziałowy)**

Wysokość płacy netto w zł (x_i)	Liczba pracowników (n_i)
1220-1266	3
1267-1313	4
1314-1360	8
1361-1407	7
1408-1454	6
1455-1501	2
Razem	30

Źródło: Dane umowne.

Górna granica przedziału nie pokrywa się z dolną granicą przedziału następnego, więc nie ma problemu z zakwalifikowaniem jednostki do odpowiedniego przedziału.

Zaokrąglenie rozpiętości przedziałów klasowych do 50, w celu łatwiejszego obliczenia mierników statystycznych.

Wysokość płacy netto w zł (x_i)	Liczba pracowników (n_i)
1200-1250	2
1250-1300	3
1300-1350	10
1350-1400	7
1400-1450	5
1450-1500	3
Razem	30

Źródło: Dane umowne.

Górna granica przedziału pokrywa się z dolną granicą przedziału następnego, więc pojawia się problem zakwalifikowania jednostki do odpowiedniego przedziału. Przyjmuje się wówczas zasadę lewostronnego domknięcia przedziału, tj. <1200; 1250), <1250; 1300) itd.

Przykład 1.8.4.

Szereg o równych przedziałach klasowych, zamknięty dołem i górą

Polskie województwa według stopy bezrobocia w roku 2009 (stan na 30 czerwca)

Stopa bezrobocia (w %)	Liczba województw
7-10	4
10-13	6
13-16	5
16-19	1
Razem	16

Źródło: Opracowanie własne na podstawie *Bezrobotni oraz stopa bezrobocia wg województw, podregionów i powiatów*, <www.stat.gov.pl>.

Szereg o równych przedziałach klasowych, otwarty dołem i górą

Ludność Polski według grup wiekowych w roku 2008 (stan na 31 grudnia)

Wiek (w latach)	Liczba osób (w tys.)
poniżej 20	8 449,7
20-40	11 841,3
40-60	10 765,8
60-80	5 878,8
80 i więcej	1 200,2
Razem	38 135,9

Źródło: Opracowanie własne na podstawie Banku Danych Regionalnych GUS, 2008.

Szereg o nierównych przedziałach klasowych, otwarty dołem i górą

Polskie miasta według liczby mieszkańców w roku 2008 (stan na 31 grudnia)

Liczba mieszkańców (w tys. osób)	Liczba miast
poniżej 10	492
10-20	180
20-50	134
50-100	47
100-200	22
200 i więcej	17
Razem	892

Źródło: Opracowanie własne na podstawie: *Ludność. Stan i struktura w przekroju terytorialnym. Stan w dniu 31 XII 2008 r.*, <www.stat.gov.pl>.

Przykład 1.8.5.

Tablica prosta

Podmioty gospodarki narodowej zarejestrowane w Polsce według wielkości w roku 2008 (stan na 31 grudnia)

Wielkość (liczba zatrudnionych)	Liczba podmiotów
0-9	3 568 137
10-49	154 833
50-249	29 323
250-999	3 996
1 000 i więcej	804
Ogółem	3 757 093

Źródło: Opracowanie własne na podstawie Banku Danych Regionalnych GUS, 2008.

Tablica złożona kombinowana

Zgony w Polsce według płci i wieku w 2008 r.

Wiek (w latach)	Ogółem	Mężczyźni	Kobiety
0	2 338	1 305	1 033
1-9	635	373	262
10-19	1 711	1 206	505
20-29	4 615	3 721	894
30-39	7 279	5 723	1 556
40-49	19 718	14 705	5 013
50-59	52 011	36 916	15 095
60 i więcej	291 092	138 394	152 698
Ogółem	379 399	202 343	177 056

Źródło: Opracowanie własne na podstawie *Rocznika Demograficznego 2009*, GUS, Warszawa 2009.

Tablica złożona zbiorcza

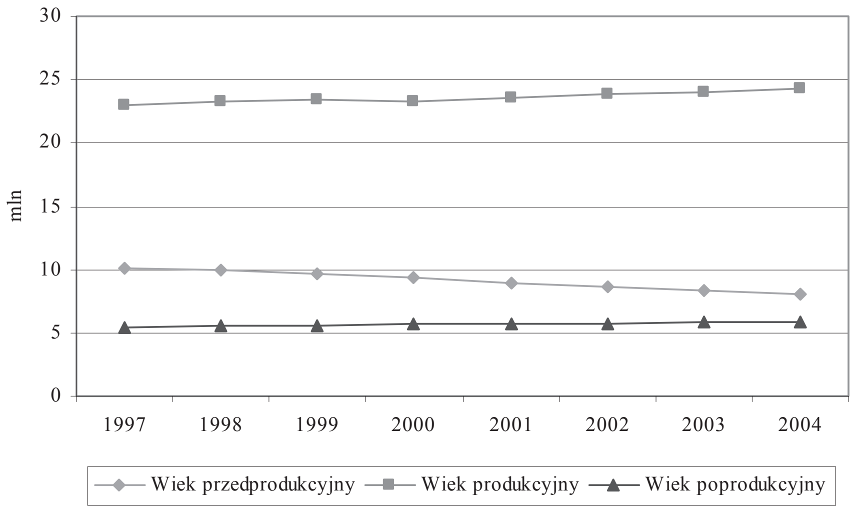
Nauczyciele akademiccy i studenci w Polsce według typów szkół wyższych w 2008 r.

Typ szkoły wyższej	Nauczyciele akademiccy	Studenci
uniwersytety	31 460	526 381
wyższe szkoły techniczne	19 527	322 111
wyższe szkoły rolnicze	5 505	87 556
wyższe szkoły ekonomiczne	9 922	356 561
wyższe szkoły pedagogiczne	4 514	107 668
akademie medyczne	10 011	10 103
wyższe szkoły morskie	607	58 015
akademie wychowania fizycznego	1 821	28 184
wyższe szkoły artystyczne	3 402	15 736
wyższe szkoły teologiczne	729	7 392
pozostałe szkoły wyższe	12 639	391 813
Ogółem	100 137	1 911 520

Źródło: Opracowanie własne na podstawie Banku Danych Regionalnych GUS, 2008.

Przykład 1.8.6.

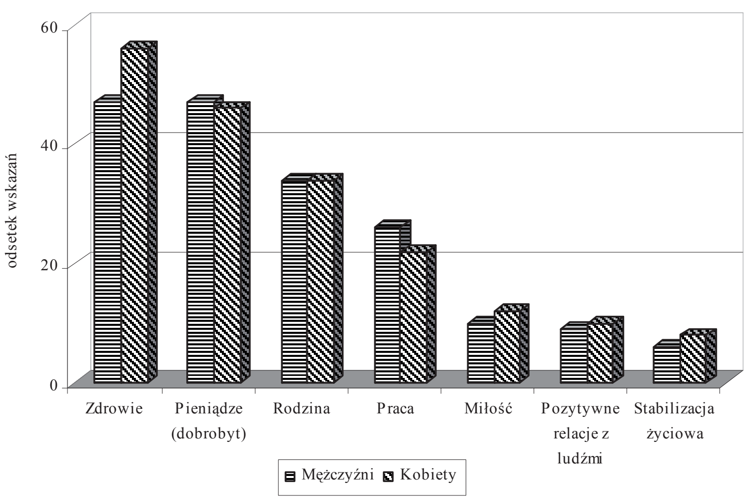
<u>Wykres liniowy</u>



Ludność w Polsce według ekonomicznych grup wiekowych w latach 2001-2008
(stan na 31 grudnia)

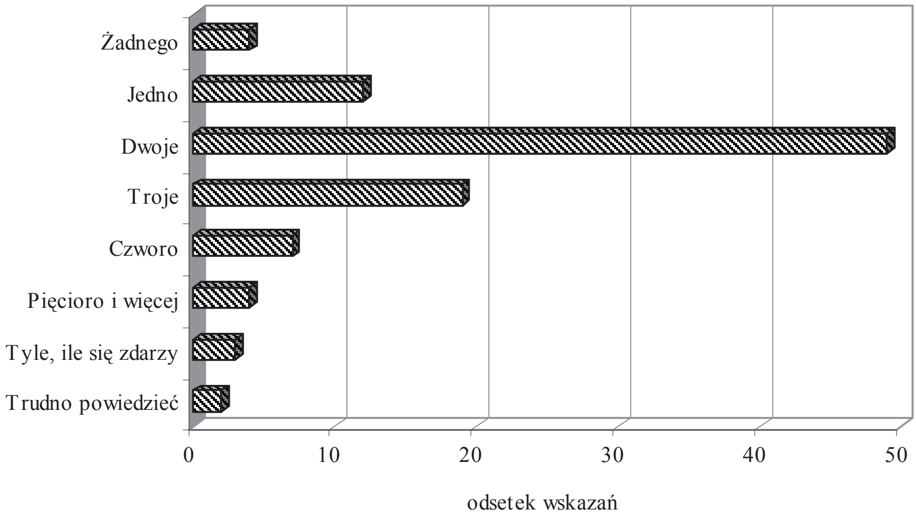
Źródło: Opracowanie własne na podstawie *Roczników Statystycznych Rzeczypospolitej Polskiej* z lat 2002-2009, GUS, Warszawa.

<u>Wykresy bryłowe</u>



Komponenty udanego życia w opinii Polaków (ze względu na płeć)

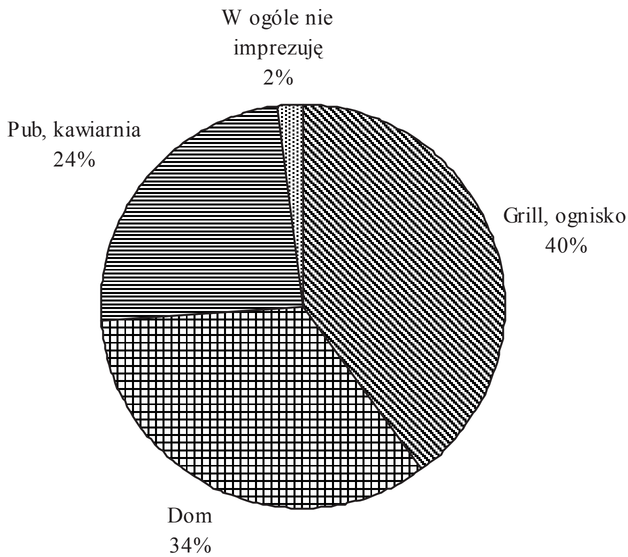
Źródło: Opracowanie własne na podstawie *Co jest w życiu najważniejsze?,* komunikat z badań CBOS, Warszawa, maj 2006, s. 6, <www.cbos.pl>.



Potrzeby prokreacyjne Polaków (ile dzieci chcieliby mieć w swoim życiu Polacy?)

Źródło: Opracowanie własne na podstawie *Potrzeby prokreacyjne oraz preferowany i realizowany model rodziny,* komunikat z badań CBOS, Warszawa, marzec 2006, s. 2, <www.cbos.pl>.

Wykres powierzchniowy



Miejsca imprez i spotkań towarzyskich Polaków

Źródło: Opracowanie własne na podstawie *Imprezy*, Instytut Badania Opinii RMF FM, maj 2005, <www.rmf.fm/instytut>.

2. ANALIZA STRUKTURY ZBIOROWOŚCI

2.1. Informacje ogólne

Zadaniem **analizy struktury zbiorowości** jest odzwierciedlenie zasadniczych właściwości w budowie badanej zbiorowości. Analizę tę przeprowadza się za pomocą tzw. parametrów opisowych, które umożliwiają dokonanie skróconego opisu struktury zbiorowości (z punktu widzenia badanej cechy zmiennej) oraz porównań między zbiorowościami.

Parametry opisowe dzieli się na:
- **parametry klasyczne**, które liczone są na podstawie wartości cechy zmiennej wszystkich jednostek badanej zbiorowości,
- **parametry pozycyjne**, które wyznaczane są na podstawie wartości cechy zmiennej wybranych jednostek badanej zbiorowości zajmujących szczególną pozycję w szeregu statystycznym.

Zakres analizy struktury zbiorowości:
- wskaźniki struktury i natężenia,
- miary położenia (przeciętne),
- miary zmienności (dyspersji),
- miary asymetrii (skośności),
- miary koncentracji.

2.2. Wskaźniki struktury i natężenia

Wskaźnik struktury (częstość, liczebność względna, frakcja, odsetek) to stosunek liczby jednostek o danej wartości cechy zmiennej do łącznej liczebności zbiorowości

$$\omega_i = \frac{n_i}{N},$$

gdzie:
n_i – liczebność cząstkowa określająca, ile jednostek zbiorowości przypada na daną wartość cechy zmiennej,
N – liczebność zbiorowości.

Wskaźnikiem struktury jest również stosunek części wartości cechy zmiennej do sumy wartości zmiennej

$$\omega_i = \frac{x_i}{\sum x_i},$$

gdzie:
x_i – wartość cechy zmiennej,
Σx_i – suma wartości cechy zmiennej.

Wskaźnik natężenia to stosunek liczby jednostek (wartości cechy) danej zbiorowości do liczby jednostek (wartości cechy) innej zbiorowości, które pozostają w przyczynowym lub logicznym związku

$$\nu_i = \frac{n_i}{m_i},$$

gdzie:
n_i – liczba jednostek jednej zbiorowości,
m_i – liczba jednostek drugiej zbiorowości.

2.3. Miary przecięte (położenia)

Miary przeciętne (położenia) charakteryzują zbiorowość statystyczną niezależnie od różnic występujących między poszczególnymi jednostkami wchodzącymi w jej skład. Dokonują one charakterystyki podobieństw zbiorowości ze względu na wyróżnioną cechę zmienną.

Podział miar położenia:

- klasyczne
  - średnia arytmetyczna (zwykła, ważona),
  - średnia chronologiczna,
  - średnia harmoniczna,
  - średnia geometryczna,
- pozycyjne
  - dominanta,
  - kwantyle,
    - kwartyle (kwartyl pierwszy, mediana, kwartyl trzeci),
    - decyle,
    - percentyle (centyle).

2.3.1. Średnie klasyczne

Średnie klasyczne liczone są na podstawie wartości cechy zmiennej wszystkich jednostek badanej zbiorowości, ukazując średni poziom tej cechy w zbiorowości. Mają one charakter abstrakcyjny, ponieważ ich wartości muszą spełniać warunek

$$x_{\min} \leq \bar{x} \leq x_{\max},$$

gdzie:
x_{min} – minimalna wartość cechy zmiennej,
x_{max} – maksymalna wartość cechy zmiennej,

ale nie muszą (choć mogą) pokrywać się z pewną wartością badanej cechy zmiennej. Najbardziej popularna jest **średnia arytmetyczna**. Charakteryzuje ona średni (przeciętny) poziom cechy zmiennej w zbiorowości. Robi to tym lepiej, im mniejsze jest zróżnicowanie między wartościami badanej zmiennej (wartości skrajne mogą bowiem zniekształcić rezultat obliczeń).

Sposób obliczania średniej arytmetycznej:

- szereg prosty (wyliczający) – **średnia arytmetyczna zwykła**

$$\bar{x}_a = \frac{\sum_{i=1}^{N} x_i}{N},$$

gdzie:
x_i – wartość cechy zmiennej,
N – liczebność zbiorowości;

- szereg rozdzielczy jednostopniowy (punktowy) – **średnia arytmetyczna ważona**

$$\bar{x}_a = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{N},$$

gdzie:
n_i – liczebność cząstkowa określająca, ile jednostek zbiorowości przypada na daną wartość cechy zmiennej;

- szereg rozdzielczy wielostopniowy (przedziałowy) – **średnia arytmetyczna ważona**

$$\bar{x}_a = \frac{\sum_{i=1}^{k} x'_i \cdot n_i}{N},$$

gdzie:
x'_i – środek przedziału klasowego,
n_i – liczebność cząstkowa określająca, ile jednostek zbiorowości przyjmuje wartość cechy zmiennej z danego przedziału klasowego.

W szeregu przedziałowym średnią arytmetyczną można obliczyć jeżeli przedziały klasowe są równe, a szereg jest zamknięty dołem i górą (jeżeli tak nie jest, to można dokonać zamknięcia szeregu pod warunkiem, że w otwartym przedziale znajduje się nie więcej niż 5% ogółu jednostek badanej zbiorowości).

2.3.2. Średnie pozycyjne

Przeciętne pozycyjne oparte są na wartościach cechy zmiennej wybranych jednostek zbiorowości charakteryzujących się szczególnym położeniem. Można je dokładnie wyznaczyć w szeregach prostych (wyliczających) i rozdzielczych jednostopniowych (punktowych), natomiast w szeregach rozdzielczych wielostopniowych (przedziałowych) można wskazać jedynie przedział, w którym znajduje się przeciętna pozycyjna, a następnie oszacować jej wartości przy wykorzystaniu wzoru interpolacyjnego.

2.3.2.1. Dominanta

Dominanta to wartość cechy zmiennej, która występuje najczęściej w badanej zbiorowości (wartość dominująca).

Sposób wyznaczania dominanty:
- szereg prosty (wyliczający) – wyznaczenie dominanty polega na wskazaniu najczęściej powtarzającej się wartości cechy zmiennej,
- szereg rozdzielczy jednostopniowy (punktowy) – wyznaczenie dominanty polega na wskazaniu wartości cechy zmiennej, której odpowiada maksymalna liczebność,

- szereg rozdzielczy wielostopniowy (przedziałowy) – wyznaczenie dominanty polega na wskazaniu przedziału, w którym znajduje się dominanta (przedział o największej liczebności), a następnie oszacowaniu jej wartości w oparciu o wzór interpolacyjny

$$D = x_0 + c_0 \cdot \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) + (n_d - n_{d+1})} ,$$

gdzie:
x_0 – dolna granica przedziału dominanty,
c_0 – rozpiętość przedziału dominanty,
n_d – liczebność przedziału dominanty,
n_{d-1} – liczebność przedziału poprzedzającego przedział dominanty,
n_{d+1} – liczebność przedziału następującego po przedziale dominanty.

W szeregu przedziałowym dominantę można oszacować tylko wtedy, gdy przedział dominanty oraz przedziały sąsiednie (poprzedzający i następujący) mają taką samą rozpiętość. Jeżeli rozkład jest symetryczny, to można skorzystać z formuły $D = \bar{x} - 3 \cdot (\bar{x} - M_e)$. W szeregu przedziałowym dominantę można wyznaczyć również graficznie, za pomocą **histogramu** (zob. przykład 2.7.3).

2.3.2.2. Kwantyle

Kwantyle to wartości cechy zmiennej, które dzielą badaną zbiorowość na określone części pod względem liczby jednostek. Wyróżnia się **kwartyle** dzielące zbiorowość na cztery części, **decyle** dzielące zbiorowość na 10 części oraz **percentyle (centyle)** dzielące zbiorowość na 100 części.

Mediana (kwartyl drugi) to wartość cechy zmiennej, która dzieli badaną zbiorowość na dwie części w taki sposób, że połowa jednostek zbiorowości charakteryzuje się wartościami nie wyższymi, a połowa nie niższymi od mediany.

Sposób wyznaczania mediany:
- szereg prosty (wyliczający):
  - nieparzysty – mediana jest wartością środkową w szeregu

$$k = \frac{N+1}{2}, \qquad M_e = x_k ,$$

gdzie:
N – liczebność zbiorowości,
x_k – wartość danej cechy zmiennej;

- parzysty – mediana jest średnią arytmetyczną dwóch wartości środkowych

$$k = \frac{N}{2}, \qquad M_e = \frac{x_k + x_{k+1}}{2};$$

- szereg rozdzielczy jednostopniowy (punktowy) – wyznaczenie mediany polega na wskazaniu w kolumnie cechy zmiennej wartości odpowiadającej cum N/2 (dla szeregu parzystego) lub cum (N+1)/2 (dla szeregu nieparzystego);
- szereg rozdzielczy wielostopniowy (przedziałowy) – wyznaczenie mediany polega na wskazaniu przedziału, w którym znajduje się mediana, a następnie oszacowaniu jej wartości w oparciu o wzór interpolacyjny

$$M_e = Q_2 = x_0 + \frac{c_0}{n_0} \cdot (\frac{N}{2} - cum_{n-1}),$$

gdzie:
x_0 – dolna granica przedziału mediany,
c_0 – rozpiętość przedziału mediany,
n_0 – liczebność przedziału mediany,
N – liczebność zbiorowości,
cum_{n-1} – skumulowana liczebność przedziału poprzedzającego przedział mediany.

Kwartyl pierwszy (dolny) to wartość cechy zmiennej, która dzieli badaną zbiorowość w taki sposób, że 25% jednostek zbiorowości charakteryzuje się wartościami nie wyższymi, a 75% jednostek nie niższymi od kwartyla pierwszego.

Sposób wyznaczania kwartyla pierwszego:
- szereg prosty (wyliczający):
  - nieparzysty

$$k = \frac{N+1}{4}, \qquad Q_1 = x_k,$$

gdzie:
N – liczebność zbiorowości,
x_k – wartość danej cechy zmiennej;

  - parzysty

$$k = \frac{N}{4}, \qquad Q_1 = \frac{x_k + x_{k+1}}{2};$$

- szereg rozdzielczy jednostopniowy (punktowy) – wyznaczenie kwartyla pierwszego polega na wskazaniu w kolumnie cechy zmiennej wartości odpowiadającej cum N/4 (dla szeregu parzystego) lub cum (N+1)/4 (dla szeregu nieparzystego);
- szereg rozdzielczy wielostopniowy (przedziałowy) – wyznaczenie kwartyla pierwszego polega na wskazaniu przedziału, w którym znajduje się kwartyl, a następnie oszacowaniu jego wartości w oparciu o wzór interpolacyjny

  $$Q_1 = x_0 + \frac{c_0}{n_0} \cdot \left(\frac{N}{4} - cum_{n-1}\right),$$

  gdzie:
  x_0 – dolna granica przedziału, w którym znajduje się kwartyl pierwszy,
  c_0 – rozpiętość przedziału, w którym znajduje się kwartyl pierwszy,
  n_0 – liczebność przedziału, w którym znajduje się kwartyl pierwszy,
  N – liczebność zbiorowości,
  cum_{n-1} – skumulowana liczebność przedziału poprzedzającego przedział kwartyla pierwszego.

Kwartyl trzeci (górny) to wartość cechy zmiennej, która dzieli badaną zbiorowość w taki sposób, że 75% jednostek zbiorowości charakteryzuje się wartościami nie wyższymi od kwartyla trzeciego, a 25% jednostek nie niższymi od kwartyla trzeciego.

Sposób wyznaczania kwartyla trzeciego:

- szereg prosty (wyliczający):
  - nieparzysty

    $$k = \frac{3(N+1)}{4}, \qquad Q_3 = x_k,$$

    gdzie:
    N – liczebność zbiorowości,
    x_k – wartość danej cechy zmiennej;
  - parzysty

    $$k = \frac{3N}{4}, \qquad Q_3 = \frac{x_k + x_{k+1}}{2};$$
- szereg rozdzielczy jednostopniowy (punktowy) – wyznaczenie kwartyla trzeciego polega na wskazaniu w kolumnie cechy zmiennej wartości odpowiadającej cum 3N/4 (dla szeregu parzystego) lub cum 3(N+1)/4 (dla szeregu nieparzystego);

- szereg rozdzielczy wielostopniowy (przedziałowy) – wyznaczenie kwartyla trzeciego polega na wskazaniu przedziału, w którym znajduje się kwartyl, a następnie oszacowaniu jego wartości w oparciu o wzór interpolacyjny

  $$Q_3 = x_0 + \frac{c_0}{n_0} \cdot \left(\frac{3N}{4} - cum_{n-1}\right),$$

  gdzie:
  x_0 – dolna granica przedziału, w którym znajduje się kwartyl trzeci,
  c_0 – rozpiętość przedziału, w którym znajduje się kwartyl trzeci,
  n_0 – liczebność przedziału, w którym znajduje się kwartyl trzeci,
  N – liczebność zbiorowości,
  cum_{n-1} – skumulowana liczebność przedziału poprzedzającego przedział kwartyla trzeciego.

W szeregu przedziałowym kwartyle można wyznaczyć graficznie, za pomocą **krzywej liczebności skumulowanych** (zob. przykład 2.7.3).

2.4. Miary zmienności (dyspersji)

Miary zmienności (dyspersji) charakteryzują zbiorowość statystyczną, uwzględniając różnice między poszczególnymi jednostkami wchodzącymi w jej skład. Dokonują one charakterystyki stopnia zróżnicowania zbiorowości ze względu na wyróżnioną cechę zmienną.

Podział miar zmienności:

- klasyczne:
  - wariancja,
  - odchylenie standardowe,
  - typowy obszar zmienności,
  - klasyczny współczynnik zmienności;
- pozycyjne:
  - rozstęp,
  - odchylenie ćwiartkowe,
  - kwartylowy obszar zmienności,
  - kwartylowy współczynnik zmienności.

2.4.1. Klasyczne miary zmienności

Klasyczne miary zmienności liczone są na podstawie wartości cechy zmiennej wszystkich jednostek badanej zbiorowości, ukazując jednocześnie różnice między wartościami badanej cechy dla poszczególnych jednostek, a wartością centralną (zazwyczaj średnią arytmetyczną).

2.4.1.1. Wariancja (drugi moment centralny)

Wariancja to średnia arytmetyczna z kwadratów odchyleń poszczególnych wartości cechy od średniej arytmetycznej tej cechy. Jest ona stosowana przy konstrukcji wielu parametrów, ale jej wyniku nie interpretuje się.

Sposób obliczania wariancji:

- szereg prosty (wyliczający)

$$s^2(x) = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N},$$

gdzie:
x_i – wartość cechy zmiennej,
$\bar{x}$ – średnia arytmetyczna cechy zmiennej,
N – liczebność zbiorowości;

- szereg rozdzielczy jednostopniowy (punktowy)

$$s^2(x) = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \cdot n_i}{N},$$

gdzie:
n_i – liczebność cząstkowa określająca, ile jednostek zbiorowości przypada na daną
wartość cechy zmiennej;

- szereg rozdzielczy wielostopniowy (przedziałowy)

$$s^2(x) = \frac{\sum_{i=1}^{N}\left(x'_i - \bar{x}\right)^2 \cdot n_i}{N},$$

gdzie:
x'_i – środek przedziału klasowego,
n_i – liczebność przedziału klasowego.

2.4.1.2. Odchylenie standardowe

Odchylenie standardowe jest bezwzględną miarą zróżnicowania, która informuje, o ile przeciętnie poszczególne jednostki badanej zbiorowości różnią się pod względem cechy zmiennej (*in plus* lub *in minus*) od średniej arytmetycznej tej zmiennej. Odchylenie jest pierwiastkiem kwadratowym z wariancji.

Sposób obliczania odchylenia standardowego:

- szereg prosty (wyliczający)

$$s(x) = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N}} \quad \text{lub} \quad s(x) = \sqrt{\frac{\sum_{i=1}^{N} x^2}{N} - (\bar{x})^2},$$

gdzie:
x_i – wartość cechy zmiennej,
$\bar{x}$ – średnia arytmetyczna cechy zmiennej,
N – liczebność zbiorowości;

- szereg rozdzielczy jednostopniowy (punktowy)

$$s(x) = \sqrt{\frac{\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 \cdot n_i}{N}},$$

gdzie:
n_i – liczebność cząstkowa określająca, ile jednostek zbiorowości przypada na daną wartość cechy zmiennej;

- szereg rozdzielczy wielostopniowy (przedziałowy)

$$s(x) = \sqrt{\frac{\sum_{i=1}^{k}\left(x'_i - \bar{x}\right)^2 \cdot n_i}{N}},$$

gdzie:
x'_i – środek przedziału klasowego,
n_i – liczebność przedziału klasowego.

2.4.1.3. Typowy obszar zmienności

Jeżeli rozkład cechy w zbiorowości jest rozkładem normalnym, to w granicach typowego obszaru zmienności mieści się około 2/3 jednostek badanej zbiorowości

$$\bar{x} - s(x) < x_{typ} < \bar{x} + s(x).$$

Z odchyleniem standardowym wiąże się tzw. **reguła 3 sigm** (twierdzenie Czebyszewa), która mówi, że jeżeli rozkład cech w zbiorowości jest rozkładem normalnym, to wyniki obserwacji układają się tak, że w przedziale:

- $\bar{x} \pm 2s(x)$ mieści się około 95% wszystkich jednostek badanej zbiorowości,
- $\bar{x} \pm 3s(x)$ mieści się około 99,74% wszystkich jednostek badanej zbiorowości.

2.4.1.4. Klasyczny współczynnik zmienności

Klasyczny współczynnik zmienności jest względną miarą zróżnicowania, która informuje o sile zróżnicowania badanej zbiorowości pod względem cechy zmiennej oraz umożliwia ocenę średniej arytmetycznej. Im wartość współczynnika jest wyższa, tym zróżnicowanie jest silniejsze, i odwrotnie

$$V_x = \frac{s(x)}{\bar{x}} \cdot 100.$$

2.4.2. Pozycyjne miary zmienności

Pozycyjne miary zmienności oparte są na wartościach cechy zmiennej wybranych jednostek zbiorowości charakteryzujących się szczególnym położeniem. Zazwyczaj są obliczane wtedy, gdy niemożliwe lub niewskazane jest wykorzystanie miar klasycznych.

2.4.2.1. Rozstęp (empiryczny obszar zmienności)

Rozstęp określa całkowitą zmienność wartości badanej cechy i tym samym służy wstępnej ocenie dyspersji

$$R = x_{\max} - x_{\min},$$

gdzie:
x_{min} – minimalna wartość cechy zmiennej,
x_{max} – maksymalna wartość cechy zmiennej.

2.4.2.2. Odchylenie ćwiartkowe

Odchylenie ćwiartkowe jest bezwzględną miarą zróżnicowania, która określa przeciętne zróżnicowanie połowy jednostek zbiorowości – jednostek środkowych, czyli skupionych wokół mediany (po odrzuceniu 25% jednostek o najniższych wartościach cechy zmiennej i 25% jednostek o wartościach najwyższych)

$$Q = \frac{Q_3 - Q_1}{2},$$

gdzie:
Q_1 – kwartyl pierwszy,
Q_3 – kwartyl trzeci.

2.4.2.3. Kwartylowy obszar zmienności

W granicach kwartylowego obszaru zmienności mieści się 50% jednostek badanej zbiorowości

M_e – Q < kwartylowy obszar zmienności < M_e + Q.

2.4.2.4. Kwartylowy współczynnik zmienności

Kwartylowy współczynnik zmienności jest względną miarą zróżnicowania, która informuje o sile zróżnicowania badanej zbiorowości pod względem cechy zmiennej oraz umożliwia ocenę mediany. Im wartość współczynnika jest wyższa, tym zróżnicowanie jest silniejsze, i odwrotnie

$$V_Q = \frac{Q}{M_e} \cdot 100 .$$

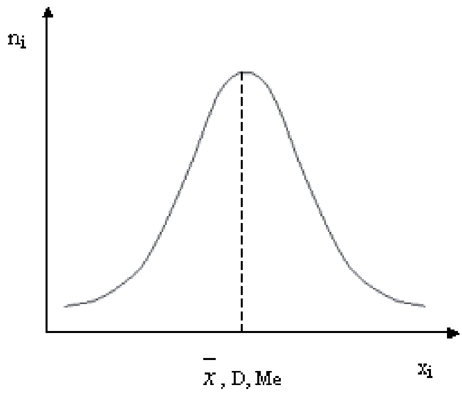
2.5. Miary asymetrii (skośności)

Miary asymetrii (skośności) określają kierunek rozkładu cech zmiennych w zbiorowości (rozkład może być symetryczny lub asymetryczny lewostronnie lub prawostronnie) oraz stopień odchylenia rozkładu cechy zmiennej od rozkładu symetrycznego[3].

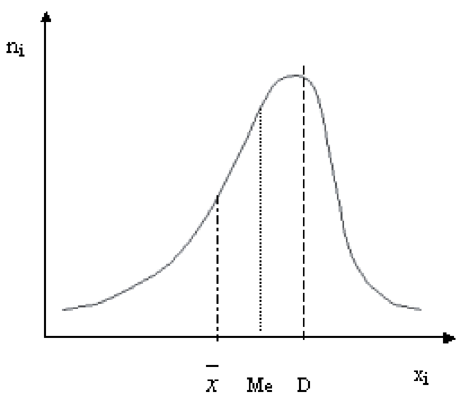
[3] Asymetria oznacza deformację rozkładu cechy zmiennej w związku z wydłużeniem ramienia krzywej liczebności w prawo lub w lewo w stosunku do dominanty. Im asymetria rozkładu jest większa, tym mniejsza jest wartość poznawcza średniej arytmetycznej oraz pozostałych miar klasycznych, i odwrotnie.

Podział miar asymetrii:

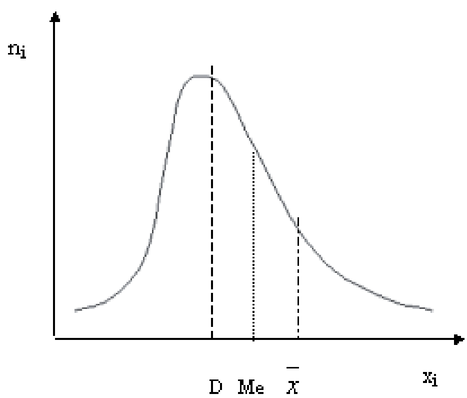
- bezwzględne
  - wskaźnik asymetrii,
  - kwartylowy wskaźnik asymetrii,
  - trzeci moment centralny,
- względne
  - klasyczno-pozycyjny współczynnik skośności,
  - kwartylowy współczynnik skośności,
  - trzeci moment centralny standaryzowany.



Rozkład symetryczny: $\bar{x}$ = Me = D



Rozkład asymetryczny lewostronnie: $\bar{x}$ < Me < D



Rozkład asymetryczny prawostronnie: $\bar{x}$ > Me > D

Schemat 2.1. Rozkłady cech zmiennych o różnej asymetrii (skośności)

Źródło: Opracowanie własne.

2.5.1. Bezwzględne miary asymetrii

Bezwzględne miary asymetrii określają kierunek asymetrii. Kierunek ten można ustalić również przez porównanie położenia średniej arytmetycznej, mediany i dominanty:

- rozkład symetryczny $\bar{x} = M_e = D$,
- rozkład asymetryczny prawostronnie (asymetria dodatnia) $\bar{x} > M_e > D$,
- rozkład asymetryczny lewostronnie (asymetria ujemna) $\bar{x} < M_e < D$.

2.5.1.1. Wskaźnik asymetrii

$$A_s = \bar{x} - D$$

Jeżeli:
- rozkład jest symetryczny, to $A_s = 0$,
- rozkład jest asymetryczny prawostronnie, to $A_s > 0$,
- rozkład jest asymetryczny lewostronnie, to $A_s < 0$.

2.5.1.2. Kwartylowy wskaźnik asymetrii

$$A_Q = Q_1 + Q_3 - 2M_e$$

Jeżeli:
- rozkład jest symetryczny, to $A_Q = 0$,
- rozkład jest asymetryczny prawostronnie, to $A_Q > 0$,
- rozkład jest asymetryczny lewostronnie, to $A_Q < 0$.

2.5.1.3. Trzeci moment centralny

Sposób obliczania trzeciego momentu centralnego:
- szereg prosty (wyliczający)

$$\mu_3 = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{N};$$

- szereg rozdzielczy jednostopniowy (punktowy)

$$\mu_3 = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x}\right)^3 \cdot n_i}{N};$$

- szereg rozdzielczy wielostopniowy (przedziałowy)

$$\mu_3 = \frac{\sum_{i=1}^{k} \left(x'_i - \overline{x}\right)^3 \cdot n_i}{N}.$$

Jeżeli:
- rozkład jest symetryczny, to $\mu_3 = 0$,
- rozkład jest asymetryczny prawostronnie, to $\mu_3 > 0$,
- rozkład jest asymetryczny lewostronnie, to $\mu_3 < 0$.

2.5.2. Względne miary asymetrii

Względne miary asymetrii określają zarówno kierunek, jak i siłę asymetrii. Im wartość bezwzględna poszczególnych miar (współczynników) jest wyższa, tym asymetria jest silniejsza, i odwrotnie.

2.5.2.1. Klasyczno-pozycyjny współczynnik skośności

$$W_{as} = \frac{\overline{x} - D}{s(x)}$$

Jeżeli:
- rozkład jest symetryczny, to $W_{as} = 0$,
- rozkład jest asymetryczny prawostronnie, to $W_{as} > 0$,
- rozkład jest asymetryczny lewostronnie, to $W_{as} < 0$.

2.5.2.2. Kwartylowy współczynnik skośności

$$W_Q = \frac{Q_1 + Q_3 - 2M_e}{2Q}$$

Jeżeli:
- rozkład jest symetryczny, to $W_Q = 0$,
- rozkład jest asymetryczny prawostronnie, to $W_Q > 0$,
- rozkład jest asymetryczny lewostronnie, to $W_Q < 0$.

2.5.2.3. Trzeci moment centralny standaryzowany

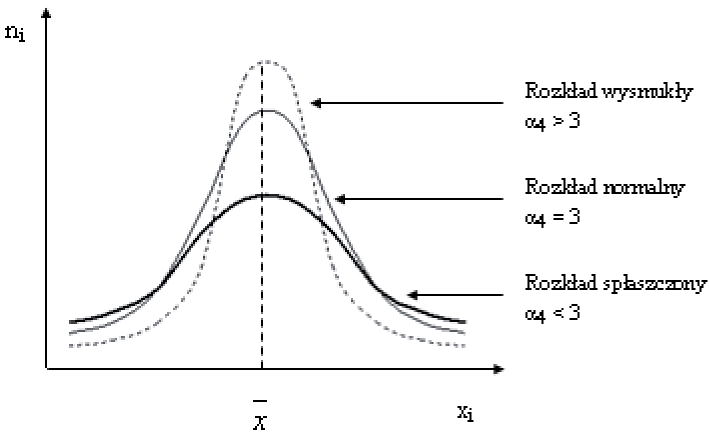
$$\alpha_3 = \frac{\mu_3}{s^3(x)}$$

Jeżeli:
- rozkład jest symetryczny, to $\alpha_3 = 0$,
- rozkład jest asymetryczny prawostronnie, to $\alpha_3 > 0$,
- rozkład jest asymetryczny lewostronnie, to $\alpha_3 < 0$.

2.6. Miary koncentracji

Miary koncentracji określają stopień skupienia poszczególnych jednostek zbiorowości ze względu na badaną cechę zmienną wokół średniej arytmetycznej tej zmiennej (kurtoza) lub stopień nierównomierności podziału zjawiska w zbiorowości.

Podział miar koncentracji:
- skupienie zbiorowości wokół średniej (kurtoza):
  - czwarty moment centralny,
  - czwarty moment centralny standaryzowany;
- nierównomierny podział zjawiska w zbiorowości:
  - wielobok koncentracji Lorenza,
  - współczynnik koncentracji.



Schemat 2.2. Rozkłady cech zmiennych o różnym skupieniu

Źródło: Opracowanie własne.

2.6.1. Kurtoza (eksces)

Koncentracja, rozumiana jako stopień skupienia poszczególnych jednostek zbiorowości ze względu na badaną cechę zmienną wokół średniej arytmetycznej tej zmiennej, oznacza deformację rozkładu w związku ze spłaszczeniem lub smukłością krzywej liczebności. Im bardziej krzywa liczebności jest wysmukła, tym koncentracja (stopień skupienia) jest silniejsza, a im bardziej krzywa liczebności jest spłaszczona, tym koncentracja (stopień skupienia) jest słabsza.

2.6.1.1. Czwarty moment centralny

Sposób obliczania czwartego momentu centralnego:

- szereg prosty (wyliczający)

$$\mu_4 = \frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^4}{N};$$

- szereg rozdzielczy jednostopniowy (punktowy)

$$\mu_4 = \frac{\sum_{i=1}^{k} \left(x_i - \bar{x}\right)^4 \cdot n_i}{N};$$

- szereg rozdzielczy wielostopniowy (przedziałowy)

$$\mu_4 = \frac{\sum_{i=1}^{k} \left(x'_i - \bar{x}\right)^4 \cdot n_i}{N}.$$

2.6.1.2. Czwarty moment centralny standaryzowany

$$\alpha_4 = \frac{\mu_4}{s^4(x)}$$

Jeżeli:

- rozkład jest wysmukły, o skupieniu silniejszym od normalnego, to $\alpha_4 > 3$,
- rozkład jest normalny ($\alpha_3 = 0$), to $\alpha_4 = 3$,
- rozkład jest spłaszczony, o skupieniu słabszym od normalnego, to $\alpha_4 < 3$.

2.6.2. Nierównomierność podziału zjawiska w zbiorowości

Koncentracja rozumiana jako nierównomierny podział zjawiska w zbiorowości oznacza nierównomierne rozłożenie wartości cechy zmiennej pomiędzy poszczególne jednostki tej zbiorowości. Jeśli wszystkie jednostki zbiorowości dysponują taką samą wartością cechy zmiennej, to koncentracja nie występuje, natomiast jeśli jedna jednostka zbiorowości dysponuje całą sumą wartości cechy zmiennej to występuje wówczas koncentracja zupełna.

2.6.2.1. Wielobok koncentracji Lorenza

Wielobok koncentracji Lorenza pozwala ocenić koncentrację zjawiska w sposób graficzny, za pomocą **krzywej Lorenza**. Koncentracja zjawiska jest tym większa, im większa jest powierzchnia zawarta między krzywą Lorenza, a linią równomiernego rozdziału (krzywa Lorenza jest bardziej wypukła w stosunku do linii równomiernego rozdziału), i odwrotnie. Jeżeli krzywa Lorenza pokrywa się z linią równomiernego rozdziału to koncentracja nie występuje (zob. przykład 2.7.4).

2.6.2.2. Współczynnik koncentracji

Współczynnik koncentracji to stosunek powierzchni zawartej między krzywą Lorenza a linią równomiernego rozdziału do powierzchni trójkąta, który powstaje z połączenia na układzie współrzędnych punktów (0;0), (100;0), (100;100). Wskaźnik przyjmuje wartość z przedziału <0; 1>. Im wartość wskaźnika jest bliższa 1, tym koncentracja jest silniejsza, a im bliżej 0, tym koncentracja jest słabsza. Jeżeli k = 0, to brak koncentracji, a jeżeli k = 1, to koncentracja jest zupełna

$$k = \frac{5000 - \sum P_i}{5000},$$

gdzie:
ΣP_i – pole powierzchni pod krzywą koncentracji Lorenza (suma pola trójkąta i pól trapezów),
pole trójkąta: (a*h)/2,
pole trapezu: [(a+b)*h]/2.

2.7. Przykłady

Przykład 2.7.1.

Wartości temperatur (w stopniach C) zaobserwowanych w dniu 18 lipca 2009 r. o godzinie 12.00 w miastach wojewódzkich były następujące (źródło: www.pogoda.onet.pl): 19, 24, 27, 27, 28, 29, 28, 29, 26, 19, 22, 25, 23, 25, 28, 26.
Na podstawie powyższych informacji należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz wskazać przeciętne pozycyjne (dominantę, medianę, kwartyle).

Rozwiązanie

<u>Zbiorowość statystyczna</u> – miasta wojewódzkie
<u>Jednostka statystyczna</u> – miasto wojewódzkie
<u>Cecha statystyczna (zmienna)</u> – temperatura (w stopniach C) zaobserwowana 18 lipca 2009 r. o godzinie 12:00

<u>Porządkowanie wariantów cechy (rosnąco)</u>: 19, 19, 22, 23, 24, 25, 25, 26, 26, 27, 27, 28, 28, 28, 29, 29

<u>Średnia arytmetyczna zwykła</u>

$$\bar{x}_a = \frac{\sum_{i=1}^{N} x_i}{N} = 405/16 = 25{,}31$$

Spełniony jest warunek $x_{\min} = 19 \leq \bar{x} = 25{,}31 \leq x_{\max} = 29$, ale średnia jest wartością abstrakcyjną.

Średnia temperatura 18 lipca 2009 r. o godzinie 12:00 w miastach wojewódzkich wynosiła 25,31°C.

Dominanta

D = 28

Wśród miast wojewódzkich dominowały te, w których temperatura 18 lipca 2009 r. o godzinie 12:00 wynosiła 28°C.

Mediana

Szereg parzysty (n = 16), więc[4]:

$$k = \frac{N}{2} = 16/2 = 8,\ M_e = \frac{x_k + x_{k+1}}{2} = \frac{x_8 + x_9}{2} = \frac{26+26}{2} = 26.$$

W połowie miast wojewódzkich temperatura 18 lipca 2009 r. o godzinie 12:00 była nie wyższa niż 26°C, a w połowie nie niższa niż 26°C.

Kwartyl pierwszy

Szereg parzysty (n = 16), więc[5]:

$$k = \frac{N}{4} = 16/4 = 4,\ Q_1 = \frac{x_k + x_{k+1}}{2} = \frac{x_4 + x_5}{2} = \frac{23+24}{2} = 23{,}5.$$

W 1/4 miast wojewódzkich temperatura 18 lipca 2009 r. o godzinie 12:00 była nie wyższa niż 23,5°C, a w 3/4 nie niższa niż 23,5°C.

Kwartyl trzeci

Szereg parzysty (n = 16), więc[6]:

[4] Gdyby szereg był nieparzysty (np. n = 15), to $k = (N+1)/2 = (15+1)/2 = 8$, $M_e = x_k = x_8 = 26$

[5] Gdyby szereg był nieparzysty (np. n = 15), to $k = (N+1)/4 = (15+1)/4 = 4$, $Q_1 = x_k = x_4 = 23$

[6] Gdyby szereg był nieparzysty (np. n = 15), to $k = 3(N+1)/4 = 3(15+1)/4 = 12$, $Q_3 = x_k = x_{12} = 28$

$$k = \frac{3N}{4} = 48/4 = 12\,,\ Q_3 = \frac{x_k + x_{k+1}}{2} = \frac{x_{12} + x_{13}}{2} = \frac{28+28}{2} = 28\,.$$

W 3/4 miast wojewódzkich temperatura 18 lipca 2009 r. o godzinie 12:00 była nie wyższa niż 28°C, a w 1/4 nie niższa niż 28°C.

Przykład 2.7.2.

W dniu 20 kwietnia 2009 r. grupie studentów PWSZ w Koninie dano do rozwiązania zadanie ze statystyki. Czas rozwiązywania tego zadania (w minutach) przez poszczególne osoby został przedstawiony w poniższej tabeli.

Czas rozwiązywania zadania (w min.)	4	5	6	7	8
Liczba studentów	6	4	3	2	2

Źródło: Dane umowne.

Na podstawie powyższych informacji należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz wskazać przeciętne pozycyjne (dominantę, medianę, kwartyle).

Rozwiązanie

Zbiorowość statystyczna – grupa studentów PWSZ w Koninie
Jednostka statystyczna – student PWSZ w Koninie
Cecha statystyczna (zmienna) – czas rozwiązywania zadania (w min.) w dniu 20 kwietnia 2009 r.

Obliczenia pomocnicze

Czas rozwiązywania zadania (x_i)	Liczba studentów (n_i)	$x_i \cdot n_i$	cum n_i
4	6	24	6
5	4	20	10
6	3	18	13
7	2	14	15
8	2	16	17
Σ	17	92	x

Średnia arytmetyczna ważona

$$\bar{x}_a = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{N} = 92/17 = 5{,}41.$$

Spełniony jest warunek $x_{\min} = 4 \le \bar{x} = 5{,}41 \le x_{\max} = 8$, ale średnia jest wartością abstrakcyjną.

Przeciętny czas rozwiązywania zadania ze statystyki przez studentów PWSZ w Koninie wynosił 5,41 minut.

Dominanta

D = 4

W grupie studentów PWSZ w Koninie dominowały osoby, które rozwiązały zadanie w ciągu 4 minut.

Mediana

Szereg nieparzysty (n = 17), więc[7]:

$$k = \frac{N+1}{2} = (17+1)/2 = 9,\ cumn_i \ge k\ \rightarrow\ cumn_i \ge 9 \rightarrow\ M_e = 5.$$

Połowa studentów rozwiązywała zadanie nie dłużej niż 5 minut, a połowa nie krócej niż 5 minut.

Kwartyl pierwszy

Szereg nieparzysty (n = 17), więc[8]:

$$k = \frac{N+1}{4} = (17+1)/4 = 4{,}5,\ cumn_i \ge k\ \rightarrow\ cumn_i \ge 4{,}5 \rightarrow\ Q_1 = 4.$$

1/4 studentów rozwiązywała zadanie nie dłużej niż 4 minuty, a 3/4 nie krócej niż 4 minuty.

[7] Gdyby szereg był parzysty (np. n = 18), to $k = N/2 = 18/2 = 9$, $cumn_i \ge k \rightarrow cumn_i \ge 9 \rightarrow M_e = 5$.

[8] Gdyby szereg był parzysty (np. n = 18), to $k = N/4 = 18/4 = 4{,}5$, $cumn_i \ge k \rightarrow cumn_i \ge 4{,}5 \rightarrow Q_1 = 4$.

Kwartyl trzeci

Szereg nieparzysty (n = 17), więc[9]:

$$k = \frac{3*(N+1)}{4} = 3*(17+1)/4 = 13{,}5 \,,\ cumn_i \geq k \ \rightarrow\ cumn_i \geq 13{,}5 \rightarrow Q_3 = 7\,.$$

3/4 studentów rozwiązywało zadanie nie dłużej niż 7 minut, a 1/4 nie krócej niż 7 minut.

Przykład 2.7.3.

Wysokość wydatków na pieczywo (w zł) poniesionych w grudniu 2009 r. przez 200 losowo wybranych gospodarstw domowych z Konina kształtowała się następująco:

Wydatki (w zł)	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Liczba gospodarstw	12	25	37	62	34	22	8

Źródło: Dane umowne.

W oparciu o powyższe informacje należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć miary położenia, dyspersji, asymetrii i koncentracji,
- wyznaczyć graficznie dominantę, medianę i kwartyle.

Rozwiązanie

Zbiorowość statystyczna – losowo wybrane gospodarstwa domowe z Konina
Jednostka statystyczna – losowo wybrane gospodarstwo domowe z Konina
Cecha statystyczna (zmienna) – wysokość wydatków na pieczywo (w zł) poniesionych w grudniu 2009 r.

[9] Gdyby szereg był parzysty (np. n = 18), to $k = 3N/4 = 54/4 = 13{,}5$, $cumn_i \geq k \ \rightarrow$ $cumn_i \geq 13{,}5 \rightarrow Q_1 = 7$.

Średnia arytmetyczna ważona

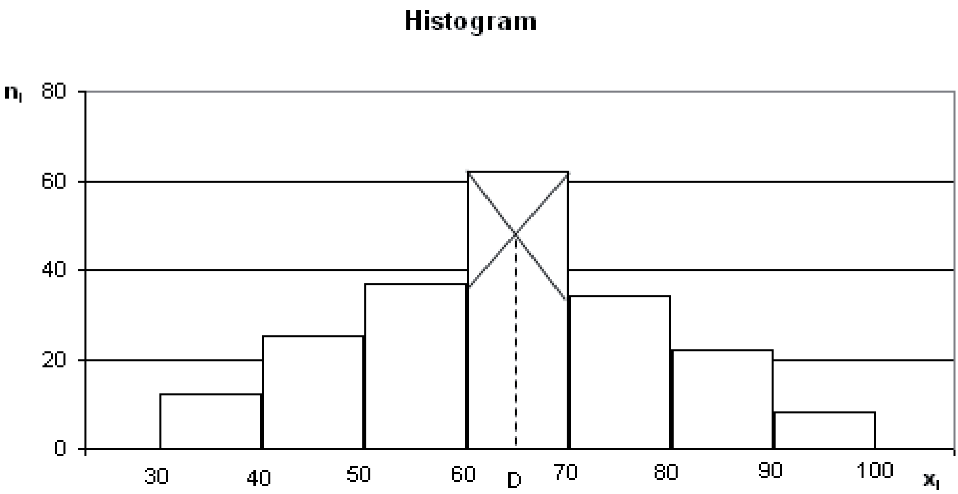
$$\bar{x}_a = \frac{\sum_{i=1}^{k} x'_i \cdot n_i}{N} = 12790/200 = 63{,}95\ .$$

Przeciętne wydatki na pieczywo poniesione w grudniu 2009 r. przez losowo wybrane gospodarstwa domowe wynosiły 63,95 zł.

Dominanta

$$D = x_0 + c_0 \cdot \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) + (n_d - n_{d+1})} = 60 + 10 * \frac{62-37}{(62-37)+(62-34)} = 64{,}72$$

Wśród losowo wybranych gospodarstw dominowały te, które w grudniu 2009 r. przeznaczyły na pieczywo 64,72 zł.



Mediana

Szereg parzysty (n = 200), więc:

$$k = \frac{N}{2} = 200/2 = 100\ ,\ cumn_i \geq k\ \rightarrow\ cumn_i \geq 100\ ,$$

$$M_e = Q_2 = x_0 + \frac{c_0}{n_0} \cdot \left(\frac{N}{2} - cum_{n-1}\right) = 60 + \frac{10}{62} \cdot (100 - 74) = 64{,}19\ .$$

Połowa gospodarstw przeznaczyła w grudniu 2009 r. na pieczywo nie więcej niż 64,19 zł, a połowa nie mniej niż 64,19 zł.

Kwartyl pierwszy

Szereg parzysty (n = 200), więc:

$$k = \frac{N}{4} = 200/4 = 50\,,\ cumn_i \geq k \ \rightarrow\ cumn_i \geq 50\,,$$

$$Q_1 = x_0 + \frac{c_0}{n_0} \cdot \left(\frac{N}{4} - cum_{n-1}\right) = 50 + \frac{10}{37} \cdot (50 - 37) = 53{,}51\,.$$

1/4 gospodarstw przeznaczyła w grudniu 2009 r. na pieczywo nie więcej niż 53,51 zł, a 3/4 nie mniej niż 53,51 zł.

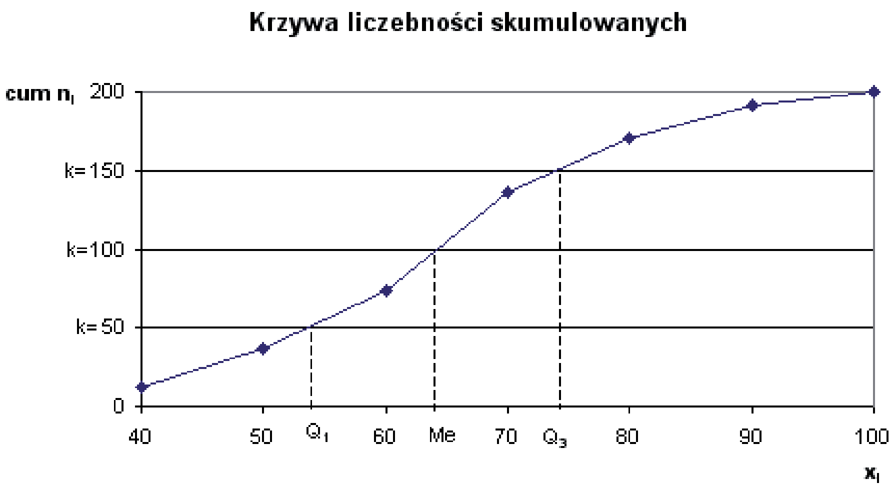
Kwartyl trzeci

Szereg parzysty (n = 200), więc:

$$k = \frac{3N}{4} = 600/4 = 150\,,\ cumn_i \geq k \ \rightarrow\ cumn_i \geq 150\,,$$

$$Q_3 = x_0 + \frac{c_0}{n_0} \cdot \left(\frac{3N}{4} - cum_{n-1}\right) = 70 + \frac{10}{34} \cdot (150 - 136) = 74{,}12\,.$$

3/4 gospodarstw przeznaczyło w grudniu 2009 r. na pieczywo nie więcej niż 74,12 zł, a 1/4 nie mniej niż 74,12 zł.



Rozstęp (empiryczny obszar zmienności)

$$R = x_{\max} - x_{\min} = 100 - 30 = 70$$

Różnica między gospodarstwem z najwyższymi wydatkami na pieczywo a gospodarstwem z najniższymi wydatkami wynosiła 70 zł.

Odchylenie standardowe

$$s(x) = \sqrt{\frac{\sum_{i=1}^{k}\left(x'_i - \bar{x}\right)^2 \cdot n_i}{N}} = \sqrt{\frac{43679}{200}} = 14,78$$

Wydatki na pieczywo w grupie losowo wybranych gospodarstw domowych odchylały się od wydatków średnich przeciętnie o 14,78 zł.

Typowy obszar zmienności

$$\bar{x} - s(x) < x_{typ} < \bar{x} + s(x)$$

$$63,95 - 14,78 < x_{typ} < 63,95 + 14,78$$

$$49,17 < x_{typ} < 78,73$$

Około 2/3 gospodarstw domowych przeznaczyło w grudniu 2009 r. na pieczywo od 49,17 do 78,73 zł.

Klasyczny współczynnik zmienności

$$V_x = \frac{s(x)}{\bar{x}} \cdot 100 = \frac{14,78}{63,95} \cdot 100 = 23,11\%$$

Gospodarstwa domowe są jednorodne pod względem wysokości poniesionych wydatków na pieczywo, a średnia arytmetyczna dobrze charakteryzuje przeciętne wydatki[10].

Wskaźnik asymetrii

$$A_s = \bar{x} - D = 63,95 - 64,72 = -0,77$$

Rozkład wydatków poniesionych na pieczywo przez losowo wybrane gospodarstwa domowe jest asymetryczny lewostronnie (ujemnie), tzn. większość gospodarstw poniosła wydatki wyższe od wydatków przeciętnych[11].

[10] Wraz ze wzrostem wartości współczynnika wzrasta zróżnicowanie zbiorowości, a tym samym maleje wartość poznawcza średniej arytmetycznej.

[11] Wartość wskaźnika większa od zera oznaczałaby, że rozkład wydatków poniesionych na pieczywo jest asymetryczny prawostronnie (dodatnio), tzn. że większość gospodarstw poniosła wydatki niższe od wydatków przeciętnych. Z kolei wartość wskaźnika równa zero oznaczałaby, że rozkład wydatków poniesionych na pieczywo jest symetryczny, tzn. że większość gospodarstw poniosła wydatki równe wydatkom przeciętnym.

Klasyczno-pozycyjny współczynnik skośności

$$W_{as} = \frac{\bar{x} - D}{s(x)} = \frac{63{,}95 - 64{,}72}{14{,}78} = -0{,}05$$

Rozkład wydatków poniesionych na pieczywo przez losowo wybrane gospodarstwa domowe jest asymetryczny lewostronnie (ujemnie), tzn. większość gospodarstw poniosła wydatki wyższe od wydatków przeciętnych, a nasilenie tej asymetrii jest bardzo małe[12].

Czwarty moment centralny standaryzowany

$$\mu_4 = \frac{\sum_{i=1}^{k}\left(x'_i - \bar{x}\right)^4 \cdot n_i}{N} = \frac{24152669{,}64}{200} = 120763{,}35$$

$$\alpha_4 = \frac{\mu_4}{s^4(x)} = \frac{120763{,}35}{47719{,}70} = 2{,}53$$

Rozkład wydatków poniesionych na pieczywo przez losowo wybrane gospodarstwa domowe ma charakter spłaszczony, tzn. koncentracja wysokości wydatków wokół wydatków przeciętnych jest słabsza niż w rozkładzie normalnym[13].

Obliczenia pomocnicze – przykład 2.7.3.

Wydatki w zł (x_i)	Liczba gospodarstw (n_i)	x_i'	$x_i' \cdot n_i$	cum_{ni}	$x_i' - \bar{x}$	$(x_i' - \bar{x})^2$	$(x_i' - \bar{x})^2 \cdot n_i$	$(x_i' - \bar{x})^4$	$(x_i' - \bar{x})^4 \cdot n_i$
30-40	12	35	420	12	-28,95	838,10	10 057,20	702 415,80	8 428 989,60
40-50	25	45	1 125	37	-18,95	359,10	8 977,50	128 954,61	3 223 865,25
50-60	37	55	2 035	74	-8,95	80,10	2 963,70	6 416,41	237 407,17
60-70	62	65	4 030	136	1,05	1,10	68,20	1,22	75,64
70-80	34	75	2 550	170	11,05	122,10	4 151,40	14 909,02	506 906,68
80-90	22	85	1 870	192	21,05	443,10	9 748,20	19 6339,83	4 319 476,26
90-100	8	95	760	200	31,05	964,10	7 712,80	929 493,63	7 435 949,04
Σ	200	x	12 790	x	x	x	43 679,00	x	24 152 669,64

[12] Wraz ze wzrostem wartości współczynnika nasilenie asymetrii wzrasta.

[13] Wartość wskaźnika większa od 3 oznaczałaby, że rozkład wydatków poniesionych na pieczywo ma charakter wysmukły, tzn. koncentracja wysokości wydatków wokół wydatków przeciętnych jest silniejsza niż w rozkładzie normalnym. Z kolei wartość wskaźnika równa 3 oznaczałaby, że rozkład wydatków poniesionych na pieczywo jest rozkładem normalnym.

Obliczenia pomocnicze – przykład 2.7.4.

Grupy województw wg liczby spółek z udziałem kapitału zagranicznego	Liczba województw	Liczba spółek	Odsetek województw	Odsetek spółek	Skumulowany odsetek	
					województw	spółek
poniżej 200	2	275	12,50	1,48	12,50	1,48
200-500	4	1 329	25,00	7,18	37,50	8,66
500-1000	3	2 121	18,75	11,46	56,25	20,12
1000-2000	5	6 812	31,25	36,79	87,50	56,91
2000 i więcej	2	7 978	12,50	43,09	100,00	100,00
Σ	16	18 515	100,00	100,00	X	X

Przykład 2.7.4.

Na podstawie danych zawartych w poniższej tabeli należy ustalić za pomocą metody graficznej i rachunkowej koncentrację spółek z udziałem kapitału zagranicznego w polskich województwach w 2007 r.

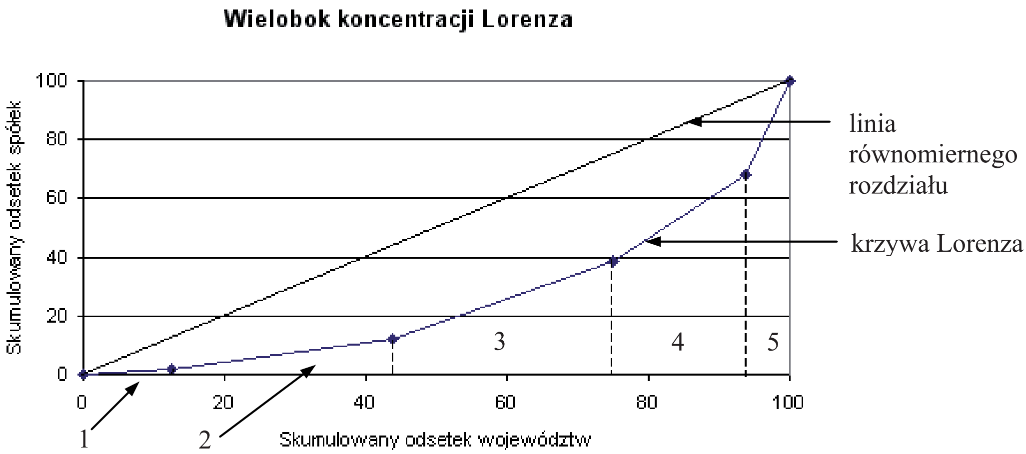
Grupy województw wg liczby spółek z udziałem kapitału zagranicznego	Liczba województw	Liczba spółek
poniżej 200	2	275
200-500	4	1 329
500-1000	3	2 121
1000-2000	5	6 812
2000 i więcej	2	7 978

Źródło: Opracowanie własne na podstawie Banku Danych Regionalnych GUS, 2007.

Rozwiązanie

O występowaniu koncentracji spółek z udziałem kapitału zagranicznego w polskich województwach świadczy zarejestrowanie w dwóch województwach (mazowieckie i dolnośląskie) ponad 40% wszystkich spółek.

<u>Metoda graficzna – wielobok koncentracji Lorenza</u>



Z wykresu wynika, że koncentracja spółek z udziałem kapitału zagranicznego w polskich województwach w 2007 r. była dość silna[14].

<u>Metoda rachunkowa – współczynnik koncentracji</u>

Wyszczególnienie	Pola figur (P_i)
pole trójkąta	(1,48*12,5)/2 = 9,25
pole I trapezu	[(8,66+1,48)*25,00]/2 = 126,75
pole II trapezu	[(20,12+8,66)*18,75]/2 = 269,81
pole III trapezu	[(56,91+20,12)*31,25]/2 = 1 203,59
pole IV trapezu	[(100+56,91)*12,50]/2 = 980,69
Σ	2 590,59

$$k = \frac{5000 - \sum P_i}{5000} = \frac{5000 - 2590{,}59}{5000} = 0{,}482$$

Współczynnik równy 0,482 świadczy o dość silnej koncentracji spółek z udziałem kapitału zagranicznego w polskich województwach w 2007 r.

[14] Koncentracja jest tym większa, im większa jest powierzchnia zawarta między krzywą Lorenza a linią równomiernego rozdziału.

2.8. Zadania

Zadanie 2.8.1.

W poniższej tabeli przedstawiono wartość aktywów i pasywów firmy „X" S.A. w latach 2005-2008 (w tys. zł).

WYSZCZEGÓLNIENIE	2005	2006	2007	2008
AKTYWA				
A. Aktywa trwałe	101 979	124 594	158 078	191 989
B. Aktywa obrotowe	153 506	175 468	200 820	213 731
Aktywa razem	255 485	300 062	358 898	405 720
PASYWA				
A. Kapitał własny	88 654	116 012	178 263	211 890
B. Zobowiązania i rezerwy na zobowiązania	166 831	184 050	180 635	193 830
Pasywa razem	255 485	300 062	358 898	405 720

Źródło: Opracowanie własne na podstawie sprawozdań finansowych firmy „X" S.A.

Na podstawie powyższych danych należy obliczyć i zinterpretować wskaźniki struktury aktywów i pasywów.

Zadanie 2.8.2.

W poniższej tabeli przedstawiono wartość aktywów i pasywów firmy „Y" sp. z o.o. w latach 2005-2008 (w tys. zł).

WYSZCZEGÓLNIENIE	2005	2006	2007	2008
AKTYWA				
A. Aktywa trwałe	4 093	5 785	4 650	4 954
B. Aktywa obrotowe	6 036	9 459	9 056	10 372
Aktywa razem	10 129	15 244	13 706	15 326
PASYWA				
A. Kapitał własny	7 693	11 382	10 124	11 201
B. Zobowiązania i rezerwy na zobowiązania	2 436	3 862	3 582	4 125
Pasywa razem	10 129	15 244	13 706	15 326

Źródło: Opracowanie własne na podstawie sprawozdań finansowych firmy „Y" S.A.

Na podstawie powyższych danych należy obliczyć i zinterpretować wskaźniki struktury aktywów i pasywów.

Zadanie 2.8.3.

Na podstawie poniższych danych dotyczących budżetu miasta Konina w 2007 r. (źródło: Bank Danych Regionalnych GUS, 2007) należy ustalić wskaźniki natężenia, tj. wysokość dochodów i wydatków przypadających na jednego mieszkańca (*per capita*):

- dochody budżetowe – 357 991,68 tys. zł,
- wydatki budżetowe – 385 376,12 tys. zł,
- liczba mieszkańców w roku 2007 (stan na 30 VI) – 80 195 osób.

Zadanie 2.8.4.

Na podstawie poniższych danych dotyczących budżetu gminy Ślesin w 2007 r. (źródło: Bank Danych Regionalnych GUS, 2007) należy ustalić wskaźniki natężenia, tj. wysokość dochodów i wydatków przypadających na jednego mieszkańca (*per capita*):

- dochody budżetowe – 31 550,98 tys. zł,
- wydatki budżetowe – 34 346,32 tys. zł,
- liczba mieszkańców w roku 2007 (stan na 30 VI) – 13 519 osób.

Zadanie 2.8.5.

W styczniu 2010 r. odnotowano następujące zużycie wody (w m^3) wśród gospodarstw domowych zamieszkałych w budynku przy ulicy Wesołej 4 w Koninie (źródło: dane umowne):

10, 12, 4, 15, 7, 6, 8, 4, 9, 7, 12, 6, 13, 3, 11, 10, 12, 5, 2, 8.

W oparciu o powyższe informacje należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz wskazać przeciętne pozycyjne (dominantę, medianę, kwartyle).

Zadanie 2.8.6.

Cena kilograma truskawek (w zł) na 19 stoiskach konińskiego targowiska kształtowały się 20 czerwca 2009 r. ceny kilograma truskawek (w zł) na 19 stoiskach zlokalizowanych na konińskim targowisku kształtowały się następująco (źródło: dane umowne):

3,50; 3,40; 4,60; 3,70; 3,20; 4,10; 5,00; 4,80; 3,60; 3,50; 3,90; 4,00; 4,30; 4,70; 4,50; 3,50; 3,70; 4,00; 3,50.

W oparciu o powyższe informacje należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz wskazać przeciętne pozycyjne (dominantę, medianę, kwartyle).

Zadanie 2.8.7.

W urzędzie gminy dokonano przeglądu list obecności pracowników z IV kwartału 2009 r., co pozwoliło ustalić liczbę spóźnień do pracy. Pogrupowane wyniki odpowiednich obliczeń zostały przedstawione w poniższej tabeli.

Liczba spóźnień	1	2	3	4	5
Liczba pracowników	8	6	4	4	3

Źródło: Dane umowne.

Na podstawie powyższych informacji należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz wskazać przeciętne pozycyjne (dominantę, medianę, kwartyle).

Zadanie 2.8.8.

Pracowników firmy „Igrek" zapytano o wielkość ich gospodarstw domowych mierzoną liczbą osób. Uzyskane informacje pogrupowano, tworząc poniższy szereg (stan na 31 grudnia 2009 r.).

Liczba osób w gospodarstwie	2	3	4	5	6	7
Liczba pracowników	14	25	32	17	10	2

Źródło: Dane umowne.

Na podstawie powyższych informacji należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz wskazać przeciętne pozycyjne (dominantę, medianę, kwartyle).

Zadanie 2.8.9.

Poniższy szereg przedstawia rozkład powierzchni mieszkań w bloku nr 6 przy ul. Nowej w Kole. Opierając się na dostępnych informacjach należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz oszacować przeciętne pozycyjne,
- wyznaczyć graficznie dominantę, medianę i kwartyle.

Powierzchnia (w m^2)	15-25	25-35	35-45	45-55	55-65	65-75	75-85
Liczba mieszkań	4	8	14	18	28	10	8

Źródło: Dane umowne.

Zadanie 2.8.10.

Poniższa tabela przedstawia rozkład firm zlokalizowanych w Słupcy ze względu na wartość ich aktywów (stan na 31 grudnia 2009 r.). Opierając się na dostępnych informacjach należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć średnią arytmetyczną oraz oszacować przeciętne pozycyjne,
- wyznaczyć graficznie dominantę, medianę i kwartyle.

Wartość aktywów (w tys. zł)	0-50	50-100	100-150	150-200	200-250	250-300	300-350
Liczba firm	7	10	12	9	4	2	1

Źródło: Dane umowne.

Zadanie 2.8.11.

Ceny samochodów oferowanych do sprzedaży na konińskiej giełdzie w dniu 13 września 2009 r. kształtowały się następująco:

Cena samochodu (w tys. zł)	0-4	4-8	8-12	12-16	16-20	20-24	24-28	28-32
Liczba samochodów	5	5	6	30	25	15	10	4

Źródło: Dane umowne.

W oparciu o powyższe informacje należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć miary położenia, dyspersji i asymetrii,
- wyznaczyć graficznie dominantę, medianę i kwartyle.

Zadanie 2.8.12.

Czas obsługi przy kasie 200 losowo wybranych klientów konińskiego hipermarketu „Zet" w dniu 30 września 2009 r. przedstawiał się następująco:

Czas obsługi (w sekundach)	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180
Liczba klientów	8	16	32	50	60	20	10	4

Źródło: Dane umowne.

W oparciu o powyższe informacje należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć miary położenia, dyspersji i asymetrii,
- wyznaczyć graficznie dominantę, medianę i kwartyle.

Zadanie 2.8.13.

Dysponując informacjami na temat wieku pracowników dwóch firm (A i B) zlokalizowanych w Koninie, należy przeprowadzić kompleksową analizę porównawczą badanych zbiorowości (stan na 31 grudnia 2009 r.).

Wiek (w latach)	24-26	26-28	28-30	30-32	32-34	34-36	36-38	38-40	40-42
Liczba pracowników firmy A	44	56	92	31	27	25	14	8	3
Liczba pracowników firmy B	2	6	29	35	47	56	64	79	32

Źródło: Dane umowne.

Zadanie 2.8.14.

Dysponując informacjami na temat liczby punktów uzyskanych z pracy kontrolnej ze statystyki opisowej przez studentów dwóch specjalności ekonomicznych PWSZ w Koninie w roku akademickim 2008/2009, należy przeprowadzić kompleksową analizę porównawczą badanych zbiorowości.

Liczba punktów	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18
Liczba studentów specjalność FiRP	2	5	9	18	34	27	8	5	2
Liczba studentów Specjalność ZIiN	6	7	17	20	18	11	6	3	2

Źródło: Dane umowne.

Zadanie 2.8.15.

Struktura Polaków ze względu na przynależność do poszczególnych grup wiekowych w 2008 r. przedstawiała się następująco:

Wiek (w latach)	19 i mniej	20-39	40-59	60-79	80 i więcej
Liczba osób (w tys.)	8 449,7	11 841,3	10 765,8	5 878,8	1 200,2

Źródło: Bank Danych Regionalnych GUS, 2008.

Na podstawie powyższych informacji należy obliczyć średnią arytmetyczną.

Zadanie 2.8.16.

W drodze badania ankietowego zapytano użytkowników 50 gospodarstw domowych o wysokość ich miesięcznych wydatków na żywność (w zł). Otrzymano następujące informacje (źródło: dane umowne):

- 6 gospodarstw wydaje mniej niż 800 zł,
- 16 gospodarstw wydaje mniej niż 1000 zł,
- 34 gospodarstwa wydają mniej niż 1200 zł,
- 44 gospodarstwa wydają mniej niż 1400 zł.

Biorąc po uwagę, że maksymalna wysokość wydatków wśród gospodarstw objętych badaniem wynosi nie więcej niż 1600 zł, a minimalna nie mniej niż 600 zł, wyznaczyć średnią wysokość miesięcznych wydatków na żywność.

Zadanie 2.8.17.

Znane są informacje o stażu pracy pracowników pewnego przedsiębiorstwa (źródło: dane umowne):

x_{min} = 4 lata,
x_{max} = 17 lat.

Która z podanych niżej liczb jest średnią arytmetyczną stażu pracy wszystkich pracowników tego przedsiębiorstwa?

3, 20, 9.

Zadanie 2.8.18.

Struktura wieku osób, które zawarły związek małżeński w Polsce w 2008 r., przedstawiała się następująco:

Wiek nowożeńców (w latach)	19 i mniej	20-24	25-29	30-39	40-49	50 i więcej
Liczba nowożeńców	16 173	158 165	206 194	94 643	18 198	18 158

Źródło: Bank Danych Regionalnych GUS, 2008.

Na podstawie powyższych informacji należy oszacować dominantę.

Zadanie 2.8.19.

Obserwacją objęto pracowników pewnego zakładu ze względu na ich wzrost. Rezultaty obserwacji prezentuje poniższa tabela.

Parametry	Wzrost w cm
$\bar{x}_a$	175,0
M_e	173,5
D	172,0
s(x)	9,0

Źródło: Dane umowne.

Podany zespół parametrów należy uzupełnić i zinterpretować otrzymane wyniki.

Zadanie 2.8.20.

Oszacuj medianę, dysponując następującymi danymi:

- liczebność zbiorowości jest równa 60,
- dolna granica przedziału mediany wynosi 8,
- rozpiętość przedziału mediany jest równa 2,
- suma liczebności przedziałów poprzedzających przedział mediany wynosi 27,
- suma liczebności przedziałów wraz z przedziałem mediany jest równa 39.

Zadanie 2.8.21.

Analiza czasu dojazdu do pracy (w minutach) osób zatrudnionych w pewnej firmie dostarczyła następujących informacji:

Parametry	Mężczyźni	Kobiety
$\bar{x}_a$	50	45
M_e	—	44
D	52	—
s(x)	6	4

Źródło: Dane umowne.

Wykorzystując te informacje należy:
- uzupełnić zespół parametrów w obydwu grupach,
- przeprowadzić analizę porównawczą zbiorowości.

Zadanie 2.8.22.

Analiza wysokości miesięcznych wynagrodzeń (w zł) 30 mężczyzn i 10 kobiet zatrudnionych w pewnym przedsiębiorstwie dostarczyła następujących informacji:

Parametry	Mężczyźni	Kobiety
$\bar{x}_a$	1800	1600
M_e	1900	1500
D	2000	1400
s(x)	200	300

Źródło: Dane umowne.

Wykorzystując te informacje, należy ustalić:
- która grupa pracowników uzyskuje wyższe wynagrodzenie i dlaczego,
- ile wynosi średnie wynagrodzenie w odniesieniu do wszystkich pracowników.

Zadanie 2.8.23.

Dysponując poniższymi wielkościami wyznaczonymi na podstawie szeregu przedziałowego, należy scharakteryzować za pomocą parametrów opisowych rozkład dochodów (w tys. zł) w próbie 100 gospodarstw domowych oraz naszkicować przybliżony wykres krzywej liczebności i zaznaczyć położenie miar tendencji centralnej.

$$\sum_{i=1}^{k} x_i' \cdot n_i = 760,$$

$$\sum_{i=1}^{k} (x_i' - \bar{x})^2 \cdot n_i = 484,$$

$$\sum_{i=1}^{k} (x_i' - \bar{x})^4 \cdot n_i = 7196{,}82,$$

$$D = 7{,}5,$$

$$M_e = 7{,}6.$$

Zadanie 2.8.24.

Poniższy szereg rozdzielczy przedstawia strukturę osób ze względu na wiek, które rozwiodły się w Polsce w 2008 r.

Wiek rozwodzących się osób (w latach)	24 i mniej	25-29	30-39	40-49	50 i więcej
Odsetek osób	4,21	15,95	39,21	24,34	16,29

Źródło: Bank Danych Regionalnych GUS, 2008.

Na podstawie powyższych informacji należy:
- określić zbiorowość, jednostkę i cechę statystyczną,
- obliczyć miary położenia, dyspersji i asymetrii.

Zadanie 2.8.25.

Na podstawie danych zawartych w poniższej tabeli należy ustalić za pomocą metody graficznej i rachunkowej koncentrację ludności w polskich miastach w 2008 r.

Grupy miast wg liczby ludności (w tys. osób)	Liczba miast	Liczba ludności (w tys. osób)
poniżej 10	492	2 273,4
10-20	180	2 643,9
20-50	134	4 192,3
50-100	47	3 211,0
100-200	22	3 044,4
200 i więcej	17	7 923,1

Źródło: *Ludność. Stan i struktura w przekroju terytorialnym. Stan w dniu 31 XII 2008 r.*, <www.stat.gov.pl>.

Zadanie 2.8.26.

Na podstawie danych zawartych w poniższej tabeli należy ustalić za pomocą metody graficznej i rachunkowej koncentrację powierzchni gospodarstw rolnych w Polsce w 2007 r.

Grupy gospodarstw wg powierzchni (w ha)	Liczba gospodarstw	Powierzchnia gospodarstw (w ha)
poniżej 2	1 193 583	944 139
2-5	613 978	1 989 940
5-10	399 868	2 836 174
10-20	243 909	3 352 972
20-50	102 315	2 955 545
50 i więcej	21 461	2 339 429

Źródło: *Charakterystyka gospodarstw rolnych w 2007 r.*, <www.stat.gov.pl>.

3. ANALIZA WSPÓŁZALEŻNOŚCI ZJAWISK

3.1. Informacje ogólne

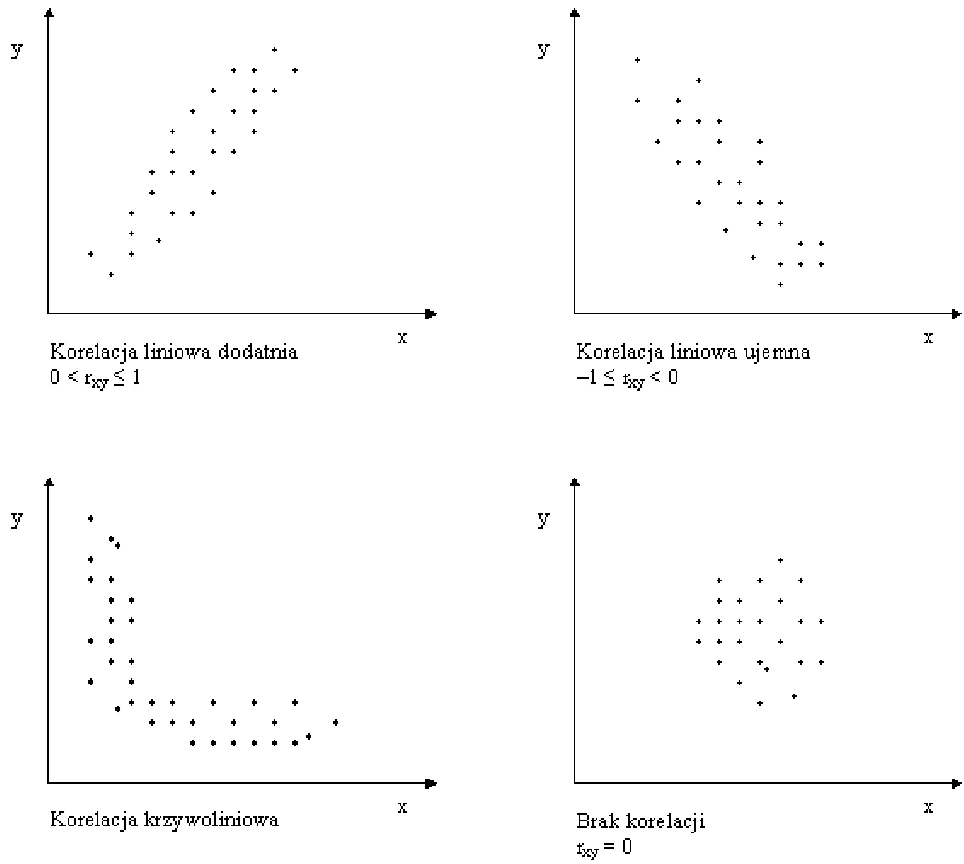
Jednostki tworzące zbiorowość statystyczną charakteryzowane są zazwyczaj za pomocą wielu cech zmiennych, które nierzadko pozostają ze sobą w pewnym związku. Określenie siły, kierunku oraz kształtu tego związku możliwe jest dzięki **analizie współzależności zjawisk**.

Zakres analizy współzależności zjawisk:
- analiza korelacji,
- analiza regresji.

3.2. Analiza korelacji

Analiza korelacji pozwala określić siłę zależności między zmiennymi, a w przypadku zależności liniowej dwóch zmiennych – także kierunek tej zależności.

Siłę związku między zmiennymi określa się za pomocą szeregu miar, których wybór zależy od tego, czy zmienne mają charakter mierzalny, czy też niemierzalny. W pierwszym przypadku można wykorzystać **współczynnik korelacji liniowej Pearsona**, natomiast w drugim m.in. **skorygowany współczynnik kontyngencji**, **współczynnik Czuprowa** i **współczynnik zbieżności V-Cramera**.



Schemat 3.1. Rodzaje korelacji dwóch zmiennych mierzalnych

Źródło: Opracowanie własne.

3.2.1. Współczynnik korelacji liniowej Pearsona

Współczynnik korelacji Pearsona określa kierunek i siłę zależności dwóch zmiennych mierzalnych. Przyjmuje on wartości z przedziału <–1; 1>, przy czym im jego wartość jest bliższa 1 lub –1, tym zależność jest silniejsza, a im bliższa 0, tym zależność jest słabsza.

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{n \cdot s(x) \cdot s(y)} \quad \text{lub} \quad r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x} \cdot \overline{y}}{s(x) \cdot s(y)},$$

gdzie:

$$s(x) = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n}} \quad \text{lub} \quad s(x) = \sqrt{\frac{\sum_{i=1}^{n} x^2}{n} - (\overline{x})^2},$$

$$s(y) = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n}} \quad \text{lub} \quad s(y) = \sqrt{\frac{\sum_{i=1}^{n} y^2}{n} - (\overline{y})^2},$$

x_i – wartość zmiennej x,
y_i – wartość zmiennej y,
$\overline{x}$ – średnia arytmetyczna zmiennej x,
$\overline{y}$ – średnia arytmetyczna zmiennej y,
n – liczba par obserwacji.

Współczynnik korelacji można wyznaczyć również na podstawie współczynników kierunkowych liniowej funkcji regresji[15]:

$$r_{xy} = \pm\sqrt{b_y \cdot b_x},$$

gdzie:

b_x – współczynnik kierunkowy funkcji regresji $\hat{x} = a + by$,

b_y – współczynnik kierunkowy funkcji regresji $\hat{y} = a + bx$.

Własności współczynnika korelacji liniowej Pearsona:

- $r_{xy} = 0$, gdy brak współzależności,
- $r_{xy} = \pm 1$, gdy zależność funkcyjna (korelacja doskonała),
- $-1 \le r_{xy} < 0$, gdy korelacja ujemna, tzn. wraz ze wzrostem wartości jednej zmiennej maleją wartości drugiej zmiennej, i odwrotnie,
- $0 < r_{xy} \le 1$, gdy korelacja dodatnia, tzn. wraz ze wzrostem wartości jednej zmiennej rosną wartości drugiej zmiennej, i odwrotnie,
- $\left|r_{xy}\right| \le 0{,}3$, gdy korelacja niewyraźna,
- $0{,}3 < \left|r_{xy}\right| \le 0{,}5$, gdy korelacja średnia,
- $\left|r_{xy}\right| > 0{,}5$, gdy korelacja wyraźna.

[15] Zob. szerzej: rozdz. 3.3.1. Metoda Najmniejszych Kwadratów (MNK).

3.2.2. Korelacja cech jakościowych

Siłę zależności między zmiennymi niemierzalnymi można określić za pomocą skorygowanego współczynnika kontyngencji, współczynnika Czuprowa lub współczynnika zbieżności V-Cramera, które w swej konstrukcji opierają się na teście niezależności chi-kwadrat[16]. Współczynniki te przyjmują wartości z przedziału <0; 1>, przy czym ich wartość bliższa 1 oznacza, że zależność jest silniejsza, natomiast wartość bliższa 0 informuje o słabszej zależności.

3.2.2.1. Skorygowany współczynnik kontyngencji

$$C_{skor} = \frac{C_{xy}}{C_{\max}},$$

gdzie:

$$C_{xy} = \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

$$C_{\max} = \frac{\sqrt{\frac{r-1}{r}} + \sqrt{\frac{k-1}{k}}}{2},$$

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}, \qquad \hat{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n},$$

r – liczba wierszy w tablicy wielodzielczej,
k – liczba kolumn w tablicy wielodzielczej,
n – liczebność próby,
n_{ij} – liczba jednostek na przecięciu i-tego wiersza i j-tej kolumny tablicy wielodzielczej (liczebność empiryczna),
$\hat{n}_{ij}$ – liczebność teoretyczna,

[16] Przy obliczaniu testu niezależności chi-kwadrat należy oprzeć się na dostatecznie dużej próbie, oraz w taki sposób podzielić obszar wartości X i Y na grupy, aby w każdej kratce tablicy wielodzielczej znalazła się dostatecznie duża liczebność teoretyczna (≥5).

$n_{i.} = \sum_j n_{ij}$ – liczebność brzegowa liczona dla i-tego wiersza po wszystkich kolumnach,

$n_{.j} = \sum_i n_{ij}$ – liczebność brzegowa liczona dla j-tej kolumny po wszystkich wierszach.

3.2.2.2. Współczynnik Czuprowa

$$T_{xy} = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)(k-1)}}}$$

3.2.2.3. Współczynnik zbieżności V-Cramera

$$V_c = \sqrt{\frac{\chi^2}{n \cdot g}},$$

gdzie:
g = min (r – 1, k – 1).

3.3. Analiza regresji

Analiza regresji pozwala określić powiązania między zmiennymi, tj. zmienną niezależną (objaśniającą) a zmienną zależną (objaśnianą), za pomocą odpowiedniej funkcji matematycznej (funkcji regresji).

3.3.1. Metoda Najmniejszych Kwadratów (MNK)

Rozkład punktów na diagramie korelacyjnym (zob. schemat 3.1) wskazuje, czy między zmiennymi występuje związek, a jeśli tak, to umożliwia wybór typu funkcji, której wykres najlepiej opisywałaby ten związek. Jeżeli zależność ma charakter liniowy, to wówczas funkcje regresji przyjmują postać

$$\hat{y} = a_y + b_y x, \qquad \hat{x} = a_x + b_x y,$$

gdzie:
a_x, a_y – wyraz wolny,
b_x – współczynnik regresji zmiennej x względem y,
b_y – współczynnik regresji zmiennej y względem x.

Parametry powyższych funkcji można oszacować posługując się metodą najmniejszych kwadratów. Pozwala ona znaleźć takie współczynniki równań regresji, aby suma kwadratów odchyleń poszczególnych wartości empirycznych od wartości teoretycznych zmiennej była jak najmniejsza

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - a_y - b_y x)^2 = \min,$$

$$\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 = \sum_{i=1}^{n}(x_i - a_x - b_x y)^2 = \min.$$

Powyższe wyrażenia są funkcjami dwóch zmiennych a i b, więc problem sprowadza się do znalezienia minimum tych funkcji. Obliczając pochodne cząstkowe względem a i b oraz przyrównując je do zera (w celu sprawdzenia warunku koniecznego istnienia ekstremum funkcji dwóch zmiennych) otrzymujemy tzw. układ równań normalnych:

- dla funkcji regresji zmiennej y względem x:

$$\begin{cases} \sum_{i=1}^{n} y_i = n a_y + b_y \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = a_y \sum_{i=1}^{n} x_i + b_y \sum_{i=1}^{n} x_i^2 \end{cases},$$

- dla funkcji regresji zmiennej x względem y:

$$\begin{cases} \sum_{i=1}^{n} x_i = n a_x + b_x \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i = a_x \sum_{i=1}^{n} y_i + b_x \sum_{i=1}^{n} y_i^2 \end{cases}.$$

Rozwiązując powyższe układy równań otrzymujemy odpowiednio

$$b_y = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \qquad a_y = \overline{y} - b_y \cdot \overline{x}$$

oraz

$$b_x = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}, \qquad a_x = \overline{x} - b_x \cdot \overline{y}\,.$$

Współczynniki równań regresji można oszacować również pośrednio, wykorzystując relację między współczynnikiem korelacji linowej i odchyleniami standardowymi badanych zmiennych

$$b_y = r_{xy} \frac{s(y)}{s(x)}, \qquad b_x = r_{xy} \frac{s(x)}{s(y)}.$$

3.3.2. Ocena oszacowanej funkcji regresji

Ocena oszacowanej funkcji regresji (ocena stopnia „dobroci”) pozwala określić, czy funkcja regresji jest dobrze dopasowana do danych empirycznych, tzn. czy dobrze opisuje ilościową stronę zależności między badanymi zmiennymi. W celu dokonania oceny oblicza się m.in.:

- **odchylenie standardowe składnika resztowego (błąd standardowy szacunku)**, które informuje, o ile średnio wartości empiryczne zmiennej odchylają się od wartości teoretycznych obliczonych na podstawie oszacowanej funkcji regresji:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{n-2}} = s(x) \cdot \sqrt{1 - r_{xy}^2}\,,$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = s(y) \cdot \sqrt{1 - r_{xy}^2} \;;$$

- **współczynnik determinacji** określający stopień, w jakim oszacowana funkcja regresji wyjaśnia zmienność zmiennej y lub x; współczynnik przyjmuje wartości z przedziału <0, 1>; im bliżej 1, tym oszacowana funkcja regresji jest lepiej dopasowana do danych empirycznych, i odwrotnie:

$$R^2 = r_{xy}^2 \;;$$

- **współczynnik zgodności**, który wskazuje, jaka część zmienności zmiennej y lub x nie jest objaśniona za pomocą oszacowanej funkcji regresji; współczynnik przyjmuje wartości z przedziału <0, 1>; im bliżej 0, tym dopasowanie funkcji jest lepsze, i odwrotnie:

$$\varphi^2 = 1 - R^2 \;;$$

- **współczynnik zmienności resztowej**, który informuje, jaki odsetek poziomu zmiennej objaśnianej stanowią wahania losowe; im bliżej 0, tym dopasowanie jest lepsze:

$$V = \frac{S_y}{\overline{y}} \cdot 100 \;.$$

3.4. Przykłady

Przykład 3.4.1.

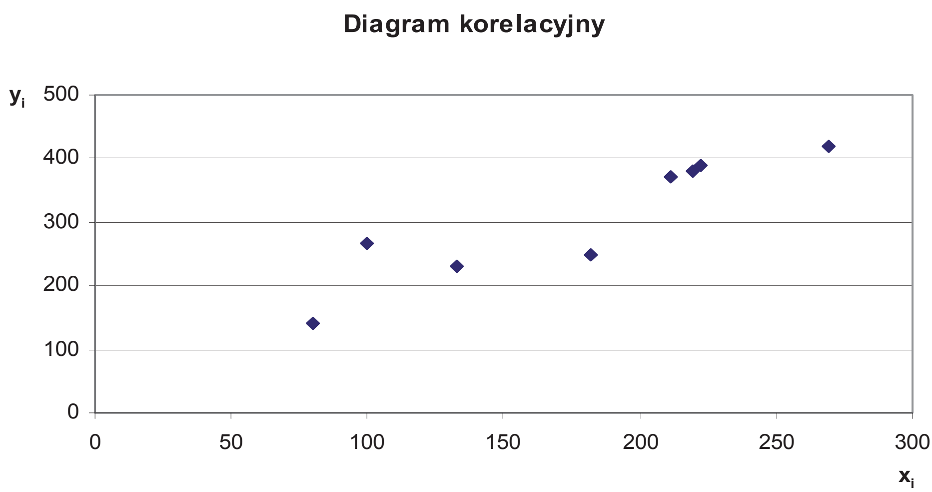
Zbadano zależność między powierzchnią nieruchomości zlokalizowanych w okolicach Konina a ich wartością. Uzyskane informacje przedstawiono w poniższej tabeli.

Powierzchnia nieruchomości (w m^2)	219	80	222	100	211	133	182	269
Wartość (w tys. zł)	380	140	390	265	370	230	248	420

Źródło: Dane umowne.

Na podstawie powyższych informacji należy:
- sporządzić diagram korelacyjny i zinterpretować rozkład punktów,
- określić kierunek i siłę związku korelacyjnego; zinterpretować otrzymany wynik,
- wyznaczyć linie regresji,
- oszacować wartość nieruchomości o powierzchni 150 m^2,
- obliczyć jaka będzie powierzchnia nieruchomości, której wartość wynosi 200 tys. zł.

Rozwiązanie



Z diagramu korelacyjnego wynika, że między powierzchnią nieruchomości a ich wartością zachodzi dodatni związek korelacyjny, tzn. wraz ze wzrostem powierzchni wzrasta wartość nieruchomości, i odwrotnie.

<u>Obliczenia pomocnicze</u>

Powierzchnia w m^2 (x_i)	Wartość w tys. zł (y_i)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})$ $*(y_i - \bar{y})$
219	380	42,00	74,62	1 764,00	5 568,14	3 134,04
80	140	-97,00	-165,38	9 409,00	27 350,54	16 041,86
222	390	45,00	84,62	2 025,00	7 160,54	3 807,90
100	265	-77,00	-40,38	5 929,00	1 630,54	3 109,26
211	370	34,00	64,62	1 156,00	4 175,74	2 197,08
133	230	-44,00	-75,38	1 936,00	5 682,14	3 316,72
182	248	5,00	-57,38	25,00	3 292,46	-286,90
269	420	92,00	114,62	8 464,00	13 137,74	10 545,04
1416	2443	x	x	30 708,00	67 997,84	41 865,00

Średnia arytmetyczna zwykła

$$\bar{x}_a = \frac{\sum_{i=1}^{N} x_i}{N} = 1416/8 = 177, \qquad \bar{y}_a = \frac{\sum_{i=1}^{N} y_i}{N} = 2443/8 = 305{,}38$$

Odchylenie standardowe

$$s(x) = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n}} = \sqrt{\frac{30708}{8}} = 61{,}96,$$

$$s(y) = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}{n}} = \sqrt{\frac{67997{,}84}{8}} = 92{,}19$$

Współczynnik korelacji liniowej Pearsona

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n \cdot s(x) \cdot s(y)} = \frac{41865}{8 \cdot 61{,}96 \cdot 92{,}19} = 0{,}92$$

Między powierzchnią nieruchomości a ich wartością zachodzi bardzo silna, dodatnia zależność[17].

Współczynnik determinacji

$$R^2 = r_{xy}^2 = 0{,}92^2 = 0{,}85$$

Zmiany wartości nieruchomość są w 85% uwarunkowane zmianami ich powierzchni, a w 15% zmianami innych czynników o charakterze losowym lub pozalosowym.

Linie regresji

$$\hat{y} = a_y + b_y x$$

[17] Zarówno w przedstawionym przykładzie, jak i zaproponowanych zadaniach oparto się na małych próbach. W związku z tym, należałoby zweryfikować istotność zależności między zmiennymi za pomocą testu dla współczynnika korelacji liniowej. Kwestia ta została jednak pominięta w niniejszym opracowaniu, bo stanowi element wnioskowania statystycznego, a nie statystyki opisowej.

$$\hat{y} - \bar{y} = r_{xy} \cdot \frac{s(y)}{s(x)} \cdot (x - \bar{x})$$

$$\hat{y} - 305,38 = 0,92 \cdot \frac{92,19}{61,96} \cdot (x - 177)$$

$$\hat{y} = 1,37x + 63,09$$

Przy wzroście powierzchni nieruchomości o 1 m^2 można spodziewać się wzrostu wartości średnio o 1,37 tys. zł.

$$\hat{x} = a_x + b_x y$$

$$\hat{x} - \bar{x} = r_{xy} \cdot \frac{s(x)}{s(y)} \cdot (y - \bar{y})$$

$$\hat{x} - 177 = 0,92 \cdot \frac{61,96}{92,19} \cdot (y - 305,38)$$

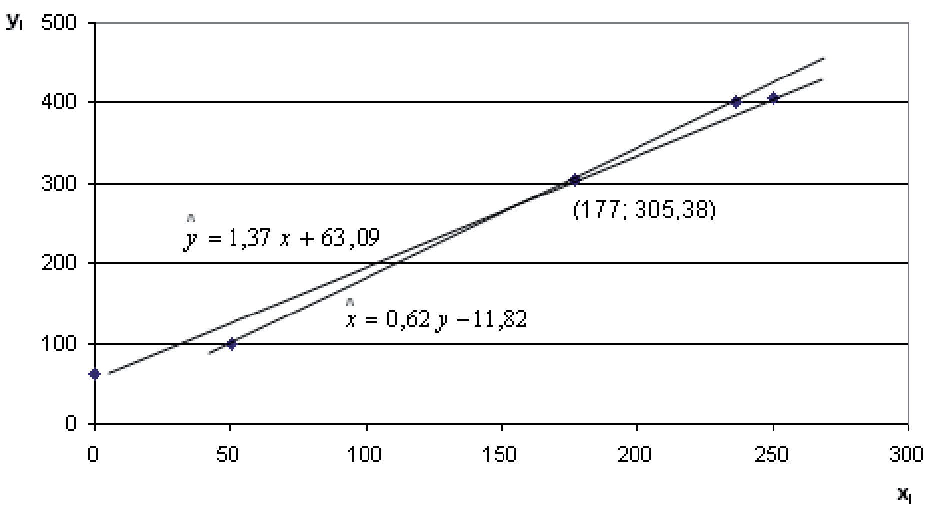
$$\hat{x} = 0,62y - 11,82$$

Wzrost wartości nieruchomości o 1 tys. zł oznacza, że powierzchnia zwiększyła się średnio o 0,62 m^2.

<u>Linie regresji – graficznie</u>

$\hat{y} = 1,37x + 63,09$ (0; 63,09) (250; 405,59)

$\hat{x} = 0,62y - 11,82$ (50,18; 100) (236,18; 400)



Wartość nieruchomości o powierzchni 150 m^2

$$\hat{y} = 1{,}37x + 63{,}09$$
$$x = 150$$
$$\hat{y} = 1{,}37 \cdot 150 + 63{,}09 = 268{,}59$$

Wartość nieruchomości o powierzchni 150 m^2 wynosiłaby 268,59 tys. zł, gdyby korelacja była doskonała (zależność funkcyjna, linie regresji pokrywałyby się, r = 1). Ponieważ tak nie jest, dokonany szacunek należy skorygować o błąd standardowy szacunku (odchylenie standardowe składnika resztowego):

$$S_y = s(y) \cdot \sqrt{1 - r_{xy}^2} = 92{,}19 \cdot \sqrt{1 - 0{,}92^2} = 36{,}13\,,$$

$$268{,}59 - 36{,}13 < \hat{y}_{x=150} < 268{,}59 + 36{,}13\,,$$

$$232{,}46 < \hat{y}_{x=150} < 304{,}72\,.$$

Wartość nieruchomości o powierzchni 150 m^2 powinna wahać się w granicach od 232,46 do 304,72 tys. zł.

Powierzchnia nieruchomości, której wartość wynosi 200 tys. zł

$$\hat{x} = 0{,}62y - 11{,}82$$
$$y = 200$$
$$\hat{x} = 0{,}62 \cdot 200 - 11{,}82 = 112{,}18$$

Powierzchnia nieruchomości, której wartość wynosi 200 tys. zł byłaby równa 112,18 m^2, gdyby korelacja była doskonała (zależność funkcyjna, linie regresji pokrywałyby się, r = 1). Ponieważ tak nie jest, dokonany szacunek należy skorygować o błąd standardowy szacunku (odchylenie standardowe składnika resztowego):

$$S_x = s(x) \cdot \sqrt{1 - r_{xy}^2} = 61{,}96 \cdot \sqrt{1 - 0{,}92^2} = 24{,}28\,,$$

$$112{,}18 - 24{,}28 < \hat{x}_{y=200} < 112{,}18 + 24{,}28\,,$$

$$87{,}9 < \hat{x}_{y=200} < 136{,}46\,.$$

Powierzchnia nieruchomości, której wartość wynosi 200 tys. zł powinna wahać się w granicach od 87,9 do 136,46 m^2.

Współczynnik zmienności resztowej

$$V = \frac{S_y}{\overline{y}} \cdot 100 = \frac{36,13}{305,38} \cdot 100 = 11,83\,\%$$

Wahania losowe stanowią 11,83% średniej wartości nieruchomości.

Przykład 3.4.2.

Jednostkowe koszty produkcji oraz wielkość produkcji pewnego dobra w siedmiu przedsiębiorstwach należących do tej samej branży zostały przedstawione w poniższej tabeli.

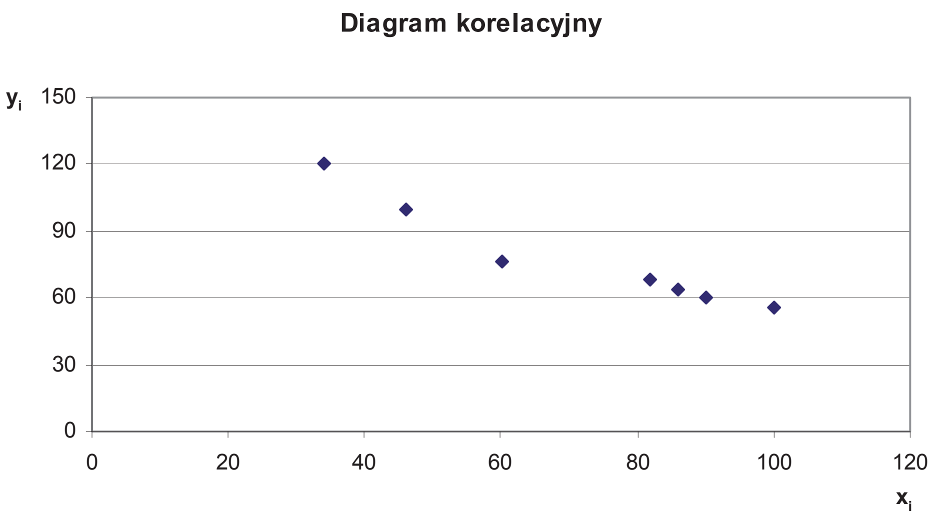
Wielkość produkcji (w tys. szt.)	34	46	60	82	86	90	100
Jednostkowe koszty produkcji (w zł)	120	100	76	68	64	60	56

Źródło: Dane umowne.

Na podstawie powyższych informacji należy:

- sporządzić diagram korelacyjny i zinterpretować rozkład punktów,
- określić kierunek i siłę związku korelacyjnego; zinterpretować otrzymany wynik,
- wyznaczyć linie regresji,
- oszacować koszt jednostkowy w przedsiębiorstwie produkującym 70 tys. sztuk,
- obliczyć ile powinna wynosić wielkość produkcji w przedsiębiorstwie, w którym koszt jednostkowy kształtuje się na poziomie 80 zł.

Rozwiązanie



Z diagramu korelacyjnego wynika, że między wielkością produkcji a jednostkowym kosztem produkcji pewnego dobra zachodzi ujemny związek korelacyjny, tzn. wraz ze wzrostem wielkości produkcji maleje koszt jednostkowy, i odwrotnie.

Obliczenia pomocnicze

Wielkość produkcji w tys. szt. (x_i)	Jednostkowe koszty produkcji w zł (y_i)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})$ $*(y_i - \bar{y})$
34	120	-37,14	42,29	1 379,59	1 788,08	-1 570,61
46	100	-25,14	22,29	632,16	496,65	-560,33
60	76	-11,14	-1,71	124,16	2,94	19,10
82	68	10,86	-9,71	117,88	94,37	-105,47
86	64	14,86	-13,71	220,73	188,08	-203,76
90	60	18,86	-17,71	355,59	313,80	-334,04
100	56	28,86	-21,71	832,73	471,51	-626,61
498	544	x	x	3 662,84	3 355,43	-3 381,72

Średnia arytmetyczna zwykła

$$\bar{x}_a = \frac{\sum_{i=1}^{N} x_i}{N} = 498/7 = 71{,}14 \qquad \bar{y}_a = \frac{\sum_{i=1}^{N} y_i}{N} = 544/7 = 77{,}71$$

Odchylenie standardowe

$$s(x) = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}} = \sqrt{\frac{3662{,}84}{7}} = 22{,}87 \qquad s(y) = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n}} = \sqrt{\frac{3355{,}43}{7}} = 21{,}89$$

Współczynnik korelacji liniowej Pearsona

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n \cdot s(x) \cdot s(y)} = \frac{-3381{,}72}{7 \cdot 22{,}87 \cdot 21{,}89} = -0{,}97$$

Między wielkością produkcji a jednostkowym kosztem produkcji pewnego dobra zachodzi bardzo silna, ujemna zależność.

Współczynnik determinacji

$$R^2 = r_{xy}^2 = (-0{,}97)^2 = 0{,}94$$

Zmiany jednostkowych kosztów produkcji pewnego dobra są w 94% uwarunkowane zmianami wielkości produkcji, a w 6% zmianami innych czynników o charakterze losowym lub pozalosowym.

Linie regresji

$$\hat{y} = a_y + b_y x$$

$$\hat{y} - \bar{y} = r_{xy} \cdot \frac{s(y)}{s(x)} \cdot (x - \bar{x})$$

$$\hat{y} - 77{,}71 = (-0{,}97) \cdot \frac{21{,}89}{22{,}87} \cdot (x - 71{,}14)$$

$$\hat{y} = -0{,}93x + 143{,}76$$

Przy wzroście wielkości produkcji o 1 tys. sztuk, można spodziewać się spadku kosztu jednostkowego średnio o 0,93 zł.

$$\hat{x} = a_x + b_x y$$

$$\hat{x} - \bar{x} = r_{xy} \cdot \frac{s(x)}{s(y)} \cdot (y - \bar{y})$$

$$\hat{x} - 71{,}14 = (-0{,}97) \cdot \frac{22{,}87}{21{,}89} \cdot (y - 77{,}71)$$

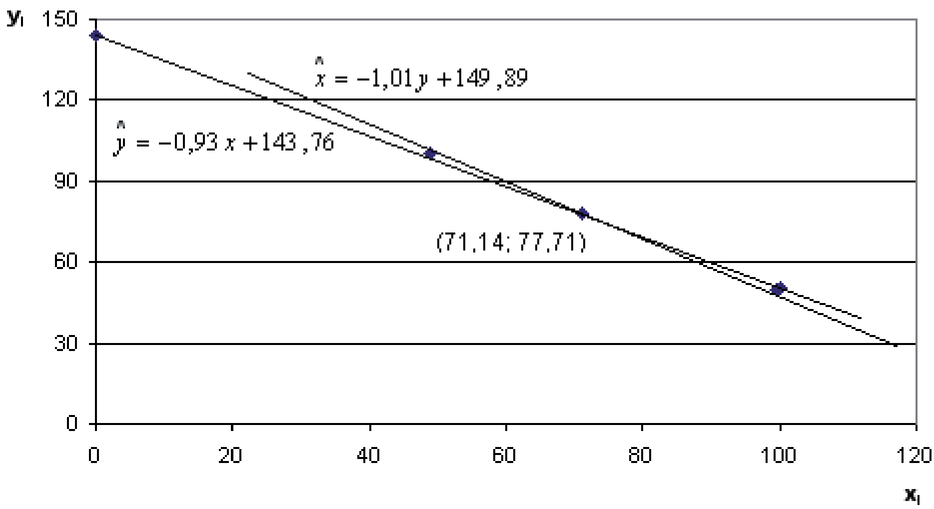
$$\hat{x} = -1{,}01\, y + 149{,}89$$

Wzrost jednostkowego kosztu produkcji o 1 zł oznacza, że wielkość produkcji zmniejszyła się średnio o 1,01 tys. sztuk.

<u>Linie regresji – graficznie</u>

$\hat{y} = -0{,}93\, x + 143{,}76$ (0; 143,76) (100; 50,76)

$\hat{x} = -1{,}01\, y + 149{,}89$ (48,89; 100) (99,39; 50)



Koszt jednostkowy w przedsiębiorstwie produkującym 70 tys. sztuk

$\hat{y} = -0{,}93\,x + 143{,}76$

x = 70

$\hat{y} = -0{,}93 \cdot 70 + 143{,}76 = 78{,}66$

Koszt jednostkowy w przedsiębiorstwie produkującym 70 tys. sztuk pewnego dobra, wynosiłby 78,66 zł, gdyby korelacja była doskonała (zależność funkcyjna, linie regresji pokrywałyby się, r = –1). Ponieważ tak nie jest, dokonany szacunek należy skorygować o błąd standardowy szacunku (odchylenie standardowe składnika resztowego):

$$S_y = s(y) \cdot \sqrt{1 - r_{xy}^2} = 21{,}89 \cdot \sqrt{1 - (-0{,}97)^2} = 5{,}32 \ ,$$

$$78{,}66 - 5{,}32 < \hat{y}_{x=70} < 78{,}66 + 5{,}32 \ ,$$

$$73{,}34 < \hat{y}_{x=70} < 83{,}98 \ .$$

Koszt jednostkowy w przedsiębiorstwie produkującym 70 tys. sztuk powinien wahać się w granicach od 73,34 do 83,98 zł.

Wielkość produkcji w przedsiębiorstwie, w którym koszt jednostkowy kształtuje się na poziomie 80 zł

$\hat{x} = -1{,}01\,y + 149{,}89$

y = 80

$\hat{x} = -1{,}01 \cdot 80 + 149{,}89 = 69{,}09$

Wielkość produkcji w przedsiębiorstwie, w którym koszt jednostkowy kształtuje się na poziomie 80 zł, wynosiłaby 69,09 tys. sztuk, gdyby korelacja była doskonała (zależność funkcyjna, linie regresji pokrywałyby się, r = –1). Ponieważ tak nie jest, dokonany szacunek należy skorygować o błąd standardowy szacunku (odchylenie standardowe składnika resztowego):

$$S_x = s(x) \cdot \sqrt{1 - r_{xy}^2} = 22{,}87 \cdot \sqrt{1 - (-0{,}97)^2} = 5{,}56 \ ,$$

$$69{,}09 - 5{,}56 < \hat{x}_{y=80} < 69{,}09 + 5{,}56 \ ,$$

$$63{,}53 < \hat{x}_{y=80} < 74{,}65 \ .$$

Wielkość produkcji w przedsiębiorstwie, w którym koszt jednostkowy kształtuje się na poziomie 80 zł, powinna wahać w granicach od 63,53 do 74,65 tys. sztuk.

<u>Współczynnik zmienności resztowej</u>

$$V = \frac{S_y}{\overline{y}} \cdot 100 = \frac{5{,}32}{77{,}71} \cdot 100 = 6{,}85\,\%$$

Wahania losowe stanowią 6,85% średniej wysokości jednostkowych kosztów produkcji.

3.5. Zadania

Zadanie 3.5.1.

Liczba zawartych małżeństw i orzeczonych rozwodów w siedmiu województwach wschodniej Polski (mazowieckie, małopolskie, lubelskie, podkarpackie, podlaskie, świętokrzyskie, warmińsko-mazurskie) kształtowała się w 2008 roku następująco:

Liczba zawartych małżeństw (w tys.)	34,0	22,0	14,8	13,9	7,8	8,6	10,0
Liczba orzeczonych rozwodów (w tys.)	9,2	4,4	1,9	2,3	1,8	1,3	2,8

Źródło: Bank Danych Regionalnych GUS, 2008.

Na podstawie danych zawartych w powyższej tabeli należy:
- sporządzić diagram korelacyjny i zinterpretować rozkład punktów,
- określić kierunek i siłę związku korelacyjnego; zinterpretować otrzymany wynik,
- wyznaczyć linie regresji,
- oszacować liczbę rozwodów w województwie, w którym zostałoby zawartych 15 tys. małżeństw,
- obliczyć ile powinna wynosić liczba zawartych małżeństw w województwie, w którym liczba orzeczonych rozwodów wynosiłaby 5 tys.

Zadanie 3.5.2.

W siedmiu województwach zachodniej Polski (lubuskie, wielkopolskie, zachodniopomorskie, dolnośląskie, opolskie, kujawsko-pomorskie, pomorskie) odnotowano w 2007 r. następującą liczbę zarejestrowanych samochodów osobowych oraz liczbę wypadków drogowych:

Liczba samochodów (w tys. sztuk)	405,9	1 483,1	570,6	1 097,2	429,6	733,8	852,9
Liczba wypadków drogowych (w tys.)	0,9	4,9	1,9	3,1	1,1	2,0	3,1

Źródło: *Transport – wyniki działalności w 2007 r.*, GUS, Warszawa 2008 oraz Bank Danych Regionalnych GUS.

Na podstawie danych zawartych w powyższej tabeli należy:
- sporządzić diagram korelacyjny i zinterpretować rozkład punktów,
- określić kierunek i siłę związku korelacyjnego; zinterpretować otrzymany wynik,
- wyznaczyć linie regresji,
- oszacować liczbę wypadków drogowych w województwie, w którym byłoby zarejestrowanych 1000 tys. samochodów osobowych,
- obliczyć ile powinno być zarejestrowanych samochodów osobowych w województwie, w którym liczba wypadków wynosiłaby 4 tys.

Zadanie 3.5.3.

W serwisie internetowym „motoAllegro” odnaleziono kilka ofert sprzedaży samochodów marki Volkswagen Golf, odnotowując jednocześnie rok produkcji aut oraz ich cenę. Informacje te przedstawiono w poniższej tabeli.

Wiek samochodu (w latach)	2	3	5	5	7	8
Cena (w tys. zł)	55	38	36	33	25	20

Źródło: Serwis internetowy „motoAllegro” z 9 lipca 2009 r., <moto.allegro.pl>.

Na podstawie powyższych informacji należy:
- sporządzić diagram korelacyjny i zinterpretować rozkład punktów,
- określić kierunek i siłę związku korelacyjnego; zinterpretować otrzymany wynik,
- wyznaczyć linie regresji,
- oszacować cenę 6-letniego samochodu marki Volkswagen Golf,
- obliczyć ile powinien wynosić wiek samochodu oferowanego za 30 tys. zł.

Zadanie 3.5.4.

Przeprowadzono badanie zależności między ceną biletów oferowanych przez sześć przedsiębiorstw transportowych za przejazd 100 km a liczbą pasażerów korzystających z usług tych przedsiębiorstw w ciągu tygodnia. Uzyskane informacje przedstawiono w poniższej tabeli.

Cena biletu za przejazd 100 km (w zł)	15	20	25	30	35	40
Liczba pasażerów	440	430	430	380	360	350

Źródło: Dane umowne.

W oparciu o powyższe informacje należy:
- sporządzić diagram korelacyjny i zinterpretować rozkład punktów,
- określić kierunek i siłę związku korelacyjnego; zinterpretować otrzymany wynik,
- wyznaczyć linie regresji,
- oszacować liczbę pasażerów przedsiębiorstwa oferującego bilety w cenie 17 zł,
- oszacować, ile będzie wynosiła cena biletu w przedsiębiorstwie, z którego usług korzysta 400 osób.

Zadanie 3.5.5.

Dane są dwie liniowe funkcje regresji:

$\hat{y} = 15{,}6 - 1{,}4x$,

$\hat{x} = 11 - 0{,}7y$,

gdzie:
x – staż pracy robotników (w latach),
y – odsetek braków powstałych przy wytwarzaniu przez robotników pewnego wyrobu (w szt.).

Określ kierunek i siłę związku korelacyjnego.

Zadanie 3.5.6.

Oszacowano liniowe funkcje regresji w postaci:

$\hat{y} = 1{,}66 + 0{,}3x$,

$\hat{x} = -4{,}6 - 2{,}9y$.

Czy funkcje regresji zostały oszacowane poprawnie?

Zadanie 3.5.7.

Oceń korelację między cechami x i y w przypadku, gdy równania regresji mają postać:

$\hat{y} = 3$,

$\hat{x} = 5$.

Zadanie 3.5.8.

W pewnym zakładzie ustalono zależność między czasem nieprzerwanej pracy (w h) a wydajnością pracy (w sztukach/h) w odniesieniu do 10 pracowników zatrudnionych bezpośrednio przy produkcji. Otrzymano współczynnik korelacji na poziomie r = -1,08. Zinterpretuj wynik.

Zadanie 3.5.9.

Poniższe równanie regresji przedstawia zależność pomiędzy czasem eksploatacji kserokopiarek należących do pewnej firmy (x – w latach) a czasem poświęcanym w ciągu miesiąca na ich konserwacje i naprawy (y – w godz.):

$\hat{y} = 0{,}98x + 2{,}64$.

Ile powinien wynosić czas przeznaczony na konserwację i naprawy 4-letniej kserokopiarki, przy założeniu, że r = 1?

Zadanie 3.5.10.

Poniższe równanie regresji przedstawia zależność pomiędzy powierzchnią sklepów obuwniczych w pewnym mieście (x – w m^2) a wysokością ich dziennego utargu (y – w tys. zł):

$$\hat{x} = 3{,}5y + 27{,}2\ .$$

Ile powinna wynosić powierzchnia sklepu, który osiąga dzienny utarg na poziomie 800 zł, przy założeniu, że r = 1.

Zadanie 3.5.11.

Zbadano zależność między wysokością dochodów własnych uzyskanych w roku 2007 przez 11 gmin miejsko-wiejskich subregionu konińskiego (x – w zł *per capita*) a odsetkiem mieszkańców tych gmin obsługiwanych przez komunalne oczyszczalnie ścieków (y – w %). Otrzymano następujące wyniki:

$$r_{xy} = 0{,}672$$

$$\overline{x} = 886{,}57$$

$$\overline{y} = 32{,}26$$

$$s(x) = 676{,}23$$

$$s(y) = 14{,}32$$

Wyznacz równanie regresji przedstawiające zależność odsetka mieszkańców obsługiwanych przez komunalne oczyszczalnie ścieków od wysokości dochodów własnych uzyskiwanych przez gminy oraz oszacuj odsetek mieszkańców obsługiwanych przez oczyszczalnie ścieków w gminie, która osiąga dochody własne na poziomie 1 tys. zł w przeliczeniu na mieszkańca.

Zadanie 3.5.12.

Badanie wybranej grupy studentów PWSZ w Koninie ze względu na odległość ich miejsca zamieszkania od uczelni (x – w km) oraz czas dojazdu do szkoły (y – w minutach) dało następujące wyniki:

- związek między zmiennymi opisuje następująca funkcja regresji:

  $$\hat{y} = 2{,}7x - 19{,}12\,,$$

- 99,74% studentów mieszkało w odległości od 1,48 do 37,35 km od uczelni,
- zróżnicowanie czasu dojazdu do szkoły, mierzone odchyleniem standardowym, wynosiło 17,84 minut.

Mając te dane należy:

- określić siłę i kierunek badanego związku korelacyjnego,
- oszacować czas dojazdu studenta mieszkającego 20 km od uczelni.

Zadanie 3.5.13.

Badanie przeprowadzone w 10 spółkach pomiędzy ceną ich akcji (x – w zł) a wysokością dywidendy wypłaconej akcjonariuszom (y – w zł) dostarczyło następujących wyników:

- przeciętna cena akcji wynosiła 65,5 zł , a średnia wysokość dywidendy wypłaconej od jednej akcji była równa 6,8 zł,
- współczynnik zmienności cen akcji wynosił 37,78%, natomiast względna dyspersja wysokości dywidendy była równa 48,24%,
- współczynnik korelacji miedzy wyróżnionymi zmiennymi wynosił 0,99.

Na podstawie powyższych informacji należy:

- wyznaczyć linię regresji prezentującą zależność wysokości dywidendy od ceny akcji,
- oszacować wysokość dywidendy dla akcji, której cena wynosiłaby 200 zł,
- określić, w jakim stopniu zmienność cen akcji wyjaśnia zmienność wysokości dywidendy.

Zadanie 3.5.14.

Zbadano zależność między wysokością dziennego obrotu (y – w zł) a liczbą zatrudnionych ekspedientek (x – w osobach) w 10 sklepach spożywczych. Liczba ekspedientek w sklepach objętych badaniem wynosiła ogółem 50 osób, a suma ich kwadratów 310. Dzienne obroty sklepów natomiast wynosiły łącznie 20880 zł, a połowa sklepów osiągnęła obroty nie wyższe niż 2072 zł. Wiadomo też, że rozkład dziennych obrotów jest asymetryczny prawostronnie, a natężenie tej asymetrii wynosi 0,05. Ponadto, okazało się, że przy wzroście liczby ekspedientek o 1 osobę można spodziewać się wzrostu dziennego obrotu średnio o 381 zł. Na podstawie tych informacji należy:

- określić siłę badanego związku i jego kierunek,
- oszacować dzienny obrót w sklepie zatrudniającym 12 ekspedientek,
- wyjaśnić, czy jest możliwe, by więcej niż 10% zmienności dziennego obrotu nie zostało wyjaśnione przez oszacowaną funkcję regresji.

Zadanie 3.5.15.

Analizując zależność między stażem pracy osób zatrudnionych w pewnej firmie (x – w latach) a wysokością ich miesięcznych zarobków (y – w tys. zł), uzyskano następujące informacje:

- przeciętny staż pracy wynosił 7,75 lat, a przeciętny zarobek 2,68 tys. zł,
- współczynnik zmienności stażu pracy wynosił 48,6%, a współczynnik dla płac 38,54%,
- współczynnik korelacji między stażem a wysokością płac był równy 0,75.

Na podstawie powyższych informacji należy:

- wyznaczyć w sposób rachunkowy linię regresji przedstawiającą zależność wysokości zarobków od stażu pracy,
- oszacować wysokość płacy dla 10-letniego stażu pracy,
- wyjaśnić, czy prawdą jest, że zmienność stażu pracy wyjaśnia zmienność zarobków zatrudnionych osób w 75%.

Zadanie 3.5.16.

W pewnym przedsiębiorstwie zebrano z lat 2000-2009 informacje o wielkości produkcji (w tys. sztuk) i całkowitych kosztach produkcji (w tys. zł) w celu określenia zależności między nimi. Okazało się, że:

- przeciętna wielkość produkcji wynosiła 149,9 tys. sztuk, a jej odchylenie standardowe 89,45 tys. sztuk,
- łączne koszty produkcji w badanym okresie wynosiły 924 tys. zł,
- współczynnik korelacji liniowej kształtował się na poziomie 0,95,
- odchylenie standardowe składnika resztowego (błąd standardowy) przy szacowaniu całkowitych kosztów produkcji względem wielkości produkcji było równe 9,84 tys. zł.

Mając powyższe dane należy:

- określić, czy przy wzroście wielkości produkcji o 1 tys. sztuk można spodziewać się zwiększenia kosztów całkowitych przeciętnie o 500 zł,
- ustalić, czy jest możliwe, aby 15% zmienności kosztów zależało od innych czynników niż wielkość produkcji,
- oszacować przy jakiej wielkości produkcji koszty wyniosą 100 tys. zł.

Zadanie 3.5.17.

Na podstawie informacji o częstotliwości korzystania z internetu przez studentów specjalności ekonomicznych PWSZ w Koninie (studia dzienne, rok akademicki 2004/2005) oraz ich miejsca zamieszkania, zbudowano poniższą tabelę korelacyjną.

Miejsce zamieszkania	Częstotliwość korzystania z internetu			
	codziennie	2-3 razy w tygodniu	raz w tygodniu	rzadziej
Konin	68	26	19	31
miasta powiatowe	13	17	13	22
inne miejscowości	28	41	56	108

Źródło: Opracowanie własne na podstawie badań ankietowych SKN „Młodych Ekonomistów".

Posługując się powyższymi informacjami należy ustalić, czy o częstotliwości korzystania z internetu decyduje miejsce zamieszkania studentów (wykorzystać skorygowany współczynnik kontyngencji, współczynnik Czuprowa oraz współczynnik zbieżności V-Cramera).

4. ANALIZA DYNAMIKI ZJAWISK

4.1. Informacje ogólne

Analiza struktury zbiorowości i analiza współzależności zjawisk pozwalają scharakteryzować zbiorowość statystyczną pod względem jednej lub kilku cech zmiennych (w izolacji lub powiązaniu), ale tylko w ujęciu statycznym. Z kolei zmiany poziomu cechy zmiennej w czasie można ocenić dzięki **analizie dynamiki zjawisk.**

Zakres analizy dynamiki zjawisk:

- metody indeksowe,
- dekompozycja szeregu czasowego.

4.2. Metody indeksowe

Zmiany, jakie zaszły w poziomie i strukturze badanego zjawiska w czasie można określić za pomocą wskaźników dynamiki. W zależności od przyjętej podstawy porównań wyróżnia się:

- **wskaźniki dynamiki o podstawie stałej (jednopodstawowe)**, które określają zmiany (wzrosty, spadki), jakie nastąpiły w poziomie badanego zjawiska w kolejnych okresach (momentach) w porównaniu z okresem (momentem) podstawowym,
- **wskaźniki dynamiki o podstawie zmiennej (łańcuchowe)**, które określają zmiany (wzrosty, spadki), jakie nastąpiły w poziomie badanego zjawiska w kolejnych okresach (momentach) w porównaniu z okresem (momentem) bezpośrednio poprzedzającym okres (moment) badany.

Podział wskaźników dynamiki:

- przyrosty:
  - absolutne,
  - względne;
- indeksy dynamiki:
  - indywidualne,
  - agregatowe:
    - wielkości absolutnych,
    - wielkości stosunkowych.

4.2.1. Przyrosty absolutne

Przyrosty absolutne (bezwzględne) są wyrażone w takich samych jednostkach miary, jak badane zjawisko. W zależności od przyjętej podstawy porównań wyróżnia się:

- **przyrosty absolutne o podstawie stałej**, które określają, o ile wzrósł lub spadł poziom badanego zjawiska w okresie badanym w porównaniu z okresem bazowym:

  $$P_{ab/0} = y_t - y_0 ,$$

  gdzie:
  y_t – poziom zjawiska w badanym okresie,
  y_0 – poziom zjawiska w okresie bazowym,

  jeżeli:
  - nastąpił spadek poziomu zjawiska, to $P_{ab/0} < 0$,
  - nastąpił wzrost poziomu zjawiska, to $P_{ab/0} > 0$,
  - zjawisko pozostało na niezmienionym poziomie, to $P_{ab/0} = 0$;

- **przyrosty absolutne o podstawie zmiennej**, które określają, o ile wzrósł lub spadł poziom badanego zjawiska w okresie badanym w porównaniu z okresem bezpośrednio poprzedzającym okres badany:

  $$P_{ab/t-1} = y_t - y_{t-1} ,$$

  gdzie:
  y_t – poziom zjawiska w badanym okresie,
  y_{t-1} – poziom zjawiska w okresie bezpośrednio poprzedzającym okres badany,

  jeżeli:
  - nastąpił spadek poziomu zjawiska, to $P_{ab/t-1} < 0$,
  - nastąpił wzrost poziomu zjawiska, to $P_{ab/t-1} > 0$,
  - zjawisko pozostało na niezmienionym poziomie, to $P_{ab/t-1} = 0$.

4.2.2. Przyrosty względne

Przyrosty względne są wyrażone w procentach lub w postaci ułamka dziesiętnego. W zależności od przyjętej podstawy porównań wyróżnia się:

- **przyrosty względne o podstawie stałej**, które określają, o ile procent wzrósł lub spadł poziom badanego zjawiska w okresie badanym w porównaniu z okresem bazowym:

$$P_{wz/0} = \frac{y_t - y_0}{y_0} \qquad \text{lub} \qquad P_{wz/0} = \frac{y_t - y_0}{y_0} \cdot 100,$$

gdzie:
y_t – poziom zjawiska w badanym okresie,
y_0 – poziom zjawiska w okresie bazowym,

jeżeli:
- nastąpił spadek poziomu zjawiska, to $P_{wz/0} < 0$,
- nastąpił wzrost poziomu zjawiska, to $P_{wz/0} > 0$,
- zjawisko pozostało na niezmienionym poziomie, to $P_{wz/0} = 0$;

- **przyrosty względne o podstawie zmiennej**, które określają, o ile procent wzrósł lub spadł poziom badanego zjawiska w okresie badanym w porównaniu z okresem bezpośrednio poprzedzającym okres badany:

$$P_{wz/t-1} = \frac{y_t - y_{t-1}}{y_{t-1}} \qquad \text{lub} \qquad P_{wz/t-1} = \frac{y_t - y_{t-1}}{y_{t-1}} \cdot 100,$$

gdzie:
y_t – poziom zjawiska w badanym okresie,
y_{t-1} – poziom zjawiska w okresie bezpośrednio poprzedzającym okres badany,

jeżeli:
- nastąpił spadek poziomu zjawiska, to $P_{wz/t-1} < 0$,
- nastąpił wzrost poziomu zjawiska, to $P_{wz/t-1} > 0$,
- zjawisko pozostało na niezmienionym poziomie, to $P_{wz/t-1} = 0$.

4.2.3. Indywidualne indeksy dynamiki

Indywidualne indeksy dynamiki określają stosunek poziomu badanego zjawiska, wyrażonego w takich samych jednostkach miary (pieniężnych lub naturalnych), w dwóch różnych okresach (momentach). W zależności od przyjętej podstawy porównań wyróżnia się:

- **indywidualne indeksy dynamiki o podstawie stałej**, które określają, o ile wzrósł lub spadł poziom badanego zjawiska w okresie badanym w porównaniu z okresem bazowym:

$$i_{t/0} = \frac{y_t}{y_0} \qquad \text{lub} \qquad i_{t/0} = \frac{y_t}{y_0} \cdot 100,$$

gdzie:
y_t – poziom zjawiska w badanym okresie,
y_0 – poziom zjawiska w okresie bazowym,

jeżeli:
- nastąpił spadek poziomu zjawiska, to $i_t/0 < 1$ lub $i_t/0 < 100\%$,
- nastąpił wzrost poziomu zjawiska, to $i_t/0 > 1$ lub $i_t/0 > 100\%$,
- zjawisko pozostało na niezmienionym poziomie, to $i_t/0 = 1$ lub $i_t/0 = 100\%$;

- **indywidualne indeksy dynamiki o podstawie zmiennej**, które określają, o ile wzrósł lub spadł poziom badanego zjawiska w okresie badanym w porównaniu z okresem bezpośrednio poprzedzającym okres badany:

$$i_{t/t-1} = \frac{y_t}{y_{t-1}} \quad \text{lub} \quad i_{t/t-1} = \frac{y_t}{y_{t-1}} \cdot 100,$$

gdzie:
y_t – poziom zjawiska w badanym okresie,
y_{t-1} – poziom zjawiska w okresie bezpośrednio poprzedzającym okres badany,

jeżeli:
- nastąpił spadek poziomu zjawiska, to $i_{t/t-1} < 1$ lub $i_{t/t-1} < 100\%$,
- nastąpił wzrost poziomu zjawiska, to $i_{t/t-1} > 1$ lub $i_{t/t-1} > 100\%$,
- zjawisko pozostało na niezmienionym poziomie, to $i_{t/t-1} = 1$ lub $i_{t/t-1} = 100\%$.

Średniookresowe tempo zmian to wskaźnik, który informuje, o ile procent, przeciętnie biorąc, zmieniał się poziom zjawiska z okresu na okres w całym przedziale czasowym objętym obserwacją. Wskaźnik ten można wyznaczyć tylko wtedy, gdy zmiany w poziomie badanego zjawiska miały charakter jednokierunkowy (albo wzrosty, albo spadki).

Sposób obliczania średniookresowego tempa zmian:
- dla danych w postaci indeksów łańcuchowych

$$\overline{S}_{t/t-1} = \sqrt[n-1]{\prod_{t=2}^{n} i_{t/t-1}} \cdot 100,$$

gdzie:
$i_{t/t-1}$ – indywidualny indeks dynamiki o podstawie zmiennej (w postaci ułamka dziesiętnego),
n – liczba badanych okresów;

- dla danych w postaci wartości absolutnych:

$$\overline{S}_{t/t-1} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100 ,$$

gdzie:
y_n – poziom zjawiska w ostatnim badanym okresie,
y_1 – poziom zjawiska w pierwszym badanym okresie,

jeżeli:
- następował spadek poziomu zjawiska z okresu na okres, to $\overline{S}_{t/t-1} < 100\%$,
- następował wzrost poziomu zjawiska z okresu na okres, to $\overline{S}_{t/t-1} > 100\%$.

Na podstawie danych o średniookresowym tempie zmian można **prognozować** poziom badanego zjawiska w przyszłości wykorzystując w tym celu formułę:

$$y_t^p = y_n \cdot \overline{S}_{t/t-1}^{(p-n)} ,$$

gdzie:
$\overline{S}_{t/t-1}$ – średniookresowe tempo zmian (wyrażone w postaci ułamka dziesiętnego),
y_n – poziom zjawiska w ostatnim badanym okresie,
n – liczba badanych okresów,
p – numer okresu, którego dotyczy prognoza.

4.2.4. Agregatowe indeksy dynamiki

Indeksy agregatowe (zespołowe) umożliwiają badanie zmian zachodzących w zjawiskach, których wielkość wyrażona jest w różnych jednostkach miary. Wśród agregatowych indeksów dynamiki wyróżnia się:

- **agregatowy indeks wartości**, który informuje, o ile wzrosła lub spadła wartość badanej grupy dóbr w okresie badanym w porównaniu z okresem podstawowym

$$I_w = \frac{\sum_{i=1}^{n} q_i \cdot p_i}{\sum_{i=1}^{n} q_0 \cdot p_0} \cdot 100,$$

gdzie:
q_i – ilość w okresie badanym,
q_0 – ilość w okresie podstawowym,
p_i – cena w okresie badanym,
p_0 – cena w okresie podstawowym;

- **agregatowe indeksy ilości**, które informują, o ile wzrosła lub spadła ilość badanej grupy dóbr w okresie badanym w porównaniu z okresem podstawowym, przy unieruchomieniu cen na poziomie z okresu:
  - podstawowego – **agregatowy indeks ilości Laspeyresa**

$$^{L}I_q = \frac{\sum_{i=1}^{n} q_i \cdot p_0}{\sum_{i=1}^{n} q_0 \cdot p_0} \cdot 100,$$

  - badanego – **agregatowy indeks ilości Paaschego**

$$^{P}I_q = \frac{\sum_{i=1}^{n} q_i \cdot p_i}{\sum_{i=1}^{n} q_0 \cdot p_i} \cdot 100.$$

Agregatowe indeksy ilości Laspeyresa i Paaschego określają granice przedziału, w którym zawarta jest „prawdziwa" wartość indeksu, natomiast za dobre przybliżenie indeksu poprawnie mierzącego zmiany ilości uznaje się **agregatowy indeks ilości Fishera**. Informuje on, o ile średnio zmieniła się ilość badanej grupy dóbr w okresie badanym w porównaniu z okresem podstawowym:

$$^{F}I_q = \sqrt{^{L}I_q \cdot {}^{P}I_q};$$

- **agregatowe indeksy cen**, które informują, o ile wzrosły lub spadły ceny badanej grupy dóbr w okresie badanym w porównaniu z okresem podstawowym, przy unieruchomieniu ilości na poziomie z okresu:
  - podstawowego – **agregatowy indeks cen Laspeyresa**

$$
{}^{L}I_p = \frac{\sum_{i=1}^{n} p_i \cdot q_0}{\sum_{i=1}^{n} p_0 \cdot q_0} \cdot 100,
$$

  - badanego – **agregatowy indeks cen Paaschego**

$$
{}^{P}I_p = \frac{\sum_{i=1}^{n} p_i \cdot q_i}{\sum_{i=1}^{n} p_0 \cdot q_i} \cdot 100.
$$

Agregatowe indeksy cen Laspeyresa i Paaschego określają granice przedziału, w którym zawarta jest „prawdziwa" wartość indeksu, natomiast za dobre przybliżenie indeksu poprawnie mierzącego zmiany cen uznaje się **agregatowy indeks cen Fishera**. Informuje on, o ile średnio zmieniły się ceny badanej grupy dóbr w okresie badanym w porównaniu z okresem podstawowym:

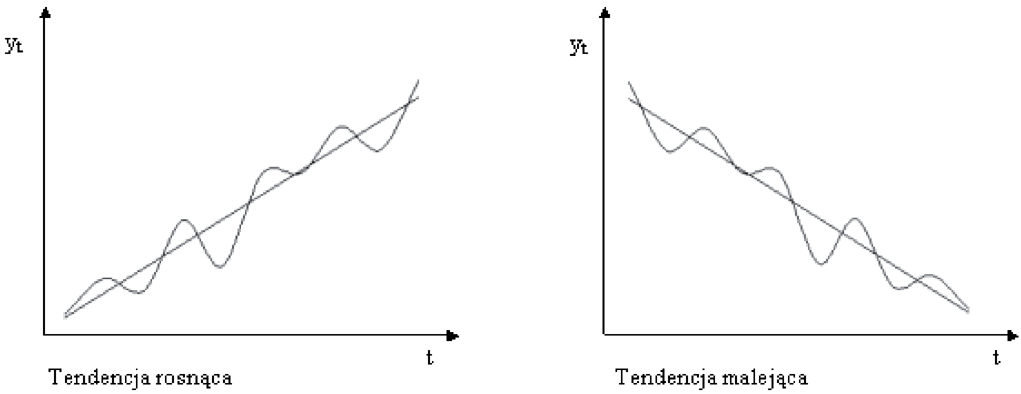
$$
{}^{F}I_p = \sqrt{{}^{L}I_p \cdot {}^{P}I_p}.
$$

4.3. Dekompozycja szeregu czasowego

Dekompozycja szeregu czasowego polega na wyodrębnieniu poszczególnych składników szeregu czasowego i dokonaniu ich pomiaru. Do składników tych należą tendencja rozwojowa (trend), wahania okresowe (sezonowe) oraz wahania przypadkowe (losowe).

4.3.1. Wyodrębnianie tendencji rozwojowej (trendu)

Tendencja rozwojowa to powolne, regularne i systematyczne zmiany (wzrosty lub spadki) badanego zjawiska obserwowane w dostatecznie długim okresie, będące rezultatem działania przyczyn głównych.



Schemat 4.1. Rodzaje tendencji rozwojowej

Źródło: Opracowanie własne.

Wyodrębnienia tendencji rozwojowej (wygładzenia szeregu dynamicznego) można dokonać, wykorzystując m.in. metody mechaniczne i analityczne.

4.3.1.1. Metoda mechaniczna – średnia ruchoma

Często stosowana mechaniczna metoda wyrównywania szeregów czasowych polega na obliczeniu tzw. **średnich ruchomych**, czyli średnich arytmetycznych z określonej liczby wyrazów szeregu. Wyznaczenie średnich ruchomych polega na zastępowaniu danych empirycznych z kolejnych okresów średnimi poziomami z okresu badanego i kilku okresów sąsiednich, dzięki czemu eliminuje się występujące wahania.

Jeżeli kolejne wartości szeregu dynamicznego (chronologicznego) oznaczymy jako y_1, y_2 … y_n, to średnią ruchomą oblicza się w następujący sposób:

- **średnia ruchoma trzyokresowa**

$$\bar{y}_{n-1} = \frac{y_{n-2} + y_{n-1} + y_n}{3},$$

- **średnia ruchoma pięciookresowa**

$$\bar{y}_{n-2} = \frac{y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1} + y_n}{5}.$$

Jeżeli w szeregu występują wahania sezonowe, należy obliczyć średnie ruchome obejmujące wszystkie wyrazy cyklu wahań (np. dla rocznego cyklu wahań wyrażonego szeregiem danych miesięcznych stosuje się średnie ruchome dwunastookresowe, a dla cyklu wyrażonego szeregiem danych kwartalnych – czterookresowe). Ponieważ liczba okresów jest wtedy zazwyczaj parzysta, pojawia się problem z przyporządkowaniem uzyskanych średnich do odpowiednich okresów. Oblicza się wówczas tzw. **średnie ruchome scentrowane**.

Jeżeli kolejne wartości szeregu dynamicznego (chronologicznego) oznaczymy jako y_1, y_2 … y_n , to średnią ruchomą scentrowaną czterookresową oblicza się w następujący sposób:

$$\bar{y}_{n-2} = \frac{\frac{1}{2}y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1} + \frac{1}{2}y_n}{4}.$$

4.3.1.2. Metoda analityczna – metoda najmniejszych kwadratów

Najczęściej stosowana analityczna metoda wyrównywania szeregów czasowych polega na dopasowaniu odpowiedniej funkcji matematycznej (funkcji trendu) do danych empirycznych według kryterium minimalizacji sumy kwadratów odchyleń między wartościami teoretycznymi i empirycznymi. Stąd nazwa metody (metoda najmniejszych kwadratów – MNK)[18].

Etapy wyznaczania funkcji trendu:

- **wybór klasy aproksymanty trendu** na podstawie wykresu (oś odciętych ilustruje czas, natomiast oś rzędnych poziom badanego zjawiska); rozkład punktów wskazuje na postać funkcji, która swym kształtem najlepiej pasuje do naniesionych wartości empirycznych, np. funkcja liniowa, wykładnicza, paraboliczna itd.;
- **oszacowanie parametrów funkcji trendu** – w przypadku funkcji liniowej należy oszacować parametry równania

  $$\hat{y}_t = a + bt,$$

[18] Por. rozdz. 3.3.1. Metoda Najmniejszych Kwadratów (MNK).

z tego:

$$a = \frac{\sum_{t=1}^{n} y_t - b\sum_{t=1}^{n} t}{n} = \overline{y_t} - b \cdot \bar{t}$$

$$b = \frac{n\sum_{t=1}^{n} y_t \cdot t - \sum_{t=1}^{n} y_t \sum_{t=1}^{n} t}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2},$$

gdzie:
a – poziom badanego zjawiska w okresie wyjściowym (tzn. dla t = 0),
b – okresowy wzrost (b > 0) lub spadek (b < 0) wielkości badanego zjawiska,
t – numer okresu, gdzie t = 1, 2,...,n,
n – liczba okresów;

- **ocena stopnia „dobroci”**, czyli dopasowania oszacowanej funkcji trendu do danych empirycznych; oceny tej dokonuje się za pomocą, m.in.:
  - **odchylenia standardowego składnika resztowego**, które informuje, o ile średnio wartości empiryczne odchylają się od wartości teoretycznych wyznaczonych na podstawie funkcji trendu:

$$S_y = \sqrt{\frac{1}{n-k}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2},$$

    gdzie:
    k – liczba szacowanych parametrów (w przypadku funkcji liniowej k = 2),
    n–k – liczba stopni swobody,

  - **współczynnika zbieżności** wskazującego, jaka część zmienności badanego zjawiska nie została wyjaśniona przez skonstruowaną funkcję trendu; współczynnik przyjmuje wartości z przedziału <0,1>, przy czym uznaje się, że funkcja jest dobrze dopasowana do danych empirycznych, gdy współczynnik nie przekracza wartości 0,2:

$$\varphi^2 = \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{\sum_{t=1}^{n}\left(y_t - \overline{y_t}\right)^2},$$

- **współczynnika determinacji** określającego stopień, w jakim skonstruowana funkcja trendu wyjaśnia zmienność badanego zjawiska; współczynnik przyjmuje wartości z przedziału <0,1>:

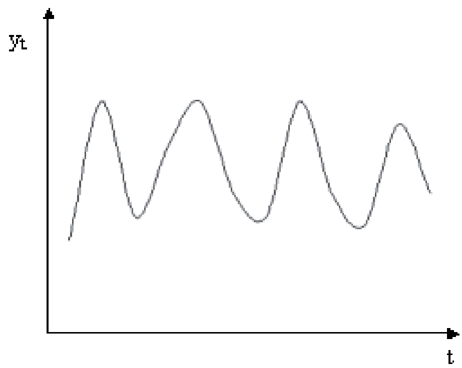
$$R^2 = 1 - \varphi^2 ,$$

- **współczynnika zmienności resztowej** informującego, jaki odsetek średniego poziomu zjawiska w badanym okresie stanowią wahania przypadkowe:

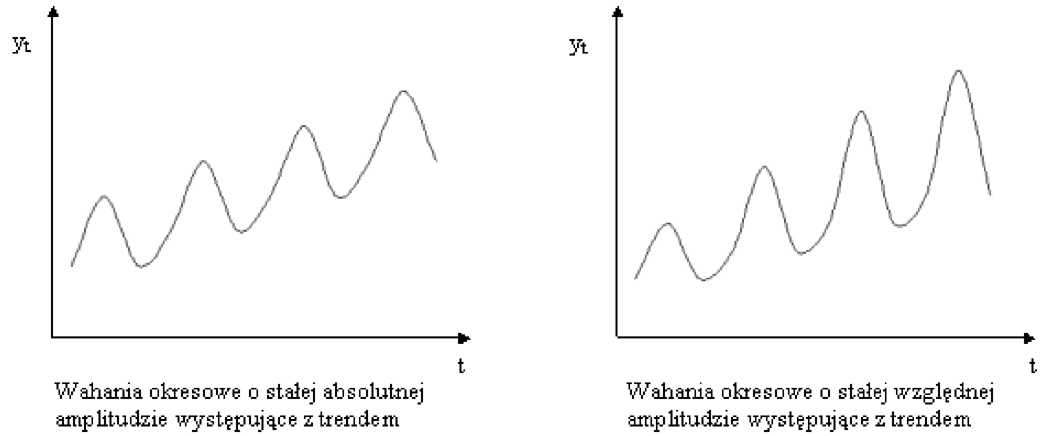
$$V_{Sy} = \frac{S_y}{\overline{y}_t} \cdot 100 .$$

4.3.2. Wyodrębnianie wahań okresowych (sezonowych)

Wahania okresowe to systematyczne wahania powtarzające się w ściśle określonych odstępach czasu. Na ich występowanie wpływ ma najczęściej pora dnia, pora roku, zwyczaje, unormowania prawne itp.



Wahania okresowe
występujące bez trendu



Schemat 4.2. Rodzaje wahań okresowych

Źródło: Opracowanie własne.

Wyodrębnienia wahań sezonowych można dokonać za pomocą różnych metod, których wybór zależy od tego, czy wahania te współwystępują z trendem, czy też nie, oraz od amplitudy wahań. Najpopularniejsze metody to:

- metoda średnich jednoimiennych okresów,
- metoda wskaźników.

4.3.2.1. Metoda średnich jednoimiennych okresów

Metoda średnich jednoimiennych okresów jest stosowana w przypadku niewystępowania trendu. Polega na ustaleniu **względnych wskaźników sezonowości** dla kolejnych podokresów, które informują, o ile procent poziom badanego zjawiska w danym podokresie różni się od przeciętnego poziomu tego zjawiska w całym badanym okresie. Względny wskaźnik sezonowości wyznacza się za pomocą formuły:

$$_{w}g_i = \frac{\bar{y}_i}{\bar{y}},$$

gdzie:
$\bar{y}_i$ – średni poziom zjawiska dla i-tego podokresu w całym badanym okresie,
$\bar{y}$ – średni poziom zjawiska dla całego badanego okresu.

W oparciu o względne wskaźniki sezonowości można wyznaczyć **bezwzględne wskaźniki sezonowości**, które informują, o ile poziom badanego zjawiska w danym podokresie różni się (w wartościach absolutnych) od przeciętnego poziomu tego zjawiska w całym badanym okresie:

$$ {}_b g_i = \overline{y} \cdot ({}_w g_i - 1) \,. $$

4.3.2.2. Metoda wskaźników

Metoda wskaźników jest stosowana w przypadku współwystępowania wahań sezonowych z trendem. Polega na wyznaczeniu uśrednionych wskaźników sezonowości dla poszczególnych faz cyklu wahań. Procedura wyznaczania wskaźników różni się w zależności od amplitudy wahań (rodzaju sezonowości):

- **sezonowość addytywna** – charakteryzuje się stałą absolutną amplitudą wahań; w celu wyodrębnienia wahań oblicza się tzw. **surowe wskaźniki sezonowości addytywnej**, które informują, o ile poziom badanego zjawiska w danym podokresie różni się (w wartościach absolutnych) od poziomu tego zjawiska wynikającego z trendu, na skutek oddziaływania zarówno wahań sezonowych, jak i przypadkowych; wskaźniki wyznacza się za pomocą następującej formuły:

  $$ g_i = \frac{1}{n_i} \sum_{i=1}^{n_i} (y_t - \hat{y}_t) \,, $$

  gdzie:
  y_t – faktyczny poziom badanego zjawiska,
  $\hat{y}_t$ - teoretyczny poziom badanego zjawiska wynikający z funkcji trendu,
  n_i – liczba obliczonych różnic dla jednoimiennych podokresów.

  Jeżeli suma obliczonych wskaźników dla wszystkich podokresów nie jest równa 0, wówczas należy wyznaczyć **czyste wskaźniki sezonowości addytywnej**, które informują, o ile poziom badanego zjawiska w danym podokresie różni się (w wartościach absolutnych) od poziomu tego zjawiska wynikającego z trendu, na skutek oddziaływania wyłącznie wahań sezonowych:

  $$ S_i = g_i - w_{kor} \,, $$

  gdzie:

  $$ w_{kor} = \frac{\sum_{i=1}^{n} g_i}{n} \,, $$

  n – liczba podokresów;

- **sezonowość multiplikatywna** – charakteryzuje się stałą względną amplitudą wahań; w celu wyodrębnienia wahań oblicza się tzw. **surowe wskaźniki sezonowości multiplikatywnej**, które informują, o ile procent poziom badanego zjawiska w danym podokresie różni się od poziomu tego zjawiska wynikającego z trendu, na skutek oddziaływania zarówno wahań sezonowych, jak i przypadkowych; wskaźniki wyznacza się za pomocą następującej formuły:

$$g_i = \frac{1}{n_i}\sum_{i=1}^{n_i}\frac{y_t}{\hat{y}_t},$$

gdzie:
y_t – faktyczny poziom badanego zjawiska,
$\hat{y}_t$ – teoretyczny poziom badanego zjawiska wynikający z funkcji trendu,
n_i – liczba obliczonych ilorazów dla jednoimiennych podokresów.

Jeżeli suma obliczonych wskaźników dla wszystkich podokresów nie jest równa liczbie podokresów, wówczas należy wyznaczyć **czyste wskaźniki sezonowości multiplikatywnej**, które informują, o ile procent poziom badanego zjawiska w danym podokresie różni się od poziomu tego zjawiska wynikającego z trendu, na skutek oddziaływania wyłącznie wahań sezonowych:

$$S_i = g_i \cdot w_{kor},$$

gdzie:

$$w_{kor} = \frac{n}{\sum_{i=1}^{n} g_i},$$

n – liczba podokresów.

4.3.3. Wyodrębnianie wahań przypadkowych (losowych)

Wahania przypadkowe są wywołane przyczynami ubocznymi lub katastroficznymi. Nie mają one wpływu na tendencję rozwojową, gdyż powodując zarówno spadki, jak i wzrosty badanego zjawiska, w dłuższym okresie znoszą się wzajemnie.

Dysponując funkcją trendu oraz informacjami na temat wahań sezonowych o stałej absolutnej amplitudzie (sezonowość addytywna), można wyodrębnić wahania przypadkowe za pomocą formuły

$$z_t = y_t - \hat{y}_t - S_i ,$$

a przeciętne zróżnicowanie siły ich oddziaływania ustalić przy wykorzystaniu formuły

$$S_z = \sqrt{\frac{\sum_{t=1}^{n} z_t^2}{n - n_i - 2}} ,$$

gdzie:
n – liczba podokresów,
n_i – liczba jednoimiennych podokresów.

4.4. Przykłady

Przykład 4.4.1.

Na podstawie danych dotyczących skupu zbóż podstawowych w Polsce w latach 2004-2008 należy ustalić przyrosty absolutne, względne oraz indywidualne indeksy dynamiki. Otrzymane wyniki należy zinterpretować.

Lata	Skup zbóż podstawowych (w kg na 1 ha użytków rolnych)
2004	371
2005	433
2006	398
2007	352
2008	350

Źródło: *Rolnictwo* z lat 2004-2008, GUS, Warszawa.

Rozwiązanie

<u>Przyrosty absolutne</u>

- jednopodstawowe (rok 2004 = 100)

$$P_{ab/0} = y_t - y_0$$

rok 2004 → 371 – 371 = 0
rok 2005 → 433 – 371 = 62
...
rok 2008 → 350 – 371 = -21

Skup zbóż podstawowych w Polsce wzrósł w 2005 r. o 62 kg na 1 ha użytków rolnych w porównaniu z rokiem 2004. Z kolei w 2008 r. skup zmalał o 21 kg w odniesieniu do roku 2004.

- łańcuchowe (rok poprzedni = 100)

$$P_{ab/t-1} = y_t - y_{t-1}$$

rok 2004 → 371 – X = X
rok 2005 → 433 – 371 = 62
…
rok 2008 → 350 – 352 = -2

Skup zbóż podstawowych w Polsce wzrósł w 2005 r. o 62 kg na 1 ha użytków rolnych w porównaniu z rokiem 2004. Z kolei w 2008 r. skup zmalał o 2 kg w odniesieniu do roku 2007.

Przyrosty względne

- jednopodstawowe (rok 2004 = 100)

$$P_{wz/0} = \frac{y_t - y_0}{y_0} \cdot 100$$

rok 2004 → [(371– 371) / 371] * 100 = 0
rok 2005 → [(433 – 371) / 371] * 100 = 16,71%
…
rok 2008 → [(350 – 371) / 371] * 100 = -5,66%

- łańcuchowe (rok poprzedni = 100)

$$P_{wz/t-1} = \frac{y_t - y_{t-1}}{y_{t-1}} \cdot 100$$

rok 2004 → [(371 – X) / X] * 100 = X
rok 2005 → [(433 – 371) / 371] * 100 = 16,71%
…
rok 2008 → [(350 – 352) / 352] * 100 = -0,57%

Indywidualne indeksy dynamiki

- jednopodstawowe (rok 2004 = 100)

$$i_{t/0} = \frac{y_t}{y_0} \cdot 100$$

rok 2004 → (371 / 371) * 100 = 100%
rok 2005 → (433 / 371) * 100 = 116,71%
…
rok 2008 → (350 / 371) * 100 = 94,34%

Skup zbóż podstawowych w Polsce wzrósł w 2005 r. o 16,71% w porównaniu z rokiem 2004. Z kolei w 2008 r. skup zmalał o 5,66% w stosunku do roku 2004.

- łańcuchowe (rok poprzedni = 100)

$$i_{t/t-1} = \frac{y_t}{y_{t-1}} \cdot 100$$

rok 2004 → (371 / X) * 100 = X
rok 2005 → (433 / 371) * 100 = 116,71%
…
rok 2008 → (350 / 352) * 100 = 99,43%

Skup zbóż podstawowych w Polsce wzrósł w 2005 r. o 16,71% w porównaniu z rokiem 2004. Z kolei w 2008 r. skup zmalał o 0,57% w stosunku do roku 2007.

Przykład 4.4.2.

Poniższe zestawienie obrazuje wielkość sprzedaży trzech artykułów gospodarstwa domowego w pewnej hurtowni oraz ich ceny w latach 2005 i 2009.

Artykuł	Wielkość sprzedaży (w tys. szt.)		Ceny jednostkowe (w zł)	
	2005	2009	2005	2009
kuchnia	3,0	4,5	1300	900
lodówka	4,5	4,0	1000	1200
pralka	1,5	3,5	1100	800

Źródło: Dane umowne.

Posługując się indeksami zespołowymi, należy przeanalizować dynamikę wartości i wielkości sprzedaży oraz cen w 2009 r. w stosunku do roku 2005 dla trzech artykułów łącznie. Otrzymane wyniki zinterpretować.

Rozwiązanie

Agregatowy indeks wartości

$$I_w = \frac{\sum_{i=1}^{n} q_i \cdot p_i}{\sum_{i=1}^{n} q_0 \cdot p_0} \cdot 100 = \frac{11650}{10050} \cdot 100 = 115{,}92\%$$

Wartość sprzedaży trzech artykułów gospodarstwa domowego w pewnej hurtowni wzrosła w roku 2009 o 15,92% w porównaniu z rokiem 2005.

Agregatowe indeksy ilości

- Laspeyresa

$${}^{L}I_q = \frac{\sum_{i=1}^{n} q_i \cdot p_0}{\sum_{i=1}^{n} q_0 \cdot p_0} \cdot 100 = \frac{13700}{10050} \cdot 100 = 136{,}32\%$$

Wielkość sprzedaży trzech artykułów gospodarstwa domowego w pewnej hurtowni wzrosła w roku 2006 o 36,32% w porównaniu z rokiem 2002, przy założeniu stałych cen z roku 2002.

- Paaschego

$${}^{P}I_q = \frac{\sum_{i=1}^{n} q_i \cdot p_i}{\sum_{i=1}^{n} q_0 \cdot p_i} \cdot 100 = \frac{11650}{9300} \cdot 100 = 125{,}27\%$$

Wielkość sprzedaży trzech artykułów gospodarstwa domowego w pewnej hurtowni wzrosła w roku 2009 o 25,27% w porównaniu z rokiem 2005, przy założeniu stałych cen z roku 2005.

- Fishera

$$ {}^{F}I_q = \sqrt{{}^{L}I_q \cdot {}^{P}I_q} = \sqrt{136,32 \cdot 125,27} = 130,68\% $$

Wielkość sprzedaży trzech artykułów gospodarstwa domowego w pewnej hurtowni wzrosła w roku 2009 w porównaniu z rokiem 2005 średnio o 30,68%.

Agregatowe indeksy cen

- Laspeyresa

$$ {}^{L}I_p = \frac{\sum_{i=1}^{n} p_i \cdot q_0}{\sum_{i=1}^{n} p_0 \cdot q_0} \cdot 100 = \frac{9300}{10050} \cdot 100 = 92,54\% $$

Ceny trzech artykułów gospodarstwa domowego w pewnej hurtowni zmalały w 2009 r. o 7,46% w porównaniu z 2005 r., przy założeniu stałych wielkości sprzedaży z 2005 r.

- Paaschego

$$ {}^{P}I_p = \frac{\sum_{i=1}^{n} p_i \cdot q_i}{\sum_{i=1}^{n} p_0 \cdot q_i} \cdot 100 = \frac{11650}{13700} \cdot 100 = 85,04\% $$

Ceny trzech artykułów gospodarstwa domowego w pewnej hurtowni zmalały w 2009 r. o 14,96% w porównaniu z 2005 r., przy założeniu stałych wielkości sprzedaży z 2009 r.

- Fishera

$$ {}^{F}I_p = \sqrt{{}^{L}I_p \cdot {}^{P}I_p} = \sqrt{92,54 \cdot 85,04} = 88,71\% $$

Ceny trzech artykułów gospodarstwa domowego w pewnej hurtowni zmalały w roku 2009 w porównaniu z rokiem 2005 średnio o 11,29%

Obliczenia pomocnicze – przykład 4.4.1.

Lata	Skup zbóż podstawowych (w kg na 1 ha użytków rolnych)	Przyrosty absolutne		Przyrosty względne (w %)		Indeksy dynamiki (w %)	
		2004 = 100	Rok poprzedni = = 100	2004 = 100	Rok poprzedni = = 100	2004 = 100	Rok poprzedni = = 100
2004	371	0	X	0,00	X	100,00	X
2005	433	62	62	16,71	16,71	116,71	116,71
2006	398	27	-35	7,28	-8,08	107,28	91,92
2007	352	-19	-46	-5,12	-11,56	94,88	88,44
2008	350	-21	-2	-5,66	-0,57	94,34	99,43

Obliczenia pomocnicze – przykład 4.4.2.

Artykuł	Wielkość sprzedaży (w tys. szt.)		Ceny jednostkowe (w zł)		$q_0 * p_0$	$q_i * p_i$	$q_i * p_0$	$q_0 * p_i$
	2005 (q_0)	2009 (q_i)	2005 (p_0)	2009 (p_i)				
Kuchnia	3,0	4,5	1300	900	3 900	4 050	5 850	2 700
Lodówka	4,5	4,0	1000	1200	4 500	4 800	4 000	5 400
Pralka	1,5	3,5	1100	800	1 650	2 800	3 850	1 200
Σ	x	x	x	x	10 050	11 650	13 700	9 300

Przykład 4.4.3.

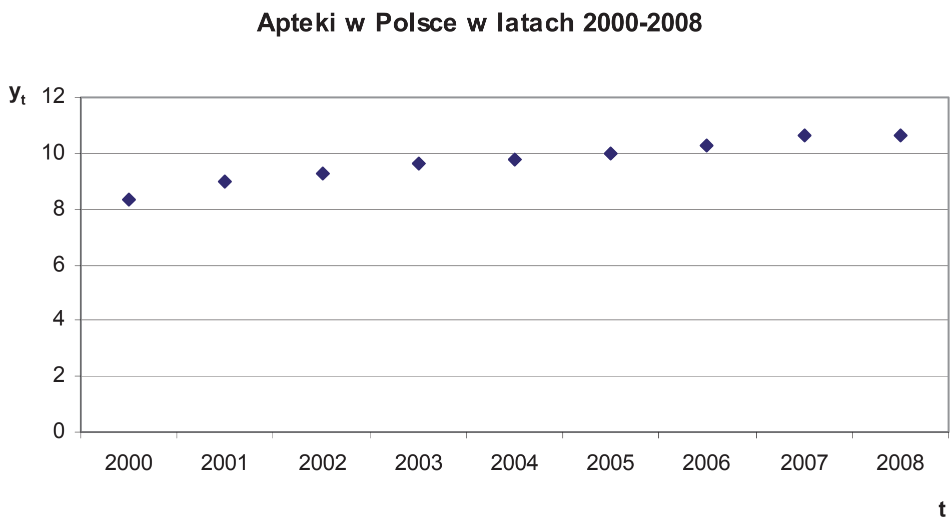
Liczba aptek w Polsce w latach 2000-2008 przedstawiała się następująco:

Lata	Liczba aptek (w tys.)
2000	8,3
2001	9,0
2002	9,3
2003	9,6
2004	9,8
2005	10,0
2006	10,3
2007	10,6
2008	10,6

Źródło: *Rocznik Statystyczny Rzeczypospolitej Polskiej* z lat 2001-2009, GUS, Warszawa.

Wyznacz tendencję rozwojową w sposób analityczny. Oszacuj liczbę aptek w roku 2011 oraz oceń stopień dopasowania funkcji trendu do danych empirycznych.

Rozwiązanie



Z wykresu wynika, że liczba stacji aptek w Polsce w latach 2000-2008 wzrastała z roku na rok (wyjątek stanowił jedynie rok 2004). Rozkład punktów wskazuje, że tendencję tę można opisać za pomocą funkcji liniowej w postaci $\hat{y}_t = a + bt$.

Okresowy wzrost/spadek badanego zjawiska

$$b = \frac{n\sum_{t=1}^{n} y_t \cdot t - \sum_{t=1}^{n} y_t \sum_{t=1}^{n} t}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2} = \frac{9 \cdot 453{,}9 - 87{,}5 \cdot 45}{9 \cdot 285 - 45^2} = 0{,}27$$

Liczba aptek w Polsce w latach 2000-2008 wzrastała z roku na rok średnio o 0,27 tys.

Poziom zjawiska w okresie wyjściowym tzn. w roku 1999 (t = 0)

$$a = \frac{\sum_{t=1}^{n} y_t - b\sum_{t=1}^{n} t}{n} = \frac{87{,}5 - 0{,}27 \cdot 45}{9} = 8{,}37$$

Liczba aptek w roku 1999 wynosiła teoretycznie 8,37 tys.; teoretycznie, tzn. gdyby na zjawisko nie oddziaływały wahania przypadkowe (gdyby przebiegało ono zgodnie z oszacowaną funkcją trendu).

Funkcja trendu

$$\hat{y}_t = a + bt$$

$$\hat{y}_t = 0,27t + 8,37$$

Liczba aptek w roku 2011, tj. t = 12

$$\hat{y}_t = 0,27t + 8,37$$

$$\hat{y}_{12} = 0,27 \cdot 12 + 8,37$$

$$\hat{y}_{12} = 11,61$$

Uwzględniając przeciętny roczny wzrost liczby aptek w Polsce można szacować, że w roku 2011 liczba ta wyniesie 11,61 tys. Jest to jednak szacunek obarczony pewnym błędem i w związku z tym należy obliczyć błąd szacunku, czyli odchylenie standardowe składnika resztowego.

$$S_y = \sqrt{\frac{1}{n-k}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}$$

$$S_y = \sqrt{\frac{1}{9-2} \cdot 0,21} = 0,17$$

$$11,61 - 0,17 < \hat{y}_{12} < 11,61 + 0,17$$

$$11,44 < \hat{y}_{12} < 11,78$$

Faktyczna liczba aptek w Polsce w latach 2000-2008 odchylała się od liczby teoretycznej wyznaczonej na podstawie funkcji trendu średnio o 0,17 tys. W roku 2011 liczba aptek powinna wahać się w granicach od 11,44 do 11,78 tys.

Współczynnik zbieżności

$$\varphi^2 = \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{\sum_{t=1}^{n}\left(y_t - \overline{y}_t\right)^2} = \frac{0,21}{4,70} = 0,04$$

Tylko 4% zmienności liczby aptek w Polsce w latach 2000-2008 nie zostało wyjaśnione przez skonstruowaną funkcję trendu.

Współczynnik determinacji

$$R^2 = 1 - \varphi^2 = 1 - 0{,}04 = 0{,}96$$

Około 96% zmienności liczby aptek w Polsce w latach 2000-2008 zostało wyjaśnione przez skonstruowaną funkcję trendu.

Współczynnik zmienności resztowej

$$V_{Sy} = \frac{S_y}{\overline{y}_t} \cdot 100 = \frac{0{,}17}{9{,}72} \cdot 100 = 1{,}75\,\%$$

Wahania losowe stanowiły 1,75% średniej liczby aptek w Polsce w latach 2000-2008.

Obliczenia pomocnicze – przykład 4.4.3.

Lata	Liczba aptek (w tys.)	t	y_t * t	t^2	$\hat{y}_t$	$y_t - \hat{y}_t$	$(y_t - \hat{y}_t)^2$	$(y_t - \overline{y}_t)^2$
2000	8,3	1	8,30	1	8,64	-0,34	0,12	2,02
2001	9,0	2	18,00	4	8,91	0,09	0,01	0,52
2002	9,3	3	27,90	9	9,18	0,12	0,01	0,18
2003	9,6	4	38,40	16	9,45	0,15	0,02	0,01
2004	9,8	5	49,00	25	9,72	0,08	0,01	0,01
2005	10,0	6	60,00	36	9,99	0,01	0,00	0,08
2006	10,3	7	72,10	49	10,26	0,04	0,00	0,34
2007	10,6	8	84,80	64	10,53	0,07	0,00	0,77
2008	10,6	9	95,40	81	10,80	-0,20	0,04	0,77
Σ	87,5	45	453,90	285	X	X	0,21	4,70

4.5. Zadania

Zadanie 4.5.1.

Na podstawie danych dotyczących wysokości dochodów budżetowych wielkopolskich gmin w latach 1997-2001 należy ustalić przyrosty absolutne, względne oraz indywidualne indeksy dynamiki. Otrzymane wyniki zinterpretować.

Lata	Dochody budżetowe w mln zł
1997	3 181
1998	3 775
1999	2 709
2000	2 938
2001	3 176

Źródło: *Rocznik Statystyczny Województwa Wielkopolskiego* z lat 2000-2002, US w Poznaniu, Poznań oraz Bank Danych Regionalnych GUS.

Zadanie 4.5.2.

Na podstawie danych dotyczących wysokości wydatków budżetowych wielkopolskich gmin w latach 1997-2001 należy ustalić przyrosty absolutne, względne oraz indywidualne indeksy dynamiki. Otrzymane wyniki zinterpretować.

Lata	Wydatki budżetowe w mln zł
1997	3 287
1998	3 917
1999	2 788
2000	3 108
2001	3 326

Źródło: *Rocznik Statystyczny Województwa Wielkopolskiego* z lat 2000-2002, US w Poznaniu, Poznań oraz Bank Danych Regionalnych GUS.

Zadanie 4.5.3.

Liczba mieszkańców miasta Konina w latach 2003-2008 przedstawiała się następująco (stan na 31 grudnia):

Lata	2003	2004	2005	2006	2007	2008
Liczba mieszkańców	81 774	81 266	80 838	80 471	80 140	79 829
Indeksy (rok poprzedni = 100)	X	99,38	99,47	99,55	99,59	99,61

Źródło: *Ludność. Stan i struktura w przekroju terytorialnym* z lat 2003-2008, GUS, Warszawa.

Na podstawie powyższych informacji:

- ustalić, o ile, przeciętnie biorąc, liczba mieszkańców malała z roku na rok,
- dokonać prognozy na rok 2011.

Zadanie 4.5.4.

Poniższe zestawienie obrazuje wielkość spożycia trzech artykułów nabiałowych przez czteroosobową rodzinę oraz ich ceny w latach 2004 i 2009.

Artykuł	Wielkość spożycia		Ceny jednostkowe (w zł)	
	2004	2009	2004	2009
mleko (w litrach)	440	400	1,20	1,50
masło (w kg)	12	10	6,20	7,10
jaja (w szt.)	350	325	0,35	0,25

Źródło: Dane umowne.

Posługując się indeksami zespołowymi, przeanalizować dynamikę wartości i wielkości spożycia oraz cen w 2009 r. w stosunku do roku 2004 dla trzech artykułów łącznie. Otrzymane wyniki zinterpretować.

Zadanie 4.5.5.

Poniższe zestawienie obrazuje wielkość sprzedaży trzech artykułów spożywczych w pewnym sklepie oraz ich ceny w latach 2005 i 2008.

Artykuł	Wielkość sprzedaży		Ceny jednostkowe (w zł)	
	2005	2008	2005	2008
kawa (w opak.)	900	1 100	4,2	3,8
dżem (w szt.)	660	620	2,5	2,2
cukier (w kg)	860	800	3,0	3,3

Źródło: Dane umowne.

Posługując się indeksami zespołowymi przeanalizować dynamikę wartości i wielkości sprzedaży oraz cen w 2008 r. w stosunku do roku 2005 dla trzech artykułów łącznie. Otrzymane wyniki zinterpretować.

Zadanie 4.5.6.

Wykorzystując dane zawarte w poniższej tabeli, zbadać dynamikę wartości i wielkości sprzedaży oraz cen czterech produktów rolnych oferowanych przez pewnego rolnika w sierpniu 2008 r. w porównaniu z sierpniem 2007 r. Otrzymane wyniki zinterpretować.

Produkt	Ceny sprzedaży (w zł/kg)		Wielkość sprzedaży (w kg)	
	VIII 2007	VIII 2008	VIII 2007	VIII 2008
ziemniaki	0,5		2 000	
ogórki	3,0		500	
pomidory	4,5		1 000	
papryka	5,0		200	

Źródło: Dane umowne.

Brakujące miejsca w tabeli uzupełnić w oparciu o następujące informacje:

- absolutne przyrosty wielkości sprzedaży w sierpniu 2008 r. w porównaniu z sierpniem 2007 r. ukształtowały się na następującym poziomie:
    - ziemniaki: 500 kg,

   - ogórki: –200 kg,
   - pomidory: 500 kg,
   - papryka: 50 kg;
- indywidualne indeksy dynamiki cen ukształtowały się w sierpniu 2008 r. w porównaniu z sierpniem 2007 r. na następującym poziomie (VIII 2007 = = 100%):
   - ziemniaki: 260%,
   - ogórki: 50%,
   - pomidory: 44,44%,
   - papryka: 90%.

Zadanie 4.5.7.

Wykorzystując dane zawarte w poniższej tabeli zbadać dynamikę wartości i wielkości importu oraz cen czterech wyrobów papierniczych w pewnej hurtowni w lutym 2009 r. w porównaniu z lutym 2006 r. Otrzymane wyniki zinterpretować.

Wyrób	Ceny zakupu (w euro/opak.)		Wielkość importu (w tys. opakowań)	
	II 2006	II 2009	II 2006	II 2009
papier	5		13	
karton	9		15	
tektura	11		8	
bibuła	7		12	

Źródło: Dane umowne.

Brakujące miejsca w tabeli uzupełnić w oparciu o następujące informacje:
- absolutne przyrosty wielkości importu w lutym 2009 r. w porównaniu z lutym 2006 r. ukształtowały się na następującym poziomie:
   - papier: –1 tys. opakowań,
   - karton: –8 tys. opakowań,
   - tektura: +3 tys. opakowań,
   - bibuła: –2 tys. opakowań;
- indywidualne indeksy dynamiki cen ukształtowały się w lutym 2009 r. w porównaniu z lutym 2006 r. na następującym poziomie (II 2006 = 100%):
   - papier: 120%,
   - karton: 133,33%,
   - tektura: 90,90%,
   - bibuła: 114,29%.

Zadanie 4.5.8.

Liczba osób pracujących w Polsce w ochronie zdrowia i pomocy społecznej w latach 1999-2007 przedstawiała się następująco:

Lata	Liczba pracujących (w tys.)
1999	967,0
2000	908,2
2001	869,0
2002	851,7
2003	704,8
2004	703,8
2005	707,1
2006	715,4
2007	737,9

Wyznacz tendencję rozwojową w sposób mechaniczny, wykorzystując średnią ruchomą trzy- i pięciookresową. Sporządź wykres szeregu pierwotnego (danych empirycznych) i szeregów średnich ruchomych (danych teoretycznych).

Źródło: Bank Danych Regionalnych GUS.

Zadanie 4.5.9.

Liczba mieszkań oddanych do użytkowania w Polsce w latach 1999-2008 kształtowała się następująco:

Lata	Liczba mieszkań oddanych do użytkowania (w tys.)
1999	82,0
2000	87,8
2001	106,0
2002	97,6
2003	162,7
2004	108,1
2005	114,1
2006	115,4
2007	133,7
2008	165,2

Wyznacz tendencję rozwojową w sposób mechaniczny, wykorzystując średnią ruchomą trzy- i pięciookresową. Sporządź wykres szeregu pierwotnego (danych empirycznych) i szeregów średnich ruchomych (danych teoretycznych).

Źródło: *Warunki życia ludności* z lat 1999-2008, GUS, Warszawa.

Zadanie 4.5.10.

Lata	Pora roku	Liczba turystów (w tys.)
2003	wiosna	649
	lato	999
	jesień	721
	zima	332
2004	wiosna	801
	lato	1275
	jesień	886
	zima	424
2005	wiosna	910
	lato	1389
	jesień	956
	zima	467

W niniejszej tabeli przedstawiona została liczba turystów zagranicznych korzystających z obiektów hotelowych w Polsce w latach 2003-2005.
Na podstawie dostępnych informacji wyznacz tendencję rozwojową w sposób mechaniczny. Sporządź wykres szeregu pierwotnego (danych empirycznych) i szeregów średnich ruchomych (danych teoretycznych).

Źródło: Opracowanie własne na podstawie *Turystyka* z lat 2003-2005, GUS, <www.stat.gov.pl>.

Zadanie 4.5.11.

Lata	Kwartał	Stopa bezrobocia (w %)
2003	I	20,6
	II	19,4
	III	19,4
	IV	19,3
2004	I	20,7
	II	19,1
	III	18,2
	IV	18,0
2005	I	18,9
	II	18,1
	III	17,4
	IV	16,7

W niniejszej tabeli przedstawiono kształtowanie się stopy bezrobocia (wg metodologii BAEL) w Polsce w latach 2003-2005.
Na podstawie dostępnych informacji wyznacz tendencję rozwojową w sposób mechaniczny. Sporządź wykres szeregu pierwotnego (danych empirycznych) i szeregów średnich ruchomych (danych teoretycznych).

Źródło: *Kwartalna informacja o aktywności ekonomicznej ludności* z lat 2003-2005, GUS, <www.stat.gov.pl>.

Zadanie 4.5.12.

Liczba uczniów w gimnazjach w Polsce w latach 2001-2008 kształtowała się następująco:

Lata	Liczba uczniów (w tys.)
2001	1 743,1
2002	1 709,0
2003	1 681,2
2004	1 648,7
2005	1 596,8
2006	1 528,8
2007	1 453,2
2008	1 381,4

Źródło: Bank Danych Regionalnych GUS.

Wyznacz tendencję rozwojową w sposób analityczny. Oszacuj liczbę gimnazjalistów w roku 2011 oraz oceń stopień dopasowania funkcji trendu do danych empirycznych.

Zadanie 4.5.13.

Długość czynnej sieci kanalizacyjnej w Polsce w latach 1999-2008 przedstawiała się następująco:

Lata	Długość sieci kanalizacyjnej (w tys. km)
1999	46,8
2000	51,1
2001	55,5
2002	61,0
2003	68,9
2004	73,9
2005	80,1
2006	84,9
2007	89,5
2008	94,7

Źródło: *Infrastruktura komunalna* z lat 2000-2009, GUS, Warszawa.

Wyznacz tendencję rozwojową w sposób analityczny. Oszacuj długość sieci kanalizacyjnej w roku 2012 oraz oceń stopień dopasowania funkcji trendu do danych empirycznych.

Zadanie 4.5.14.

W poniższej tabeli przedstawiono dzienną wartość sprzedaży pieczywa (w zł) w pewnym sklepie w ciągu czterech kolejnych tygodni.

Tydzień	Wartość sprzedaży (w zł)						
	Pn	Wt	Śr	Cz	Pt	So	Nd
1	540	366	426	402	660	234	106
2	558	350	448	396	678	248	98
3	532	372	434	390	638	252	94
4	544	364	444	410	646	240	102

Źródło: Dane umowne.

Na podstawie powyższych informacji należy wyodrębnić wahania okresowe.

Zadanie 4.5.15.

Poniższa tabela przedstawia liczbę zakładów sektora prywatnego w Polsce, które w poszczególnych kwartałach lat 2005-2007 zgłosiły zwolnienia grupowe pracowników.

Lata	Kwartał	Liczba zakładów
2005	I	167
	II	304
	III	305
	IV	335
2006	I	93
	II	118
	III	154
	IV	153
2007	I	61
	II	81
	III	102
	IV	107

Wiedząc, że liniowa funkcja trendu ma postać $\hat{y}_t = -18{,}18t + 283{,}18$, wyodrębnij wahania sezonowe.

Źródło: *Kwartalna informacja o rynku pracy* z lat 2005-2007, GUS, <www.stat.gov.pl>.

Zadanie 4.5.16.

Poniższa tabela przedstawia liczbę ofert pracy zgłoszonych do urzędów pracy w Polsce w poszczególnych kwartałach lat 2005-2007.

Lata	Kwartał	Liczba ofert (w tys.)
2005	I	192
	II	240
	III	260
	IV	193
2006	I	237
	II	326
	III	313
	IV	242
2007	I	300
	II	366
	III	323
	IV	253

Wiedząc, że liniowa funkcja trendu ma postać $\hat{y}_t = 9{,}43t + 209{,}12$, wyodrębnij wahania sezonowe.

Źródło: *Kwartalna informacja o rynku pracy* z lat 2005-2007, GUS, <www.stat.gov.pl>.

5. PODSUMOWANIE

Zadanie 5.1.

Rezultaty badania dotyczącego wysokości miesięcznych wydatków na czasopisma (w zł) ponoszonych przez emerytów z dwóch miast – Konina i Koła, przedstawiają się następująco:

Parametry	Emeryci	
	z Konina	z Koła
$\bar{x}_a$	15	12
M_e	14	13
D	13	14
x_{typ}	–	–
V_x	–	55%
W_{as}	0,4	–

Źródło: Dane umowne.

Wykorzystując podane informacje należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- uzupełnić podany zespół parametrów,
- przeprowadzić analizę porównawczą obydwu zbiorowości.

Zadanie 5.2.

Rezultaty badania dotyczącego powierzchni gospodarstw rolnych (w ha) w dwóch gminach – Kramsk i Krzymów, przedstawiają się następująco:

Parametry	Gospodarstwa	
	w Kramsku	w Krzymowie
$\bar{x}_a$	8,0	6,0
M_e	8,6	5,4
D	9,0	5,0
x_{typ}	–	–
V_x	–	20%
W_{as}	-0,2	–

Źródło: Dane umowne.

Wykorzystując podane informacje należy:

- określić zbiorowość, jednostkę i cechę statystyczną,
- uzupełnić podany zespół parametrów,
- przeprowadzić analizę porównawczą obydwu zbiorowości.

Zadanie 5.3.

Poniższe równania regresji przedstawiają zależność pomiędzy powierzchnią mieszkań (x – w m^2) a ich ceną (y – w tys. zł) w dwóch miastach – A i B:

- miasto A: $\hat{y} = 3{,}5x - 5$,
- miasto B: $\hat{y} = 2{,}5x - 5$.

Wskaż i uzasadnij, w którym mieście są tańsze mieszkania. Oszacuj cenę 40-metrowego mieszkania w mieście A, wiedząc, że:

$r_{xy} = 0{,}95$,

$s(y) = 32{,}05$.

Zadanie 5.4.

Poniższe równania regresji przedstawiają zależność pomiędzy wysokością rocznych dochodów użytkowników gospodarstw domowych (x – w tys. euro) a spożyciem ryb w ciągu roku (y – w kg/osobę) w dwóch krajach – A i B:

- kraj A: $\hat{y} = 0{,}5x + 0{,}5$,
- kraj B: $\hat{y} = 0{,}8x + 0{,}5$.

Wskaż i uzasadnij, w którym kraju spożywa się więcej ryb. Oszacuj spożycie ryb (w przeliczeniu na osobę) w gospodarstwie z kraju B uzyskującym roczne dochody na poziomie 20 tys. euro, wiedząc, że:

$r_{xy} = 0{,}93$,

$s(y) = 10$.

Zadanie 5.5.

Współzależność między miesięcznymi dochodami członków gospodarstw domowych (x – w tys. zł) a ich wydatkami na zakup kaszanki (y – w zł) opisują następujące linie regresji:

$\hat{y} = -3x + 23{,}5$

$\hat{x} = -0{,}25y + 6{,}7$

Na podstawie powyższych informacji należy:

- zinterpretować równanie regresji opisujące zależność wysokości wydatków na kaszankę od wysokości miesięcznych dochodów,
- określić kierunek oraz siłę związku korelacyjnego i zinterpretować otrzymany wynik.

Zadanie 5.6.

Współzależność między wielkością sprzedaży pewnego artykułu (y – w kg) a jego ceną (x – w zł) w kolejnych miesiącach 2009 r. opisują następujące linie regresji:

$\hat{y} = -20x + 277$,

$\hat{x} = -0{,}045y + 12$.

Na podstawie powyższych informacji należy:

- zinterpretować równanie regresji opisujące zależność wielkości sprzedaży artykułu od jego ceny,
- określić kierunek oraz siłę związku korelacyjnego i zinterpretować otrzymany wynik.

Zadanie 5.7.

Dynamika liczby studentów uczelni publicznych w Polsce w latach 2004-2008 kształtowała się następująco:

Lata	2004	2005	2006	2007	2008
Indeksy (rok 2004 = 100)	100	99,13	96,74	94,85	94,09

Źródło: Bank Danych Regionalnych GUS.

Zinterpretuj indeks liczby studentów w 2008 r. w porównaniu z 2004 r. Wiedząc, że w 2005 r. liczba studentów uczelni publicznych wynosiła 1319,10 tys. osób, ustal:

- absolutną liczbę studentów w poszczególnych latach,
- przyrost absolutny liczby studentów w 2008 r. w porównaniu z rokiem 2004.

Zadanie 5.8.

Dynamika liczby zarejestrowanych bezrobotnych w Polsce w latach 2004-2008 kształtowała się następująco:

Lata	2004	2005	2006	2007	2008
Indeksy (rok 2004 = 100)	100	92,45	76,99	58,23	49,13

Źródło: Bank Danych Regionalnych GUS.

Zinterpretuj indeks liczby zarejestrowanych bezrobotnych w 2008 r. w porównaniu z 2004 r. wiedząc, że w 2005 r. liczba bezrobotnych wynosiła 2 mln 773 tys. osób, ustal:

- absolutną liczbę zarejestrowanych bezrobotnych w poszczególnych latach,
- przyrost absolutny liczby bezrobotnych w 2008 r. w porównaniu z rokiem 2004.

Zadanie 5.9.

Liniowa funkcja trendu dla liczby osób w wieku poprodukcyjnym (w tys.) w Polsce w latach 1995-2008 ma następującą postać (źródło: obliczenia własne na podstawie Banku Danych Regionalnych GUS):

$$\hat{y}_t = 5280{,}96 + 60{,}51t$$

Zinterpretuj parametry równania oraz oszacuj liczbę osób w wieku poprodukcyjnym w 2011 r., wiedząc, że $S_y = 33{,}10$.

Zadanie 5.10.

Liniowa funkcja trendu dla liczby podmiotów sektora prywatnego (w tys.) zarejestrowanych w Polsce w latach 1998-2008 ma następującą postać (źródło: obliczenia własne na podstawie Banku Danych Regionalnych GUS):

$$\hat{y}_t = 2846{,}98 + 77{,}41t$$

Zinterpretuj parametry równania oraz oszacuj liczbę podmiotów zarejestrowanych w 2012 r., wiedząc, że $S_y = 92{,}14$.

ROZWIĄZANIA ZADAŃ

2.8.5. $\bar{x}_a = 8{,}2$; D=12; $M_e = 8$; $Q_1 = 5{,}5$; $Q_3 = 11{,}5$

2.8.6. $\bar{x}_a = 3{,}97$; D=3,50; $M_e = 3{,}90$; $Q_1 = 3{,}50$; $Q_3 = 4{,}50$

2.8.7. $\bar{x}_a = 2{,}52$; D=1; $M_e = 2$; $Q_1 = 1$; $Q_3 = 4$

2.8.8. $\bar{x}_a = 3{,}9$; D=4; $M_e = 4$; $Q_1 = 3$; $Q_3 = 5$

2.8.9. $\bar{x}_a = 53{,}33$; $D = 58{,}57$; $M_e = 55{,}36$; $Q_1 = 42{,}50$; $Q_3 = 63{,}39$

2.8.10. $\bar{x}_a = 128{,}33$; $D = 120$; $M_e = 122{,}92$; $Q_1 = 71{,}25$; $Q_3 = 176{,}39$

2.8.11. $\bar{x}_a = 16{,}8$; $D = 15{,}31$; $M_e = 16{,}64$; $Q_1 = 13{,}2$; $Q_3 = 21{,}07$; $R = 32$;
$s(x) = 6{,}46$; $10{,}34 < x_{typ} < 23{,}26$; $V_x = 38{,}45\%$; $A_s = 1{,}49$; $W_{as} = 0{,}23$

2.8.12. $\bar{x}_a = 95{,}8$; $D = 104$; $M_e = 97{,}6$; $Q_1 = 76{,}25$; $Q_3 = 114{,}67$; $R = 160$;
$s(x) = 29{,}84$; $65{,}96 < x_{typ} < 125{,}64$; $V_x = 31{,}15\%$; $A_s = -8{,}2$; $W_{as} = -0{,}27$

2.8.13.a. $\bar{x}_a = 29{,}87$; $D = 28{,}74$; $M_e = 29{,}09$; $Q_1 = 27{,}11$; $Q_3 = 32{,}15$; $R = 18$;
$s(x) = 3{,}78$; $26{,}09 < x_{typ} < 33{,}65$; $V_x = 12{,}65\%$; $W_{as} = 0{,}3$; $\alpha_4 = 3{,}05$

2.8.13.b. $\bar{x}_a = 35{,}46$; $D = 38{,}48$; $M_e = 36$; $Q_1 = 32{,}66$; $Q_3 = 38{,}59$; $R = 18$;
$s(x) = 3{,}77$; $31{,}69 < x_{typ} < 39{,}23$; $V_x = 10{,}63\%$; $W_{as} = -0{,}8$; $\alpha_4 = 2{,}31$

2.8.14.a. $\bar{x}_a = 9{,}13$; $D = 9{,}39$; $M_e = 9{,}24$; $Q_1 = 7{,}28$; $Q_3 = 11{,}07$; $R = 18$;
$s(x) = 3{,}11$; $6{,}02 < x_{typ} < 12{,}24$; $V_x = 34{,}06\%$; $W_{as} = -0{,}08$; $\alpha_4 = 3{,}35$

2.8.14.b. $\bar{x}_a = 7{,}69$; $D = 7{,}2$; $M_e = 7{,}5$; $Q_1 = 5{,}12$; $Q_3 = 9{,}94$; $R = 18$;
$s(x) = 3{,}67$; $4{,}02 < x_{typ} < 11{,}36$; $V_x = 47{,}72\%$; $W_{as} = 0{,}13$; $\alpha_4 = 2{,}82$

2.8.16. $\bar{x}_a = 1100$

2.8.19. $166 < x_{typ} < 184$; $V_x = 5{,}14\%$; $W_{as} = 0{,}33$

2.8.20. $M_e = 8{,}5$

2.8.21. mężczyźni: Me = 50,67; $44 < x_{typ} < 56$; $V_x = 12\%$; $W_{as} = -0{,}33$
kobiety: D = 42; $41 < x_{typ} < 49$; $V_x = 8{,}89\%$; $W_{as} = 0{,}75$

2.8.22. $\bar{x}_a = 1750$

2.8.23. $\bar{x}_a = 7{,}6$, $s(x) = 2{,}2$; $5{,}4 < x_{typ} < 9{,}8$; $V_x = 28{,}95\%$; $W_{as} = 0{,}045$; $\alpha_4 = 3{,}07$

2.8.24. $M_e = 37{,}61$; $Q_1 = 31{,}23$; $Q_3 = 46{,}42$; $Q = 7{,}60$;
30,01 < kwartylowy obszar zmienności < 45,21; $V_Q = 20{,}21\%$; $A_Q = 2{,}43$; $W_Q = 0{,}16$

2.8.25. $k = 0{,}684$

2.8.26. $k = 0{,}64$

3.5.1. $r_{xy} = 0{,}95$; $R^2 = 0{,}9$; $\hat{y} = 0{,}28x - 1{,}05$; $\hat{x} = 3{,}22y + 4{,}95$;
$3{,}75 < \hat{y}_{x=20} < 5{,}35$; $18{,}35 < \hat{x}_{y=5} < 23{,}75$; $V = 23{,}60\%$

3.5.2. $r_{xy} = 0{,}97$; $R^2 = 0{,}94$; $\hat{y} = 0{,}003x - 0{,}34$; $\hat{x} = 270{,}69y + 138{,}38$;
$2{,}35 < \hat{y}_{x=1000} < 2{,}97$; $1133{,}62 < \hat{x}_{y=4} < 1308{,}66$; $V = 12{,}76\%$

3.5.3. $r_{xy} = -0{,}95$; $R^2 = 0{,}9$; $\hat{y} = -5{,}07x + 59{,}83$; $\hat{x} = -0{,}18y + 11{,}15$;
$25{,}95 < \hat{y}_{x=6} < 32{,}87$; $5{,}1 < \hat{x}_{y=30} < 6{,}4$; $V = 10{,}03\%$

3.5.4. $r_{xy} = -0{,}96$; $R^2 = 0{,}92$; $\hat{y} = -4{,}07x + 510{,}4$; $\hat{x} = -0{,}23y + 117{,}59$;
$431{,}06 < \hat{y}_{x=17} < 451{,}36$; $23{,}2 < \hat{x}_{y=400} < 27{,}98$; $V = 2{,}55\%$

3.5.5. $r_{xy} = -0{,}99$

3.5.9. $\hat{y} = 6{,}56$

3.5.10. $\hat{x} = 30$

3.5.11. $\hat{y} = 0{,}014x + 19{,}64$; $23{,}04 < \hat{y}_{x=1000} < 44{,}24$

3.5.12. $r_{xy} = 0{,}91$; $27{,}48 < \hat{y}_{x=20} < 42{,}28$

3.5.13. $\hat{y} = 0{,}13x - 1{,}72$; $23{,}82 < \hat{y}_{x=200} < 24{,}74$; $R^2 = 0{,}98$

3.5.14. $r_{xy} = 0{,}97$; $4521{,}62 < \hat{y}_{x=12} < 4988{,}38$; $\varphi^2 = 0{,}06$

3.5.15. $\hat{y} = 0{,}2x + 1{,}13$; $2{,}45 < \hat{y}_{x=10} < 3{,}81$; $R^2 = 0{,}56$

3.5.16. $b_y = 0{,}33$; $\varphi^2 = 0{,}1$; $142{,}49 < \hat{x}_{y=100} < 198{,}35$

3.5.17. $C_{skor} = 0{,}434$; $T_{xy} = 0{,}25$; $V_c = 0{,}277$

4.5.3. $\overline{S}_{t/t-1} = 99{,}52\%$; $y_{2011}^{9} = 78685$

4.5.4. $I_w = 103{,}77\%$; ${}^{L}I_q = 90{,}46\%$; ${}^{P}I_q = 90{,}34\%$; ${}^{F}I_q = 90{,}4\%$;
${}^{L}I_p = 114{,}87\%$; ${}^{P}I_p = 114{,}72\%$; ${}^{F}I_p = 114{,}79\%$

4.5.5. $I_w = 102{,}17\%$; ${}^{L}I_q = 106{,}99\%$; ${}^{P}I_q = 106{,}15\%$; ${}^{F}I_q = 106{,}57\%$;
${}^{L}I_p = 96{,}25\%$; ${}^{P}I_p = 95{,}5\%$; ${}^{F}I_p = 95{,}87\%$

4.5.6. $I_w = 97{,}81\%$; ${}^{L}I_q = 126{,}88\%$; ${}^{P}I_q = 125{,}2\%$; ${}^{F}I_q = 126{,}04\%$;
${}^{L}I_p = 78{,}13\%$; ${}^{P}I_p = 77{,}09\%$; ${}^{F}I_p = 77{,}61\%$

4.5.7. $I_w = 93{,}01\%$; ${}^{L}I_q = 84{,}41\%$; ${}^{P}I_q = 79{,}72\%$; ${}^{F}I_q = 82{,}03\%$;
${}^{L}I_p = 116{,}67\%$; ${}^{P}I_p = 110{,}19\%$; ${}^{F}I_p = 113{,}38\%$

4.5.12. $\hat{y}_t = -51{,}43t + 1824{,}21$; $1231{,}91 < \hat{y}_{11} < 1285{,}05$ $\varphi^2 = 0{,}04$;
$R^2 = 0{,}96$; $V_{Sy} = 1{,}67\%$

4.5.13. $\hat{y}_t = 5{,}51t + 40{,}34$; $116{,}46 < \hat{y}_{14} < 118{,}50$; $\varphi^2 = 0{,}003$; $R^2 = 0{,}997$;
$V_{Sy} = 1{,}44\%$

4.5.15. $S_I = 0{,}53$; $S_{II} = 0{,}88$; $S_{III} = 1{,}16$; $S_{IV} = 1{,}42$

4.5.16. $g_I = -13{,}27$; $g_{II} = 44{,}97$; $g_{III} = 23{,}53$; $g_{IV} = -55{,}23$; $S_i = g_i$

5.1. Emeryci z Konina: $s(x) = 5$; $10 < x_{typ} < 20$; $V_x = 33{,}33\%$,
emeryci z Koła: $s(x) = 6{,}6$; $5{,}4 < x_{typ} < 18{,}6$; $W_{as} = -0{,}3$

5.2. Gospodarstwa w Kramsku: $s(x) = 5$; $3 < x_{typ} < 13$; $V_x = 62{,}5\%$,
gospodarstwa w Krzymowie: $s(x) = 1{,}2$; $4{,}8 < x_{typ} < 7{,}2$; $W_{as} = 0{,}83$

5.3. $125 < \hat{y}_{x=40} < 145$

5.4. $12{,}82 < \hat{y}_{x=20} < 20{,}18$

5.5. $r_{xy} = -0{,}87$

5.6. $r_{xy} = -0{,}95$

5.7. $P_{ab/0} = -78{,}64$

5.8. $P_{ab/0} = -1525{,}83$

5.9. $6276{,}53 < \hat{y}_{17} < 6342{,}73$

5.10. $3915{,}99 < \hat{y}_{15} < 4100{,}27$

BIBLIOGRAFIA

Chromińska M., Ignatczyk W., *Statystyka. Teoria i zastosowanie*, WSB, Poznań 2004.

Jóźwiak J., Podgórski J., *Statystyka od podstaw*, PWE, Warszawa 2009.

Kukuła K., *Elementy statystyki w zadaniach*, PWN, Warszawa 2003.

Metody statystyczne. Teoria i zadania, red. C. Domański, UŁ, Łódź 2001.

Metody statystyczne. Zarys teorii, przykłady i zadania, red. J. Suchecka, Wydział Zarządzania Politechniki Częstochowskiej, Częstochowa 2002.

Podgórski J., *Statystyka dla studiów licencjackich*, PWE, Warszawa 2005.

Pułaska-Turyna B., *Statystyka dla ekonomistów*, Difin, Warszawa 2005.

Sobczyk M., *Statystyka*, PWN, Warszawa 2007.

Statystyka pod red. J. Paradysza, AE, Poznań 2005.

Statystyka. Zbiór zadań, red. H. Kassyk-Rokicka, PWE, Warszawa 2005.

Trzpiot G., Kończak G., *Statystyka opisowa w przykładach i zadaniach*, GWSH, Katowice 2009.

Zeliaś A., Pawełek B., Wanat S,, *Metody statystyczne. Zadania i sprawdziany*, PWE, Warszawa 2002.

Wysocki F., Lira J., *Statystyka opisowa*, AR, Poznań 2007.