STATYSTYKA

PODRĘCZNIK. STUDIA PRZYPADKÓW

ANNA TATARCZAK

9
TOM

Tom 9 Kompetencje, Wiedza, Innowacje – Zintegrowany Program Rozwoju WSEI w Lublinie – Etap II

STATYSTYKA

Część pierwsza
Podręcznik

Anna Tatarczak

Lublin 2021

WYŻSZA SZKOŁA EKONOMII I INNOWACJI W LUBLINIE

Seria wydawnicza:
Kompetencje, Wiedza, Innowacje – Zintegrowany Program Rozwoju WSEI w Lublinie – Etap II
Tom 9

STATYSTYKA
Cz. I - Podręcznik Cz. II - Studia przypadków

Autor:
Anna Tatarczak
ORCID: 0000-0001-8573-5791

Recenzenci:
dr Anna Makarewicz
dr Adam Kiersztyn

Korekta:
Magdalena Grela-Tokarczyk

Skład, łamanie:
Marta Krysińska-Kudlak

Projekt okładki:
Patrycja Kaczmarek

Grafika na okładce:
Mediamodifier/Pixabay.com
rawpixel.com/Freepik.com

@Copyright by
Innovatio Press, Lublin 2021

Wszelkie prawa zastrzeżone. W sprawie kopiowania lub rozpowszechniania fragmentów albo całości niniejszej pracy należy kontaktować się z wydawcą.
Publikacja oraz recenzje zrealizowano w ramach Programu Operacyjnego Wiedza Edukacja Rozwój 2014-2020 współfinansowanego ze środków Europejskiego Funduszu Społecznego. Proces wydawniczy (korekta, skład) sfinansowano z dotacji na zadania związane z zapewnieniem osobom niepełnosprawnym warunków do pełnego udziału w procesie kształcenia.

Printed in Poland
Innovatio Press Wydawnictwo Naukowe
Wyższej Szkoły Ekonomii i Innowacji
20-209 Lublin, ul. Projektowa 4
tel.: +48 81 749 17 77, fax: + 48 81 749 32 13
www.wsei.lublin.pl

ISBN – wersja elektroniczna: 978-83-66159-83-9

Spis treści

Wstęp 5

1. Statystyka i podstawowe pojęcia statystyczne 7
   1.1. Co to jest statystyka? 8
   1.2. Podstawowe pojęcia statystyczne 11
   1.3. Metody badań statystycznych 17
   1.4. Etapy badania statystycznego 21
   1.5. Zastosowanie programów komputerowych w statystyce 24
   1.6. Pytania sprawdzające 30

2. Dane statystyczne 36
   2.1. Rodzaje danych i ich źródła 37
   2.2. Metody badań w naukach ekonomicznych 40
   2.3. Statystyka publiczna 41
   2.4. Prezentacja graficzna danych 43
   2.5. Zastosowanie pakietu EXCEL 49
   2.6. Pytania sprawdzające 51

3. Klasyfikacja danych 55
   3.1. Szereg szczegółowy 59
   3.2. Szereg rozdzielczy przedziałowy 63
   3.3. Pytania sprawdzające 68

4. Analiza struktury 70
   4.1. Ocena położenia rozkładu 72
      4.1.1. Miary położenia klasyczne 73
      4.1.2. Miary położenia pozycyjne 76
   4.2. Ocena dyspersji 83
   4.3. Ocena asymetrii 91
   4.4. Ocena koncentracji 93
   4.5. Zastosowanie pakietu EXCEL 96
   4.6. Pytania sprawdzające 98

5. Analiza współzależności cech 103
5.1. Wprowadzenie do badania współzależności cech 104
5.2. Analiza korelacji.....105
5.2.1. Analiza korelacji cech mierzalnych 106
5.2.2. Analiza korelacji cech niemierzalnych.....113
5.3. Analiza regresji 114
5.3.1. Liniowa funkcja regresji.....114
5.3.2. Ocena liniowej funkcji regresji.....117
5.4. Zastosowanie pakietu EXCEL 120
5.5. Pytania sprawdzające 121

6. Analiza dynamiki.....125
6.1. Definicja i klasyfikacja metod analizy dynamiki 126
6.2. Przyrosty bezwzględny.....128
6.3. Przyrosty względne 131
6.4. Indeksy indywidualne 134
6.5. Indeksy agregatowe 138
6.6. Średnie tempo zmian 143
6.7. Pytania sprawdzające 145

7. Wzory 149
7.1. Załącznik 1 149
7.2. Załącznik 2 154

8. Literatura 155

Wstęp

Głównym celem podręcznika jest uporządkowanie i opis podstawowych **metod i technik stosowanych na wykładzie ze statystyki opisowej**. Z wykładem związane są ćwiczenia laboratoryjne, w trakcie których są praktycznie ilustrowane zagadnienia przedstawiane na wykładzie. Podstawowym narzędziem programowym, stosowanym do opisu poszczególnych procedur obliczeniowych, jak i do obliczeń w laboratorium komputerowym, jest program EXCEL.

Podręcznik ten jest głównie przeznaczony dla studentów Wyższej Szkoły Ekonomii i Innowacji poznających podstawy statystyki. Mogą z niej również korzystać praktycy prowadzący badania własne lub kierujący pracami badawczymi.

Podręcznik składa się z sześciu rozdziałów, jego układ jest dostosowany do programu statystyki obowiązującego na kierunkach ekonomicznych. Rozdział pierwszy przedstawia podstawowe pojęcia statystyki oraz etapy badania statystycznego. Rozdział drugi jest poświęcony danym i ich graficznej reprezentacji. Rozdział trzeci wprowadza czytelnika w pojęcie klasyfikacji danych w postaci szeregu. Rozdział czwarty szeroko opisuje statystyki opisowe, a kolejny rozdział dotyczy analizy współzależności. Ostatni szósty rozdział dotyczy dynamiki zjawisk.

Każdy rozdział pracy zamyka szeroki **zestaw pytań do samodzielnego rozwiązania**. Integralną częścią większości rozdziałów są informacje dotyczące korzystania z pakietu komputerowego EXCEL podczas rozwiązywania problemów z zakresu statystyki.

Dołączona bibliografia przedmiotu umożliwi zainteresowanym pogłębienie wiedzy Czytelnikom dotarcie do innych podręczników ze statystyki.

1 2 3 4 5 6

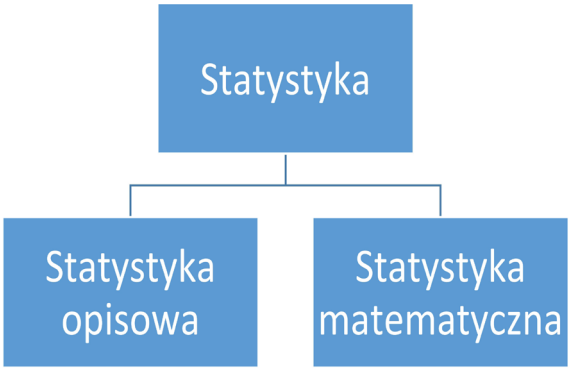
1. Statystyka i podstawowe pojęcia statystyczne

1.1. Co to jest statystyka?

Statystyka pochodzi z łacińskiego słowa *status* (stan rzeczy). Jest nauką, która zajmuje się metodami badania prawidłowości występujących w zbiorowościach oraz charakteryzowaniem tych prawidłowości za pomocą liczb.

Głównym celem analizy statystycznej jest:

- przetworzenie dużej liczby danych do przejrzystej postaci,
- interpretacja danych,
- uogólnienie danych.



Rysunek 1. Dziedziny statystyki.

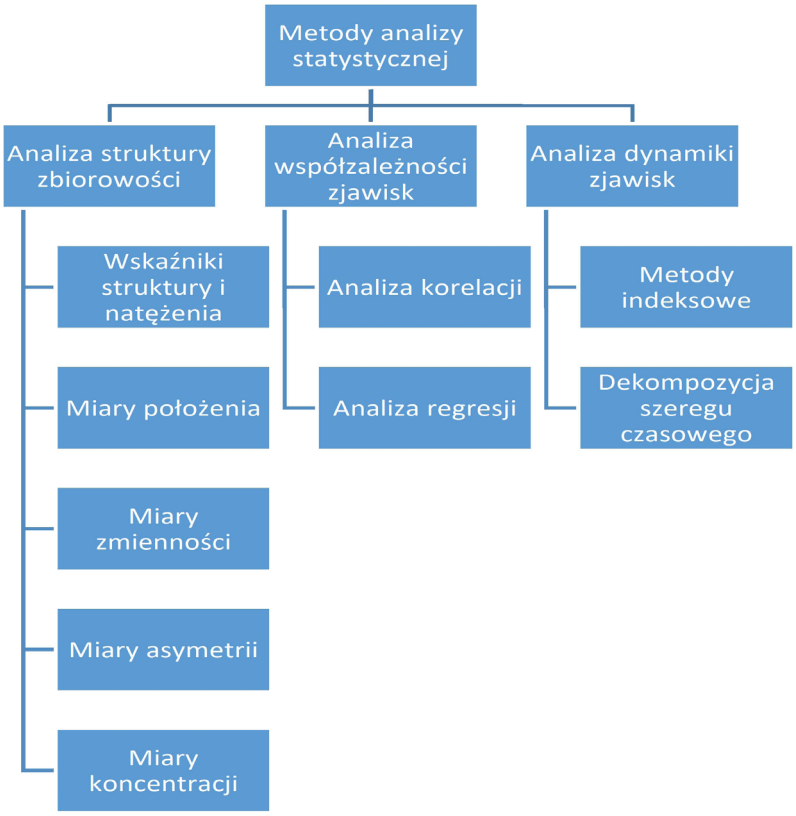
Statystykę z punktu widzenia dziedziny naukowej możemy podzielić na statystykę opisową i matematyczną (patrz Rysunek 1).

Statystyka matematyczna dostarcza metod wyboru prób losowych i reguł wnioskowania, czyli pozwala na uogólnienia wniosków wynikających z obserwacji części zbiorowości, tak aby ryzyko popełnienia błędu było małe (rachunek prawdopodobieństwa).

Statystyka opisowa dostarcza metod i procedur gromadzenia, opracowania i prezentacji danych statystycznych, jej celem jest zwięzły opis materiału statystycznego.

W ramach statystyki opisowej można wyróżnić trzy podstawowe działy analizy (patrz Rysunek 2):

- 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.



Rysunek 2. Podział metod analizy statystycznej.

Tabela 1. Podział statystyki.

	Kryterium formalno-statystyczne	
	Opis	**Wnioskowanie**
Analiza struktury zjawisk	Miary tendencji centralnej Miary rozproszenia	
	Miary asymetrii Miary spłaszczenia Miary koncentracji	
Analiza współzależności zjawisk	Miary regresji i korelacji dwóch zmiennych Miary regresji i korelacji wielorakiej	
Analiza dynamiki zjawisk	Miary tendencji rozwojowej Miary okresowości Miary wahań resztowych	
	Indeksy statystyczne	

Źródło: A. Luszniewicz, *Statystyka ogólna*, PWE, Warszawa 1987, s. 28.

Statystyka jest obecnie jedną z najpowszechniej wykorzystywanych dziedzin matematyki. Przykładowe obszary zastosowań statystyki to:

- badania rynku, np. prognozowanie sprzedaży, segmentacji rynku, analizy migracji klientów,
- badania marketingowe, np. skuteczność kampanii reklamowej, analizy sprzedażowe, badanie wartości marki,
- badania satysfakcji klientów, np. poziom jakości obsługi klienta,
- badania farmaceutyczne, np. substytucja leków w aptekach.

Statystyka wykorzystywana jest w wielu innych obszarach, i tak wyróżniamy kilka rodzajów statystyki stosowanej:

- ekonometria,
- demografia,
- fizyka statystyczna,
- biometria,
- psychometria,
- statystyka gospodarcza,
- socjologia statystyczna.

1.2. Podstawowe pojęcia statystyczne

Zbiorowości statystyczne (populacje statystyczne) – zbiór dowolnych elementów (osób, przedmiotów, faktów), podobnych pod względem określonych cech (ale nie identycznych) i poddanych badaniom statystycznym.

Podział zbiorowości statystycznych jest następujący:

- skończenie lub nieskończenie liczne,
- statyczne (elementy obserwowane są w danym momencie) lub dynamiczne (w określonym przedziale czasu),
- jednowymiarowe (badanie pod względem jednej cechy), wielowymiarowe (wielu cech),
- jednorodne lub niejednorodne – oznacza, że wszystkie jednostki zbiorowości statystycznej muszą posiadać co najmniej jedną wspólną cechę.

Przykład

Zbiorowościami statystycznymi są: pracownicy danego przedsiębiorstwa, sklepy spożywcze w danym mieście, środki transportu danego przedsiębiorstwa, spółki notowane na Warszawskiej Giełdzie Papierów Wartościowych.

Jednostka statystyczna – elementy składowe zbiorowości poddawane bezpośredniej obserwacji lub pomiarowi (obiekt badania). W celu poprawnego i jednoznacznego określenia jednostki statystycznej należy dokładnie określić ją pod względem:

- rzeczowym (kto lub co jest przedmiotem badania),
- czasowym (jakiego momentu lub okresu dotyczy badanie),
- terytorialnym (gdzie znajduje się zbiorowość).

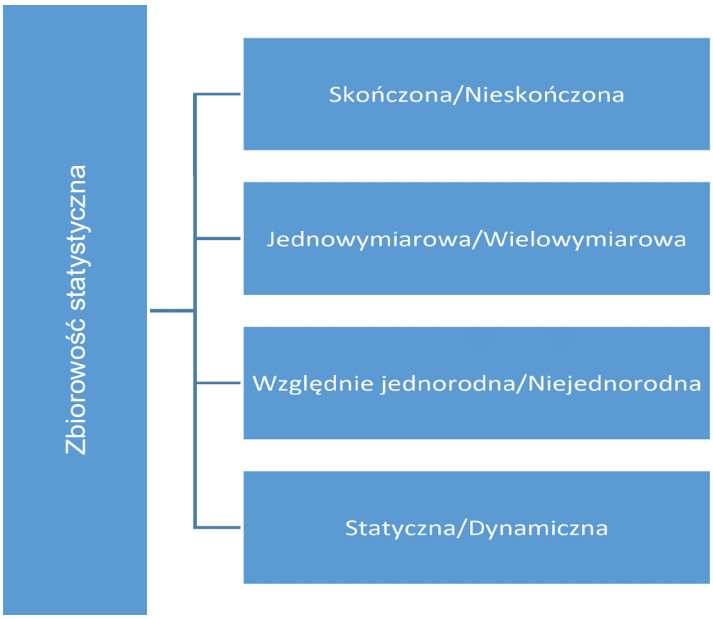
Przykład

Jednostkami statystycznymi mogą być studenci opisani następującym zdaniem: studenci pierwszego roku Wyższej Szkoły Ekonomii i Innowacji w Lublinie w dniu 1 września 2021 roku według oceny końcowej z przedmiotu Statystyka.

Zbiorowość (populacja) generalna – wszystkie elementy, będące przedmiotem badania, co do których formułowane są wnioski ogólne.

Zbiorowość próbna (próba) – podzbiór populacji generalnej, obejmujący część jej elementów – wybranych w określony sposób. Próba podlega badaniu statystycznemu, a wynik jest uogólniany na zbiorowość generalną.

Zbiorowość statystyczna i cecha statystyczna to dwa podstawowe pojęcia statystyki. Poniżej przedstawiamy różne sposoby klasyfikacji zbiorowości statystycznej w zależności od przyjętego kryterium.



Rysunek 3. Klasyfikacja zbiorowości statystycznej.

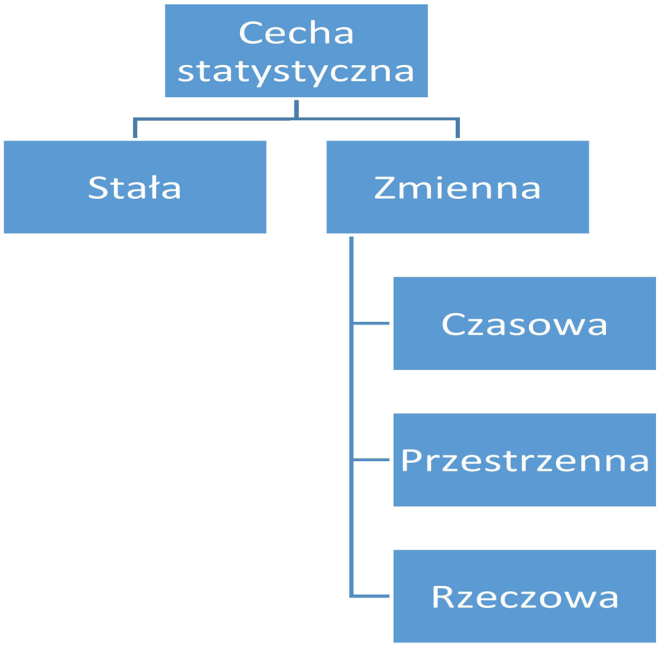
Cecha statystyczna – właściwość jednostki statystycznej podlegająca bezpośredniemu badaniu statystycznemu. Cecha może przyjmować różne wartości, które nazywamy odmianami lub wariantami.

Przykład

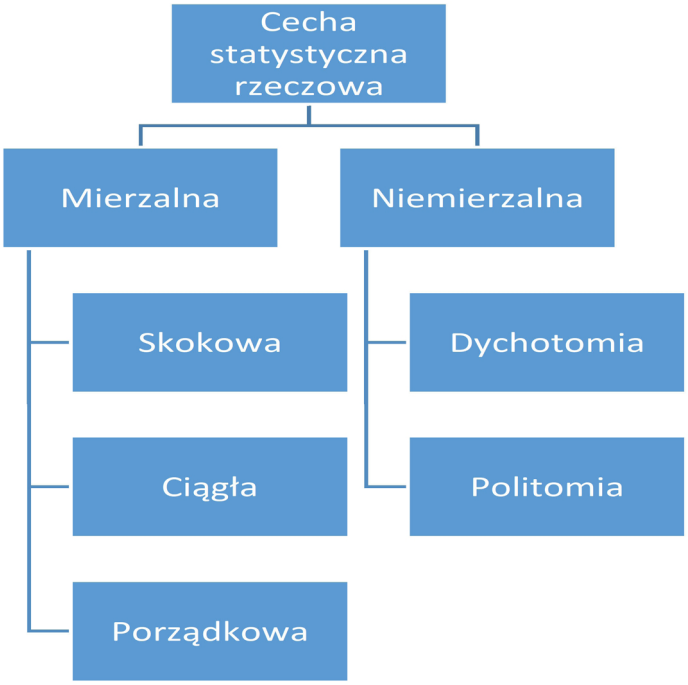
Cecha płeć ma dwa warianty: kobieta i mężczyzna.
Cecha ocena ma cztery warianty: bardzo dobry, dobry, dostateczny i niedostateczny.

Cechy statystyczne dzielimy na stałe i zmienne. Z kolei cechy zmienne dzielimy na:

- czasowe, np. rok urodzenia,
- przestrzenne, np. miejsce zamieszkania,
- rzeczowe:
  - ilościowe (mierzalne) – np. wzrost, masa, wiek,
    - ciągłe – np. wzrost, masa, wiek (w rozumieniu ilości dni między datą urodzin a datą badania),
    - porządkowe (quasi-ilościowe) – np. klasyfikacja wzrostu: (niski, średni, wysoki),
    - skokowe (dyskretne) – np. ilość posiadanych dzieci, ilość gospodarstw domowych, wiek (w rozumieniu ilości skończonych lat),
  - jakościowe (niemierzalne) – np. kolor oczu, płeć, grupa krwi,
    - dychotomia, np. płeć,
    - politomia, np. wykształcenie.



Rysunek 4. Podział cech statystycznych.



Rysunek 5. Podział cech statycznych rzeczowych.

Przykład

Przykładowe cechy to:
- jakościowe: wykształcenie, zawód, miejsce zamieszkania, płeć, kolor oczu,
- skokowe: liczba studentów na danym roku, liczba dzieci w rodzinie, rozmiar obuwia,
- ciągłe: temperatura na dworze, staż pracy, zyski firmy, masa danego produktu.

W przypadku badań zbiorowości wielowymiarowych zmienne (cechy mierzalne) dzielimy na:
- stymulanty – te cechy, których wyższe wartości pozwalają zakwalifikować daną jednostkę statystyczną jako lepszą z punku widzenia realizowanego badania,
- dominanty – cechy, których wysokie wartości świadczą o niskiej pozycji jednostki w zbiorze.

Przykład

W 100 zakładach produkcyjnych przeprowadzono badania w celu ustalenia zależności między wydajnością mierzoną w sztukach X a płacą w PLN. W analizowanym przykładzie zbiorowość to: 100 zakładów produkcyjnych, a jednostką jest jeden zakład produkcyjny.

Badanie statystyczne – ogół prac mających na celu poznanie struktury określonej zbiorowości statystycznej.

Próba – podzbiór populacji generalnej podlegający bezpośredniemu badaniu ze względu na ustaloną cechę w celu wyciągnięcia wniosków na temat rozkładu wartości tej cechy w całej populacji generalnej.

Obserwacja statystyczna – proces zbierania informacji statystycznej.

Pomiar – czynność przyporządkowania liczb przedmiotom (obiektom) lub wydarzeniom zgodnie z pewnym zbiorem reguł. Wynikiem pomiaru są dwa rodzaje wielkości, te, które mówią o liczebności zbioru obiektów, i te, które charakteryzują stopień nasilenia zjawiska.

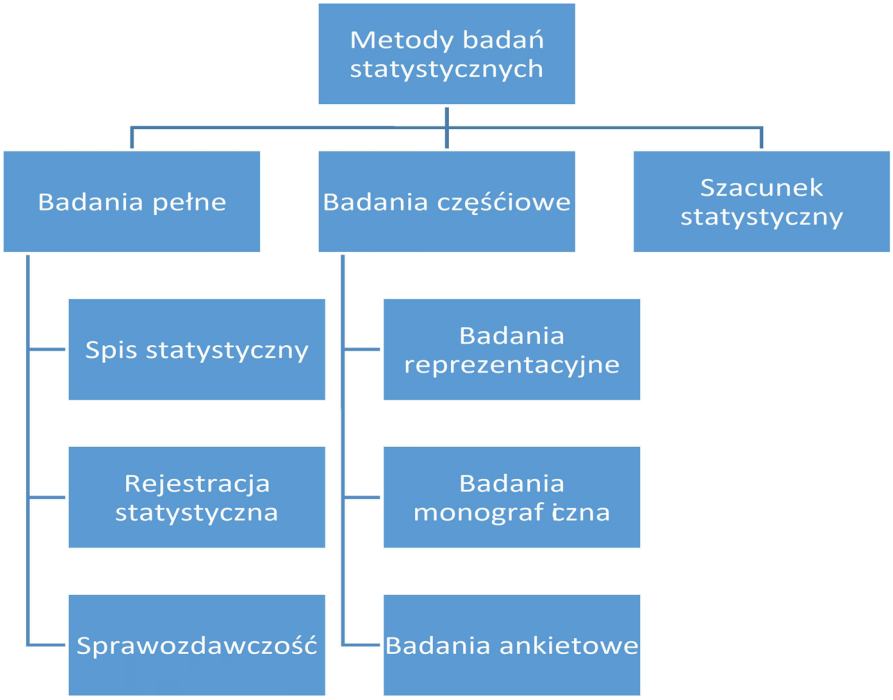
Skale pomiarowe ze względu na relacje dzielimy na:

- **nominalne** – relacja: równe różne; pomiar polega na zastosowaniu liczby jako nazwy, czyli grupowaniu jednostek w klasy (kategorie), którym przypisuje się nazwy czy liczby, np. studenci według rodzaju studiów, szczególny przypadek – skala dychotomiczna (dwupunktowa);
- **porządkowe** – relacja: większe lub mniejsze; pomiar polega na grupowaniu jednostek w klasy (kategorie), którym przypisuje się nazwy lub liczby i porządkuje się te klasy ze względu na stopień natężenia, w jakim posiadają one badaną cechę;
- **przedziałowe** – relacja: większe o tyle; pomiar występuje wtedy, gdy uporządkowany zbiór wartości cechy składa się z liczb rzeczywistych. Zero w tej skali ustalone jest dowolnie, np. skala Celsjusza i Fahrenheita, skala pozwala stwierdzić, tylko o ile jest coś wyższe;
- **stosunkowe** (ilorazowe) – relacja: tyle razy większe; spełnia wszystkie aksjomaty liczb, pomiary w tej skali charakteryzują się stałymi ilorazami i zerem bezwzględnym, tylko w tej skali możliwe jest porównywanie jednostek za pomocą względnych charakterystyk: np. jeden obiekt jest dwa razy cięższy od drugiego.

Rysunek 6. Podział skal pomiarowych.

1.3. Metody badań statystycznych

Każde badanie statystyczne wymaga ustalenia właściwej metody. Wybór metody jest uzależniony od celu badania, rodzaju zbiorowości statystycznej, tematu badania, jego szczegółowości, środków, jakimi dysponujemy. Jedną z metod badań statystycznych są badania statystyczne pełne, częściowe oraz szacunek statystyczny – patrz Rysunek 7.



Rysunek 7. Podział metod badań statystycznych ze względu na liczbę jednostek objętych badaniem.

Spis statystyczny to jeden z rodzajów badania statystycznego pełnego. Stosuje się go do określenia zachowań poszczególnych zjawisk masowych w określonym z góry czasie (momencie przeprowadzenia spisu statystycznego). Na bazie spisów statystycznych jest możliwe wyciąganie konkretnych wniosków dotyczących danego zjawiska na przestrzeni dłuższego czasu. Spisy dostarczają konkretnych informacji dotyczących ludności, tj. wieku, zawodu, pozio-

mu wykształcenia itp., ale nie tylko, bo mogą też charakteryzować gospodarkę (wielkość przedsiębiorstw, sekcje PKD itp.), środowisko naturalne (czystość rzek, zalesienie itp.), czy gospodarstwa rolne.

Rejestracja statystyczna polega na systematycznym zbieraniu określonych faktów, ustalonych wcześniej jako przedmiot badania statystycznego. Dotyczy przede wszystkim zjawisk, które stale występują w całym społeczeństwie bądź w jego grupach i nie należą do zjawisk typowo losowych. Dzięki nim można określić zachowania społeczeństwa lub jego grup na przestrzeni wielu lat.

Sprawozdawczość statystyczna obejmuje informacje statystyczne przekazywane obligatoryjnie na jednolitych formularzach przez osoby fizyczne, podmioty gospodarcze i instytucje do władz nadrzędnych lub organów statystycznych (np. zeznania podatkowe).

Szacunek statystyczny polega na ustaleniu nieznanych wielkości lub właściwości zbiorowości na podstawie innych zbiorowości pozostających w określonym związku przyczynowo-skutkowym z interesującą nas nieznaną populacją.

Badania ankietowe przeprowadzane są wówczas, gdy informacje o zbiorowości są gromadzone za pomocą ankiety rozsyłanej do ściśle określonych osób lub instytucji. Badania monograficzne polegają na szczegółowym jednym z wielu punktów widzenia opisie i analizie wybranej jednostki statystycznej. Warunkiem koniecznym jest to, że badana jednostka powinna być typowa, powszechnie występująca lub wskazująca kierunek rozwoju, a nie przodująca.

Podstawowym narzędziem badawczym statystyki jest kwestionariusz ankietowy. Ankieta powinna być tak skonstruowana, aby stwarzała ankietowanemu możliwość formułowania własnych opinii, spostrzeżeń i wniosków. Odpowiedzi ujmowane są częściej werbalnie (opisowo) niż liczbowo. Powinna być w miarę krótka. Jedną z podstawowych zasad stosowanych w badaniach tego rodzaju jest zachowanie anonimowości ankietowanych.

Kwestionariusz składa się z:

- tytułu,
- wyjaśnienia celu badania,

- instrukcji (sposobu wypełnienia ankiety, określenia ilości wyboru odpowiedzi oraz czasu na wypełnienie),
- wyboru takich pytań, aby dokładnie odnaleźć odpowiedź na swoje pytanie – metryczki (wiek, płeć, miejsce zamieszkania).

Ze względu na formę w kwestionariuszach ankietowych stosowane są następujące rodzaje pytań:

- **pytania otwarte** – nie posiadają ustalonego zestawu odpowiedzi. Respondent otrzymuje możliwość swobodnego wypowiedzenia się na dany temat;
- **pytania półotwarte** – łączą w sobie cechy pytania zamkniętego i otwartego. Respondent dokonuje wyboru tylko jednej odpowiedzi z przewidzianego zestawu. W przypadku gdy żadna z proponowanych odpowiedzi nie oddaje rzeczywistej sytuacji, ma on mimo to możliwość udzielenia odpowiedzi, wybierając podpunkt „inne”;
- **pytania zamknięte** – zawierają pewną liczbę przygotowanych odpowiedzi, spośród których respondent dokonuje wyboru. W pytaniach jednokrotnego wyboru wybiera się tylko jedną odpowiedź, tę, którą uzna za najlepiej oddającą stan jego wiedzy lub sytuację, w której się znajduje. Natomiast pytania wielokrotnego wyboru posiadają ustaloną liczbę wariantów odpowiedzi, z których respondent może wybrać jedną bądź kilka odpowiedzi;
- **pytania dychotomiczne** proponują respondentowi do wyboru dwie wykluczające się wzajemnie odpowiedzi, zwykle „tak – nie”, bez możliwości udzielenia odpowiedzi neutralnej typu „nie wiem”;
- **pytania filtrujące** służą autorom badania do podzielenia respondentów na tych, którym na podstawie uzyskanej odpowiedzi będą zadawane określone, dalsze pytania, oraz na tych, których nie dotyczy problematyka poruszana w tych pytaniach.

Przykład

Czy uważa Pan(i), że ruch feministyczny w Polsce jest skuteczny?

☐ tak
☐ nie
☐ nie wiem

Rysunek 8. Kwestionariusz – pytanie dychotomiczne.

Czy czujesz się związany z wymienionymi niżej osobami?

	Tak	Nie	Nie wiem
Rodzina	☐	☐	☐
Przyjaciele	☐	☐	☐
Sąsiedzi	☐	☐	☐

Rysunek 9. Kwestionariusz – pytania zamknięte.

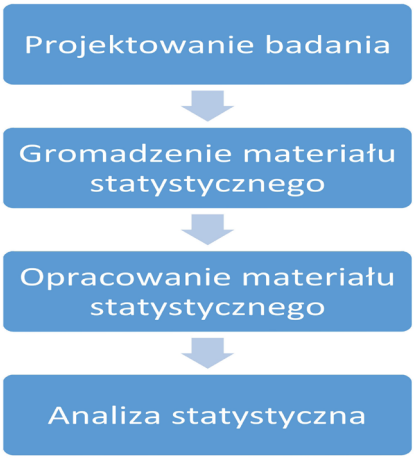
Metoda reprezentacyjna jest częściowym badaniem statystycznym opartym na próbie pobranej ze zbiorowości generalnej w sposób losowy.

Wybór jednej z wyżej opisanych metod badania statystycznego jest uzależniony od następujących czynników:

- celu badania,
- zbiorowości statystycznej,
- kosztów danego badania.

1.4. Etapy badania statystycznego

Badanie statystyczne to zespół czynności, których celem jest określenie w zbiorowości statystycznej prawidłowości dotyczących badanych cech statystycznych. Celem każdego badania statystycznego jest wyznaczenie wartości wybranych parametrów badanej zbiorowości statystycznej oraz analiza tej zbiorowości. Wybierając metodę badania, uwzględniamy różnorodne czynniki. Najważniejsze z nich to: cel badania, rodzaj zbiorowości statystycznej, tematyka badania, stopień szczegółowości badania.



Rysunek 10. Etapy badania statystycznego.

Badanie statystyczne prowadzone jest w kilku etapach (patrz Rysunek 10):

1. **Projektowanie badania statystycznego (przygotowanie badania statystycznego) obejmuje między innymi:**
   - określenie celu badania i sprecyzowanie problemu, który ma być rozwiązany dzięki uzyskaniu informacji pochodzących z badań statystycznych;
   - hipotezy badawcze, określa się je na ogół w badaniach naukowych, hipotezy muszą być możliwe do zweryfikowania, nie mogą dotyczyć np. przyszłości;
   - określenie przedmiotu badania, czyli określenie zbiorowości i jednostki statystycznej. Jednostki statystyczne powinny być

określone pod względem rzeczowym, czasowym i terytorialnym, np. przedsiębiorstw, gospodarstw domowych, grupy osób.

2. Obserwacja, czyli zebranie materiału statystycznego.

W zależności od źródła zbieranych informacji wyróżniamy materiał pierwotny i wtórny informacji. **Materiał pierwotny** jest rezultatem zaplanowanych specjalnie do badania dochodzeń statystycznych, zaś **materiałem wtórnym** nazywamy dane gromadzone dla innych celów i wykorzystywane wtórnie do określonych badań. Techniki zebrania materiału statystycznego to np. obserwacja, wywiad, ankieta, eksperyment.

Zebrany materiał statystyczny stanowi surowy materiał statystyczny. Z reguły jest on obarczony błędami: systematycznymi oraz losowymi. **Błędy systematyczne** wynikają z jednokierunkowych tendencji do zniekształcania rzeczywistości. **Błędy losowe** (przypadkowe) mogą wynikać z nieświadomości, braku wiedzy, pobłażliwego podejścia do przekazywania danych itd. Błędy systematyczne z reguły kumulują się w wynikowej informacji statystycznej. Błędy przypadkowe, co do zasady, powinny się znosić.

Surowy materiał statystyczny może zawierać błędy logiczne oraz rachunkowe. **Błędy logiczne** widoczne są w danych sprzecznych z przepisami prawa, np. wynagrodzenie mniejsze od minimalnego. **Błędy rachunkowe** wynikają na ogół z niewłaściwego zastosowania operacji arytmetycznych (np. sumowania).

W automatycznej kontroli danych stosuje się **błędy twarde** – błędy, które muszą zostać poprawione, oraz **błędy uznaniowe** – sytuacje, w których algorytmy kontrolne sygnalizują odstępstwo od założonych wartości, ale dane mogą być poprawne.

3. Opracowanie i prezentacja materiału statystycznego.

W trzecim etapie wykonujemy następujące czynności:
- kontrolę formalną i merytoryczną,
- grupowanie materiału statystycznego,
- podliczanie (zliczanie) materiału statystycznego,
- prezentację materiału statystycznego.

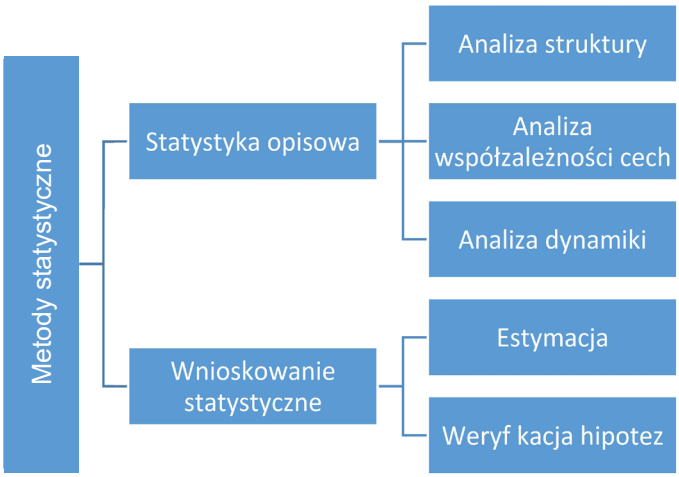
Do opracowania materiału statystycznego zaliczamy między innymi: porządkowanie, grupowanie i zliczanie. **Grupowanie** to podział zbiorowości statystycznej na jednorodne lub względnie jednorodne części. Podstawowe **metody prezentacji danych statystycznych** to przedstawienie w formie szeregów statystycznych, tablic i wykresów.

Rodzaje grupowania:

- typologiczne – np. według cech terytorialnych, rzeczowych, czasowych, mające na celu wyodrębnienie grup różnych jakościowo;
- wariacyjne – mające na celu uporządkowanie badanej zbiorowości i poznanie jej struktury, które polega na łączeniu w klasy jednostek statystycznych o odpowiednich wartościach cech statystycznych.

4. Analiza statystyczna.

Analiza statystyczna polega na ocenie własności zbiorowości statystycznych, przy wykorzystaniu właściwych metod statystycznych. Klasyfikacja metod statystycznych jest zaprezentowana na poniższym rysunku.



Rysunek 11. Klasyfikacja metod statystycznych.

Przykład

Przykładem celu badania może być badanie jakości obsługi klienta.
Tego typu badanie przeprowadza się na podstawie przygotowanej ankiety, gdzie klient odpowiada na pytania dotyczące obsługi.

Celem badania jest określenie poziomu zadowolenia klienta.

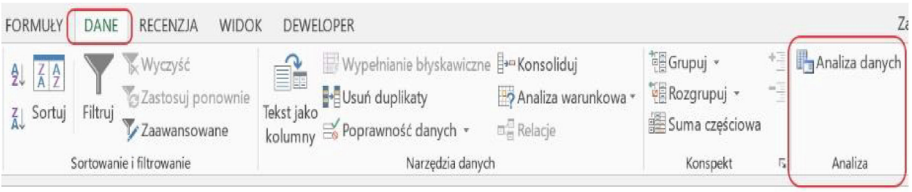
1.5. Zastosowanie programów komputerowych w statystyce

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 (R, Gretl, Statistica, SPSS, SAS** i innych). Pakiety te są stosunkowo proste w obsłudze, jednak wymagają pewnej wiedzy ze statystyki, aby można było poprawnie używać zawartych w nich procedur oraz interpretować uzyskane wyniki. Wymienione oprogramowanie jest omówione poniżej.

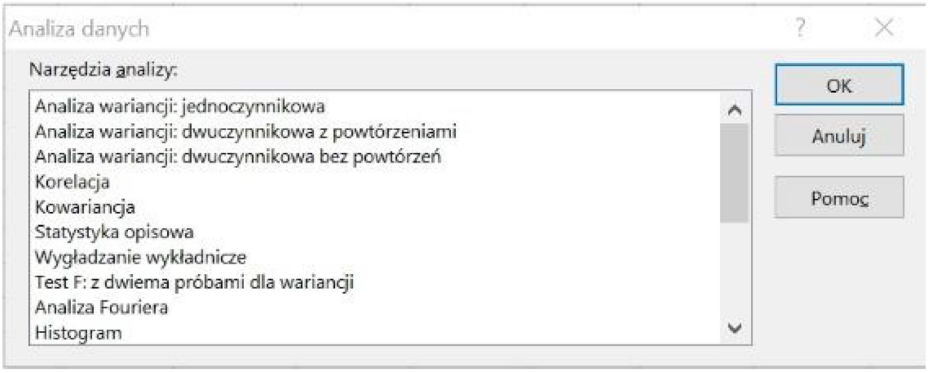
Funkcje statystyczne w istotnym stopniu wykorzystywane są w różnych programach komputerowych. Przykładem może być tutaj **arkusz kalkulacyjny Excel**. Wybór arkusza kalkulacyjnego jako narzędzia prezentacji metod statystycznych jest podyktowany wieloma względami. Znajomość arkuszy kalkulacyjnych jest coraz powszechniejsza ze względu na ich uniwersalność zastosowań. Każda nowsza wersja Excela ma coraz więcej i coraz bardziej rozbudowane funkcje statystyczne. Arkusze kalkulacyjne mogą zastąpić w wielu przypadkach specjalistyczne pakiety statystyczne. Arkusz kalkulacyjny umożliwia także prezentację metod statystycznych w sposób tradycyjny, za pomocą rozbudowanej tablicy roboczej, z ukazaniem etapów obliczeń. Tej właściwości nie mają profesjonalne pakiety statystyczne.

Dużym udogodnieniem dla użytkujących arkusz kalkulacyjny do analizy statystycznej jest system funkcji, między innymi matematycznych i statystycznych, oraz **opcja Analiza danych**. Opcja Analiza danych jest wybierana z zakładki Dane/Analiza – patrz rysunek poniżej.



Rysunek 12. Moduł Analiza danych pakietu Excel.

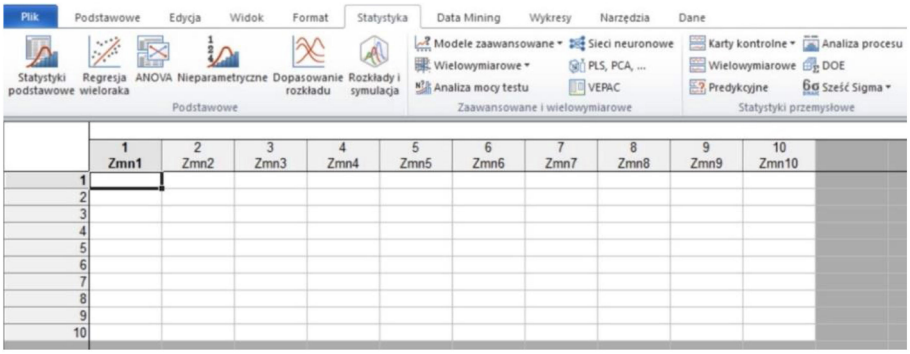
Funkcje statystyczne umożliwiają szybkie znajdowanie wielkości najczęściej charakteryzujących rozkłady empiryczne (średnia, wariancja, odchylenie standardowe, częstość itd.). Bardzo duże znaczenie mają także **wbudowane w arkusze kalkulacyjne najczęściej wykorzystywane rozkłady teoretyczne**: dwumianowy, Poissona, normalny, t- studenta. Opcja Analiza danych, może w wielu przypadkach zastąpić profesjonalny pakiet statystyczny. Dzięki niej można m.in. dokonać analizy wariancji, korelacji, regresji, testy statystyczne oraz statystykę opisową.



Rysunek 13. Dostępne funkcje Analizy danych.

Znaczenie komputerowych pakietów statystycznych jest istotne szczególnie w rozbudowanych firmach przy przeprowadzaniu różnych analiz statystycznych. Wówczas korzysta się z takich fachowych pakietów statystycznych, jak Statistica.

Statistica to wysokiej jakości system do analizy danych, graficznej prezentacji i zarządzania danymi, dla korporacji i pojedynczych analityków. Jest to szeroki wybór najbardziej zaawansowanych metod spośród wszystkich dostępnych na rynku narzędzi do modelowania i prognozowania, z automatyzacją wyboru modeli i interaktywnymi narzędziami wizualizacji. Dodatkowo korzysta się również z tzw. systemów korporacyjnych, czyli zintegrowanych systemów analitycznych.



Rysunek 14. Interfejs programu Statistica.

Kolejne oprogramowanie to SPSS. Jest to zintegrowana rodzina produktów przydatnych na wszystkich etapach procesu analitycznego – od planowania i zbierania danych po analizowanie, raportowanie i wdrożenie. Podobnie jak Statistica, SPSS jest oprogramowaniem płatnym.

R to nazwa języka programowania, platformy programistycznej, jak i całego projektu, w ramach którego rozwijane jest zarówno środowisko, jak i język. R bardzo często jest nazywany pakietem statystycznym, ponieważ dostępnych jest szereg pakietów i funkcji wykorzystywanych do zastosowań statystycznych. Okazuje się, że R znajduje znacznie więcej zastosowań: automatyczne generowanie raportów, wysyłanie maili, renderowanie trójwymiarowych animacji. R jest projektem opartym o licencję GNU GPL, wyposażonym w bardzo szczegółową dokumentację dostępną w Internecie. Język R jest językiem interpretowanym, korzystanie z niego sprowadza się do podania szeregu komend, które mają zostać kolejno wykonane. Komendy mogą być zebrane w formie skryptu, który zostaje następnie wykonywany krok po kroku**.**

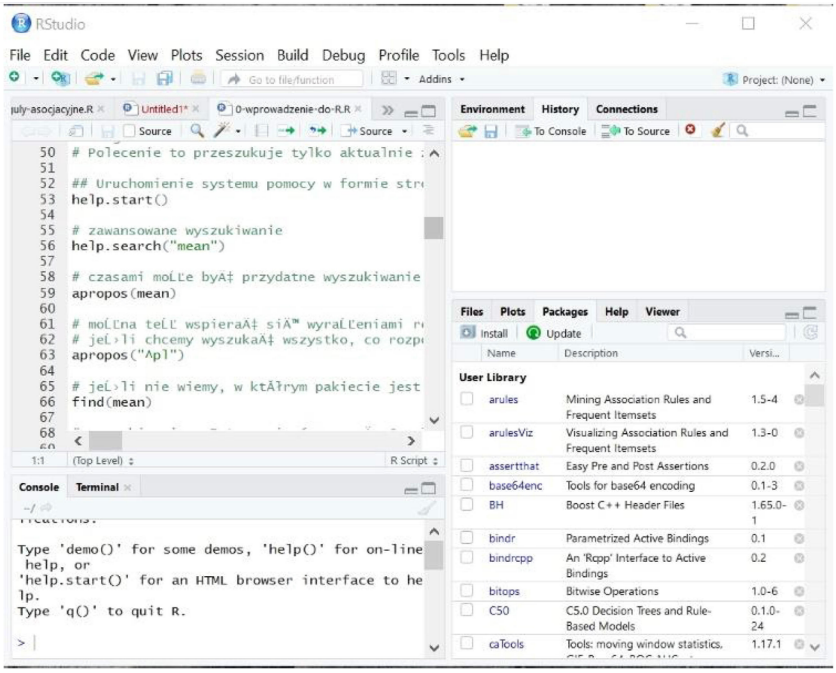
Główne elementy charakterystyczne tego języka programowania:

- Jest to bardzo elastyczne (darmowe) środowisko analityczne o bogatej funkcjonalności, stosowane w wielu badawczych i praktycznych pracach dotyczących analizy danych i odkrywania wiedzy.
- R jako język programowania dostarcza wbudowanych operacji ułatwiających przetwarzanie tabelarycznych zbiorów danych, mechanizmów graficznego opisu danych, bogatych bibliotek funkcji analitycznych, obejmujących szeroki zakres metod statystycznych i metod odkrywania wiedzy oraz – co niezwykle ważne – interaktywny interpreter poleceń i (dla niektórych platform) graficzny interfejs użytkownika.
- Wszystkie niezbędne informacje o języku R znajdziemy na stronach CRAN (The Comprehensive R Archive Network), gdzie dostępne są pakiety źródłowe, dokumentacja oraz obszerne zasoby bibliotek.

Dla większości systemów operacyjnych pakiet R jest dostępny zarówno w postaci źródłowej, jak i skompilowanej. Najlepiej skorzystać ze skompilowanego pliku instalacyjnego, który można

pobrać z jednego z serwerów, których lista znajduje się na stronie: http://cran.r-project.org/mirrors.html.

Do tworzenia i edytowania skryptów w R służą specjalne edytory, przykładem może być RStudio. Jego instalacja jest możliwa na systemie Windows, Linux oraz MacOS. Można go pobrać ze strony: www.rstudio.com. Edytor umożliwia wykonywanie pojedynczych komend wpisywanych prosto do konsoli albo wykonywanie całych skryptów przygotowanych uprzednio w oknie skryptowym.



Rysunek 15. Okno RStudio.

Całe okno RStudio podzielone jest na 4 panele główne, które można konfigurować, wybierając z menu: *Tools* i *Global Options*. W ten sposób wywołamy okno *Options*, widoczne na powyższym rysunku. W powyższej konfiguracji prawy górny panel służy do pisania kodu programu. Linijkę kod możemy przekazać do konsoli R (znajdującej się w lewym górnym panelu) za pomocą kombinacji klawiszy: CTRL+Enter. Jeśli chcemy, aby tych linii było więcej, wtedy należy je zaznaczyć i użyć tego skrótu. Z innymi dostępnymi skrótami klawiaturowymi można się zapoznać, wybierając z menu Help i Keyboard Shortcuts.

Ostatnie omawiane oprogramowanie to **program GRETL**, autorstwa Allina Cottrella z Uniwersytetu Wake Forest w Północnej Karolinie w Stanach Zjednoczonych. Jest rozwijanym od kilku lat pakietem ekonometrycznym. Należy do oprogramowań Powszechnej Licencji Publicznej (GNU), czyli o swobodnym dostępie dla wszystkich użytkowników. Program ten, niezwykle popularny i powszechnie stosowany na świecie, zawiera podstawowe procedury i metody ekonometryczne, przygotowany jest tak, aby można go było wykorzystać na zajęciach z ekonometrii.

1.6. Pytania sprawdzające

1. Wyjaśnij, czym zajmuje się statystyka.

……
……
……
……
……
……
……
……

2. Opisz obszary zastosowań statystyki.

……
……
……
……
……
……
……
……

3. Wyjaśnij podstawowe pojęcia – zbiorowość statystyczna, cecha statystyczna.

……
……
……
……
……
……
……
……
……

4. Podaj przykłady zbiorowości statystycznej.

……
……
……
……
……
……
……
……
……

5. Omów etapy badania statystycznego.

……
……
……
……
……
……
……
……
……
……

6. Omów klasyfikację metod statystycznych.

……
……
……
……
……
……
……
……
……
……

7. Zdefiniuj jednostkę oraz trzy przykładowe cechy wraz z rodzajem dla podanych zbiorowości:

- użytkownicy YouTuba,
- gospodarstwa domowe,
- przedsiębiorstwa,
- klienci sklepu,
- drzewa w Ogrodzie Saskim.

.....
.....
.....
.....
.....
.....
.....
.....
.....

8. Poniżej znajduje się fragment przykładowej ankiety. Jaki rodzaj pytań zawiera poniższa ankieta?

Prosimy o wypełnienie kilku dodatkowych rubryk, które posłużą wyłącznie do celów statystycznych:

1. Płeć: M. Mężczyzna K. Kobieta

2. Wiek: ..... lat

3. Wykształcenie: podstawowe, zasadnicze, średnie, pomaturalne, wyższe, inne

4. Zawód: .....

5. Wielkość gospodarstwa – nieruchomości: .....

6. Sołectwo (nazwa wsi): .....

7. Główne źródło utrzymania: .....

8. Jak długo mieszka Pani/Pan na terenie gminy: ..... lat

9. Miesięczny dochód na 1 osobę w rodzinie
   - poniżej 400 zł
   - 4400 - 699 zł
   - 7700 - 999 zł
   - 1 1000 – 1 999 zł
   - powyżej 2 000 zł

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

10. Uzupełnij poniższą tabelkę.

Skala	Własności skali	Przykłady skali
Nominalna		
Porządkowa		
Interwałowa		
Ilorazowa		

11. Na podstawie danych zawartych w tablicy poniżej określ typ skali do poszczególnych zmiennych.

Lp.	Imię i nazwisko	Wykształce-nia	Temperatura ciała	Wzrost	Waga	Lata pracy	Kolor włosów	Płeć
1.	Pracownik 1	średnie	36,3	190	67	4	Czarne	K
2.	Pracownik 2	wyższe	36,9	182	55	8	Rude	K
3.	Pracownik 3	średnie	36,5	173	87	2	Brązowe	M
4.	Pracownik 4	wyższe	36,3	179	90	14	Blond	M
5.	Pracownik 5	średnie	37,3	160	120	7	Czarne	M
6.	Pracownik 6	średnie	36,2	170	59	11	Blond	K
7.	Pracownik 7	wyższe	36,8	170	68	4	Czarne	K
8.	Pracownik 8	średnie	36,9	155	54	21	Rude	M
9.	Pracownik 9	wyższe	36,3	169	61	7	Brązowe	M
10.	Pracownik 10	średnie	36,6	170	85	14	Blond	K

.....

.....

.....
.....
.....
.....
.....
.....
.....

12. W urzędzie gminy, zatrudniającym 150 osób, dokonano badania pracowników pod względem wieku, płci, zajmowanego stanowiska, wykształcenia i wynagrodzenia. Określ zbiorowość statystyczną tego badania. Ustal, czy jest to populacja generalna czy próba. Określ jednostkę statystyczną oraz zdefiniuj cechy statystyczne.

.....
.....
.....
.....
.....
.....
.....
.....
.....

13. Przeprowadzając okresowe badania lekarskie w pewnej lubelskiej firmie, zmierzono między innymi wzrost i wagę oraz spytano o liczbę przebytych szczepień. Określ zbiorowość statystyczną tego badania. Ustal, czy jest to populacja generalna czy próba. Określ jednostkę statystyczną oraz zdefiniuj cechy statystyczne.

.....
.....
.....
.....
.....

1 2 3 4 5 6

2. Dane statystyczne

2.1. Rodzaje danych i ich źródła

Dane to surowe fakty, liczby lub inne szczegóły dotyczące pewnych zdarzeń. Na podstawie danych tworzone są informacje. Czyli dane poddane analizie (przetworzeniu) stają się informacją.

Następnie informacje, które mogą być pomocne w procesie zarządzania firmą, tworzą **informację zarządczą**, która daje podstawę do efektywnego podejmowania decyzji.

Rysunek 16. Informacja w procesie decyzyjnym.

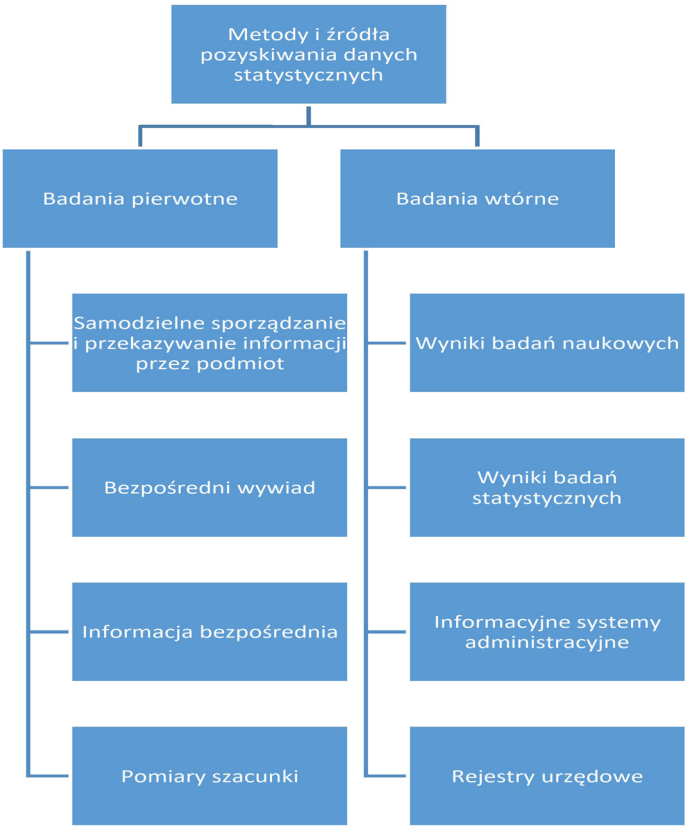
Źródło: opracowanie własne na podstawie J. Stoner, Ch. Wankel, *Kierowanie*, PWE, Warszawa 1992.

Rozróżnienie danych pierwotnych i wtórnych jest jednym z podstawowych podziałów w metodologii badań. **Dane pierwotne** to dane zbierane bezpośrednio od respondentów w trakcie wywiadów, ankiet. Mogą być one zbierane przez:

- instytucje i spółki państwa,
- agencje badawcze i badaczy,
- historyczne (z baz danych zapisanych na serwerach).

Dane wtórne są natomiast wynikiem dokonanych wcześniej analiz lub pochodzą z dokumentów źródłowych. Zasadniczo zaleca się łączyć w badaniu analizy danych pierwotnych i wtórnych, ponieważ konfrontując te dwa źródła – otrzymuje się pełniejszy obraz. Dodatkową zaletą danych wtórnych jest to, że ich pozyskanie jest zdecydowanie tańsze od danych pierwotnych. Są to też często dane wysokiej jakości.

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 na stronie internetowej urzędu.



Rysunek 17. Metody i źródła pozyskiwania danych statystycznych.

Tabela 2. Decyzje podejmowane na podstawie informacji.

Kryterium	Rodzaj	Przykład
Granice organizacyjne przedsiębiorstwa	wewnętrzne	raporty sprzedażowe, finansowe, bilans, budżet, raporty roczne, wewnętrzne gazetki, szkolenia, zebrania
	zewnętrzne	prasa ogólna i specjalistyczna, publikacje instytucji badawczych, rządowych i samorządowych
Cel gromadzenia informacji	wtórne	raporty powstające dla innych celów niż rozwiązanie danego problemu badawczego firmy
	pierwotne	raporty powstałe dla innych celów niż rozwiązanie danego problemu
Etap procesu decyzyjnego	analityczne	raporty dotyczące analizy sprzedaży według poszczególnych produktów, rynków, klientów, oceny wizerunku, pozycji
	prognostyczne	prognozy dotyczące potencjału danego rynku, możliwości wejścia lub wyjścia konkurentów
	kontrolne	raporty dotyczące skuteczności i efektywności promocji, rentowności kanałów dystrybucji
Kraj pochodzenia	krajowe	raporty dotyczące skuteczności i efektywności promocji, rentowności kanałów dystrybucji
	zagraniczne	opracowania statystyczne urzędów poszczególnych krajów oraz organizacji, czasopisma zagraniczne

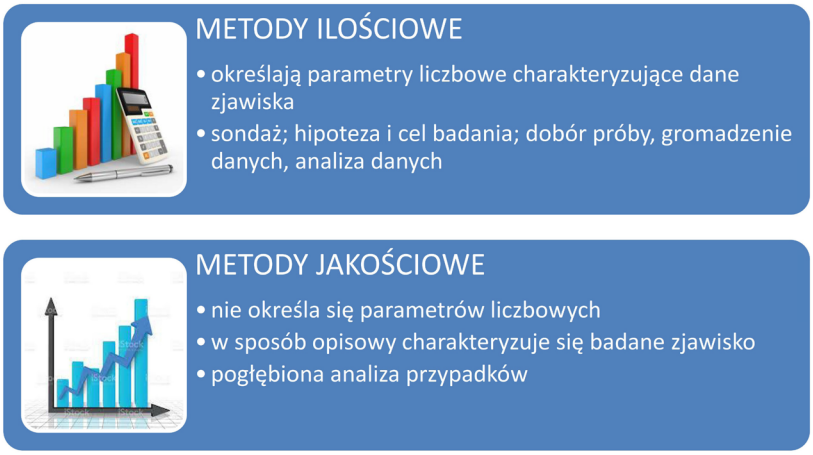
Źródło: E Frąckiewicz, *Informacyjne uwarunkowania podejmowania decyzji marketingowych*, „Przegląd Organizacji” 2006, nr 3, s. 41.

Dane pochodzące z różnych źródeł mogą być też różnej jakości. Należy oceniać je pod kątem ich[1]:

- prawdziwości,
- jednoznaczności,
- identyfikowalności zjawiska przez zmienną/zmienne,
- kompletności,
- aktualności w przyszłości, kosztu zbierania i opracowywania,
- porównywalności.

2.2. Metody badań w naukach ekonomicznych

Podział na metody ilościowe i jakościowe to podstawowy podział metod w naukach badawczych. Najkrócej ujmując, metody ilościowe zbierają wartości liczbowe, które można zliczać, a jakościowe zbierają opinie, pozwalające na pogłębioną interpretację. Metody ilościowe dają możliwość obróbki narzędziami statystycznymi, ale wyniki są bardziej ogólne. Metody jakościowe dają możliwość pogłębionej obserwacji szczególnych przypadków, ale nie można generalizować danych na populację. Metod ilościowych i jakościowych nie można stosować zamiennie. To metody, które służą innym celom. Zazwyczaj optymalnym rozwiązaniem jest łączenie obu metod w ramach jednego badania.

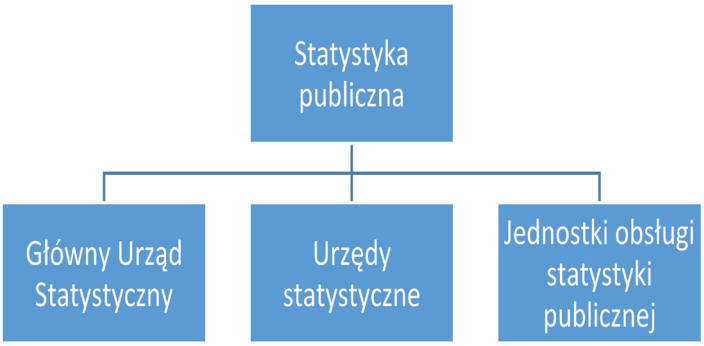


Rysunek 18. Metody badań.

[1] P. Dittman, *Metody prognozowania sprzedaży w przedsiębiorstwie*, Wyd. AE im. O Langego, Wrocław 2000, s. 38.

2.3. Statystyka publiczna

Statystyka publiczna zapewnia rzetelne, obiektywne i systematyczne informowanie społeczeństwa, organów państwa i administracji publicznej oraz podmiotów gospodarki narodowej o sytuacji społeczno-ekonomicznej kraju. Podstawowe zadanie, jakim jest dostarczanie rzetelnej i obiektywnej informacji statystycznej, statystyka spełnia, prowadząc badania statystyczne.



Rysunek 19. Organy statystyki publicznej.

Przyjmuje się, że statystyka publiczna powinna pełnić następujące funkcje[2]:

- Wspomagać rozwój społeczeństwa informacyjnego i gospodarki opartej na wiedzy, dostarczając informacji, metainformacji i wiedzy statystycznej o zjawiskach i procesach objętych badaniami statystycznymi oraz prowadząc edukację statystyczną społeczeństwa.
- Przyczyniać się do podniesienia sprawności i efektywności funkcjonowania państwa poprzez zapewnienie zaplecza informacyjnego, analitycznego i metodycznego władzom państwowym i samorządowym, wspomagając te organy w procesie podejmowania decyzji.
- Inicjować obserwację statystyczną nowych zjawisk i procesów w społeczeństwie i gospodarce oraz opracowywać metody ich identyfikacji i pomiaru, wykorzystując najlepsze metody i technologie statystyczne.

[2] A. Brzezińska, *Zasoby informacyjne polskiej statystyki publicznej*, Urząd Statystyczny w Szczecinie, Szczecin 2012.

- Opracowywać standardy informacyjne dla systemów informacji społecznej i gospodarczej kraju, wspomagać ich wdrażanie i stosowanie, oddziałując tym samym na jakość i integralność informacji w badaniach statystycznych prowadzonych poza systemem statystyki publicznej oraz w administracyjnych systemach informacyjnych.
- Aktywnie i systematycznie badać potrzeby informacyjne społeczeństwa, gospodarki, organów państwa wszystkich szczebli i innych jednostek sektora publicznego w zakresie zapotrzebowania na informacje statystyczne. Na tej bazie tworzyć ofertę informacyjną i kreować potrzeby informacyjne.

Całość systemu informacji statystycznej można podzielić na dwa bloki tematyczne: statystyka społeczna oraz statystyka gospodarcza – patrz Rysunek 20.

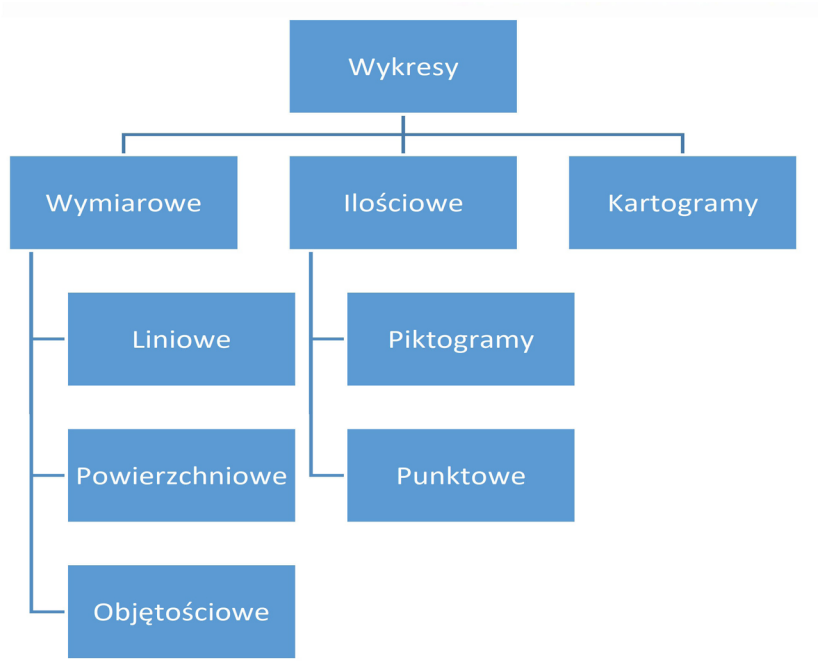
Rysunek 20. System informacji statystycznej.

2.4. Prezentacja graficzna danych

Formami prezentacji materiału statystycznego są: tablice, wykresy, charakterystyki opisowe oraz wzory analityczne. W tym rozdziale omówimy graficzną formę prezentacji danych statystycznych.

Niezmiernie ważną rolę w analizie statystycznej odgrywa graficzna prezentacja danych statystycznych. Dzięki wykresom można przeprowadzić nie tylko wstępną analizę zjawiska czy zbiorowości, ale także dokonać doboru odpowiednich parametrów opisu rozkładu cechy. Graficzna prezentacja danych jest mniej dokładna niż prezentacja tabelaryczna (patrz rozdział Klasyfikacja danych), jednak pozwala ona na szybką wizualizację dotyczącą kształtowania się poziomu danego zjawiska. Istnieje wiele różnych typów wykresów. W literaturze przedmiotu można spotkać **podział na wykresy wymiarowe, wykresy ilościowe i kartogramy**.

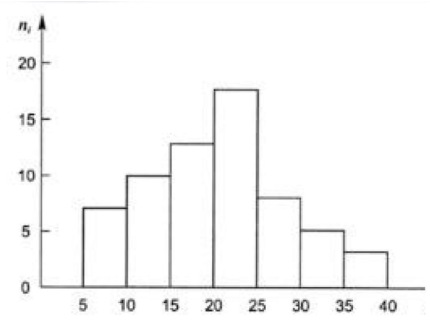
Podział ten został przedstawiony schematycznie na rysunku poniżej.



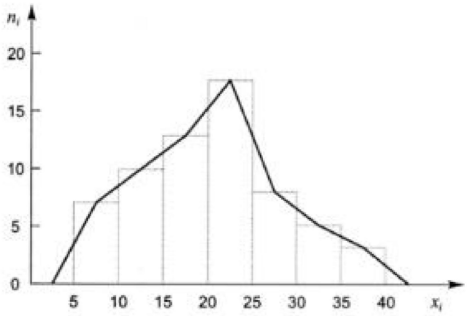
Rysunek 21. Rodzaje wykresów.

Do prezentacji graficznej opracowanego materiału statystycznego z zastosowaniem wykresów wymiarowych stosujemy najczęściej:

- wykres słupkowy;
- wykres kołowy;
- wykres obrazkowy (piktogram);
- diagram szeregów czasowych – umieszcza się w układzie współrzędnych. Na osi poziomej (odciętych) umieszcza się skalę czasu, a na osi pionowej (rzędnych) częstości odpowiadające danemu okresowi;
- histogram – umieszcza się również w układzie współrzędnych, oś rzędnych reprezentuje częstości, a oś odciętych reprezentuje różne wartości zmiennej. Zbudowanie histogramu wymaga wykreślenia słupków dla każdego przedziału prezentowanej wartości cechy, których wysokość będzie proporcjonalna do częstości (Rysunek 22);
- wielobok liczebności – powstaje z histogramu przez łączenie liniami prostymi punktów o współrzędnych: środek przedziału i liczebność w danej klasie;
- wykresy częstości (diagramy) – polegają na umieszczeniu w układzie współrzędnych częstości na osi rzędnych oraz wartości prezentowanej zmiennej na osi odciętych;
- diagram częstości i procentów skumulowanych – wykreśla się go w układzie współrzędnych na podstawie tablicy częstości skumulowanych, wyznaczając punkt odpowiadający każdej wartości. Następnie łączy się ze sobą punkty. Podobnie można zbudować diagram częstości procentowych. Wtedy na osi pionowej częstości wyrażone będą w procentach.

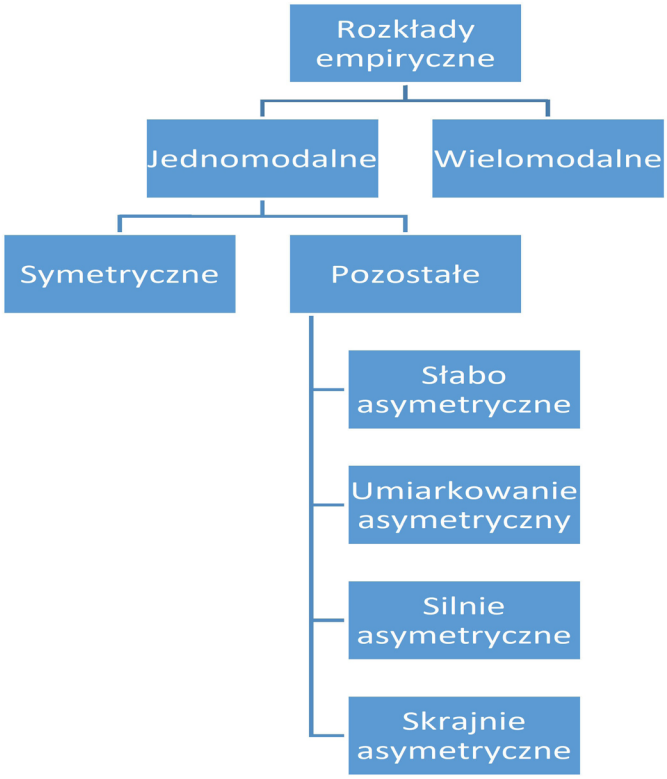


Rysunek 22. Przykładowy histogram.

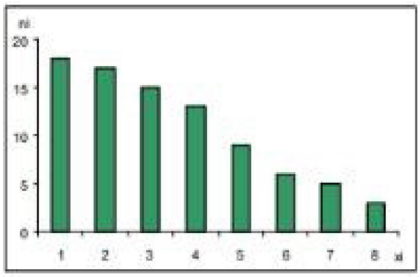


Rysunek 23. Przykładowy wielobok liczebności.

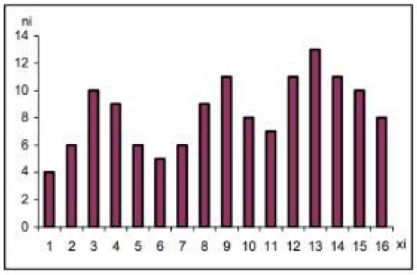
Z rysowaniem wieloboków liczebności związane jest zagadnienie rozróżniania głównych kształtów krzywych. **Rozkład empiryczny** to uzyskany na podstawie badania statystycznego opis wartości przyjmowanych przez cechę statystyczną przy pomocy częstości ich występowania. Rozkład taki można przedstawić w postaci **tabeli liczności (patrz rozdział Klasyfikacja danych)** lub w postaci wykresu (histogramu).



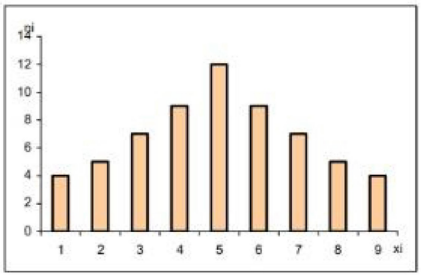
Rysunek 24. Rodzaje rozkładów empirycznych.



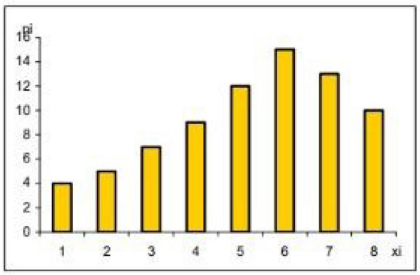
Rysunek 25. Rozkład skrajnie symetryczny.



Rysunek 26. Rozkład wielomodalny.



Rysunek 27. Rozkład symetryczny.



Rysunek 28. Rozkład umiarkowanie symetryczny.

Rozkłady mają pewne charakterystyki (patrz rozdział Analiza struktury), **które określają:**

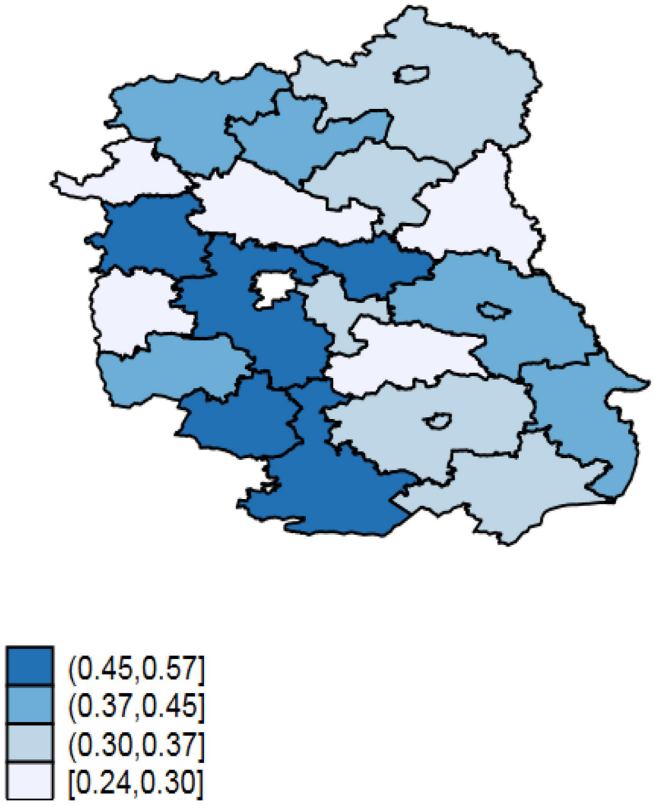
- położenie wykresu lub określonych jego części wzdłuż osi x (miary położenia),
- rozmieszczenie wartości zmiennej i jej przeciętne wartości (miary centralne),
- zróżnicowanie wartości zmiennej, skupienie, spłaszczenie (miary rozproszenia i koncentracji),
- symetrię rozkładu (miary symetrii).

Wykresy ilościowe powstają poprzez wielokrotne użycie odpowiedniego symbolu reprezentującego przyjętą jednostkę, zatem tworząc go, należy najpierw dobrać jej wielkość. Jako symbol najczęściej wybiera się figurę geometryczną lub obrazek przedstawiający poglądowo dane zjawisko (np. sylwetka samochodu jako jednostka produkcji samochodów w kolejnych latach w pewnej firmie). Jest to bardzo wygodne zwłaszcza przy prezentacji cech jakościowych.

Przykładowy wykres ilościowy został przedstawiony na Rysunku 29.

Rysunek 29. Liczba ludności w milionach.

Kartogramy to rodzaj wykresu wykonywanego na mapach, obrazującego terytorialne występowanie danego zjawiska. Powstają one poprzez umieszczenie figur, symboli graficznych w odpowiedniej liczbie lub wielkości na poszczególnych jednostkach terytorialnych lub też poprzez nadanie barw lub zakreskowanie obrazujące natężenie zjawiska. Przykładowy kartogram został przedstawiony na Rysunku 30.



Rysunek 30. Kartogram.

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:

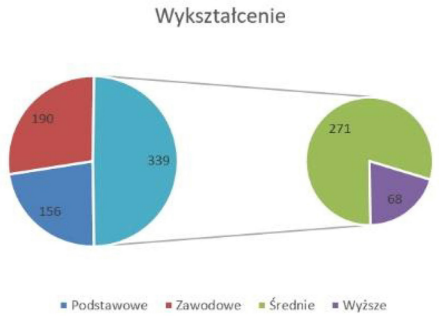
- tytułu,
- pola wykresu,
- skali,
- legendy,
- źródła danych statystycznych i ewentualnych uwag wyjaśniających.

2.5. Zastosowanie pakietu EXCEL

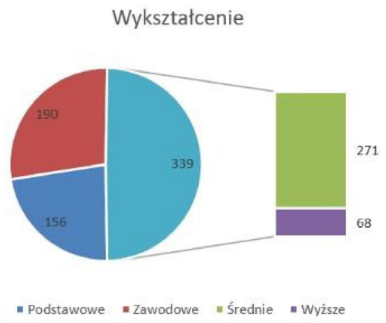
Dostępne są następujące **typy wykresów kołowych, pierścieniowych oraz kolumnowych w Excelu 2016 (**patrz Rysunki 31 – 38):

- kołowy (standardowy),
- kołowy – kołowy,
- kołowy – słupkowy,
- kołowy 3D (3-W),
- pierścieniowy,
- kolumnowy,
- kolumnowy-grupowany,
- skumulowany-grupowany,
- kolumnowy-grupowany 3-D.

Rysunek 31. Wykres kołowy 2-W.



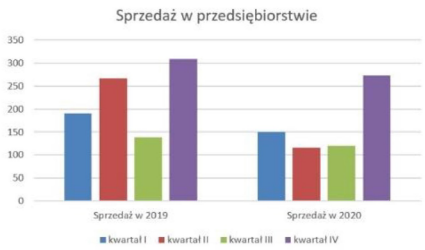
Rysunek 32. Wykres kołowy kołowego.



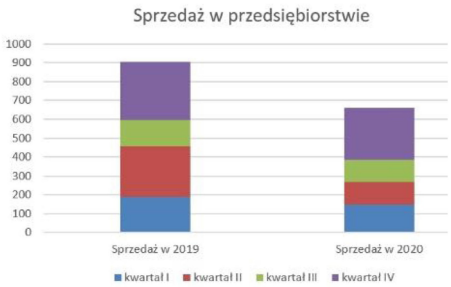
Rysunek 33. Wykres słupkowy kołowego.

Rysunek 34. Wykres pierścieniowy.

Rysunek 35. Wykres kolumnowy grupowany.



Rysunek 36. Wykres kolumnowy grupowany.



Rysunek 37. Skumulowany kolumnowy.

Rysunek 38. Kolumnowy grupowany 3-W.

Aby wstawić tego typu wykres, należy:

1. Zaznaczyć tabelę z danymi.
2. Wybrać w kategorii WSTAWIANIE z sekcji WYKRESY ikonkę wykresu kołowego.
3. Wybrać odpowiedni podtyp wykresu (kołowy, pierścieniowy, 3-W).

2.6. Pytania sprawdzające

1. Co to są dane pierwotne i wtórne?

.....
.....
.....
.....
.....
.....
.....
.....

2. Podaj przykład danych pierwotnych.

.....
.....
.....
.....
.....
.....
.....
.....

3. Omów metody i źródła pozyskiwania danych statystycznych.

.....
.....
.....
.....
.....
.....
.....
.....

4. Omów metody ilościowe oraz jakościowe badań.

……
……
……
……
……
……
……
……
……

5. Jakie są zadania statystyki publicznej?

……
……
……
……
……
……
……
……
……

6. Wymień organy statystyki publicznej.

……
……
……
……
……
……
……
……
……
……

7. Omów system informacji statystycznej.

……
……
……
……
……
……
……
……
……

8. Co to jest rozkład empiryczny zmiennej?

……
……
……
……
……
……
……
……
……

9. Jaki jest sens wyróżniania kształtów rozkładów empirycznych?

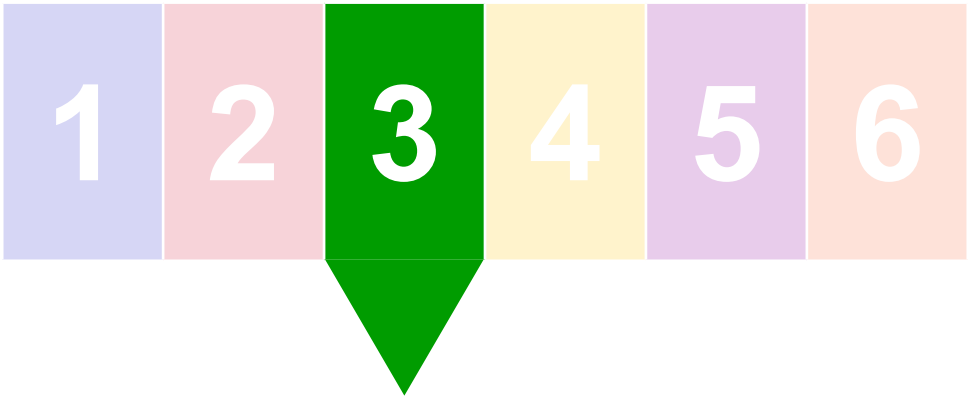
……
……
……
……
……
……
……
……
……
……

10. Jakie są najczęściej spotykane w praktyce rodzaje wykresów empirycznych?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

11. Wymień podstawowe rodzaje wykresów.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....



3. Klasyfikacja danych

Dane charakteryzujące rozwój zjawisk gospodarczych lub społecznych są odnotowywane i gromadzone wraz z upływem czasu, tworząc ciągi liczbowe określane mianem szeregów czasowych.

Szereg czasowy to uporządkowany zbiór obserwacji statystycznych (zbiór stanów zmiennych), charakteryzujących zmiany poziomu określonego zjawiska w czasie.

Szereg czasowy momentów przedstawia dane charakteryzujące zjawisko (stan obiektu) w określonych **momentach** czasu, którymi są najmniejsze (dalej niepodzielne) jednostki czasu, jak np. kurs euro na zakończenie sesji dla kolejnych dni miesiąca.

Szereg czasowy okresów przedstawia dane charakteryzujące zjawisko (stan obiektu) w określonych **okresach** czasu. Okres czasu to pewien odcinek, np. dzień, tydzień, miesiąc czy rok. Taki szereg czasowy okresów tworzy na przykład średni kurs euro dla kolejnych dni miesiąca. W tym wypadku każda wartość szeregu jest średnią z poziomów kursów odnotowanych w danym dniu w trakcie trwającej zwykle kilka godzin sesji giełdowej, a więc wartość ta jest uśrednionym poziomem dla całego (kilkugodzinnego) odcinka czasu w danym dniu notowań.

Przykład

W tabeli poniżej przedstawiono szereg czasowy momentów dotyczący liczby rowerów dziecięcych będących w magazynie sklepu rowerowego, w momencie pierwszego dnia każdego miesiąca.

Data	01.10.2020	01.11.2020	01.12.2020	01.01.2021	01.02.2021
Liczba rowerów dziecięcych	20	50	10	30	20

Przykład

W tabeli poniżej przedstawiono szereg czasowy okresów dotyczący liczby rowerów dziecięcych będących w magazynie sklepu rowerowego. W szeregu dynamicznym okresami są poszczególne kwartały 2020 roku, stanowią łącznie pewien przedział czasowy.

Kwartał	I	II	III	IV
Liczba rowerów dziecięcych	221	153	219	230

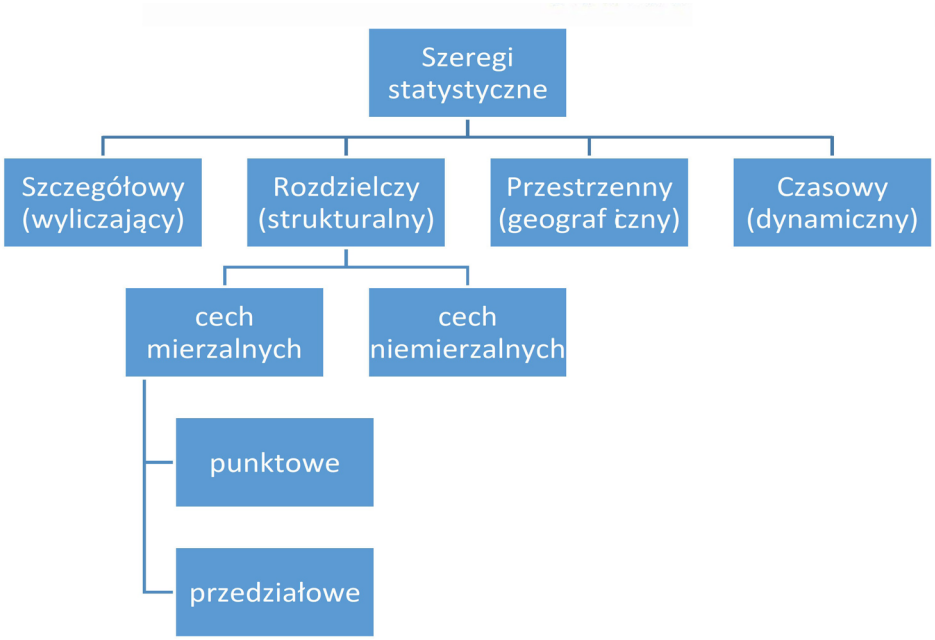
Jednowymiarowy szereg czasowy jest to zbiór stanów jednej zmiennej Y, uporządkowanych według wartości zmiennej czasowej t (t = 1, 2, n), charakteryzujących zjawisko (stan obiektu), np. kurs jednej waluty dla kolejnych dni miesiąca.

Wielowymiarowy szereg czasowy jest to zbiór stanów m zmiennych $Y1$, $Y2$,…,Ym, uporządkowanych według wartości zmiennej czasowej t (t = 1, 2,…, n), charakteryzujących zjawisko (stan obiektu), np. kurs kilku walut dla kolejnych dni miesiąca.

Rysunek 39. Rodzaje szeregów czasowych.

Źródło: *Prognozowanie w zarządzaniu przedsiębiorstwem*, t. 2, *Prognozowanie na podstawie szeregów czasowych*, red. J. Nazarko, Wyd. Politechniki Białostockiej, Białystok 2004, s. 20.

Szereg statystyczny to odpowiednio usystematyzowany i uporządkowany surowy materiał statystyczny. Różne rodzaje szeregów statystycznych przedstawia Rysunek 40. W dalszych rozdziałach omówimy kolejno każdy z wymienionych typów szeregu statystycznego.



Rysunek 40. Rodzaje szeregów statystycznych.

3.1. Szereg szczegółowy

Szereg rozdzielczy szczegółowy to ciąg liczbowych wielkości statystycznych uporządkowanych według badanej cechy (rosnąco lub malejąco). Tworzy się poprzez grupowanie powtarzających się wartości badanej cechy w próbie. Buduje się go przeważnie dla cechy skokowej.

Przykład

Zbadano liczbę dzieci dla 6 rodzin. Otrzymano wyniki: 2, 4, 3, 1, 2, 3
Szereg szczegółowy to: 1, 2, 2, 3, 3, 4.

Przykład

Dziesięciu studentów, zapytanych o liczbę rodzeństwa, dało następujące odpowiedzi:

0, 2, 1, 4, 2, 3, 1, 2, 2, 4.

Przedstawione dane są nieuporządkowane. Zapisano je w postaci szeregu szczegółowego.

Przykład

Dwudziestu uczniów zapytano, ile godzin spędzają dziennie przed komputerem. Udzielili oni następujących odpowiedzi:
2,5; 3; 4; 1; 4,5; 0,5; 5; 6; 3,5; 12; 2; 3,5; 2; 2,5; 1; 3,5; 2; 3,5; 2; 6,5.
W badaniu mamy do czynienia z cechą statystyczną: liczba godzin spędzonych dziennie przed komputerem. Jest to cecha mierzalna ciągła. Zbiorowość (populację) statystyczną tworzą poszczególne jednostki badania, czyli poszczególni uczniowie. Informacje statystyczne o wartości cechy dla każdej z jednostek tworzą indywidualny szereg wartości cech (surowy materiał statystyczny). Po uporządkowaniu otrzymujemy szereg szczegółowy:
0,5; 1; 1; 2; 2; 2; 2; 2,5; 2;5; 3; 3,5; 3,5; 3,5; 3,5; 3,5; 2,5; 4; 4,5; 5; 6; 12.

Szeregi rozdzielcze dla cech mierzalnych dzielą się na:

- szeregi punktowe – stosowane, gdy liczba wariantów cechy jest niewielka,
- szeregi z przedziałami klasowymi – stosowane, gdy liczba wariantów badanej cechy jest duża.

Klasy są przedziałami, najczęściej jednakowej długości, które tworzy się, przyjmując upraszczające założenie, że wszystkie wartości znajdujące się w danej klasie są identyczne z tzw. środkiem klasy.

Przykład

Tabela 3. Szereg rozdzielczy punktowy.

Liczba miesięcznych wizyt w kinie	Liczba studentów
0	15
1	19
2	13
3	3
4	5
5	1

Przykład

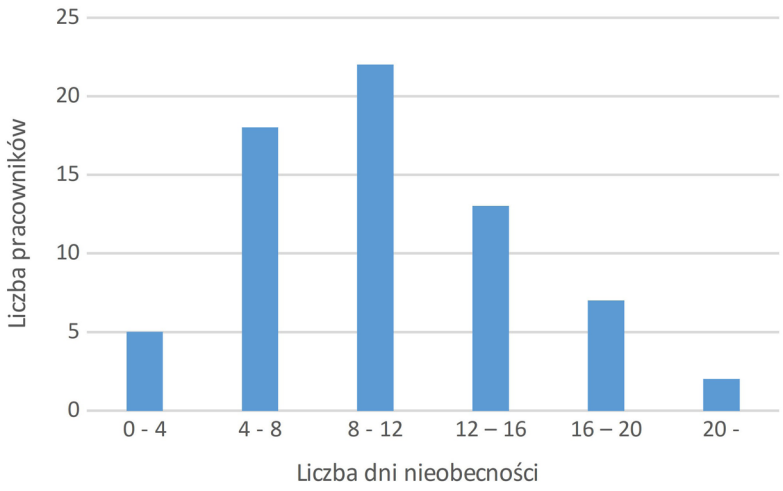
Tabela 4. Szereg rozdzielczy.

Poziom wykształcenia	podstawowe	zawodowe	średnie	wyższe
Liczba pracowników	5	14	23	14

Przykład

Tabela 5. Szereg rozdzielczy przedziałowy.

Liczba dni nieobecności	Liczba pracowników
0 – 4	5
4 – 8	18
8 – 12	22
12 – 16	13
16 – 20	7
20 –	2

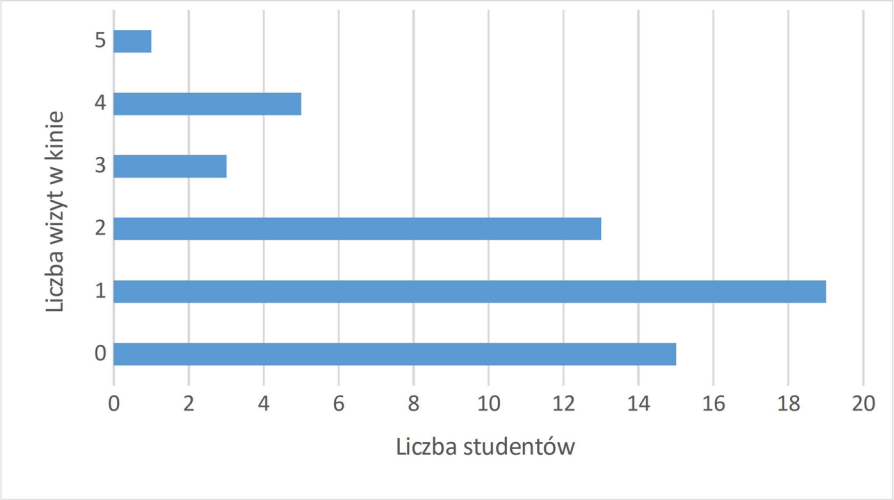


Rysunek 41. Graficzna reprezentacja szeregu.

Szereg rozdzielczy składa się z dwóch kolumn:

- w pierwszej kolumnie znajdują się warianty badanej cechy,
- w drugiej kolumnie liczba jednostek zbiorowości statystycznej, która posiada dany wariant cechy.

Dane zawarte w szeregu dynamicznym mogą być zaprezentowane w postaci wykresu, który w bardziej obrazowy sposób przedstawia dynamikę badanego zjawiska.



3.2. Szereg rozdzielczy przedziałowy

Tworzenie szeregu rozdzielczego przedziałowego zaleca się budować w sytuacji, gdy badana cecha jest ciągła albo jest cechą skokową o wielu wariantach. Wówczas na wstępie warianty cechy grupuje się w przedziały klasowe, a zadaniem badacza jest zakwalifikowanie każdej jednostki statystycznej do odpowiedniego przedziału. Ostatecznie otrzymujemy liczebności cząstkowe poszczególnych przedziałów klasowych, wskazujące na to, ile bądź jaki odsetek jednostek badania przyjmuje wartości cechy z danego przedziału.

Niech $x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n$ będą wartościami cechy n-elementowej próby.

Przy większej liczebności próby (n > 30) w celu ułatwienia analizy próbę grupuje się w klasy, tj. przedziały najczęściej jednakowej długości, przyjmując uproszczenie, że wszystkie wartości należące do klasy mają wartość równą wartości środka przedziału. Następnie oblicza się liczebności przedziałów, tzn. liczby – ilość elementów próby, które mają wartości cechy z określonego przedziału. Ustalenie tych elementów tworzy szereg rozdzielczy danych próby.

Próbę nieprzedstawioną za pomocą szeregu rozdzielczego nazywamy **próbą nieuporządkowaną** (danymi nieuporządkowanymi).

Szereg rozdzielczy przedziałowy tworzymy, gdy mamy dużą liczbę (n > 30) nieuporządkowanych danych i chcemy wykonywać dalsze obliczenia. Poniżej przedstawiamy procedurę tworzenia szeregu przedziałowego.

UWAGA

Tworzenie szeregu przedziałowego

Krok 1
Oblicza się tzw. rozstęp danych $R = x_{max} - x_{min}$, gdzie x_{max}, x_{min} jest odpowiednio największą i najmniejszą wartością cechy w próbie.

Krok 2
Ustalamy liczbę klas. Istnieje kilka reguł ustalenia orientacyjnego liczby klas k i długości przedziału. Liczba klas nie może być za mała (traci się szczegóły danych) i nie za duża (traci się przejrzystość danych).

Tabela 7. Liczba klas.

Liczba pomiarów n	Liczba klas k
30 – 60	6 – 8
60 – 100	7 – 10
100 – 200	9 – 12
200 – 500	11 – 17
500 – 1500	16 – 25

Krok 3

Wyznaczamy długość przedziałów. Długość przedziału przyjmuje się $h \approx \frac{R}{k}$ z nadmiarem, tak aby $k \geq R$

Punkty stanowiące granice poszczególnych klas ustala się z dokładnością ½*α* gdzie *α* jest dokładnością pomiaru wartości cechy w próbie. Po ustaleniu tych elementów tworzy się szereg rozdzielczy próby.

Krok 4

Wyznaczamy pierwszy przedział, kolejne przedziały i liczebności

Przedziały klasowe cechy $a_i – b_i$	Liczebność klasy n_i	Środki klasy $x_i = (a_i + b_i) / 2$	Częstość cechy w klasie $g_i = n_i / n$	Łamana rozkładu empirycznego $w_i - g_i / h$
$a_1 – b_1$	n_1	x_1	g_1	w_1
$a_2 – b_2$	n_2	x_2	g_2	w_2
...	...	...	...	...
...	...	...	...	...
$a_k – b_k$	n_k	x_k	g_k	w_k

$n = n_1 + n_2 + ... + n_k$ liczebność próby

$h = b_i – a_i$ długość przedziału klasowego

k liczba przedziałów klasowych

a_i, b_i końce i-tego przedziału klasowego

Analiza szeregów czasowych sprowadza się do następujących zagadnień:

- analiza opisowa szeregu czasowego (patrz rozdział Analiza struktury),
- porównanie poziomów zjawiska w czasie (patrz rozdział Analiza dynamiki),
- dekompozycja szeregu czasowego.

WAŻNE

Etapy tworzenia szeregu rozdzielczego przedziałowego:

1. wyznaczamy x_{max} oraz x_{min}
2. wyznaczamy rozstęp z próby $R = x_{max} - x_{min}$
3. wyznaczamy ilość przedziałów klasowych K
4. wyznaczamy długość przedziału klasowego h, $h \approx \frac{R}{K}$
5. wyznaczamy lewy koniec pierwszego przedziału klasowego

   $\alpha = x_{min} - \frac{a}{2}$, gdzie α jest dokładnością pomiaru

Przy wyznaczaniu liczby i szerokości przedziałów należy brać pod uwagę następujące kryteria:

1. Liczba przedziałów nie może być ani zbyt mała, ani zbyt duża. Opracowanie danych zawsze wiąże się z pewną stratą informacji (uogólnieniem), która będzie poważna, jeżeli utworzymy małą liczbę przedziałów o dużej rozpiętości. Natomiast duża liczba wąskich przedziałów prowadzi do tego, że mamy ciągle do czynienia ze zbyt dużą ilością informacji, na podstawie których trudno wyobrazić sobie rozkład wartości. Rzadko stosuje się mniej niż 6 i więcej niż 15 klas.
2. Liczba obserwacji w poszczególnych przedziałach nie powinna być zbyt mała. W większości klas należy dysponować odpowiednio dużą liczbą obserwacji, aby zbudować szereg przedziałowy.

3. Ze względów praktycznych dobrze jest granice przedziałów określać liczbami zaokrąglonymi, a takie mogą ze wzorów nie wynikać. Nie będzie zatem niewłaściwe pewne przesunięcie granic w stosunku do tych, które wynikają z zaproponowanych wyżej wzorów.
4. Zazwyczaj wymagana jest równa rozpiętość wszystkich przedziałów.

3.3. Pytania sprawdzające

1. Wyjaśnij potrzebę grupowania danych statystycznych.

……
……
……
……
……
……
……
……
……
……
……
……
……
……

2. Omów zasady konstrukcji przedziałowych szeregów rozdzielczych.

……
……
……
……
……
……
……
……
……
……
……
……
……
……

3. Co wpływa na decyzję o liczbie klas szeregu rozdzielczego?

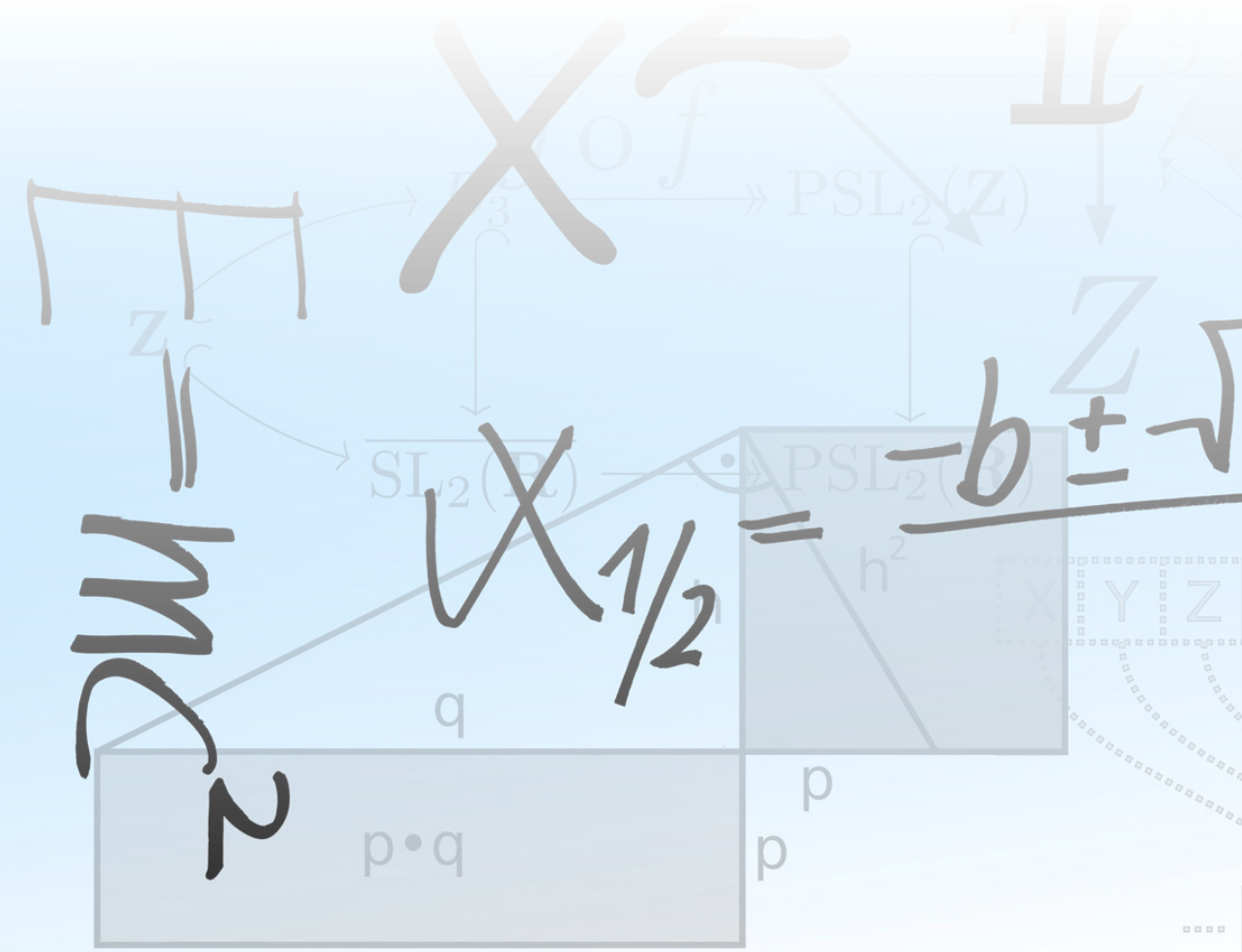
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

4. Jakie praktyczne kryteria należy stosować przy wyznaczaniu liczby i szerokości przedziałów dla szeregu rozdzielczego?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

1 2 3 **4** 5 6

4. Analiza struktury



Analiza statystyczna jest ostatnim etapem badania statystycznego (patrz rozdział Etapy badania statystycznego). Celem analizy statystycznej jest wykrycie prawidłowości i zależności zachodzących w badanej zbiorowości statystycznej.

Pod pojęciem **analizy struktury** rozumiemy badanie budowy (składu) określonej zbiorowości lub próby, tj. ustalenie, z jakich składa się elementów oraz jaką część stanowią owe elementy w całym zbiorze. Analiza struktury polega na określeniu i zinterpretowaniu prawidłowości występujących w budowie zbiorowości z punktu widzenia określonej cechy statystycznej. Podstawowym miernikiem statystycznym wykorzystywanym w analizie struktury jest wskaźnik struktury.

Wskaźnik struktury jest to stosunek części zbiorowości do całej zbiorowości (nazywany udziałem względnym – procentowym) lub części do części nazywany stosunkiem względnym (procentowym). Wskaźniki struktury, jako proste i zrozumiałe mierniki, znajdują szerokie zastosowanie w praktyce badań statystycznych. Istotną ich zaletą jest fakt, że mogą być stosowane dla dowolnych cech (mierzalnych i niemierzalnych), gdyż wykorzystują liczebności cząstkowe i łączną liczebność zbiorowości, bez udziału samych wariantów cechy.

Wskaźniki struktury możemy ogólnie podzielić na dwa typy:

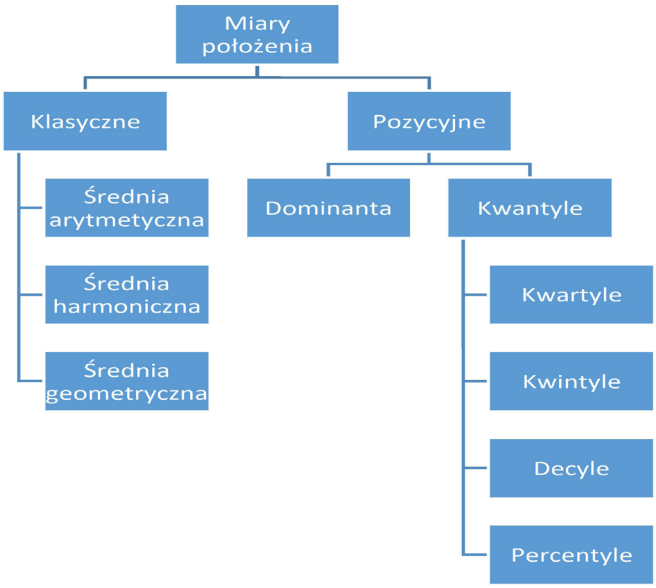
1. **parametryczne (klasyczne)** – obliczane na podstawie wszystkich obserwacji (wszelkie odmiany średniej, odchylenie standardowe, skośność, kurtoza);
2. **nieparametryczne** – wyznaczane na podstawie ich miejsca w szeregu lub częstotliwości występowania (moda, mediana, kwartyle, kwantyle, centyle).

Statystyki parametryczne wykorzystujemy najczęściej do opisu rozkładów o tendencji centralnej, czyli takich, w których punkt skupienia znajduje się w środku rozkładu (symetryczne) lub w jego pobliżu (umiarkowanie asymetryczne). Statystyki nieparametryczne można stosować do opisu każdego typu rozkładu, lecz są szczególnie przydatne w analizie szeregów silnie asymetrycznych i takich, w których nie posiadamy pełnych informacji o rozkładzie (są otwarte przedziały klasowe).

4.1. Ocena położenia rozkładu

Jeżeli cecha, którą analizujemy w zbiorowości, jest cechą mierzalną, to zbiorowość możemy scharakteryzować w sposób syntetyczny za pomocą miar wyrażających jej przeciętny poziom. Miary przeciętne charakteryzują średni lub typowy poziom wartości cechy. Są to więc takie miary, wokół których skupiają się wszystkie pozostałe wartości analizowanej cechy. Miary przeciętne dzielą się na miary klasyczne i pozycyjne. Pierwsze wyznaczane w oparciu o wszystkie wartości cechy, drugie wskazują określoną pozycję jednostek (np. środkową lub dominującą).

Miary położenia wskazują, wokół jakich wartości skupia się rozkład analizowanych zmiennych. Dzieli się je na klasyczne i pozycyjne. **Miary klasyczne** to średnie: arytmetyczna, harmoniczna i geometryczna. Do **miar pozycyjnych** należy dominanta (modalna, wartość najczęstsza) oraz kwantyle. Wśród kwantyli najczęściej stosowane są: kwartyle (dzielące zbiorowość na cztery części pod względem liczebności), kwintyle (dzielące zbiorowość na pięć części), decyle (dzielące zbiorowość na dziesięć części) oraz percentyle (dzielące zbiorowość na sto części).



Rysunek 43. Miary tendencji centralnej.

4.1.1. Miary położenia klasyczne

Jeżeli cecha, którą analizujemy w zbiorowości, jest cechą mierzalną, to zbiorowość możemy scharakteryzować w sposób syntetyczny za pomocą miar wyrażających jej przeciętny poziom.

Miary przeciętne charakteryzują średni lub typowy poziom wartości cechy. Są to więc takie miary, wokół których skupiają się wszystkie pozostałe wartości analizowanej cechy.

UWAGA

Definicja
Średnia arytmetyczna $\bar{X}$ jest ilorazem sumy wartości zmiennej i liczebności badanej zbiorowości. Z szeregów wyliczających średnią arytmetyczną obliczamy następująco:

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

gdzie:
x_i – wariant cechy mierzalnej
N – liczebność badanej cechy

Ś**rednia wy**raża **przeciętny poziom badanej cechy (zmiennej) w populacji, np. przeciętna miesięczna sprzedaż, średnia ocena na świadectwie szkolnym itp. Interpretacja średniej i metoda jej wyznaczania jest zawsze taka sama, jednak techniczny sposób obliczenia średniej zależy od typu szeregu statystycznego, z którym mamy do czynienia. Średnia jest sumą wartości cechy podzieloną przez liczbę jednostek zbiorowości.**

UWAGA

WŁASNOŚCI ŚREDNIEJ:

- Średnia jest lepsza od mediany dla rozkładów bliskich do symetrycznych, gorsza od mediany dla rozkładów dalekich od symetrycznych lub w obecności wartości odstających (oddalonych);
- na poziom średniej arytmetycznej silny wpływ wywierają wartości ekstremalne (skrajne), przy czym wpływ ten jest silniejszy w przypadku wysokich wartości zmiennej;
- średnia arytmetyczna jest miarą prawidłową tylko w odniesieniu do zbiorowości jednorodnych, o niewielkim zróżnicowaniu wartości zmiennej u poszczególnych jednostek.

WAŻNE

Wzory
Średnia z próby

- $\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N}$ - szereg wyliczający
- $\overline{X} = \frac{\sum_{i=1}^{N} x_i n_i}{N}$ - szereg rozdzielczy punktowy
- $\overline{X} = \frac{\sum_{i=1}^{N} x_i^{¿} n_i}{N}$ - szereg rozdzielczy przedziałowy,

gdzie $x_i^{¿}$ - środek przedziału klasowego

UWAGA

Definicja

Średnia harmoniczna $\bar{X}_H$ jest odwrotnością średniej arytmetycznej z odwrotności wartości zmiennych

$$\bar{X}_H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

gdzie:
x_i – wariant cechy mierzalnej
N – liczebność badanej cechy

Przy obliczaniu średniej harmonicznej z szeregów rozdzielczych (punktowych i przedziałowych) zachodzi konieczność stosowania wag (czyli uwzględniania liczebności).

$$\bar{X}_H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

UWAGA

Średnią harmoniczną stosuje się wówczas, gdy wartości zmiennej podane są w jednostkach względnych, wagi zaś – w jednostkach względnych liczników.

Przykładowo można tu wymienić takie zmienne, jak:

- prędkość pojazdu (zmienna – km/h, waga – liczba kilometrów),
- gęstość zaludnienia (zmienna – mieszkańcy/km², waga – liczba mieszkańców).

4.1.2. Miary położenia pozycyjne

Dominanta (moda) – wartość, która w danym zbiorze danych lub rozkładzie prawdopodobieństwa występuje najczęściej. W rozkładzie dyskretnym (którego odpowiednikiem w badaniach empirycznych jest cecha skokowa) jest wartością, która jest najbardziej prawdopodobna. Dla rozkładów ciągłych (cechy ciągłej) dominanta to maksimum funkcji gęstości prawdopodobieństwa. Z punktu widzenia analizy matematycznej dominanta to nic innego jak ekstremum funkcji.

W badaniach empirycznych dla szeregów indywidualnych i rozdzielczych punktowych dominanta to ta wartość cechy, której odpowiada największa liczebność. W szeregu rozdzielczym przedziałowym dominantę można wyznaczyć graficznie, wykorzystując histogram, lub analitycznie (za pomocą wzoru).

Sposób wyznaczania dominanty zależy od typu szeregu statystycznego, z którym mamy do czynienia. Dla szeregów szczegółowych lub rozdzielczych punktowych dominantę wystarczy wskazać, gdyż jest to wartość o największej liczebności.

WAŻNE

Wzory

Dominanta (moda)

- szereg wyliczający i szereg rozdzielczy punktowy: moda to wartość najczęstsza, o ile nie jest to wartość skrajna (wówczas moda jest nieokreślona)
- szereg rozdzielczy przedziałowy:

$$D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D$$

gdzie:

x_D – ¿ – lewy koniec przedziału z modą (czyli przedziału o największej liczebności, ale różnego od przedziału pierwszego i ostatniego)

h_D – ¿ – długość przedziału z modą

n_D – ¿ – liczebność przedziału z modą

n_{D-1} – ¿ – liczebność przedziału poprzedzającego przedział z modą

n_{D+1} – ¿ – liczebność przedziału następującego po przedziale z modą

Przykład

W roku akademickim 2020/2021 student otrzymał następujące oceny z przedmiotu Statystyka:

2; 2; 3; 3; 3; 3,5; 3,5; 4; 4; 4,5; 4,5; 4,5; 4,5; 5; 5.

Dominanta wynosi 4,5. Oznacza to, że student najczęściej otrzymywał ocenę 4,5.

Przykład

Wyniki kolokwium ze Statystyki w jednej z grup przedstawia poniższa tabela.

Ocena	2	3	3,5	4	4,5	5
Liczba studentó	3	10	9	5	6	1

Dominanta wynosi 3. Oznacza to, że najwięcej studentów otrzymało ocenę dostateczną 3.

UWAGA

Aby wyznaczyć dominantę w przypadku szeregu rozdzielczego przedziałowego, muszą być spełnione następujące warunki:

- musi występować jeden przedział klasowy z wyraźnie dominującą liczebnością w stosunku do pozostałych,
- przedział klasowy, w którym znajduje się dominanta oraz przedziały z nią sąsiadujące muszą mieć jednakowe rozpiętości,
- szereg nie może być skrajnie asymetryczny z otwartym przedziałem dominującym.

Mediana jest to **wartość, która jest umieszczona dokładnie w środku**, pod warunkiem, że mamy do czynienia z uporządkowaną (z punktu widzenia badanej cechy) zbiorowością według wielkości jej elementów, tzn. od ich wartości najmniejszej do największej. Mediana dzieli zbiorowość na dwie równe części w ten sposób, że połowa

jednostek ma wartości cechy niższe lub równe medianie, a połowa ma wartości cechy większe lub równe od mediany.

WAŻNE

Wzory

Mediana – wartość środkowa w uporządkowanej próbie

• szereg wyliczający i szereg rozdzielczy punktowy:

$$Me=\begin{cases} x_{\frac{n+1}{2}}, & gdy\ n\ jest\ nieparzyste \\ \frac{1}{2}\left(x_{\frac{n}{2}}+x_{\frac{n}{2}+1}\right), & gdy\ n\ jest\ parzyste \end{cases}$$

tzn. mediana jest to środkowa liczba, gdy n jest liczbą nieparzystą, albo średnia arytmetyczna dwóch środkowych liczb, gdy n jest liczbą parzystą

- szereg rozdzielczy przedziałowy:

$$Me=x_{Me}+\frac{h}{n_{Me}}\left(\frac{n}{2}-\sum_{i=1}^{k-1} n_i\right)$$

gdzie:

x_{Me} – lewy koniec przedziału z medianą

h – długość przedziału z medianą

n_{Me} – liczebność przedziału z medianą

k – numer przedziału zawierającego medianę

Mediana ma poniższe własności:

- lepsza do opisu wartości typowych niż średnia arytmetyczna w przypadku:

  - rozkładów skośnych, gdyż leży bliżej wartości występujących najczęściej niż średnia arytmetyczna,
  - populacji o bardzo rzadko występujących wartościach ekstremalnych, gdyż jest mało podatna na zmiany wartości skrajnych w losowych próbach;
- jej wadą jest mała stabilność, która jednak rośnie wraz ze wzrostem liczebności próby.

UWAGA

Własności mediany

- Może być ona wyznaczana w szeregach o otwartych przedziałach klasowych,
- można ją wyznaczać do opisania zbiorowości, których nie można określić liczbowo (do wyznaczenia mediany nie jest konieczna znajomość wszystkich wartości cechy mierzalnej),
- jest jedyną średnią, którą można wyznaczyć dla rozkładów skrajnie asymetrycznych,
- nie jest wrażliwa na wartości skrajne (w przeciwieństwie do średniej arytmetycznej),
- może być wyznaczana w szeregach o nierównych rozpiętościach przedziałów klasowych, tj. w sytuacji, kiedy niemożliwe jest wyznaczenie dominanty.

Jeżeli konieczna jest bardziej szczegółowa analiza właściwości strukturalnych, oprócz mediany, która jest kwartylem drugim, znajdują zastosowanie kwartyl pierwszy i trzeci. Kwartyle należą do miar statystycznych zwanych kwantylami, które dzielą zbiorowość statystyczną w określonej proporcji.

Kwantyle są to liczby, które dzielą zbiór wartości badanej cechy na równe (mniej więcej) części pod względem liczby obserwowanych wartości. Najczęściej używane kwantyle to: **kwartyle** (3 kwartyle, podział na 4 części; drugi kwartyl to mediana), **decyle** (9 decyli, podział na 10 części), **percentyle** (99 percentyli, podział na 100 części).

UWAGA

Kwartyle
Kwartyle wyznaczamy w sposób analogiczny jak w przypadku mediany, z tym że należy uwzględnić konkretny numer kwantyla. W przypadku danych uporządkowanych rosnąco i przedstawionych w postaci szeregu szczegółowego wartości kwartyla pierwszego i trzeciego możemy wyznaczyć przez podzielenie zbiorowości na dwie części: pierwszą, złożoną z jednostek nie większych od mediany i drugą, złożoną z jednostek nie mniejszych od mediany.
Mediana każdej z tych części jest odpowiednio kwartylem pierwszym i trzecim.

WAŻNE

Wzory
Kwartyle (dolny Q1 i górny Q3) – wartości, które dzielą uporządkowaną próbę w stosunku 1:3 i 3:1

- szereg rozdzielczy przedziałowy:

$$Q_1 = x_{Q_1} + \frac{\frac{N}{4} - cum_{Q_1 - 1}}{n_{Q_1}} \cdot h_{Q_1}$$

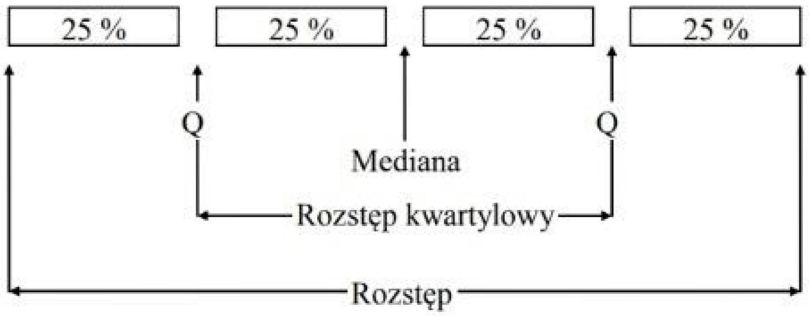
$$Q_3 = x_{Q_3} + \frac{\frac{3}{4}N - cum_{Q_3 - 1}}{n_{Q_3}} \cdot h_{Q_3}$$

gdzie:

x_{Q_1} – lewy koniec przedziału zawierającego Q1
h – długość przedziału zawierającego Q1

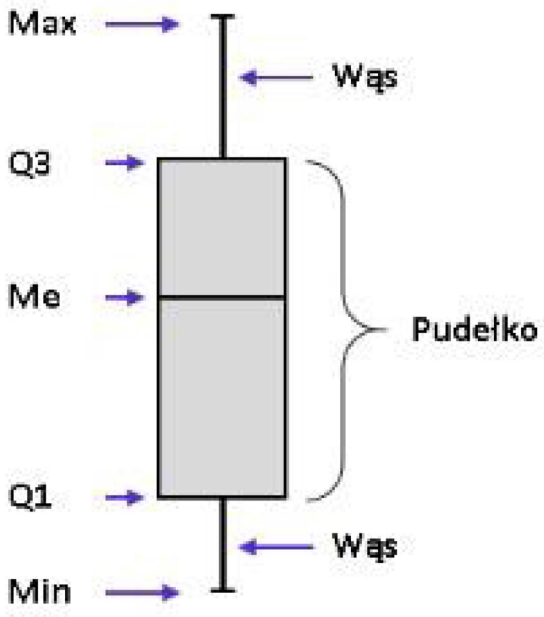
n_{Q_1} – liczebność przedziału zawierającego Q1
k – numer przedziału zawierającego Q1

Należy zauważyć, że znajomość wartości Q1 i Q3 w uporządkowanym szeregu pozwala nam na stwierdzenie, że połowa (50%) środkowych jednostek danej zbiorowości statystycznej przyjmuje wartość od Q1 do Q3.



Rysunek 44. Kwartyle.

Pożytecznym wykresem, tworzonym na podstawie kwartyli, jest wykres pudełkowy.



Rysunek 45. Wykres pudełkowy.

Przedstawiony na rysunku wykres pudełkowy składa się z następujących elementów:
Max – wartość maksymalna
Q3 – trzeci kwartyl
Me – mediana
Q1 – pierwszy kwartyl
Min – wartość minimalna

Analiza wykresu pudełkowego umożliwia nam wyciągnięcie wniosków w następujących obszarach:

- **Położenie**: o położeniu świadczy cały wykres pudełkowy, m.in możemy określić zakres danych (Min, Max).
- **Rozproszenie**: im dłuższy wykres, tym dane są bardziej rozproszone, tzn. mogą przyjmować bardziej różniące się wartości. O rozproszeniu świadczą także długie wąsy – tzn. występują obserwacje skrajne (bardzo odbiegające od pudełka).
- **Kształt:** jeżeli wykres wygląda symetrycznie względem kreski z medianą, to możemy podejrzewać, że wykres cechy jest **symetryczny**. Jeżeli pudełko nie jest równo podzielone albo/i wąsy są różnej długości, to mamy do czynienia z **rozkładem asymetrycznym** – najczęściej to, czy asymetria jest prawostronna czy lewostronna, możemy odczytać po odległości Max i Min od Mediany: jeżeli jedna z tych odległości jest znacząco większa, to mamy asymetrię prawostronną (jeżeli znacząco większa jest odległość Max od Me) lub lewostronną (jeżeli odległość Min od Me jest większa).

Jeżeli zbiorowość jest bardzo liczna, może się okazać, że podział zbiorowości na ćwiartki jest niewystarczający. Stosujemy wówczas decyle (dzielące zbiorowość na subpopulacje dziesięcioprocentowe) lub percentyle.

4.2. Ocena dyspersji

Miary średnie, chociaż reprezentują wszystkie jednostki badanej zbiorowości, nie dają wyczerpującej charakterystyki szeregu statystycznego, nie pozwalają przeniknąć w wewnętrzny układ zbiorowości. W rozdziale tym poznamy miary oceny zróżnicowania.

Elasyczne 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).

Rysunek 46. Miary rozproszenia.

Najprostszą i najbardziej intuicyjną miarą zmienności przypadków w populacji próby jest rozstęp. W związku z tym, że przy jego obliczeniu ignoruje się wszystkie dane (za wyjątkiem dwóch wartości – minimalnej i maksymalnej), nie daje on jednak informacji o zróżnicowaniu poszczególnych wartości cechy w zbiorowości.

WAŻNE

Wzory

Rozstęp

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

Rozstęp to miara prosta i łatwa do obliczenia. Jest ona jednak bardzo czuła na dwie skrajne wartości cechy, które często różnią się istotnie od wszystkich pozostałych wartości, a nierzadko są wartościami nietypowymi dla badanej zbiorowości, dlatego jest to miara o małej wartości poznawczej i wykorzystywana jest najczęściej do wstępnej oceny zróżnicowania badanej zbiorowości.

WAŻNE

Definicja

Odchylenie ćwiartkowe Q jest to parametr określający odchylenie wartości cechy od mediany. Mierzy poziom zróżnicowania tylko części jednostek; po odrzuceniu 25% jednostek o wartościach najmniejszych i 25% jednostek o wartościach największych.

$$Q = \frac{1}{2}(Q_3 - Q_1)$$

Odchylenie ćwiartkowe *Q* informuje, o ile przeciętnie wartości cechy 50% środkowych jednostek zbiorowości różnią się od mediany.

Tym samym odchylenie ćwiartkowe nie mierzy zróżnicowania całej zbiorowości, ale tylko 50% środkowych jednostek. 25% jednostek o najniższych wartościach cechy i 25% o najwyższych wartościach cechy jest odrzucana, nieuwzględniana w obliczeniach. Na wartość odchylenia ćwiartkowego nie mają wpływu skrajne, często przypadkowe wartości szeregu statystycznego. Odchylenie ćwiartkowe ma przejrzystą interpretację i można je obliczyć nawet wtedy, gdy w szeregu rozdzielczym występują otwarte przedziały klasowe.

WAŻNE

Definicja

Współczynnik zmienności jest ilorazem bezwzględnej mary zmienności cechy i średniej wartości tej cechy, jest wielkością niemianowaną, najczęściej podawaną w procentach

$$V = \frac{s}{\bar{x}} \cdot 100\,\%$$

Współczynnik zmienności stosuje się w porównaniach zróżnicowania:

- kilku zbiorowości pod względem tej samej cechy,
- tej samej zbiorowości pod względem kilku różnych cech.

Tabela 8. Klasyczny współczynnik zmienności.

Wartość klasycznego współczynnika zmienności	Zróżnicowanie cechy
0 – 20%	słabe
20 – 40%	umiarkowane
40 – 60%	silne
Powyżej 60%	bardzo silne

Odchylenie przeciętne jest miarą rozrzutu. Mówi o tym, o jaką wartość różnią się przeciętnie wartości cechy (zmiennej) od średniej. Im większa wartość odchylenia, tym większe zróżnicowanie wartości zmiennej.

WAŻNE

Definicja

Odchylenie przeciętne d jest to średnia arytmetyczna bezwzględnych odchyleń wartości cechy od średniej arytmetycznej. Określa, o ile jednostki danej zbiorowości różnią się średnio, ze względu na wartość cechy, od średniej arytmetycznej.

Wybór wzoru na odchylenie przeciętne, podobnie jak dla średniej arytmetycznej, uzależnione jest od rodzaju szeregu statystycznego, a więc od przedstawienia danych.

WAŻNE

Wzory

Odchylenie przeciętne od średniej

$d_1 = \frac{1}{n}\sum_{i=1}^{n} ¿x_i - \bar{x} \vee ¿¿$ – szereg wyliczający

$d_1 = \frac{1}{n}\sum_{i=1}^{n} ¿x_i - \bar{x} \vee n_i$ – szereg rozdzielczy punktowy

$d_1 = \frac{1}{n}\sum_{i=1}^{n} ¿x_i^{¿} - \bar{x} \vee ¿n_i ¿$ – szereg rozdzielczy przedziałowy,

gdzie: $x_i^{¿}$ – środek przedziału klasowego

Wzory

Odchylenie przeciętne od mediany

$d_1 = \frac{1}{n}\sum_{i=1}^{n} ¿x_i - Me \vee ¿¿$ – szereg wyliczający

$d_1 = \frac{1}{n}\sum_{i=1}^{n} ¿x_i - Me \vee n_i$ – szereg rozdzielczy punktowy

$d_1 = \frac{1}{n}\sum_{i=1}^{n} ¿x_i^{¿} - Me \vee ¿n_i ¿$ – szereg rozdzielczy przedziałowy,

gdzie: $x_i^{¿}$ – środek przedziału klasowego

Odchylenie standardowe jest kolejną miarą rozrzutu. Mówi o tym, jak wartości cechy (zmiennej) są rozrzucone wokół średniej. Im większa wartość odchylenia, tym większe zróżnicowanie wartości zmiennej, im mniejsze odchylenie, tym mniejsze zróżnicowanie. Im mniejsza wartość odchylenia, tym obserwacje są bardziej skupione wokół średniej, im większa wartość odchylenia, tym obserwacje są bardziej oddalone od średniej.

WAŻNE

Definicja

Odchylenie standardowe s jest to pierwiastek kwadratowy z wariancji. Stanowi miarę zróżnicowania o mianie zgodnym z mianem badanej cechy, określa przeciętne zróżnicowanie poszczególnych wartości cechy od średniej arytmetycznej.

Odchylenie standardowe jest charakterystyką często wykorzystywaną w badaniach statystycznych. Ma ono następujące właściwości:

- jest wielkością obliczaną na podstawie wszystkich obserwacji w danym szeregu,
- jego wartość nie zmieni się, jeśli liczebności szeregu wyrazimy w liczbach względnych (procentach) dostatecznie dokładnie ustalonych,
- wartość odchylenia standardowego nie zmieni się, jeśli do wszystkich wartości zmiennej w szeregu dodamy pewną stałą liczbę,
- jeśli wszystkie wartości szeregu pomnożymy przez pewną stałą liczbę większą od zera, to odchylenie standardowe również będzie tylokrotnie większe.

Kwadrat odchylenia standardowego zwany jest wariancją i również mierzy stopień rozproszenia obserwacji wokół średniej arytmetycznej. Jednak z powodu podniesienia jednostki do kwadratu, traci się część interpretacji.

WAŻNE

Definicja

Wariancja jest to średnia arytmetyczna kwadratów odchyleń poszczególnych wartości cechy od średniej arytmetycznej zbiorowości.

Dane

$x_1, x_2, \ldots, x_n$ – dane indywidualne

$\overline{x}_j$ = 1, 2,…, k – wartości środkowe w klasach

n_j = 1, 2, …, k – liczebność w klasach

Tabela 9. Miary dyspersji – wzory.

Miara	Dane indywidualne	Szereg rozdzielczy
Wariancja z próby	$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$	$s^2=\frac{1}{n}\sum_{j=1}^{k}n_j(\overline{x}_j-\overline{x})^2$
Odchylenie standardowe	$s=\sqrt{s^2}$	$s=\sqrt{s^2}$
Współczynnik zmienności	$v=\frac{s}{\overline{x}}$	$v=\frac{s}{\overline{x}}$
Rozstęp	$R=x_{\max}-x_{\min}$	
Rozstęp ćwiartkowy	$Q=Q_3-Q_1$	
Odchylenie ćwiartkowe	$Q_c=\frac{Q_3-Q_1}{2}$	
Pozycyjny współczynnik zmienności	$V=\frac{Q_c}{Me}$	

Typowy obszar zmienności cechy mówi nam o tym, jakie wartości typowo przyjmuje dana cecha. Wyróżniamy dwa rodzaje **typowego obszaru zmienności**:

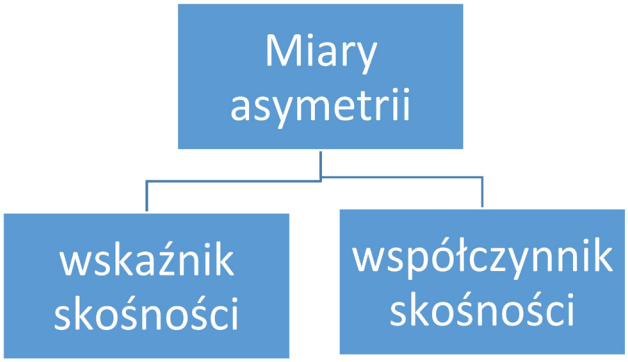
- klasyczny $X_{typ}\in(\overline{X}-s,\overline{X}+s)$
- pozycyjny $X_{poz}\in(Me-Q,Me+Q)$

UWAGA

Około 2/3 (67%) wszystkich jednostek badanej zbiorowości statystycznej posiada wartości cechy w typowym obszarze zmienności.

4.3. Ocena asymetrii

Kolejnym etapem analizy struktury jest badanie asymetrii, czyli skośności (lewostronnej bądź prawostronnej) szeregu statystycznego. Analizując szeregi strukturalne, można spotkać się z przypadkiem, gdy średni poziom badanej cechy i jej zróżnicowania nie obrazuje dostatecznie różnic między badanymi szeregami, a szczegółowa obserwacja szeregów wyklucza podobieństwo tych szeregów. W takim przypadku posługujemy się miarami asymetrii.



Rysunek 47. Miary asymetrii.

Zagadnienie asymetrii (skośności) można zbadać za pomocą miar asymetrii. Ich konstrukcja opiera się na spostrzeżeniu, że w szeregu symetrycznym wszystkie trzy miary średnie: średnia arytmetyczna, dominanta i mediana są równe.

WAŻNE

Definicja

Wskaźnik skośności

$$W_s = \bar{x} - Mo$$

jest to wielkość bezwzględna wyrażona jako różnica między średnią arytmetyczną a modalną.

Wskaźnik skośności jest miarą bezwzględną (mianowaną) i jego przydatność jest niewielka, ponieważ nie nadaje się do porównywania asymetrii cech, które mierzone są w różnych jednostkach miary. Wartość tego miernika zależy również od stopnia rozproszenia (zmienności) cechy w badanej zbiorowości.

Współczynniki skośności (asymetrii) są stosowane w porównaniach do określenia siły oraz kierunku asymetrii, są to liczby niemianowane, im większa ich wartość, tym silniejsza asymetria.

Współczynnik asymetrii jest liczbą niemianowaną. Na ogół przyjmuje wartość z przedziału od -1 do +1. Może się zdarzyć, że przy bardzo silnej asymetrii wartość bezwzględna współczynnika będzie większa od 1. Znak współczynnika informuje o kierunku asymetrii, natomiast wartość bezwzględna o sile asymetrii: im większa wartość bezwzględna, tym silniejsza asymetria.

WAŻNE

Wzory

Współczynnik asymetrii

$$A = \frac{M_3}{s^3},$$

gdzie:

$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$ – szereg wyliczający

$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 n_i$ – szereg rozdzielczy punktowy

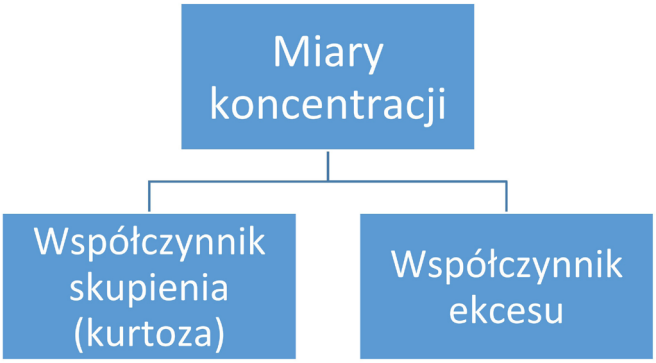
$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i^{¿} - \bar{x})^3 n_i$ – szereg rozdzielczy przedziałowy,

$x_i^{¿}$ – środek przedziału klasowego

4.4. Ocena koncentracji

Opisane w poprzednich rozdziałach miary średnie, miary zróżnicowania i miary asymetrii pozwalają w sposób wyczerpujący opisać strukturę badanej zbiorowości. W niektórych sytuacjach opis ten można uzupełnić (wzbogacić) miarami koncentracji.

Miara koncentracji rozkładu to miara wskazująca na nierównomierne rozdysponowanie cechy pomiędzy elementy próby losowej. Podstawową miarą informującą o skupieniu wartości cechy wokół średniej jest kurtoza. Określa ona stopień koncentracji w porównaniu z szablonem, jaki stanowi rozkład normalny.



Rysunek 48. Miary koncentracji.

UWAGA

Kurtoza jest miarą koncentracji wyników.
Kurtoza informuje nas o tym, na ile nasze obserwacje, wyniki są skoncentrowane wokół średniej.
Miara ta informuje nas, jak dużo naszych wyników/obserwacji jest zbliżona do wartości średniej, czy większość z zaobserwowanych wyników ma wartość podobną do średniej.

WAŻNE

Wzory

Wzór na **współczynnik kurtozy** ma postać

$$K=\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{s}\right)^4-\frac{3(n-1)^2}{(n-2)(n-3)}$$

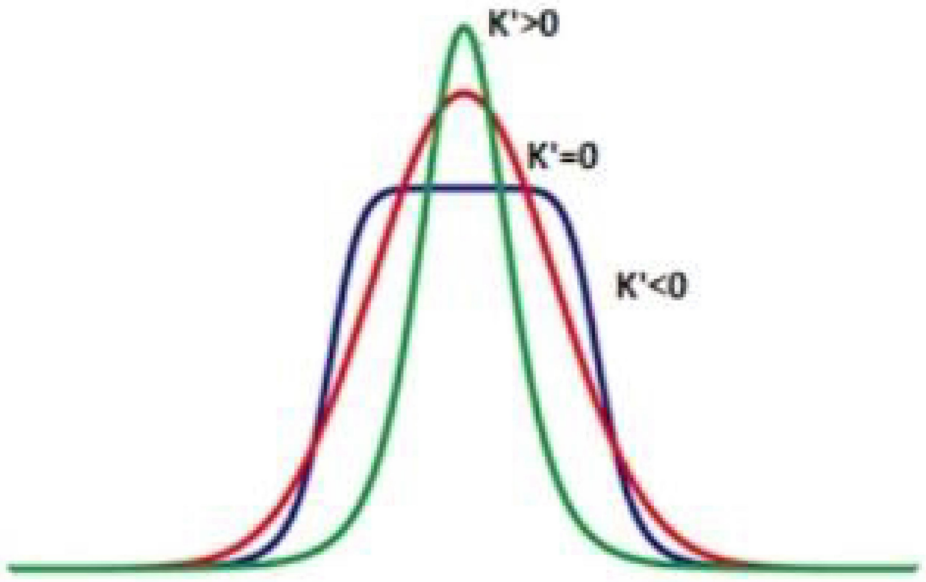
gdzie:
n – liczba obserwacji
x_i– ¿ – poszczególne obserwacje,
s – odchylenie standardowe,
$\bar{x}$– ¿ – średnia wyników z próby

Wartości współczynnika kurtozy K:

- **K = 3** – rozkład ma taką koncentrację, jak rozkład normalny,
- **K > 3** – rozkład leptokurtyczny,
- **K < 3** – rozkład platokurtyczny.

Współczynnik ekcesu K’ = K −3, gdzie K to współczynnik kurtozy. Zatem:

- **K’= 0** – dla rozkładów o takiej koncentracji, jak rozkład normalny,
- **K’ > 0** – dla rozkładów leptokurtycznych,
- **K’< 0** – dla rozkładów platokurtycznych.



Rysunek 49. Wartości współczynnika ekcesu.

UWAGA

- Im wyższa wartość K, tym bardziej wysmukła krzywa liczebności, więc większa koncentracja wartości cechy wokół średniej.
- Małe wartości K wskazują natomiast na spłaszczenie rozkładu zbiorowości względem badanej cechy.

4.5. Zastosowanie pakietu EXCEL

Do rozwiązywania problemów z zakresu statystyki możemy korzystać z szerokich możliwości oferowanych przez program EXCEL. Tabela 10 zawiera najczęściej używane funkcje statystyczne służące do analizy danych.

Tabela 10. Statystyczne funkcje dostępne w pakiecie EXCEL.

Pojęcie statystyczne	Funkcja w EXCEL-u	Opis
kurtoza	KURTOZA	Zwraca kurtozę zbioru danych. Rozkład najbardziej zbliżony do normalnego ma kurtozę zerową, najbardziej wysmukły - dodatnią, najbardziej spłaszczony - ujemną.
liczebność	ILE.LICZB	Zlicza liczby znajdujące się na liście argumentów.
maksimum	MAX	Zwraca maksymalną wartość listy argumentów.
mediana	MEDIANA	Zwraca medianę podanych liczb.
minimum	MIN	Zwraca minimalną wartość z listy argumentów.
moda	WYST.NAJCZĘŚCIEJ. WART	Zwraca wartość najczęściej występującą w zbiorze danych.
odchylenie średnie	ODCH.ŚREDNIE	Zwraca średnią wartość odchyleń absolutnych punktów danych od ich wartości średniej.
średnia arytmetyczna	ŚREDNIA	Zwraca wartość średnią argumentów.
średnia geometryczna	ŚREDNIA. GEOMETRYCZNA	Zwraca średnią geometryczną.
średnia harmoniczna	ŚREDNIA. HARMONICZNA	Zwraca średnią harmoniczną.
wariancja	WARIANCJA.POP	Oblicza wariancję na podstawie całej populacji.
wariancja	WARIANCJA.PRÓBKI	Szacuje wariancję na podstawie próbki.
współczynnik skośności	SKOŚNOŚĆ	Zwraca skośność rozkładu.

4.6. Pytania sprawdzające

1. Kiedy do opisu populacji nie można użyć średniej arytmetycznej?

.....
.....
.....
.....
.....
.....
.....

2. Jak wyznacza się średnią arytmetyczną i jakie są jej podstawowe własności?

.....
.....
.....
.....
.....
.....
.....
.....

3. Podaj interpretację mediany.

.....
.....
.....
.....
.....
.....
.....
.....
.....

4. Kiedy możemy obliczać dominantę?

.....
.....
.....
.....
.....
.....
.....
.....
.....

5. Co to jest wyraz środkowy, co to jest mediana?

.....
.....
.....
.....
.....
.....
.....
.....
.....

6. Jak wyznacza się kwartyle i jaka jest ich interpretacja?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

7. Jaki jest cel stosowania miar rozproszenia?

.....
.....
.....
.....
.....
.....
.....
.....
.....

8. Wymień miary rozproszenia.

.....
.....
.....
.....
.....
.....
.....
.....
.....

9. Wymień klasyczne miary dyspersji i podaj ich interpretację.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

10. Wymień pozycyjne miary dyspersji i podaj ich interpretację.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

11. Jaka jest różnica między miarami klasycznymi i pozycyjnymi?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

12. Co oznacza i dlaczego badamy asymetrię?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

13. Jakie znasz miary do oceny asymetrii?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

1 2 3 4 5 6

5. Analiza współzależności cech

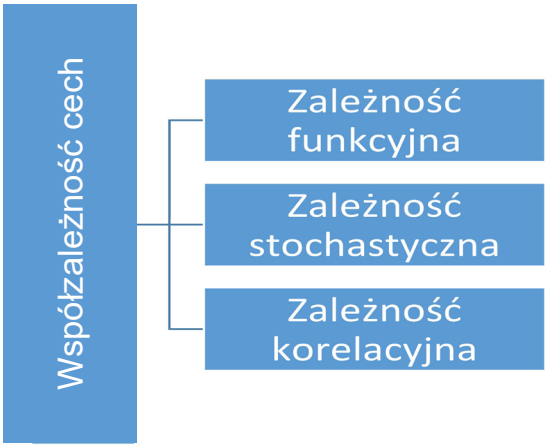
W tym rozdziale omówimy metody stosowane dla potrzeb wykrywania zależności lub współzależności między dwiema cechami.

5.1. Wprowadzenie do badania współzależności cech

W badaniach empirycznych jednostki statystyczne charakteryzowane są nie przez jedną, ale przez kilka cech. Cechy te często nie występują oddzielnie, ale mają na siebie wpływ oraz posiadają wzajemne uwarunkowania. Dlatego też często istnieje potrzeba badania zależności między tymi cechami.

Możemy wyróżnić trzy podstawowe rodzaje zależności.

1. **Zależność funkcyjna** – polega na tym, że zmiana wartości jednej cechy powoduje ściśle określoną zmianę wartości drugiej cechy. Zależność tę można opisać za pomocą równania, np. y=-4x+105.
2. **Zależność stochastyczna** (losowa) – polega na tym, że wraz ze zmianą wartości jednej zmiennej zmieniają się szanse realizacji drugiej zmiennej.
3. **Zależność korelacyjna** (statystyczna) – danej wartości zmiennej niezależnej odpowiada przeciętna wartość zmiennej zależnej.



Rysunek 50. Rodzaje zależności między cechami.

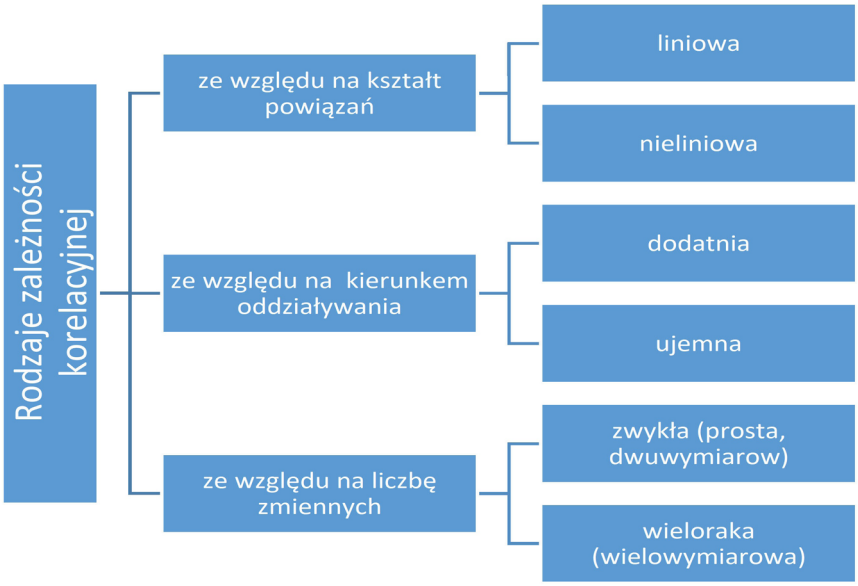
Przykład

Przykładami związków między dwoma zjawiskami mogą być związki między:
- wagą a wzrostem,
- dochodami a wydatkami,
- stażem pracy a wynagrodzeniem,
- zatrudnieniem a wielkością produkcji,
- liczbą pracowników a kosztami wyjazdów służbowych,
- wielkością produkcji a kosztami,
- ceną produktu a podażą produktu.

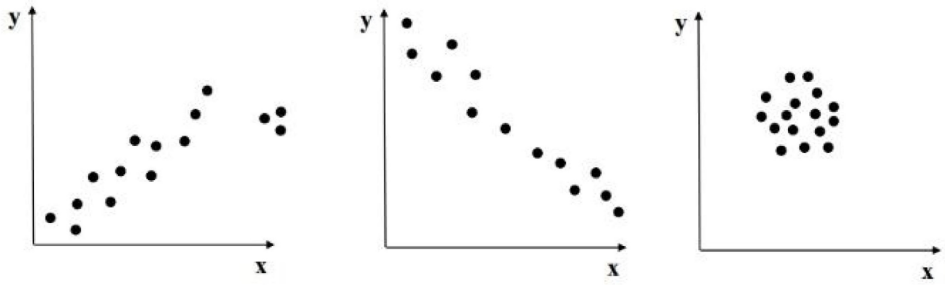
Stwierdzenia istnienia (lub braku) związku korelacyjnego między dwoma zjawiskami można dokonać na podstawie:
- porównania kształtowania się szeregów statystycznych szczegółowych charakteryzujących rozpatrywane zjawiska,
- wykresu,
- zbudowanej tablicy korelacyjnej, gdy dane o dwóch zjawiskach podane są w formie szeregów rozdzielczych.

5.2. 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.



Rysunek 51. Rodzaje zależności korelacyjnych.



Rysunek 52. Rodzaje związków korelacyjnych: dodatni, ujemny, brak korelacji.

5.2.1. Analiza korelacji cech mierzalnych

Jeżeli pomiędzy dwiema cechami zachodzi korelacja liniowa, to jej kierunek i siłę można zmierzyć za pomocą współczynnika korelacji liniowej Pearsona, który w skrócie nazwany jest współczynnikiem korelacji.

WAŻNE

Wzory

Współczynnik korelacji liniowej Pearsona mierzy związek między zmiennymi x i y, gdy są one mierzalne i podane w postaci szeregów szczegółowych.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{C(X,Y)}{S(X)\cdot S(Y)}$$

gdzie:

r_{xy} – współczynnik korelacji liniowej Pearsona

i=1,2,…, - numer jednostki statystycznej

x_i – wartość pierwszej cechy statystycznej u jednostki statystycznej o numerze i

y_i – wartość drugiej cechy statystycznej u jednostki statystycznej o numerze i

$\bar{x}$ – średnia arytmetyczna pierwszej cechy

$\bar{y}$ – średnia arytmetyczna drugiej cechy

S(X), S(Y) – odchylenia standardowe zmiennych X oraz Y

UWAGA

Abyśmy mogli liczyć współczynnik korelacji liniowej Pearsona:

- wartości obu cech nie muszą być uporządkowane,
- zmienne muszą być ilościowe,
- muszą one pochodzić z populacji o rozkładzie zbliżonym do normalnego,
- nie mogą zawierać obserwacji odstających (zaburzą one wynik).

Jeśli współczynnik korelacji jest = 0 rozważane zmienne losowe nazywanymi nieskorelowanymi. Brak korelacji nie oznacza, że między tymi zmiennymi nie istnieje jakakolwiek zależność losowa. Oznacza to jedynie, że poziomy badanych zjawisk nie ulegają zmianie na skutek zmiany wartości jednej z nich.

Jeśli współczynnik korelacji przyjmuje wartości w przedziale od -1 do 0 lub od 0 do 1, to między zmiennymi występuje odpowiednio ujemna lub dodatnia liniowo zależność korelacyjna.

UWAGA

Im jest bliższa zera, tym związek jest słabszy.
Im bliżej 1 (lub -1), tym związek silniejszy.

Znak współczynnika korelacji mówi o kierunku zależności:

+ oznacza związek dodatni. Zmiany zachodzą w tym samym kierunku, tj. wzrost (spadek) wartości jednej cechy powoduje wzrost (spadek) średniej wartości drugiej cechy.

– oznacza związek ujemny. Zmiany zachodzą w przeciwnym kierunku, tj. wzrost (spadek) wartości jednej cechy powoduje spadek (wzrost) średniej wartości drugiej cechy.

Tabela 11. Miary zależności współczynnika korelacji liniowej.

$\lvert r_{xy} \rvert$	Zależność liniowa
< 0.2	brak zależności
0.2 – 0.4	zależność niska
0.4 – 0.7	zależność uiarkowana
0.7 – 0.9	zależność znacząca
>0.9	zależność bardzo silna

UWAGA

Własności współczynnika korelacji liniowej Pearsona:

- Pozwala na określenie nie tylko siły, ale i kierunku zależności między zmiennymi.
- Dodatni znak współczynnika wskazuje na sytuację, gdy wzrostowi jednej cechy towarzyszy na ogół wzrost wartości drugiej cechy.
- Ujemny znak współczynnika wskazuje na sytuację, gdy wzrostowi jednej cechy towarzyszy na ogół spadek wartości drugiej cechy.

Na podstawie wyliczonego współczynnika korelacji obliczyć można tzw. **wskaźnik determinacji liniowej**, informujący o procencie wyjaśnionej liniowo zmienności zmiennej zależnej przez zmienną niezależną. Wskaźnik ten oblicza się na podstawie wzoru:

$$WD = r_{xy}^2 \cdot 100\,\%$$

WAŻNE

Podsumowując informacje dotyczące korelacji liniowej Pearsona:

- związek między zmiennymi musi być liniowy,
- pokazuje siłę i kierunek związku między zmiennymi,
- współczynnik korelacji przyjmuje wartości z przedziału [-1.1],
- jest najczęściej stosowanym testem statystycznym.

Współczynnik korelacji Spearmana, podobnie jak Pearsona, służy do badania zależności pomiędzy danymi, jednak w odróżnieniu od niego pozwala na badanie dowolnych monotonicznych zależności, nie tylko liniowych. Pomiar siły korelacji przy wykorzystaniu współczynnika korelacji rang jest mniej dokładny w porównaniu z pomiarem za pomocą współczynnika korelacji liniowej Pearsona, ale z kolei sposób jego obliczania jest dużo prostszy.

WAŻNE

Wzory

Współczynnik korelacji rang Spearmana

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{n\left(n^2 - 1\right)}$$

gdzie:

d_i – ¿ – różnica między rangami nadanymi wartościom zmiennej x uporządkowanym rosnąco i rangami nadanymi odpowiadającym im wartościom zmiennej y

n – liczba obserwacji

i – numer jednostki statystycznej

WAŻNE

Współczynnik korelacji rang Spearmana stosuje się do opisu siły korelacji dwóch cech, gdy:

- cechy są mierzalne, a liczebność badanej zbiorowości jest nieliczna,
- cechy mają charakter jakościowy i istnieje możliwość ich zrangowania, czyli uporządkowania od najmniejszej do największej lub odwrotnie.

Interpretacja współczynnika Spearmana:

$r > 0$ korelacja dodatnia – gdy wartość X rośnie, to Y również wzrasta
$r = 0$ brak korelacji – wzrost X nie powoduje zawsze podobnej zmiany Y
$r < 0$ korelacja ujemna – gdy X rośnie, to Y maleje

Tabela 12 . Siła korelacji dla współczynnika Spearmana.

Wartość współczynnika Spearmana	Siła korelacji
0 – 0.2	bardzo słaba zależność
0.2 – 0.4	słaba zależność
0.4 – 0.6	umiarkowana zależność
0.6 – 0.8	silna zależność
0.8 – 1	bardzo silna zależność

UWAGA

Aby obliczyć współczynnik korelacji rang, należy kolejno:

1. Nadać każdej jednostce statystycznej rangę według pierwszej cechy.
2. Nadać każdej jednostce statystycznej rangę według drugiej cechy.
3. Obliczyć dla każdej jednostki statystycznej różnicę rang.
4. Podnieść każdą różnicę rang do kwadratu.
5. Obliczyć sumę podniesionych do kwadratu różnic rang.
6. Obliczyć współczynnik korelacji rang zgodnie ze wzorem.

5.2.2. Analiza korelacji cech niemierzalnych

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.

WAŻNE

Wzory

Współczynnik kontyngencji C Pearsona

$$C=\sqrt{\frac{\chi^2}{\chi^2+n}}$$

gdzie:

n – liczba obserwacji

χ^2 – wynik testu chi-kwadrat między dwiema zmiennymi

WAŻNE

- **Współczynnik kontyngencji C Pearsona może być stosowany przy tablicach wielodzielnych o dowolnie dużej liczbie poziomów dwóch zmiennych (minimalna liczba wynosi 2, czyli tabela 2x2).**
- **Współczynnik przyjmuje wartości od 0 (co oznacza niezależność dwóch zmiennych, brak zależności pomiędzy nimi) do prawie 1 (co oznacza bardzo silną zależność pomiędzy nimi).**

5.3. Analiza regresji

Przy wyznaczaniu współczynnika korelacji liniowej Pearsona interesowało nas zagadnienie siły i kierunku zależności między cechami X i Y. W wielu analizach warto scharakteryzować bardziej szczegółowo związek między nimi i wskazać na tę cechę, która jest przyczyną kształtowania się drugiej, czyli skutku. Mówimy wówczas o występowaniu związku przyczynowo-skutkowego. Do jego analizy posłużymy się metodą regresji prostej.

Analiza regresji jest bardzo popularną i chętnie stosowaną techniką statystyczną pozwalającą opisywać związki zachodzące pomiędzy zmiennymi wejściowymi (objaśniającymi) a wyjściowymi (objaśnianymi).

5.3.1. Liniowa funkcja regresji

Celem regresji jest zbudowanie na podstawie dostępnych informacji modelu opisującego rzeczywistość. Taki model pełni funkcję poznawczą – dostarcza wiedzy na temat zjawiska, a także umożliwia prognozowanie (predykcję) nieznanej wartości analizowanej cechy.

Aby zapisać równanie regresji, należy wyznaczyć (oszacować) parametry strukturalne funkcji, czyli a i b. Współczynnik a oraz b wyznaczamy z poniższych wzorów.

WAŻNE

Wzory

I linia regresji

$$x = a_X \cdot y + b$$

$$a_X = \frac{C(X,Y)}{S_Y^{\,2}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

$$b = \overline{x} - a_X \cdot \overline{y}$$

gdzie:

x – zmienna objaśniana (zależna)

y – zmienna objaśniająca (niezależna)

a_X – parametr linii regresji (współczynnik kierunkowy prostej)

b – drugi parametr linii regresji.

UWAGA

W pierwszej linii regresji mamy sytuację, w której zmienna y ma wpływ na kształtowanie się zmiennej x. Interpretacja parametrów a i b jest następująca:

Parametr a – jeżeli zmienna y wzrośnie o jednostkę, to zmienna x wzrośnie lub spadnie o a.

Parametr b – nie interpretuje się tego parametru, gdyż nie ma on sensu ekonomicznego.

WAŻNE

Wzory

II linia regresji

$$y = a_Y \cdot x + b$$

$$a_Y = \frac{C(X,Y)}{S_X^{\,2}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$b = \overline{y} - a_Y \cdot \overline{x}$$

gdzie:
x – zmienna objaśniająca (niezależna)
y – zmienna objaśniana (zależna)
a_Y – parametr linii regresji
b – drugi parametr linii regresji

UWAGA

W drugiej linii regresji mamy sytuację, w której zmienna x ma wpływ na kształtowanie się zmiennej y. Interpretacja parametrów a i b jest następująca:

Parametr a – jeżeli zmienna x wzrośnie o jednostkę, to zmienna y wzrośnie lub spadnie o a.

Parametr b – nie interpretuje się tego parametru, gdyż nie ma on sensu ekonomicznego.

Bardzo duże znaczenie ma współczynnik regresji a. Wskazuje on bowiem, o ile przeciętnie zmieni się wartość zmiennej zależnej Y, jeśli wartość zmiennej niezależnej X wzrośnie o jednostkę.

- Ujemny współczynnik a wskazuje na to, że pod wpływem wzrostu zmiennej X o jednostkę, zmienna zależna Y zmaleje średnio o a jednostek.
- Dodatni współczynnik a wskazuje na to, że pod wpływem wzrostu zmiennej X o jednostkę, zmienna zależna Y wzrośnie średnio o a jednostek.

5.3.2. Ocena liniowej funkcji regresji

Znaczenie poznawcze funkcji regresji jest tym większe, im silniejsza jest korelacja między badanymi zmiennymi. Silny związek oznacza bowiem, że zmienna niezależna determinuje w znacznym stopniu zmiany w poziomie zmiennej zależnej, niewielki natomiast jest wpływ indywidualnych odchyleń. Wskazuje na to kwadrat współczynnika korelacji liniowej Pearsona, czyli tzw. współczynnik determinacji.

WAŻNE

Wzory

Współczynnik determinacji R^2 jest miarą unormowaną, $R^2 \in \langle 0; 1 \rangle$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Związek między współczynnikiem determinacji R^2 a współczynnikiem korelacji liniowej Pearsona r: $R^2 = r^2$ lub $r = \pm\sqrt{R^2}$

Współczynnik determinacji określa, jaka część danych jest wytłumaczona przez model – im większy, tym prosta regresji jest lepiej dopasowana do danych. Możemy go interpretować jako tę część zmienności zmiennej y, która została wyjaśniona przez regresję, czyli przez liniowy związek pomiędzy zmienną celu a zmienną objaśniającą. Im prosta regresji jest lepiej dopasowana do danych (czyli im współczynnik determinacji większy), tym możemy przewidzieć wartości Y dla bardziej oddalonych wartości X od danych, na podstawie których stworzyliśmy prostą regresji liniowej.

Tabela 13. Współczynnik determinacji.

Współczynnik determinacji	Dopasowanie
0 – 0.5	niezadawalające
0.5 – 0.6	słabe
0.6 – 0.8	zadawalające
0.8 – 0.9	dobre
0.9 – 1.0	bardzo dobre

Po znalezieniu równania funkcji regresji należy zbadać, na ile nasze oszacowanie pokrywa się z rzeczywistością. W tym celu badamy dopasowanie funkcji regresji do danych empirycznych. Do oceny dopasowania prostej regresji do punktów empirycznych wykorzystuje się tak zwane reszty, które stanowią różnicę między rzeczywistymi i teoretycznymi (wyliczonymi z równania prostej) wartościami zmiennej zależnej.

WAŻNE

Wzory

Wariancja i odchylenie standardowe składnika resztowego są to miary rozproszenia punktów empirycznych wokół prostej regresji. Są to miary nienormowane (przyjmują dowolne dodatnie wartości i nie mogą służyć do porównania).

Składnik resztowy: $z_i = y_i - \hat{y}_i$

$$S^2(z) = \frac{\sum_{i=1}^{n} z_i^2}{n-2}$$

$$S(z) = \sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n-2}}$$

5.4. Zastosowanie pakietu EXCEL

Pojęcie statystyczne	Funkcja w EXCEL-u	Opis
Kowariancja	KOWARIANCJA.POPUL	Zwraca kowariancję, czyli średnią wartość iloczynów odpowiednich odchyleń.
Kowariancja	KOWARIANCJA.PRÓBKI	Zwraca kowariancję próbki, czyli średnią iloczynów odchyleń dla każdej pary punktów danych w dwóch zbiorach danych.
Regresja liniowa	REGLINX	Zwraca wartość trendu liniowego.
Regresja liniowa – parametry	REGLINP	Zwraca parametry trendu liniowego.
Regresja liniowa – przedział ufności	REGLINX.ETS.CONFINT	Zwraca przedział ufności dla wartości prognozy dla określonej daty docelowej.
Regresja liniowa – nachylenie linii	NACHYLENIE	Zwraca nachylenie linii regresji liniowej.
Regresja liniowa – błąd	REGBŁSTD	Zwraca błąd standardowy przewidzianej wartości y dla każdej wartości x w regresji.
Współczynnik korelacji	WSP.KORELACJI	Zwraca współczynnik korelacji dwóch zbiorów danych.
Współczynnik determinacji	R.KWADRAT	Zwraca kwadrat współczynnika korelacji momentu iloczynu Pearsona.

5.5. Pytania sprawdzające

1. Na czym polega analiza współzależności?

……
……
……
……
……
……
……
……

2. Jaka jest różnica między zależnością stochastyczną a zależnością funkcyjną?

……
……
……
……
……
……
……
……

3. Dlaczego przed dokonaniem analizy korelacji należy sprawdzić, czy pomiędzy dwoma cechami występuje związek przyczynowo-skutkowy?

……
……
……
……
……
……
……

4. Jakie rodzaje współzależności mogą wystąpić między dwiema cechami?

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

5. Wyjaśnij cele i założenia analizy regresji.

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

6. W jaki sposób wyznacza się parametry funkcji regresji?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

7. Jakie znasz miary służące do oceny siły i kierunku współzależności między dwoma cechami?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

8. Jaka jest interpretacja parametrów liniowej funkcji regresji?

.....
.....
.....
.....
.....
.....
.....
.....
.....

9. Jaka jest interpretacja współczynnika determinacji liniowej?

.....
.....
.....
.....
.....
.....
.....
.....
.....

10. Jak liczymy współczynnik korelacji liniowej Pearsona?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

1 2 3 4 5 6

6. Analiza dynamiki

Zjawiska społeczno-gospodarcze zmieniają się wraz z upływem czasu. Metody statystyki przedstawione w rozdziale Analiza struktury pozwalały badać nam opisane zjawiska w ściśle określonym okresie czasu. Analiza dynamiki polega na przedstawieniu rozmiarów i kierunków rozwoju zjawisk, tzn. na ustaleniu stopnia wzrostu lub spadku badanego zjawiska w czasie.

Przykład

Liczba ludności Polski w poszczególnych latach kształtowała się następująco:

Rok	Liczba ludności
1946	24 mln
1960	30 mln
2000	38,5 mln
2010	38,5 mln
2019	38,4 mln

6.1. Definicja i klasyfikacja metod analizy dynamiki

Miary dynamiki (nazywane również miernikami dynamiki) są wykorzystywane do pomiaru wielkości i kierunku zmian zachodzących w poziomie badanego zjawiska wraz z upływem czasu. Istnieją różne mierniki dynamiki: niektóre z nich pozwalają na wyrażenie wielkości zachodzących zmian w liczbach mianowanych, a inne – w liczbach względnych.

Mierniki dynamiki mogą być obliczane dla zjawisk jednorodnych (które mogą być liczbowo wyrażone w jednakowych jednostkach fi-

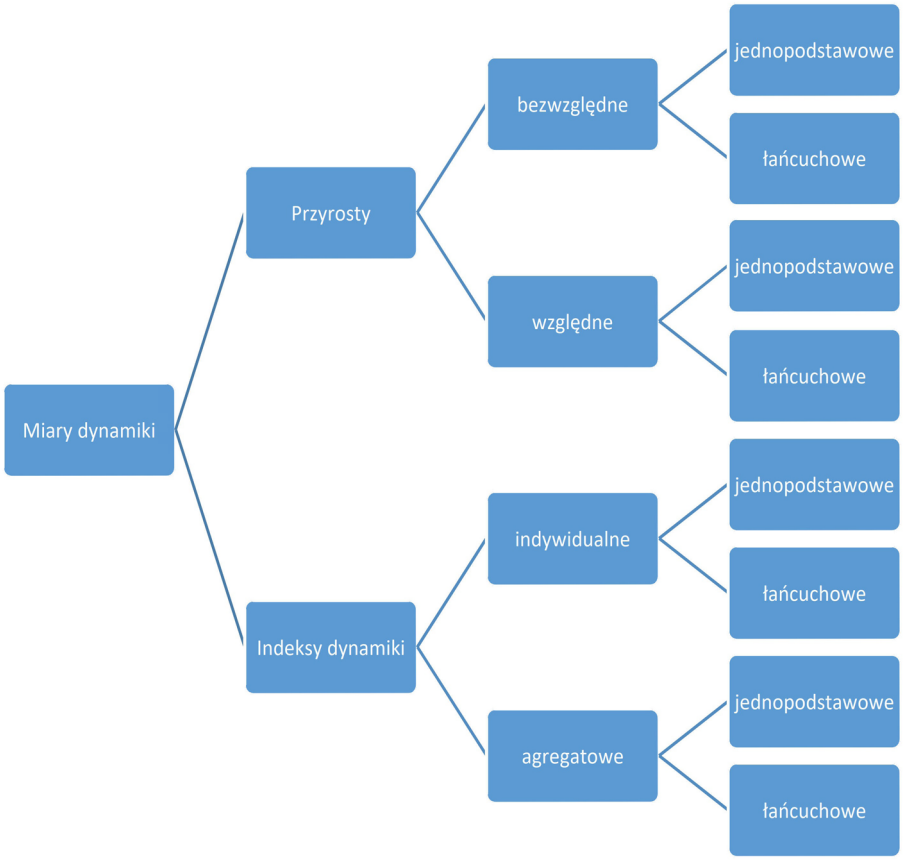
zycznych) lub dla zjawisk złożonych (wyrażonych w różnych jednostkach miary).

Miarami statystycznymi wykorzystywanymi do badania dynamiki zjawisk jednorodnych są:

- przyrost bezwzględny,
- przyrost względny i tempo wzrostu,
- indeksy indywidualne.

Miarami statystycznymi wykorzystywanymi do analizy dynamiki zjawisk złożonych są indeksy agregatowe, do których należą:

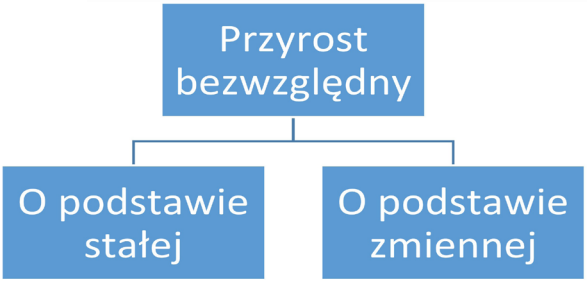
- agregatowy indeks wartości,
- agregatowy indeks ilości,
- agregatowy indeks cen.



Rysunek 53. Miary dynamiki.

6.2. Przyrosty bezwzględny

Przyrosty bezwzględne mogą być przyrostami o podstawie stałej lub przyrostami o podstawie zmiennej.



Rysunek 54. Przyrosty bezwzględne.

WAŻNE

Wzory

Przyrost bezwzględny o podstawie stałej jest różnicą między wielkością zjawiska w okresie badanym i podstawowym

$$\Delta y_s = y_n - y_t$$

gdzie:

Δy_s – przyrost absolutny o podstawie stałej

y_n – wielkość zjawiska w badanym okresie

y_t – wielkość zjawiska w okresie podstawowym

n – oznaczenie kolejnego okresu

t – oznaczenie okresu podstawowego

Wzory

Przyrost bezwzględny o podstawie zmiennej jest różnicą między wielkością zjawiska w okresie badanym i poprzedzającym okres badany

$$\Delta y_{n/(n-1)} = y_n - y_{n-1}$$

gdzie:

$\Delta y_{n/(n-1)}$ – przyrost bezwzględny o podstawie stałej

y_n – wielkość zjawiska w badanym okresie

y_{n-1} – wielkość zjawiska w okresie poprzedzającym okres badany

n – oznaczenie kolejnego okresu

Przyrosty absolutne informują o tym, ile jednostek wzrósł (gdy przyrost bezwzględny jest wyrażony liczbą dodatnią) lub zmalał (gdy przyrost bezwzględny jest wyrażony liczbą ujemną) poziom badanego zjawiska w okresie (momencie) badanym w porównaniu z okresem (momentem) przyjętym za podstawę. Przyrosty bezwzględne (zarówno jednopodstawowe, jak i łańcuchowe) są wielkościami mianowanymi, wyrażonymi w tych samych jednostkach miary co badane zjawisko.

Przykład

W pierwszych dwóch kolumnach tabeli zaprezentowany został szereg dynamiczny, który przedstawia liczbę sprzedanych produktów w kinach w pewnym lubelskim przedsiębiorstwie w latach 2008–2013.

Rok	Sprzedaż (w tys. sztuk)	Przyrost bezwzględny o podstawie stałej (w tys. sztuk)	Przyrost bezwzględny o podstawie zmiennej (w tys. sztuk)
2015	35,2	1	-
2016	28,7	-6,50	-6,50
2017	31,6	-3,60	2,90
2018	36,1	0,90	4,50
2019	34,5	-0,70	-1,60
2020	23,8	-11,40	-10,70

Trzecia kolumna zawiera informacje o wielkości przyrostów bezwzględnych o podstawie stałej – jako okres bazowy przyjęto rok 2015, dlatego yc = y2015. Zaprezentowane w tej kolumnie liczby informują, o ile tysięcy liczba sprzedanych produktów w danym roku różniła się od liczby sprzedanych produktów w roku 2015. Wybrane obliczenia i interpretacje przyrostów bezwzględnych o podstawie stałej przedstawione zostały poniżej.

- Dla roku 2015: Δys = y2015 – y2015 = 0. W tym przypadku zarówno okresem badanym, jak i okresem podstawowym jest ten sam rok – dlatego przyrost bezwzględny o podstawie stałej wynosi 0, ponieważ odejmujemy od siebie tę samą liczbę.
- Dla roku 2016: Δys = y2016 – y2015 = 28,7 – 35,2 = -6,50. W roku 2016 sprzedano o 3,9 tys. produktów mniej niż w roku 2015.

Czwarta kolumna tabeli zawiera informacje o wielkości przyrostów bezwzględnych o podstawie zmiennej. Zaprezentowane w tej kolumnie liczby informują o tym, ile wynosiła różnica między liczbą sprzedanych produktów w danym roku a liczbą sprzedanych produktów w roku poprzednim. Wybrane obliczenia i interpretacje przyrostów bezwzględnych o podstawie zmiennej przedstawione zostały poniżej.

- Dla roku 2016: $\Delta y = y2016 - y2015 = 28{,}7 - 35{,}2 = -6{,}50$. W roku 2016 sprzedano o 3,9 tys. produktów mniej niż w roku 2015.
- Dla roku 2017: $\Delta y = y2017 - y2016 = 31{,}6 - 28{,}7 = 2{,}9$. W roku 2017 sprzedano o 2,9 tys. produktów więcej niż w roku 2017.

6.3. Przyrosty względne

WAŻNE

Definicja

Przyrostem względnym nazywamy iloraz przyrostów absolutnych zjawiska do jego poziomu w okresie (momencie) przyjętym za podstawę porównań. Przyrosty względne mogą być jednopodstawowe i łańcuchowe.

WAŻNE

Wzory

Szereg łańcuchowy przyrostów względnych tworzymy, dzieląc kolejne przyrosty absolutne łańcuchowe przez wartości cechy w okresie poprzednim

$$\Delta y_{n/(n-1)} = \frac{y_n - y_{n-1}}{y_{n-1}}$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_{n-1} – wielkość zjawiska w okresie poprzedzającym badany okres

n – oznaczenie kolejnego okresu

Wzory

Szereg jednopodstawowy przyrostów względnych tworzymy, dzieląc kolejne przyrosty absolutne łańcuchowe przez wartości cechy w okresie bazowym

$$\Delta y_{n/i} = \frac{y_n - y_i}{y_i}$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_i – wielkość zjawiska w okresie podstawowym

n – oznaczenie kolejnego okresu

i – oznaczenie okresu podstawowego

Przyrost względny wyrażony w procentach jest nazywany **tempem wzrostu.** Przyrosty względne wyrażone są w ułamkach lub procentach (tempo wzrostu). **Informują one o tym, o ile wyższy lub niższy jest poziom badanego zjawiska w danym okresie w sto-**

sunku do okresu przyjętego za podstawę (indeksy względne jednopodstawowe) lub do okresu bezpośrednio poprzedzającego (indeksy względne łańcuchowe).

WAŻNE

Wzory
Tempo wzrostu o podstawie zmiennej

$$T_z = \frac{y_n - y_{n-1}}{y_{n-1}} \cdot 100\%$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_{n-1} – wielkość zjawiska w okresie poprzedzającym badany okres

n – oznaczenie kolejnego okresu

WAŻNE

Wzory
Tempo wzrostu o podstawie stałej

$$T_s = \frac{y_n - y_i}{y_i} \cdot 100\%$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_i – wielkość zjawiska w okresie podstawowym

n – oznaczenie kolejnego okresu

i – oznaczenie okresu podstawowego

6.4. Indeksy indywidualne

Kolejną grupą miar dynamiki są **indeksy**, nazywane również **wskaźnikami dynamiki.** Obliczenie indeksu polega na podzieleniu dwóch liczb, z których pierwsza przedstawia wielkość zjawiska w okresie badanym, a druga – w okresie podstawowym. Otrzymany wynik dzielenia jest zawsze dodatnią liczbą niemianowaną.

WAŻNE

Definicja

Indeksem dynamiki nazywamy stosunek wielkości badanego zjawiska w danym okresie (momencie) do wielkości tego samego zjawiska w innym okresie (momencie), przyjętym za podstawę porównań:

$$i=\frac{y_n}{y_0} \quad \text{lub} \quad i=\frac{y_n}{y_0}\cdot 100\%$$

gdzie:

y_n – poziom zjawiska w pewnym okresie

y_0 – poziom zjawiska w okresie odniesienia

UWAGA

Indeks dynamiki ma następujące właściwości:

- jest liczbą niemianowaną (tzn. nie posiada jednostki),
- jest liczbą nieujemną $(i \geq 0)$,
- równy jest jedności, gdy poziom zjawiska w obu okresach jest jednakowy $(y_n = y_0)$,
- wartość indeksu w postaci procentowej pomniejszona o 100%, tj.

$$\frac{y_n}{y_0} \cdot 100\% - 100\%$$

Rozróżniamy dwa rodzaje indeksów dynamiki: indywidualne (proste) i agregatowe. Indeksy indywidualne służą do analizy jednego szeregu czasowego $y_1, y_2, \ldots, y_n$. Przy badaniu zmian wartości cechy w czasie można posłużyć się ciągami indeksów jednopodstawowych lub indeksów łańcuchowych.

WAŻNE

Definicja

Indeksy jednopodstawowe informują, jakie zmiany nastąpiły w poziomie zjawiska w kolejnych okresach w stosunku do okresu przyjętego jako podstawa (bazowego).

$$\frac{y_1}{y_0}, \frac{y_2}{y_0}, \frac{y_3}{y_0}, \ldots, \frac{y_n}{y_0}$$

Przykład

Lata - t	Sprzedaż - y_t	Indeksy jednopodstawowe
2013	51	1
2014	52	52/51=1,02
2015	48	48/51=0,94
2016	46	46/51=0,90
2017	50	50/51=0,98
2018	53	53/51=1,03

Interpretacja: $i_{\frac{2018}{2013}} = 1{,}03$ oznacza, że w roku 2018 sprzedaż wzrosła o 3% względem roku 2013.

WAŻNE

Definicja

Indeksy łańcuchowe (o podstawie zmiennej) informują, jakie zmiany nastąpiły w poziomie zjawiska w kolejnym okresie w stosunku do okresu go poprzedzającego.

$$\frac{y_2}{y_1}, \frac{y_3}{y_2}, \frac{y_4}{y_3}, \ldots, \frac{y_n}{y_{n-1}}$$

Przykład

Lata - t	Sprzedaż - y_t	Indeksy łańcuchowe
2013	51	1
2014	52	52/51=1,02
2015	48	48/50=0,96
2016	46	46/48=0,92
2017	50	50/46=1,09
2018	53	53/50=1,06

Interpretacja: $i_{\frac{2018}{2017}} = 1{,}06$ oznacza, że w roku 2018 sprzedaż wzrosła o 6% względem roku 2017.

Zmiany indeksów łańcuchowych na jednopodstawowe dokonujemy według następujących zasad:

1. Indeks jednopodstawowy w okresie następującym bezpośrednio po okresie przyjętym za podstawę jest taki sam, jak indeks łańcuchowy.
2. Indeks jednopodstawowy w okresie przyjętym za podstawę wynosi 100%.
3. Dalsze indeksy jednopodstawowe po okresie przyjętym za podstawowy otrzymujemy, mnożąc w sposób narastający kolejne indeksy łańcuchowe, licząc od wskaźnika łańcuchowego znajdującego się po okresie podstawowym.
4. Indeksy jednopodstawowe przed okresem podstawowym są odwrotnością kolejnych indeksów łańcuchowych, licząc od okresu przyjętego za podstawę.

UWAGA

Wartości indeksów większe od 100% świadczą o wzroście poziomu zjawiska w okresie badanym w porównaniu z okresem podstawowym, a gdy są mniejsze od 100% – świadczą o spadku danego zjawiska.

6.5. Indeksy agregatowe

Czasem zachodzi potrzeba analizy dynamiki nie tyle jednego zjawiska, ile agregatu zjawisk. Przykładowo, producent kilku towarów chce poznać dynamikę ogólnej wartości swojej sprzedaży. Na ogólną wartość sprzedaży w danym okresie wpływa ilość sprzedaży poszczególnych towarów oraz ich cena w tym okresie – dynamika ogólnej wartości może więc zależeć od dynamiki ilości sprzedawanych towarów lub od dynamiki cen tych towarów lub od obu tych czynników. W tym sensie jest to dynamika zjawiska złożonego i potrzeba do niej trochę zmodyfikowanych narzędzi.

Indeksy ilości i cen można wyznaczyć według **dwóch formuł standaryzacyjnych: Laspeyresa i Paaschego**. Ich zastosowanie polega na ustaleniu stałego poziomu jednego z dwóch czynników: cen albo ilości. Dla uzyskania agregatowego indeksu ilości unieruchamiane są ceny, natomiast dla otrzymania agregatowego indeksu cen unieruchamiane są ilości. W przypadku gdy interesuje nas zbadanie dynamiki zmian rozmiarów produkcji, obliczamy agregatowe indeksy ilości według formuł Laspeyresa (przyjmują jako stałe ceny z okresu podstawowego) i Paaschego (przyjmują jako stałe ceny z okresu badanego).

Dla każdego dobra $i = 1, 2, \ldots, n$ **dane są:**

p_{i0} – cena i-tego dobra w okresie podstawowym t_0,

p_{i1} – cena i-tego dobra w okresie podstawowym t_1,

q_{i0} – cena i-tego dobra w okresie podstawowym t_0,

q_{i1} – cena i-tego dobra w okresie podstawowym t_1.

WAŻNE

Wzory

Indeks wartości produkcji

$$I_w = \frac{\sum_{i=1}^{n} p_{i1} \cdot q_{i1}}{\sum_{i=1}^{n} p_{i0} \cdot q_{i0}} \cdot 100\%$$

Indeks wartości produkcji informuje o tym, o ile procent zmieni się wartość produkcji w okresie badanym t_1 w stosunku do okresu podstawowego t_0.

WAŻNE

Wzory

Indeks cen według Laspeyresa

$$I_p^L = \frac{\sum_{i=1}^{n} p_{i1} \cdot q_{i0}}{\sum_{i=1}^{n} p_{i0} \cdot q_{i0}} \cdot 100\,\%$$

Indeks cen według Laspeyresa informuje o tym, o ile procent zmieniłaby się wartość produkcji, gdyby zmiana wartości produkcji następowała jedynie wskutek zmian cen, a wielkości produkcji byłyby niezmienne i na takim poziomie, jak w okresie podstawowym t_0.

WAŻNE

Wzory

Indeks cen według Paaschego

$$I_p^P = \frac{\sum_{i=1}^{n} p_{i1} \cdot q_{i1}}{\sum_{i=1}^{n} p_{i0} \cdot q_{i1}} \cdot 100\,\%$$

Indeks cen według Paaschego informuje o tym, o ile procent zmieniłaby się wartość produkcji, gdyby zmiana wartości produkcji

następowała jedynie wskutek zmian cen, a wielkości produkcji byłyby niezmienne i na takim poziomie, jak w okresie badanym t_1.

WAŻNE

Wzory

Indeks wielkości produkcji według Laspeyresa

$$I_p^L = \frac{\sum_{i=1}^{n} p_{i0} \cdot q_{i1}}{\sum_{i=1}^{n} p_{i0} \cdot q_{i0}} \cdot 100\%$$

Indeks wielkości produkcji według Laspeyresa informuje o tym, o ile procent zmieniłaby się wartość produkcji, gdyby zmiana wartości produkcji następowała jedynie wskutek zmian wielkości produkcji, a ceny byłyby niezmienne i na takim poziomie, jak w okresie podstawowym t_0.

WAŻNE

Wzory

Indeks wielkości produkcji według Paaschego

$$I_p^P = \frac{\sum_{i=1}^{n} p_{i1} \cdot q_{i1}}{\sum_{i=1}^{n} p_{i1} \cdot q_{i0}} \cdot 100\%$$

Indeks wielkości produkcji według Paaschego informuje o tym, o ile procent zmieniłaby się wartość produkcji, gdyby zmiana wartości produkcji następowała jedynie wskutek zmian wielkości produkcji, a ceny byłyby niezmienne i na takim poziomie, jak w okresie badanym t_1.

WAŻNE

Wzory

Indeks cen według Fishera

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

Wzory

Indeks wielkości produkcji według Fishera

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

UWAGA

- **Agregatowe indeksy cen informują o przeciętnej zmianie cen wszystkich produktów wchodzących w skład danego agregatu w okresie badanym w stosunku do okresu podstawowego, przy założeniu, że ilości poszczególnych produktów byłyby na stałym poziomie (albo z okresu podstawowego, albo z okresu badanego).**

- **Agregatowe indeksy ilości informują, o ile przeciętnie w okresie badanym w stosunku do okresu podstawowego zmieniły się ilości wszystkich produktów (zjawisk jednorodnych), wchodzących w skład agregatu (zjawiska złożonego), przy założeniu, że ich ceny były na stałym poziomie (albo z okresu podstawowego, albo z okresu badanego).**

6.6. Średnie tempo zmian

Poznane w poprzednich rozdziałach **miary dynamiki** (przyrosty, indeksy i tempo zmian) **pozwalają ocenić relacje pomiędzy wielkością badanego zjawiska w różnych okresach**. Oprócz tych miar w analizie dynamiki wykorzystuje się również **miary średnie**, które **pozwalają ustalić średni poziom zjawiska w czasie lub średnie tempo zmian** (średnie tempo dynamiki).

WAŻNE

Definicja

Średniookresowe tempo zmian w czasie określa średniookresowy wzrost lub spadek badanego zjawiska, przypadający na analizowaną jednostkę czasu:

$$\overline{T} = (\overline{i}_G - 1) \cdot 100\,\%$$

gdzie:

$\overline{T}$ – średniookresowe tempo zmian w czasie,

i_G – średniookresowy indeks łańcuchowy.

6.7. Pytania sprawdzające

1. W jaki sposób i kiedy budujemy szeregi wskaźników dynamiki?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

2. Jak wyznaczamy i interpretujemy wskaźniki dynamiki zjawisk jednorodnych?

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

3. Jaka jest różnica między względnymi a bezwzględnymi miarami dynamiki?

……
……
……
……
……
……
……
……
……

4. Jakie własności spełniają indeksy indywidualne?

……
……
……
……
……
……
……
……
……

5. Jakie miary są wykorzystywane do analizy dynamiki zjawisk złożonych?

……
……
……
……
……
……
……
……
……
……

6. Omów procedurę przekształcania szeregu indeksów indywidualnych.

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

7. Jaka jest interpretacja średniego tempa zmian?

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

8. Kiedy wykorzystujemy indeksy agregatowe?

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

9. O czym informują nas agregatowe wartości cen oraz ilości?

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

7. Wzory

7.1. Załącznik 1

miara	rodzaje szeregów		
	szczegółowy prosty	rozdzielczy punktowy	rozdzielczy przedziałowy
Wskaźnik struktury	$\omega_i = \frac{n_i}{N}$ $\omega_i = \frac{n_{i\,sk}}{N}$		
MIARY KLASYCZNE			
Klasyczne miary położenia			
średnia arytmetyczna	$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{k} x_i n_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{N} x_i^{¿} n_i}{N}$
średnia harmoniczna	$\overline{X}_H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$	$\overline{X}_H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$	
średnia geometryczna	$\overline{X}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots\ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$		
Klasyczne miary zmienności (rozproszenia, dyspersji)			
wariancja	$s^2 = \frac{\sum_{i=1}^{N} (x_i - \overline{x}_i)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x}_i)^2 \cdot n_i}{N}$	$s^2 = \frac{\sum_{i=1}^{k} (x_i^{¿} - \overline{x}_i)^2 \cdot n_i}{N}$
odchylenie standardowe	$s = \sqrt{s^2}$		

obszar zmienności	$\bar{x} - s < x_{typ} < \bar{x} + s$		
odchylenie przeciętne	$d = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{N}$	$d = \frac{\sum_{i=1}^{k} \lvert x_i - \bar{x} \rvert \cdot n_i}{N}$	$d = \frac{\sum_{i=1}^{k} \lvert x_i - \bar{x} \rvert \cdot n_i}{N}$
współczynnik zmienności	$V_s = \frac{S}{\bar{x}}$		
Klasyczne miary asymetrii			
współczynnik asymetrii	$A_s = \frac{\bar{x} - D}{S}$		
MIARY POZYCYJNE			
Pozycyjne miary położenia			
dominanta	$D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D$		
mediana	$Me = \begin{cases} x_{\frac{n+1}{2}}, & gdy\ n\ jest\ nieparzyste \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & gdy\ n\ jest\ parzyste \end{cases}$	$Me = x_{Me} + \frac{h}{n_{Me}}\left(\frac{n}{2} - \sum_{i=1}^{k-1} n_i\right)$	
kwartyl 1		$Q_1 = X_{Q_1} + \frac{\frac{N}{4} - cum_{Q_1 - 1}}{n_{Q_1}} \cdot h_{Q_1}$	
kwartyl 3		$Q_3 = X_{Q_3} + \frac{\frac{3}{4}N - cum_{Q_3 - 1}}{n_{Q_3}} \cdot h_{Q_3}$	
Pozycyjne miary zmienności			
rozstęp	$R = X_{max} - X_{min}$		
odchylenie ćwiartkowe	$Q = \frac{(Q_3 - M) + (M - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$		
typowy obszar zmienności	$M - Q < X_{typ} < M + Q$		
współczynniki zmienności	$V_Q = \frac{Q}{M} \quad V_{Q_{1_{Q_3}}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$		

Pozycyjne miary asymetrii		
współ-czynnik skośności	$A_Q = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} = \frac{Q_1 + Q_3 - 2M}{2Q} \leftarrow \left(\frac{Q_3 - Q_1}{2}\right)$	
równanie Pearsona	$2\overline{X} = 3M - D$	
KORELACJA		
miara	**rodzaje szeregów**	
	szereg szczegółowy	**tablica korelacyjna**
współ-czynnik korelacji rang Spe-armana	$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{n(n^2 - 1)}$	
współ-czynnik Korelacji Pearsona	$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{C(X,Y)}{S(X) \cdot S(Y)}$	$r_{xy} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \overline{x})(y_j - \overline{y})n_{ij}}{\sqrt{\sum_{i=1}^{r}(x_i - \overline{x})^2 \cdot n_{i.} \cdot \sum_{j=1}^{s}(y_j - \overline{y})^2 n_{.j}}} = \frac{C(X,Y)}{S(X) \cdot S(Y)}$
kowarian-cja	$C(X,Y) = \frac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$	$C(X,Y) = \frac{1}{n} \cdot \sum_{i=1}^{r}\sum_{j=1}^{s}(x_i - \overline{x})(y_j - \overline{y}) \cdot n_{ij}$
równanie regresji	$a_y = r_{xy} \cdot \frac{S(Y)}{S(X)}$ $\quad b_y = \overline{y} - a_y \cdot \overline{x}$ $a_x = r_{xy} \cdot \frac{S(X)}{S(Y)}$ $\quad b_x = \overline{x} - a_x \cdot \overline{y}$	
współ-czynnik determi-nacji	$R^2 = r_{xy}^2$	
współ-czynnik zbieżno-ści	$\varphi^2 = 1 - R^2$	
KORELACJA CECH JAKOŚCIOWYCH		
współ-czynnik asocjacji φ Julle'a	$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$	

ANALIZA DYNAMIKI ZJAWISK	
PRZYROSTY ABSOLUTNE obliczane w stosunku do	
jednego okresu:	$t^* = 1$ $Y_2 - Y_1, \ldots\ldots, Y_{n-1} - Y_1, Y_n - Y_1$ $t^* = k$ $Y_1 - Y_k, Y_2 - Y_k \ldots\ldots, Y_{n-1} - Y_k, Y_n - Y_k$ $\Delta_{t/k} = Y_t - Y_k \quad t = 1, \ldots, n$
stale zmieniającego się okresu bazowego	$Y_2 - Y_1, \ldots\ldots, Y_{n-1} - Y_{n-2}, Y_n - Y_{n-1}$ $\Delta_{t/t-1} = Y_t - Y_{t-1} \quad t = 2,3, \ldots, n$
PRZYROSTY WZGLĘDNE	
postać jednopodstawowa	$d_{t/k} = \frac{\Delta_{t/k}}{Y_k} = \frac{Y_t - Y_k}{Y_k} \quad t = 1,2, \ldots\ldots, n$
postać łańcuchowa	$d_{t/t-1} = \frac{\Delta_{t/t-1}}{Y_{t-1}} = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \quad t = 2, 3, \ldots\ldots, n$
INDYWIDUALNE INDEKSY DYNAMIKI	
jednopodstawowe	$i_{n/1} = \frac{y_n}{y_1} = \frac{y_n - y_1 + y_1}{y_1} = d_{n/1} + 1$
łańcuchowe	$i_{n/1} = \frac{y_n}{y_{n-1}} = d_{n/n-1} + 1$
średnie tempo zmian zjawiska w czasie	$i_G = \sqrt[n-1]{i_{n/n-1} \cdot i_{n-1/n-2} \cdots\cdots i_{2/1}} = \sqrt[n-1]{\frac{Y_n}{Y_{n-1}} \cdot \frac{Y_{n-1}}{Y_{n-2}} \cdots\cdots \frac{Y_2}{Y_1}} = \sqrt[n-1]{i_{n/1}}$
średnie okresowe tempo zmian zjawiska w czasie	$T_n = i_G - 1$ lub w procentach $T_n[\%] = i_G \; ¿100 - 100$ $t^* = 1$ $\frac{Y_n}{Y_0} = a \quad (1+T)^n \quad Y_n = Y_o \; ¿(1+T)^n$ $i_{n/o} = (1+T)^2 \quad T = \sqrt[n]{i_{n/o}} - 1$
zamiana indeksów jednopodstawowych na łańcuchowe	$\frac{Y_n}{Y_{n-1}} = \frac{Y_n}{Y_1} : \frac{Y_{n-1}}{Y_1} \quad i_{n/n-1} = i_{n/1} : i_{n-1/1} \; i_{n/n-1} = i_{n/k} : i_{n-1/k}$

zamiana indeksów łańcuchowych na jednopodstawowe	$\frac{Y_n}{Y_1} = \frac{Y_n}{Y_{n-1}} \cdot \frac{Y_{n-1}}{Y_{n-2}} \cdot \ldots\ldots\ldots\ldots \cdot \frac{Y_2}{Y_1}$ $i_{n/1} = i_{n/n-1} \cdot i_{n-1/n-2} \cdot \ldots\ldots\ldots \cdot i_{2/1}$ $i_{n/k} = i_{n/n-1} \cdot i_{n-1/n-2} \cdot \ldots\ldots\ldots \cdot i_{k+1/k}$ n>k

INDYWIDUALNE INDEKSY CEN, ILOŚCI I WARTOŚCI		
p - ceny	q- ilości	w- wartości
$i_p = \frac{p_n}{p_o}$	$i_q = \frac{q_n}{q_o}$	$i_w = \frac{w_n}{w_o}$
$w = p \cdot q$ $w_n = p_n \cdot q_n$ $w_o = p_o \cdot q_o$ $i_w = \frac{w_n}{w_o} = \frac{p_n \cdot q_n}{p_o \cdot q_o} = i_p \cdot i_q$ $i_w = i_p \cdot i_q \longleftarrow$ równość indeksowa		

AGREGATOWE INDEKSY DYNAMIKI		
agregatowy indeks cen formuły Laspeyresa	agregatowy indeks ilości formuły Laspeyresa	agregatowy indeks wartości
${}_L I_p = \frac{\sum_j p_{nj} q_{oj}}{\sum_j p_{oj} q_{oj}} = \frac{\sum p_n q_o}{\sum p_o q_o}$	${}_L I_q = \frac{\sum p_o q_n}{\sum p_o q_o}$	$I_w = \frac{\sum p_n q_n}{\sum p_o q_o} = \frac{\sum w_n}{\sum w_p}$
agregatowy indeks cen formuły Paaschego	agregatowy indeks ilości formuły Paaschego	$I_w = {}_P I_p \cdot {}_L I_q = {}_P I_q \cdot {}_L I_p$
${}_P I_p = \frac{\sum p_n q_n}{\sum p_o q_n}$	${}_P I_q = \frac{\sum p_n q_n}{\sum p_n q_o}$	
${}_F I_p = \sqrt{{}_L I_p \cdot {}_P I_p}$	${}_F I_q = \sqrt{{}_L I_q \cdot {}_P I_q}$	indeksy Fischera

7.2. Załącznik 2

Tabela 14. Znaczenie sygnalizacji błędów w Excelu.

Błąd	Opis
#####	Komórka zawiera liczbę, datę lub czas albo formułę daty i (lub) czasu zwracającą wynik ujemny. Zbyt mała szerokość kolumny.
#ARG!	Używany jest błędny typ argumentu lub operand. Funkcja Autokorekta formuł nie może poprawić formuły.
#DZIEL/0!	W formule była wykonywana próba dzielenia przez zero.
#NAZWA	Excel nie rozpoznaje tekstu w formule.
#N/D!	Komórki arkusza roboczego zawierają dane, które nie są jeszcze dostępne.
#ADR!	Błędne odwołanie do komórki.
#LICZBA!	Użycie nieprawidłowego argumentu w funkcji.
#ZERO!	Odwołanie do przecięcia dwóch obszarów, które nie mają części wspólnej.

8. Literatura

1. Domański, C. (2001). Metody statystyczne: teoria i zadania. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
2. Greń, J. (1982). Statystyka matematyczna: modele i zadania. Warszawa: PWN.
3. Jóźwiak, J., Podgórski, J. (2006). Statystyka od podstaw. Warszawa: PWE.
4. Kassyk-Rokicka, H. (1999). Statystyka Zbiór Zadań. Warszawa: PWE.
5. Kassyk-Rokicka, H. (1992). Statystyka nie jest trudna. Warszawa: PWE.
6. Kończak, G., Trzpiot, G. (2002): Analizy statystyczne z arkuszem kalkulacyjnym Microsoft Excel. Katowice: Wydawnictwo AE w Katowicach.
7. Kukuła, K. (2011). Elementy statystyki w zadaniach. Warszawa: Wydawnictwo Naukowe PWN.
8. Maksimowicz-Ajchel, A. (2019). Wstęp do statystyki: metody opisu statystycznego. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
9. Middleton, M.R. (2004). Microsoft Excel w analizie danych. Warszawa: Wydawnictwo RM.
10. Ostasiewicz, S., Rusnak, Z., Siedlecka, U. (2006). Statystyka: elementy teorii i zadania. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
11. Panek, T. [red.]. (2007). Statystyka społeczna. Warszawa: PWE.
12. Piłatowska, M. (2007). Repetytorium ze statystyki. Warszawa: Wydawnictwo Naukowe PWN.

13. Pułaska-Turyna, B. (2008). Statystyka dla ekonomistów. Warszawa: Difin.
14. Regel, W. (2003). Podstawy statystyki w Excelu. Warszawa: Mikom.
15. Rószkiewicz, M. (2002). Statystyka: kurs podstawowy. Warszawa: Efekt.
16. Wieczorkowska, G., Wierzbiński, J. (2007). Statystyka. Analiza badań społecznych. Warszawa: Wyd. Naukowe Scholar.
17. Zeliaś, A. (2000). Metody statystyczne. Warszawa: PWE.

Grafika użyta do rozdzielania poszczególnych rozdziałów:

1. Ad. 1 – jcomp/Freepik.com
2. Ad. 2 - freepik/Freepik.com
3. Ad. 3 – geralt/Pixabay.com
4. Ad. 4 – geralt/Pixabay.com
5. Ad. 5 – PhotoMix-Company/Pixabay.com
6. Ad. 6 – rawpixel.com/Freepik.com

STATYSTYKA

Część druga
Studia przypadków

Anna Tatarczak

Lublin 2021

Spis treści

Wstęp 161

Studium przypadku 1. Prezentacja graficzna danych 162
Przypadek 1 164
Przypadek 2 170
Przypadek 3 174
Przypadek 4 176
Zadania 180

Studium przypadku 2. Klasyfikacja danych 182
Przypadek 1 184
Przypadek 2 187
Przypadek 3 192
Przypadek 4 194
Zadania 198

Studium przypadku 3. Analiza struktury 199
Przypadek 1 205
Przypadek 2 210
Przypadek 3 215
Przypadek 4 218
Przypadek 5 220
Zadania 225

Studium przypadku 4. Analiza współzależności cech 230
Przypadek 1 236
Przypadek 2 247
Przypadek 3 249
Przypadek 4 252
Zadania 255

Studium przypadku 5. Analiza dynamiki 259
Przypadek 1 264
Przypadek 2 267
Przypadek 3 269
Przypadek 4 270
Zadania 272

Bibliografia 274

Wstęp

Głównym celem opracowania jest uporządkowanie i opis podstawowych metod i technik stosowanych na wykładzie ze statystyki opisowej. Z wykładem związane są ćwiczenia laboratoryjne, w trakcie których są praktycznie ilustrowane zagadnienia przedstawiane na wykładzie. Podstawowym narzędziem programowym, stosowanym do opisu poszczególnych procedur obliczeniowych, jak i do obliczeń w laboratorium komputerowym, jest program EXCEL.

Opracowanie jest głównie przeznaczone dla studentów Wyższej Szkoły Ekonomii i Innowacji poznających podstawy statystyki. Mogą również korzystać z niego praktycy prowadzący badania własne lub kierujący pracami badawczymi.

Zostało opisanych pięć głównych case study (każde z podziałem na mini-case), które są najczęściej wykorzystywane w pracy z obszaru ekonomii. Wszystkie rozdziały mają podobną strukturę: na początku przedstawiony jest cel ćwiczeń oraz problem, który będzie analizowany. Następnie czytelnik zapozna się z krótkim wstępem teoretycznym i znajdzie wykaz funkcji/poleceń wykorzystywanych w danym rozdziale. Główna treść każdego rozdziału to rozwiązane przykłady praktyczne ze szczegółowymi opisami w postaci komentarzy. Ponieważ najlepszą nauką języka programowania jest praktyka, więc każdy z case'ów kończy się ćwiczeniami do wykonania.

Dołączona bibliografia przedmiotu umożliwi zainteresowanym pogłębieniem wiedzy Czytelnikom dotarcie do innych podręczników ze statystyki.

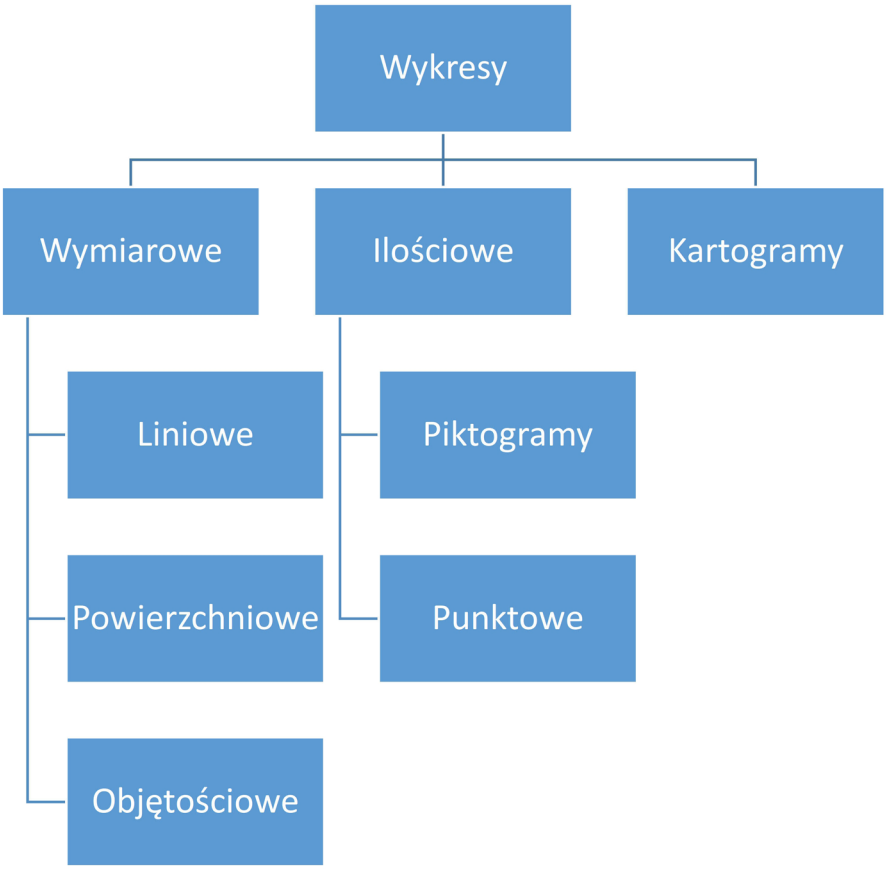
Studium przypadku 1. Prezentacja graficzna danych

Cel ćwiczenia

Celem analizowanych case'ów jest dogłębne poznanie metod i technik graficznej prezentacji danych oraz uzyskanie umiejętności stosowania nabytej wiedzy do analizy problemów ekonomicznych, a także poznanie możliwości oprogramowania EXCEL, Power BI oraz Gretl do graficznej analizy danych. Student w ramach omawianych problemów pozna różne rodzaje wykresów wykorzystywanych do graficznej prezentacji danych. Ważnym elementem poruszanych problemów jest wykształcenie krytycyzmu oraz świadomości wpływania warunków brzegowych badania statystycznego na uzyskany wynik.

Wstęp teoretyczny

Ważną rolę w analizie statystycznej odgrywa graficzna prezentacja danych statystycznych. Dzięki wykresom można przeprowadzić nie tylko wstępną analizę zjawiska czy zbiorowości, ale także dobrać odpowiednie parametry opisu rozkładu cechy. Graficzna prezentacja danych jest mniej dokładna niż prezentacja tabelaryczna (patrz rozdział Klasyfikacja danych), jednak pozwala ona na szybką wizualizację dotyczącą kształtowania się poziomu danego zjawiska. Istnieje wiele różnych typów wykresów. W literaturze przedmiotu można spotkać podział na wykresy wymiarowe, wykresy ilościowe i kartogramy. Podział ten został przedstawiony schematycznie na rysunku poniżej.



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:

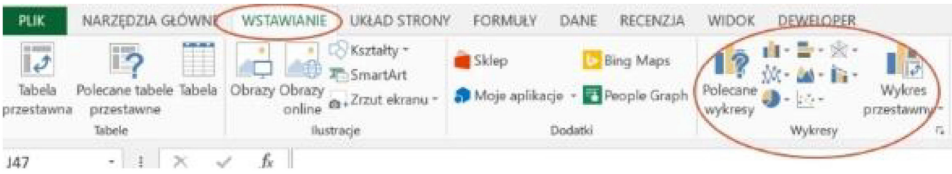
- tytułu,
- pola wykresu,
- skali,
- legendy,
- źródła danych statystycznych i ewentualnych uwag wyjaśniających.

W programie EXCEL dostępnych jest wiele typów wykresów. Aby wstawić jeden z nich, należy:

Zaznaczyć tabelę z danymi.

Wybrać w kategorii WSTAWIANIE z sekcji WYKRESY ikonkę wykresu kołowego.

Wybrać odpowiedni podtyp wykresu (kołowy, pierścieniowy, 3-W)



Przypadek 1

Polecenie:

Dla danych z tabeli należy dokonać wizualizacji danych z wykorzystaniem oprogramowania EXCEL.

Województwo	Produkcja sprzedana przemysłu	Inwestycje	Wartość Brutto środków trwałych	Płace	Stopa bezrobocia rejestrow. %	Podmioty gospodarki
	Na mieszkańca w tys. zł		w tys. zł			
Dolnośląskie	35,22	4,63	140 713	3,83	10,67	30,27
Kujawsko-pomorskie	23,66	2,96	66 028	2,25	15,04	17,60
Lubelskie	13,99	1,94	47 076	3,40	12,24	15,24
Lubuskie	26,26	2,84	37 636	3,26	12,93	10,08
Łódzkie	26,20	3,52	95 203	3,43	11,16	21,39
Małopolskie	22,78	2,76	107 796	3,55	9,09	32,03
Mazowieckie	43,45	6,01	395 519	4,72	8,81	61,95
Opolskie	22,31	3,72	37 223	3,46	11,64	7,09
Podkarpackie	17,33	2,45	56 803	3,21	14,14	14,40
Podlaskie	17,01	2,10	29 497	3,36	12,37	8,86

Województwo	Produkcja sprzedana przemysłu	Inwestycje	Wartość Brutto środków trwałych	Płace	Stopa bezrobocia rejestrow. %	Podmioty gospo-darki
	Na mieszkańca w tys. zł		w tys. zł			
Pomorskie	33,63	3,90	88 496	3,79	10,42	25,85
Śląskie	42,67	3,95	220 977	3,89	8,84	37,42
Świętokrzyskie	18,64	2,25	34 817	3,29	13,8	9,22
Warmińsko -mazurskie	18,57	2,17	34 327	3,20	18,14	11,56
Wielkopolskie	37,63	3,60	147 190	3,45	7,56	34,73
Zachodniopo-morskie	19,54	2,82	58 055	3,48	15	19,03

Rozwiązanie:

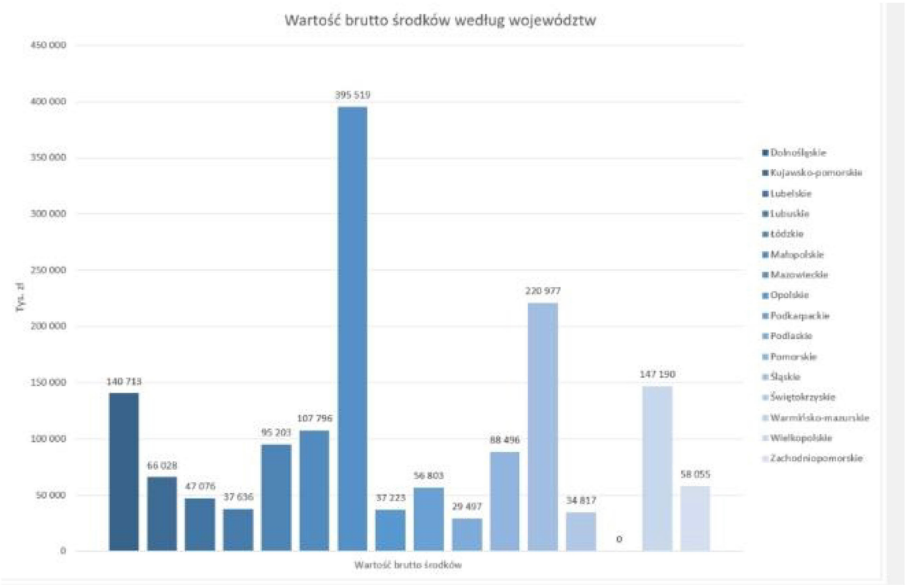
Krok 1.

W pierwszym kroku przenosimy dane do nowego arkusza EXCEL.

	A	B	C	D	E	F	G
1							
2	Wojewodstwo	Produkcja	Inwestycje	Wartość brutto środków	Place	Stopa bezrobocia	Podmioty gospodarski
3	Dolnośląskie	35,22	4,63	140 713	3,83	10,67	30,27
4	Kujawsko-pomorskie	23,66	2,96	66 028	2,25	15,04	17,6
5	Lubelskie	13,99	1,94	47 076	3,4	12,24	15,24
6	Lubuskie	26,26	2,84	37 636	3,26	12,93	10,08
7	Łódzkie	26,2	3,52	95 203	3,43	11,16	21,39
8	Małopolskie	22,78	2,76	107 796	3,55	9,09	32,03
9	Mazowieckie	43,45	6,01	395 519	4,72	8,81	61,95
10	Opolskie	22,31	3,72	37 223	3,46	11,64	7,09
11	Podkarpackie	17,33	2,45	56 803	3,21	14,14	14,4
12	Podlaskie	17,01	2,1	29 497	3,36	12,37	8,86
13	Pomorskie	33,63	3,9	88 496	3,79	10,42	25,85
14	Śląskie	42,67	3,95	220 977	3,89	8,84	37,42
15	Świętokrzyskie	18,64	2,25	34 817	3,29	13,8	9,22
16	Warmińsko-mazurskie	18,57	2,17	34 327	3,2	18,14	11,56
17	Wielkopolskie	37,63	3,6	147 190	3,45	7,56	34,73
18	Zachodniopomorskie	19,54	2,82	58 055	3,48	15	19,03

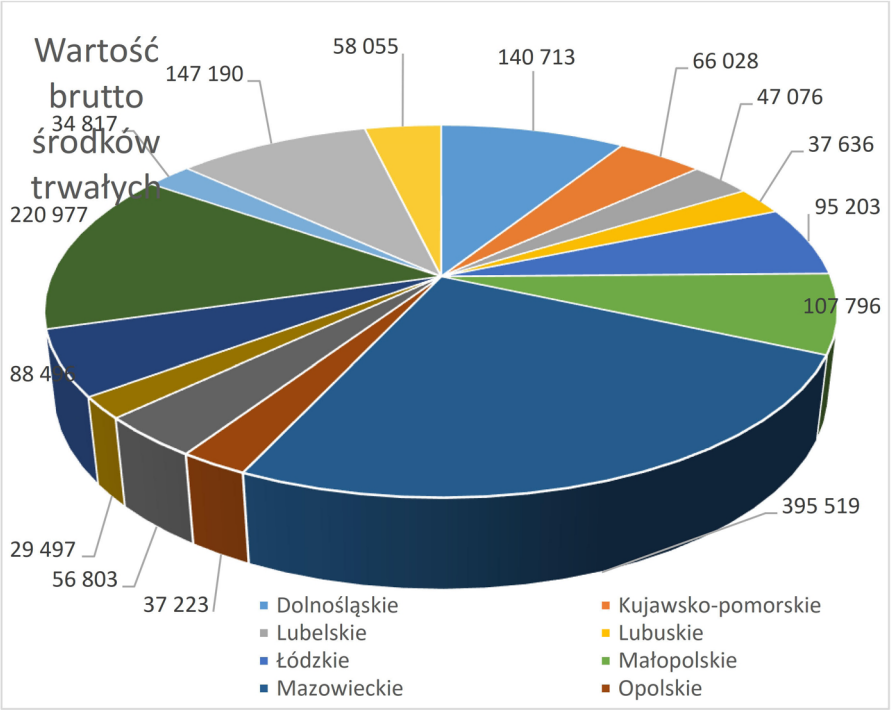
Krok 2.

Ponieważ nie mamy dokładnie sprecyzowane, które dane mamy zwizualizować, więc dokonamy kilku przykładowych prezentacji graficznych danych. Zaznaczmy kolumnę A oraz D z danymi, a następnie z Zakładki *Wstawianie* wybieramy wykres *Kolumnowy*. Otrzymany wykres formatujemy (dodajemy tytuły osi, legendę itd.)



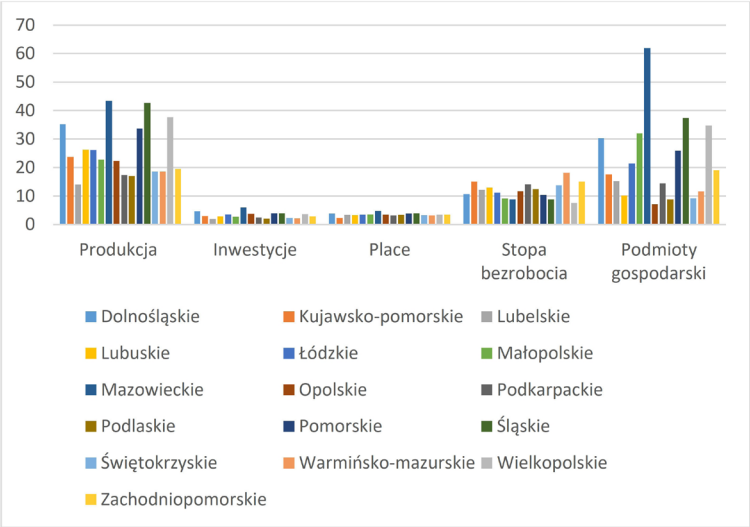
Krok 3.

Zaznaczamy kolumnę A oraz D z danymi, a następnie z zakładki *Wstawianie* wybieramy wykres *Kołowy 3-W*.



Krok 4.

Zaznaczamy wszystkie kolumny z danymi, a następnie z zakładki *Wstawianie* wybieramy wykres *Niestandardowy złożony*.

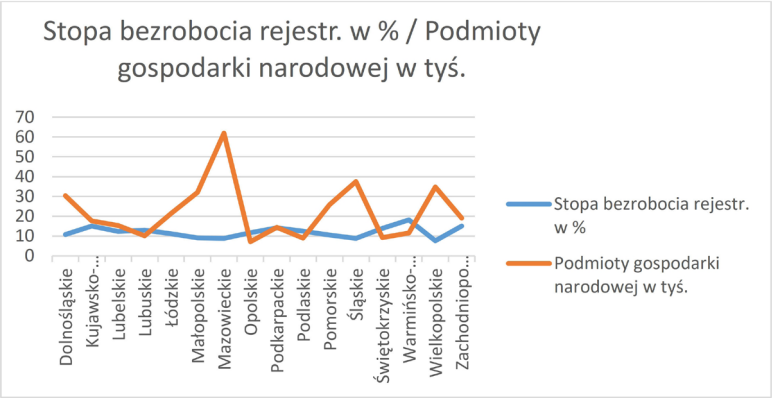


Krok 5.

Zaznaczamy kolumny A, B, C oraz E z danymi, a następnie z zakładki *Wstawianie* wybieramy wykres *Słupkowy grupowany*.

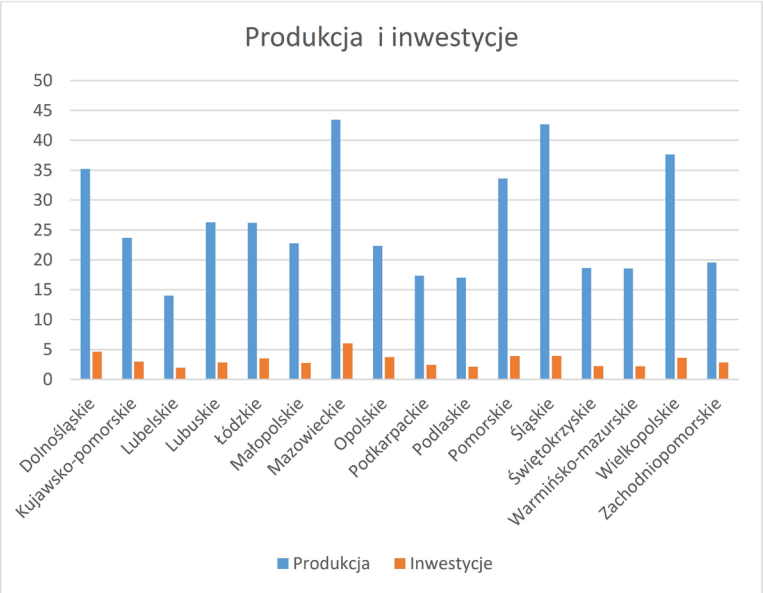
Krok 6.

Zaznaczamy kolumny A, F oraz G z danymi, a następnie z zakładki *Wstawianie* wybieramy *Wykres liniowy*.



Krok 7.

Zaznaczamy kolumny A, B oraz C z danymi, a następnie z zakładki *Wstawianie* wybieramy wykres *Kolumnowy grupowany*.



Przypadek 2

Polecenie:

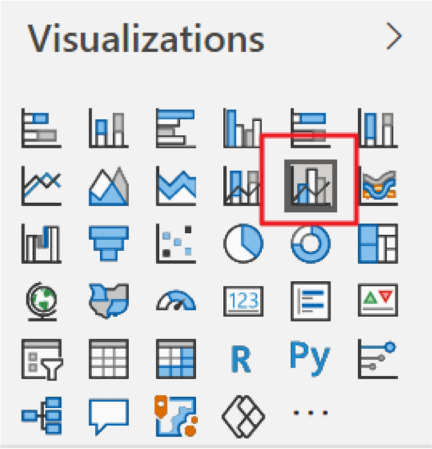
Dla danych z tabeli należy dokonać wizualizacji danych z wykorzystaniem oprogramowania POWER BI. Typ zastosowanego wykresu według uznania.

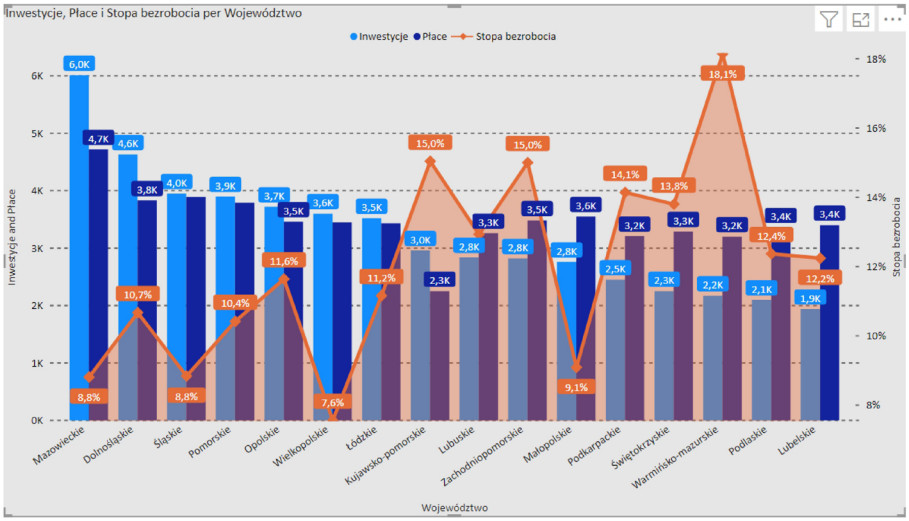
Województwo	Produkcja sprzedana przemysłu	Inwest.	Wartość Brutto Środków trwałych	Płace	Stopa Bezrobocia Rejestrow. %	Podmioty gospodarki
	Na mieszkańca W tys. zł		W tys. zł			
Dolnośląskie	35,22	4,63	140 713	3,83	10,67	30,27
Kujawsko-pomorskie	23,66	2,96	66 028	2,25	15,04	17,60
Lubelskie	13,99	1,94	47 076	3,40	12,24	15,24
Lubuskie	26,26	2,84	37 636	3,26	12,93	10,08
Łódzkie	26,20	3,52	95 203	3,43	11,16	21,39
Małopolskie	22,78	2,76	107 796	3,55	9,09	32,03
Mazowieckie	43,45	6,01	395 519	4,72	8,81	61,95
Opolskie	22,31	3,72	37 223	3,46	11,64	7,09
Podkarpackie	17,33	2,45	56 803	3,21	14,14	14,40
Podlaskie	17,01	2,10	29 497	3,36	12,37	8,86
Pomorskie	33,63	3,90	88 496	3,79	10,42	25,85
Śląskie	42,67	3,95	220 977	3,89	8,84	37,42
Świętokrzyskie	18,64	2,25	34 817	3,29	13,8	9,22
Warmińsko -mazurskie	18,57	2,17	34 327	3,20	18,14	11,56
Wielkopolskie	37,63	3,60	147 190	3,45	7,56	34,73
Zachodniopo-morskie	19,54	2,82	58 055	3,48	15	19,03

Rozwiązanie:

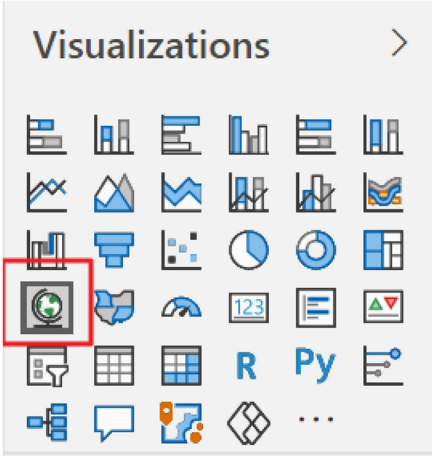
Krok 1.

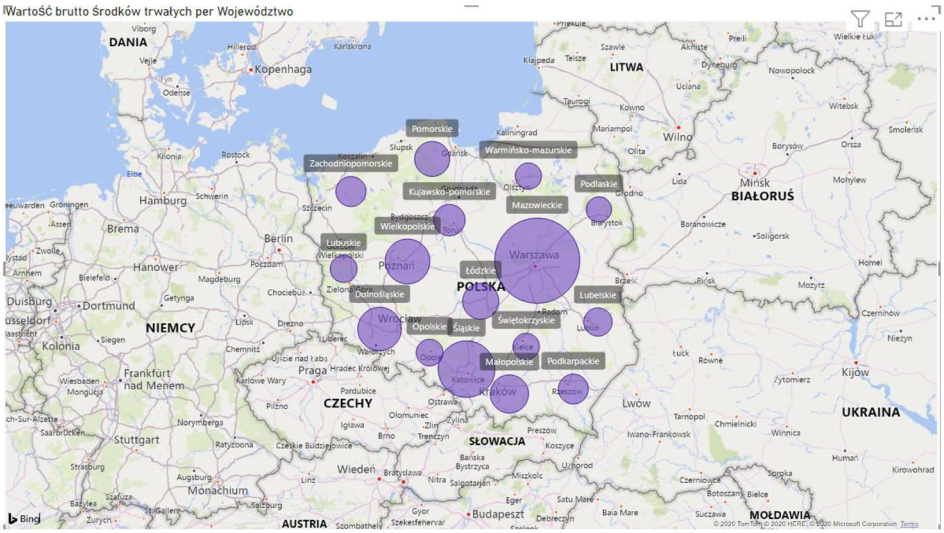
Korzystając z poniższych przykładowych typów wykresów, otrzymamy poniższe wizualizacje.



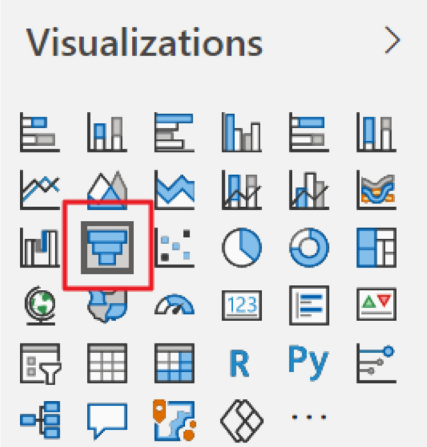


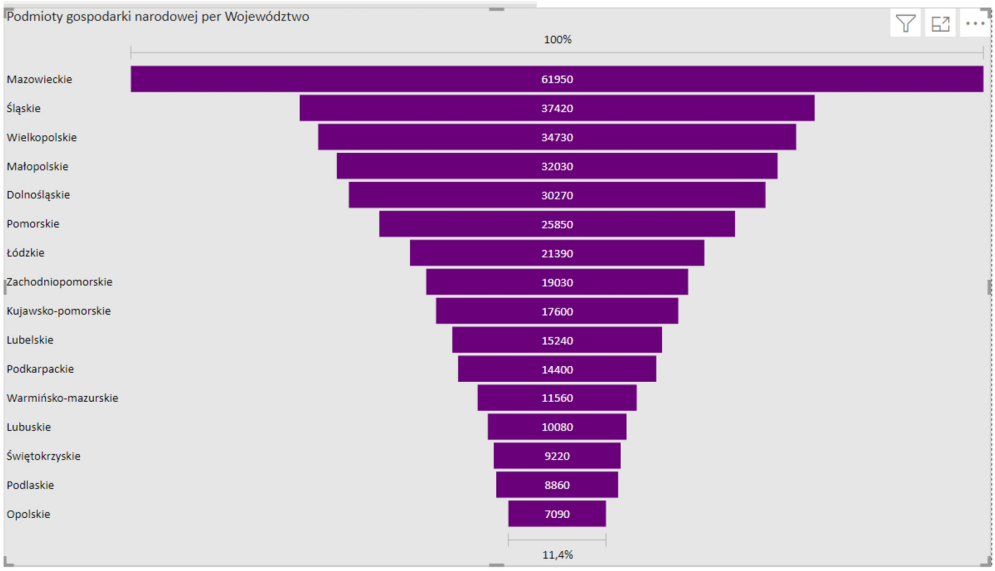
Krok 2.


Visualizations
123
R
Py

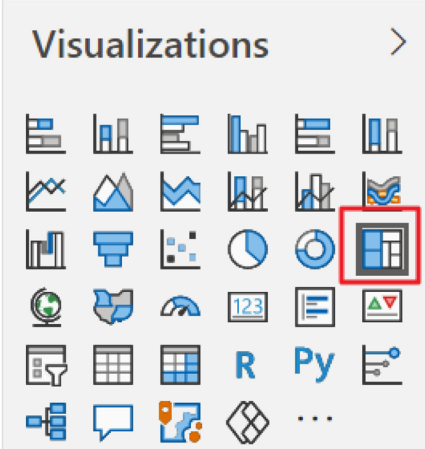

Wartość brutto środków trwałych per Województwo
Pomorskie
Warmińsko-mazurskie
Zachodniopomorskie
Kujawsko-pomorskie
Podlaskie
Mazowieckie
Wielkopolskie
Lubuskie
Łódzkie
Warszawa
Poznań
Dolnośląskie
Lubelskie
Opolskie
Śląskie
Świętokrzyskie
Małopolskie
Podkarpackie
Kraków
POLSKA
DANIA
LITWA
BIAŁORUŚ
NIEMCY
CZECHY
SŁOWACJA
UKRAINA
MOŁDAWIA
AUSTRIA
Bing

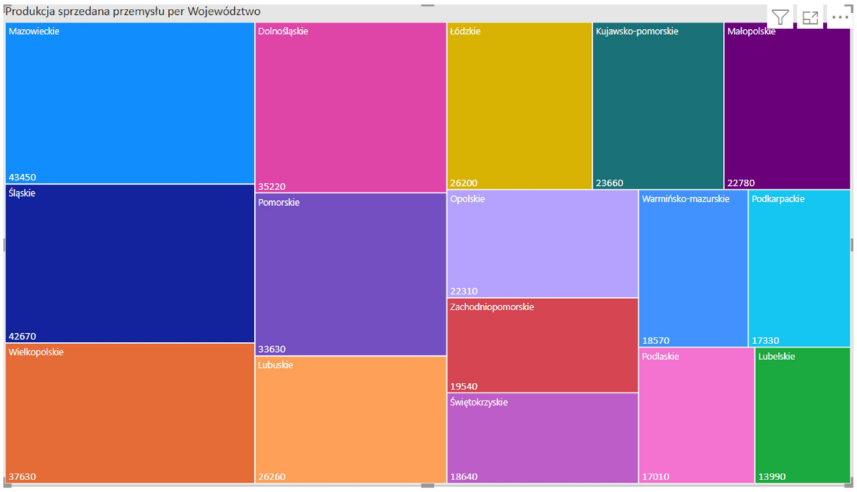
Krok 3.


Visualizations
123
R
Py


Podmioty gospodarki narodowej per Województwo
100%
Mazowieckie 61950
Śląskie 37420
Wielkopolskie 34730
Małopolskie 32030
Dolnośląskie 30270
Pomorskie 25850
Łódzkie 21390
Zachodniopomorskie 19030
Kujawsko-pomorskie 17600
Lubelskie 15240
Podkarpackie 14400
Warmińsko-mazurskie 11560
Lubuskie 10080
Świętokrzyskie 9220
Podlaskie 8860
Opolskie 7090
11,4%

Krok 4.





Przypadek 3

Polecenie:

Załącznik3.xlsx zawiera dane dotyczące rynku nieruchomości w Polsce. Dane są publikowane kwartalnie na stronach internetowych Narodowego Banku Polskiego. Opis zawartych danych znajduje się w zakładce *Opis danych*. Dla miasta Lublin (ceny ofertowe) dokonaj prezentacji graficznej danych.

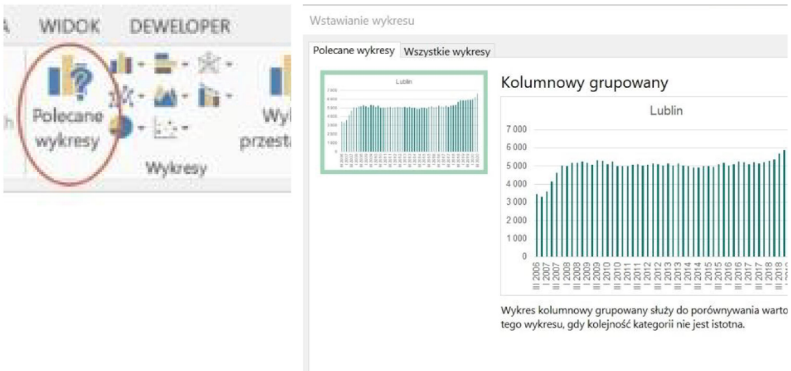
Rozwiązanie:

Krok 1.

Zaznaczamy kolumnę z zakresem czasowym (A8:A64) oraz z cenami mieszkań dla miasta Lublin (I8:I64).

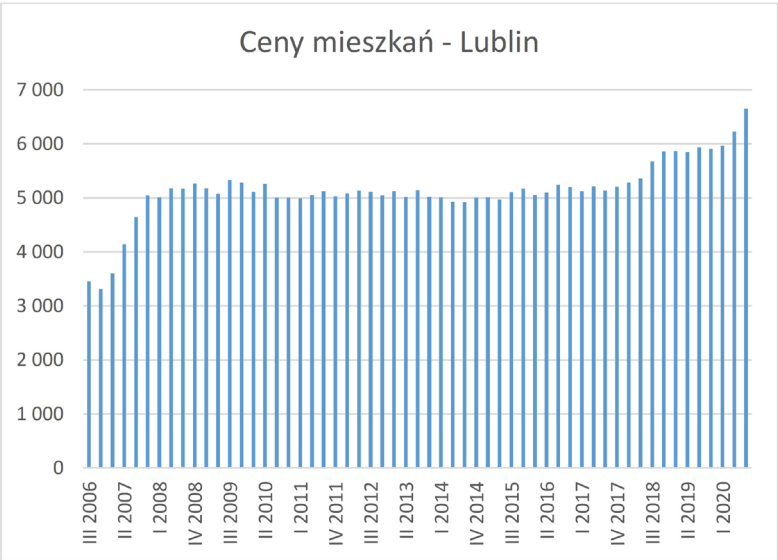
Krok 2.

Z zakładki *Wstawianie*, sekcja *Wykresy* wybieramy *Polecane wykresy*, a następnie wykres *Kolumnowy grupowy*.



Krok 3.

W otrzymanym wykresie formatujemy tytuł wykresu.



Przypadek 4

Polecenie:

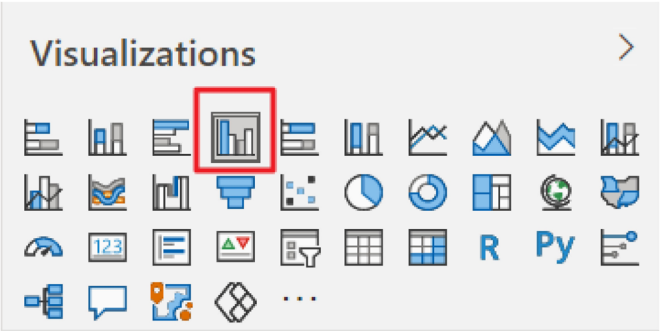
Dla danych z pliku Załącznik8.xlsx dokonaj prezentacji graficznej danych z wykorzystaniem programu Power BI.

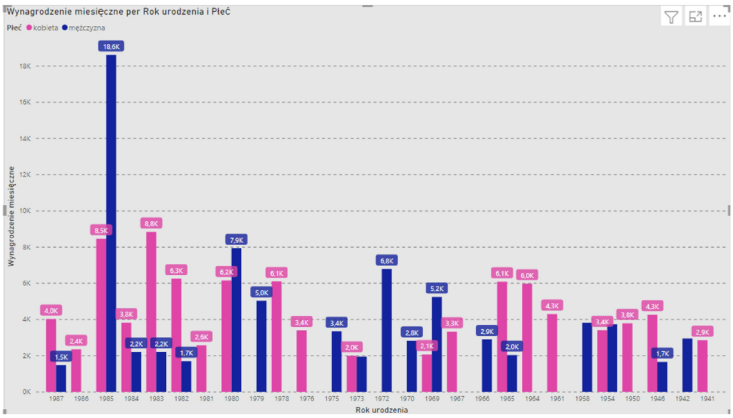
	A	B	C	D	E	F	G	H
1	lp	płeć	rok urodzenia	wykształcenie	ulubiony typ filmu	wynagrodzenie miesięczne	liczba dzieci	czas dojazdu do pracy [min]
2	1	kobieta	1978	podstawowe	horror	3169	0	23
3	2	mężczyzna	1979	zawodowe	przyrodniczy	3061	1	45
4	3	kobieta	1965	średnie	kryminał	2748	2	65
5	4	kobieta	1964	licencjat	film akcji	3763	1	92
6	5	mężczyzna	1980	inżynierskie	dramat	1585	2	31
7	6	kobieta	1964	magisterskie	komedia	2219	0	15
8	7	mężczyzna	1985	podstawowe	romans	2874	1	63
9	8	kobieta	1986	zawodowe	melodramat	2359	2	14
10	9	kobieta	1987	licencjat	horror	1007	2	88
11	10	mężczyzna	1985	licencjat	kryminał	2919	1	66
12	11	mężczyzna	1982	inżynierskie	przyrodniczy	1699	2	34

Rozwiązanie:

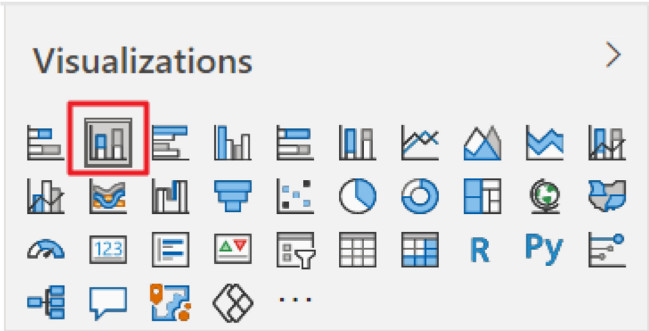
Wybierając wskazane poniżej metody prezentacji danych, możemy dokonać przykładowych wizualizacji danych.

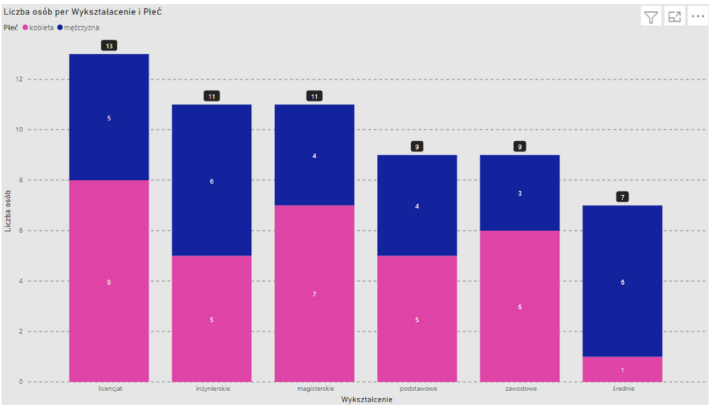
Krok 1.



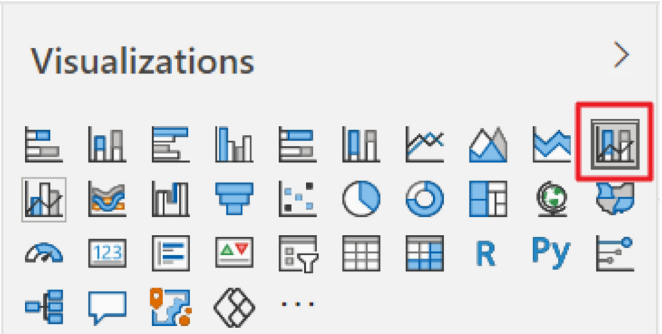

Wynagrodzenie miesięczne per Rok urodzenia i Płeć
Płeć kobieta mężczyzna
Wynagrodzenie miesięczne
Rok urodzenia

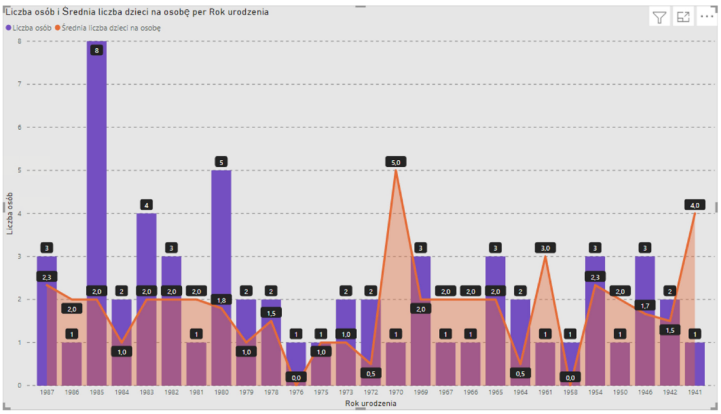
Krok 2.


Visualizations
R
Py

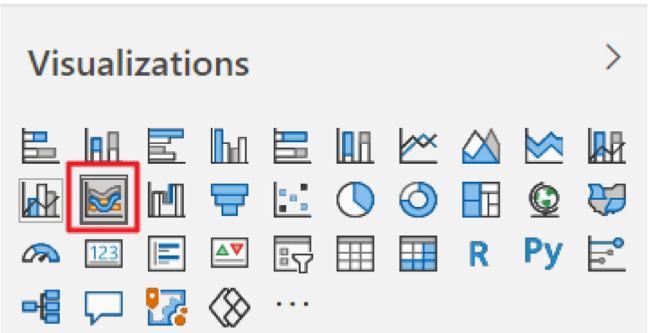

Liczba osób per Wykształcenie i Płeć
Płeć kobieta mężczyzna
Liczba osób
licencjat
inżynierskie
magisterskie
podstawowe
zawodowe
średnie
Wykształcenie

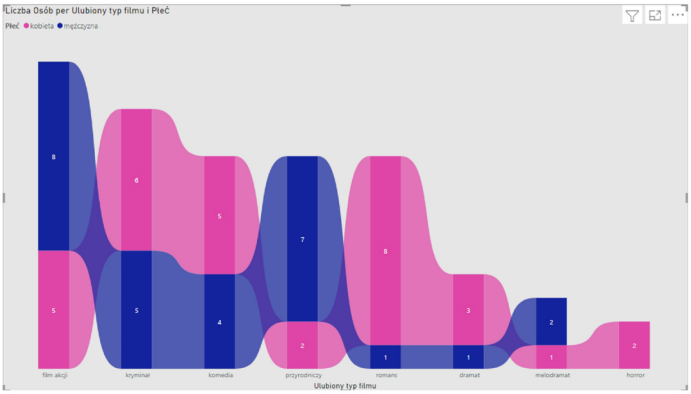
Krok 3.


Visualizations
R
Py

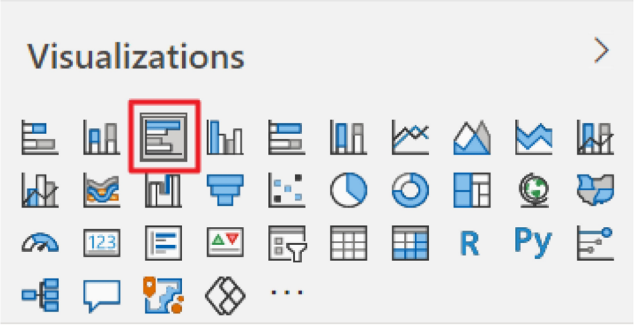

Liczba osób i Średnia liczba dzieci na osobę per Rok urodzenia
Liczba osób
Średnia liczba dzieci na osobę
Liczba osób
Rok urodzenia

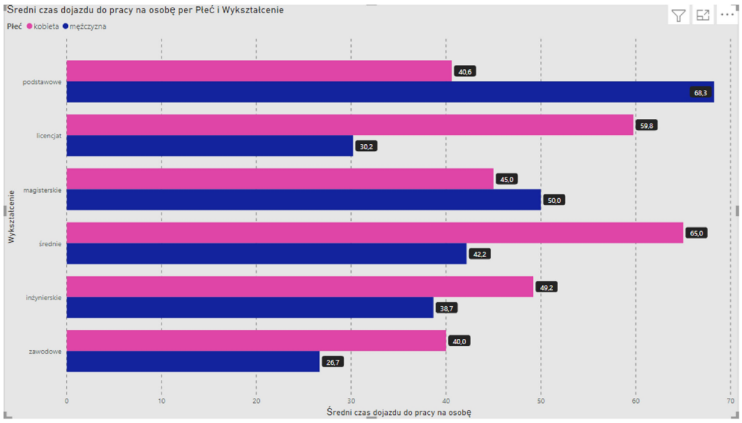
Krok 4.


Visualizations
R
Py


Liczba Osób per Ulubiony typ filmu i Płeć
Płeć kobieta mężczyzna
film akcji
kryminał
komedia
przygodniczy
romans
dramat
melodramat
horror
Ulubiony typ filmu

Krok 5.


Visualizations
R
Py


Średni czas dojazdu do pracy na osobę per Płeć i Wykształcenie
Płeć kobieta mężczyzna
podstawowe
licencjat
magisterskie
średnie
inżynierskie
zawodowe
Wykształcenie
40,6
68,3
59,6
30,2
45,0
50,0
65,0
42,2
49,2
38,7
40,0
26,7
0
10
20
30
40
50
60
70
Średni czas dojazdu do pracy na osobę

Zadania

Zadanie 1.

Zmierzono wzrost 80 studentek pewnego kierunku i otrzymano dane:

169, 153, 148, 154, 170, 157, 175, 159, 160, 165, 156, 160, 151, 174, 169, 186, 170, 159, 161, 173, 163, 164, 159, 158, 169, 155, 159, 148, 163, 161, 171, 171, 167, 180, 183, 162, 163, 177, 166, 159, 172, 162, 172, 161, 145, 162, 168, 155, 145, 154, 153, 167, 166, 153, 179, 162, 163, 162, 166, 168, 165, 163, 161, 177, 170, 149, 160, 164, 148, 154, 161, 172, 163, 149, 159, 155, 171, 156, 170, 154.

Zaproponuj graficzną prezentację powyższych danych.

.....
.....
.....
.....
.....
.....
.....

Zadanie 2.

Dane są obserwacje dotyczące zysku pewnego sklepu spożywczego (w tys. zł) w 20 kolejnych dniach roboczych:

16,4; 12,5; 14,4; 13,0; 14,4; 14,4; 13,9; 11,1; 16,5; 15,8; 13,0; 15,8; 18,5; 10,5; 16,4; 15,7; 16,5; 16,9; 11,1; 16,4.

Zaproponuj graficzną prezentację powyższych danych.

.....
.....
.....
.....
.....
.....
.....
.....
.....

Zadanie 3.

Przeprowadzono badanie sklepów branży meblarskiej ze względu na poziom rocznych zysków z działalności handlowej. Rozkład tych zysków ze względu na badaną cechę prezentuje tabelka:

Roczny zysk (w mln zł)	Liczba sklepów
0,5	10
1	17
1,5	12
2	8
2,5	3

Zaproponuj graficzną prezentację powyższych danych.

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Studium przypadku 2. Klasyfikacja danych

Cel ćwiczenia

Celem analizowanych case'ów jest dogłębne poznanie metod klasyfikacji danych oraz uzyskanie umiejętności stosowania nabytej wiedzy do analizy problemów ekonomicznych.

Wstęp teoretyczny

Szereg rozdzielczy szczegółowy to ciąg liczbowych wielkości statystycznych uporządkowanych według badanej cechy (rosnąco lub malejąco). Tworzy się poprzez grupowanie powtarzających się wartości badanej cechy w próbie.

Szeregi rozdzielcze dla cech mierzalnych dzielą się na:

- szeregi punktowe – stosowane, gdy liczba wariantów cechy jest niewielka,
- szeregi z przedziałami klasowymi – stosowane, gdy liczba wariantów badanej cechy jest duża.

Szereg rozdzielczy składa się z dwóch kolumn:

- w pierwszej kolumnie znajdują się warianty badanej cechy,
- w drugiej kolumnie liczba jednostek zbiorowości statystycznej, która posiada dany wariant cechy.

Tworzenie szeregu rozdzielczego przedziałowego zaleca się budować w sytuacji, gdy badana cecha jest ciągła albo jest cechą skokową o wielu wariantach. Wówczas na wstępie warianty cechy grupuje się w przedziały klasowe, a zadaniem badacza jest zakwalifikowanie każdej jednostki statystycznej do odpowiedniego przedziału.

Etapy tworzenia szeregu rozdzielczego przedziałowego:

1. wyznaczamy oraz
2. wyznaczamy rozstęp z próby
3. wyznaczamy ilość przedziałów klasowych K
4. wyznaczamy długość przedziału klasowego .
5. wyznaczamy lewy koniec pierwszego przedziału klasowego gdzie α jest dokładnością pomiaru

Tabela 1. Reguły ustalania liczby przedziałów klasowych.

Liczba pomiarów n	Liczba przedziałów klasowych K
30 – 60	6 – 8
60 – 100	7 – 10
100 – 200	9 – 12
200 – 500	11 – 17
500 – 1500	16 – 25

Wykorzystywane funkcje

Funkcja	Opis
CZĘSTOŚĆ (tablica_dane; tablica_przedziały)	Funkcja CZĘSTOŚĆ oblicza częstość występowania wartości w zakresie wartości, a następnie zwraca pionową tablicę liczb. Ponieważ funkcja CZĘSTOŚĆ zwraca tablicę, musi być wprowadzona jako formuła. W składni funkcji CZĘSTOŚĆ występują następujące argumenty: tablica_dane – tablica lub odwołanie do zbioru wartości, dla których mają być zliczane częstości. Jeśli argument tablica_dane nie ma żadnych wartości, funkcja CZĘSTOŚĆ zwraca tablicę zer tablica_przedziały – tablica lub odwołanie do przedziałów, w których mają być grupowane wartości argumentu tablica_dane. Jeśli argument tablica_przedziały nie zawiera żadnych wartości, funkcja CZĘSTOŚĆ zwraca liczbę elementów w argumencie tablica_dane.
MIN(liczba1; [liczba2];...)	Zwraca najmniejszą liczbę w zbiorze wartości. W składni funkcji MIN występują następujące argumenty: Liczba1; liczba2;... Argument liczba1 jest wymagany, pozostałe są opcjonalne. Od 1 do 255 argumentów, dla których należy wyznaczyć wartość minimalną.
MAX(liczba1; [liczba2];...))	Zwraca największą wartość w zbiorze wartości. W składni funkcji MAX występują następujące argumenty: Liczba1; liczba2;... Argument liczba1 jest wymagany, pozostałe są opcjonalne. Od 1 do 255 argumentów, dla których należy wyznaczyć wartość minimalną.

Przypadek 1

Polecenie:

Studenci WSEI zostali poproszeni o podanie swojej wagi. Wyniki ankiety znajdują się w pliku Załącznik2.xlsx. Pogrupuj dane w postaci szeregu rozdzielczego przedziałowego oraz przedstaw interpretację graficzną grupowania danych.

Rozwiązanie:
Do rozwiązania tego zadania wykorzystamy jedną z funkcji Excela CZĘSTOŚĆ(). Aby przedstawić dane w postaci szeregu, będziemy wykorzystywać procedurę opisaną we Wstępie teoretycznym.

Krok 1.

Obliczamy wartość minimalną z danych, wpisując w komórce F2 formułę =MIN(A2:A143)

Krok 2.

Obliczamy wartość maksymalną z danych, wpisując w komórce G2 formułę =MAX(A2:A143)

Krok 3.

Obliczamy wartość rozstępu z danych, wpisując w komórce G2 formułę =G2-F2

Krok 4.

Zgodnie z Tabelą 1 przyjmij K=10

Krok 5.

Obliczamy długość przedziału klasowego, wpisując w komórce J2 formułę =H2/I2. Wyniki dotychczasowych analiz są widoczne na rysunku poniżej.

F	G	H	I	J
Min	Max	Rozstęp	K	h
37	150	113	10	11

Krok 6.

Wyznacz końce przedziałów.

Krok 7.

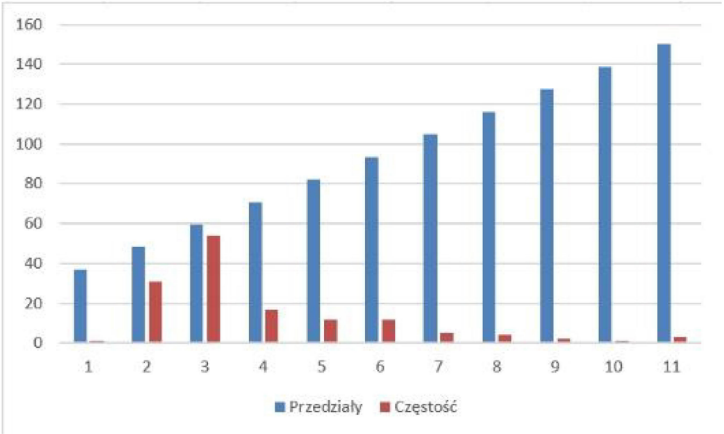
Zaznacz zakres komórek, który mają zająć wyniki. W tym przypadku jest to zakres komórek D2:D12. Wpisz formułę (ale jeszcze jej nie zatwierdzaj): =CZĘSTOŚĆ(A2:A101;C2:C8).

Po wpisaniu formuły przytrzymaj równocześnie naciśnięte klawisze *Ctrl* oraz *Shift* i naciśnij klawisz *Enter*. Następnie zwolnij wszystkie trzy klawisze. Ta sekwencja klawiaturowa powiadamia Excel, że chcesz, aby interpretował tę formułę jako formułę tablicową. Wyniki dotychczasowych analiz są widoczne na poniższym rysunku.

C	D
Przedziały	Częstość
37	1
48	31
60	54
71	17
82	12
94	12
105	5
116	4
127	2
139	1
150	3

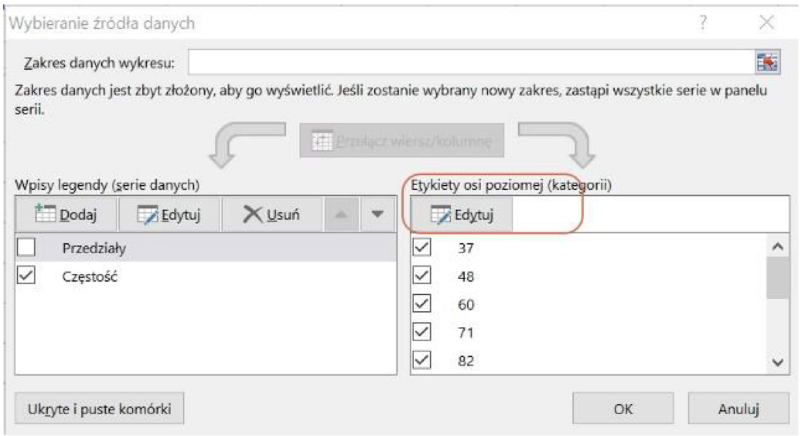
Krok 8.

Teraz przejdziemy do graficznej prezentacji danych. Zaznacz dane, które chcesz wykreślić — czyli zakres C1:D12. Kliknij kartę *Wstawianie*, a następnie przycisk *Polecane wykresy* w grupie *Wykresy*. Wybierz typ wykresu *Kolumnowy grupowany* z wykresów *2-W*. Reprezentacja graficzna jest widoczna na rysunku poniżej.



Krok 9.

Popraw wykres, klikając przycisk *Zaznacz dane* na karcie *Projektowanie*, która jest widoczna, gdy wykres jest aktywny. Pojawi się okno dialogowe pokazane na rysunku poniżej.



Kliknij przycisk *Edytuj* w sekcji *Etykiety osi poziomej (kategorii)*. Pojawi się nowe okno dialogowe *Etykiety osi*. Zaznacz komórki C2:C12, aby ustanowić ten zakres jako źródło etykiet dla osi poziomej. Kliknij przycisk *OK*. Kliknij etykietę Przedziały w lewym polu listy.

Następnie kliknij przycisk *Usuń*, aby usunąć ją jako serię z wykresu, i przycisk *OK*, aby zamknąć okno. Teraz otrzymasz interpretację graficzną jak na rysunku poniżej.

Przypadek 2

Polecenie:

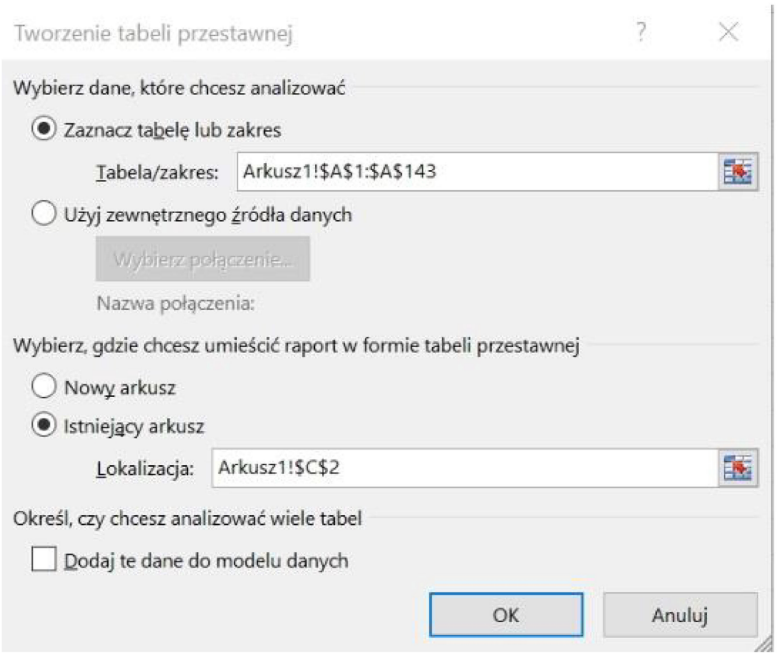
Studenci WSEI zostali poproszeni o podanie swojej wagi. Wyniki ankiety znajdują się w pliku Załącznik2.xlsx. Pogrupuj dane w postaci szeregu rozdzielczego przedziałowego oraz przedstaw interpretację graficzną grupowania danych. Do analiz zastosuj narzędzie Tabela przestawna programu EXCEL.

Rozwiązanie:

Krok 1.

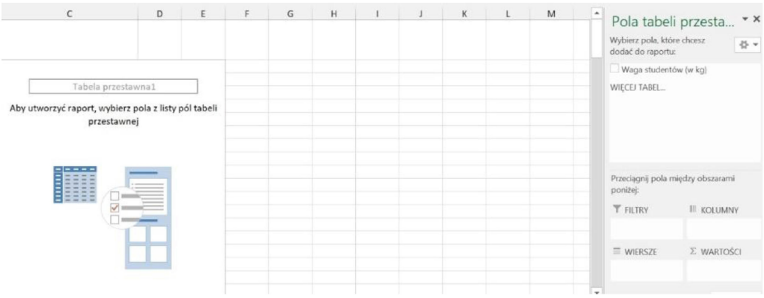
Zaznacz dowolną komórkę w zakresie A2:A143. Kliknij kartę *Wstawianie*, a następnie *Tabela przestawna.*

Pojawi się okno dialogowe widoczne na rysunku poniżej.



Krok 2.

Kliknij przycisk opcji *Istniejący arkusz*. Następnie kliknij pole edycji zakresu *Lokalizacja*, a potem pustą komórkę w arkuszu, która po prawej stronie i poniżej ma inne puste komórki. Kliknij przycisk *OK*. Teraz arkusz będzie wyglądać jak na rysunku poniżej.



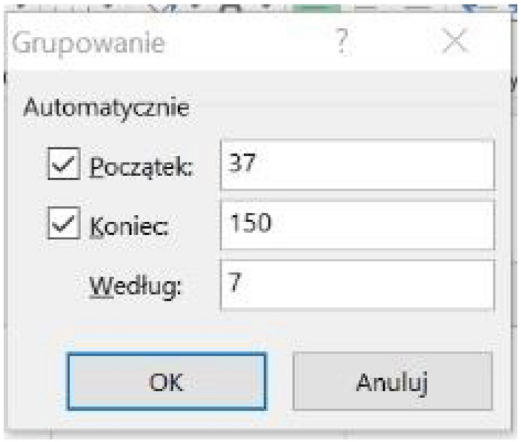
Krok 3.

Kliknij pole *Waga studentów* na liście pól tabeli przestawnej i przeciągnij je do obszaru *Wiersze*. Tabela przestawna będzie wyglądała jak na rysunku poniżej.

Etykiety wierszy	Liczba z Waga studentów (w kg)
37-46	22
47-56	54
57-66	21
67-76	8
77-86	11
87-96	15
97-106	3
107-116	2
117-126	2
127-136	1
137-146	2
147-156	1
Suma końcowa	**142**

Krok 4.

Kliknij prawym przyciskiem myszy dowolną komórkę zawierającą etykietę wiersza. Wybierz polecenie *Grupuj* z menu podręcznego – patrz Rysunek poniżej.

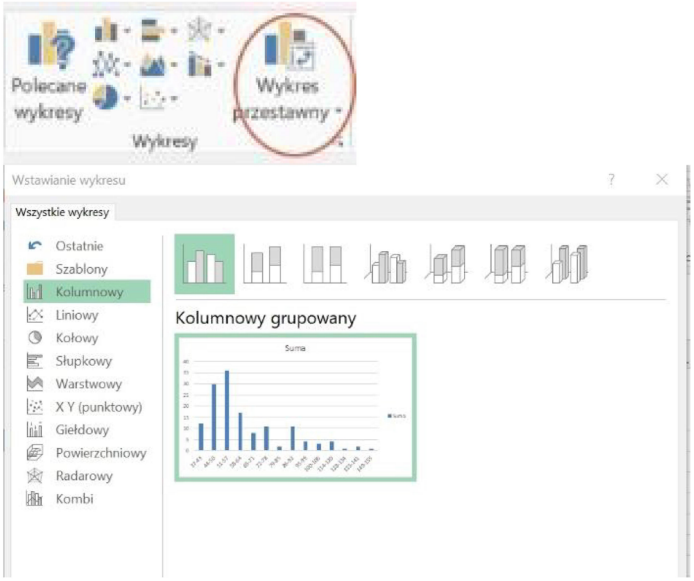


W tym kroku możemy dokonać grupowania szeregu ze względu na długość klasy. W naszym zadaniu ustawiamy długość klasy równą 7. Wyniki grupowania są widoczne na rysunku poniżej.

Etykiety wierszy	Liczba z Waga studentów (w kg)
37-43	12
44-50	30
51-57	36
58-64	17
65-71	8
72-78	11
79-85	2
86-92	11
93-99	4
100-106	3
114-120	4
128-134	1
135-141	2
149-155	1
Suma końcowa	**142**

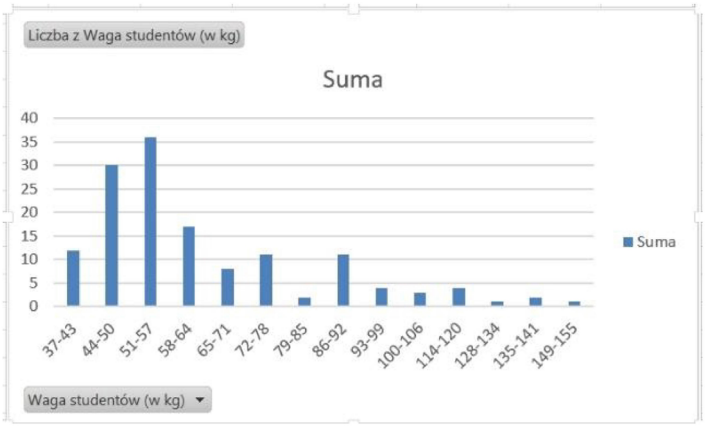
Krok 5.

W dalszym etapie przejdziemy do wizualizacji danych. Z zakładki *Wstawianie/Wykresy* wybieramy *Wykres przestawny*, a następnie *Kolumnowy grupowany*.



Krok 6.

Wykres widoczny jest na rysunku poniżej.



Przypadek 3

Polecenie:

Dwudziestoosobowa grupa studentów została poproszona o podanie liczby dzieci. Dane znajdują się w tabeli oraz pliku Załącznik6.xlsx.

Nr	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
Liczba dzieci	0	2	1	4	2	0	3	4	2	2	3	3	5	1	3	3	5	0	4	3

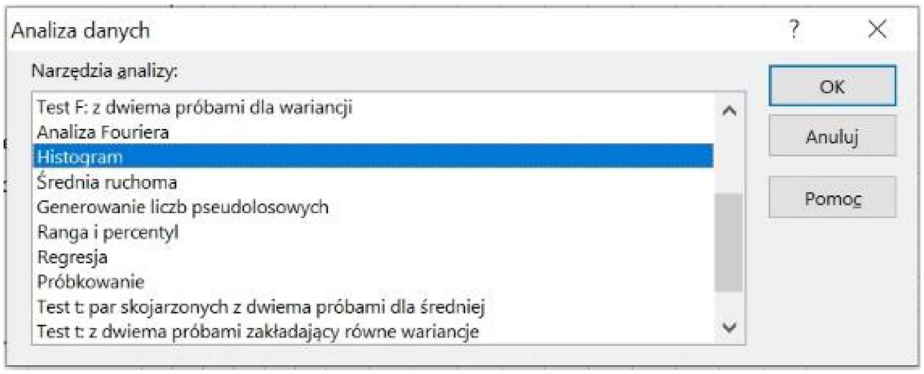
Utwórz szereg rozdzielczy, oblicz liczebności skumulowane, częstości względne i częstości względne skumulowane. Utwórz histogram.

Rozwiązanie:

Poniżej sposób wykonania zadania w programie EXCEL, z wykorzystaniem dodatku *Analiza danych.*

Krok 1.

Wczytaj dane z pliku Załącznik6.xlsx. Wybierz z menu *Dane/Analiza/Histogram*.



Krok 2.

W oknie zaznacz zakres danych. Zaznacz opcję Łączny udział procentowy oraz *Wykres wyjściowy*.

Krok 3.

Zostanie utworzony szereg rozdzielczy zawierający: wartości zmiennych, liczebności oraz częstości względne skumulowane.

Zbiór danych (koszyk)	*Częstość*	*Łączna wartość %*
0	3	15,00%
1	2	25,00%
3	4	45,00%
4	6	75,00%
Więcej	5	100,00%

Krok 4.

Dodatkowo zostanie utworzony wykres składający się z: *histogramu liczebności* oraz *wykresu liniowego częstości względnych skumulowanych*.

Krok 5.

Aby obliczyć liczebności względne, częstości względne oraz częstości skumulowane, należy dodać do tabeli kolejne kolumny (patrz rysunek poniżej).

Zbiór danych (koszyk)	*Częstość*	*Łączna wartość %*	*Liczebność skumulowana*	*Częstość względna*	*Częstość względna skumulowana*
0	3	15,00%	3	0,15	0,15
1	2	25,00%	5	0,10	0,25
3	4	45,00%	9	0,20	0,45
4	6	75,00%	15	0,30	0,75
Więcej	5	100,00%	20	0,25	1,00

Przypadek 4

Polecenie:

Dane są obserwacje dotyczące zysku pewnego sklepu meblowego (w tys. zł) w 20 kolejnych dniach roboczych:

16,4; 12,5; 14,4; 13,0; 14,4; 14,4; 13,9; 11,1; 16,5; 15,8; 13,0; 15,8; 18,5; 10,5; 16,4; 15,7; 16,5; 16,9; 11,1; 16,4.

Przedstaw obserwacje statystyczne w postaci: szeregu rozdzielczego punktowego oraz szeregu rozdzielczego przedziałowego.

Rozwiązanie:

Krok 1.

Rozpoczynamy od graficznej reprezentacji danych.

Wyznaczamy wartość minimalną i maksymalną z danych podanych w tabeli za pomocą funkcji *min()* i *max()*, wpisując odpowiednią formułę w wierszu.

Krok 2.

Wyznaczamy *Rozstęp*, odejmując wartość minimalną od maksymalnej.

Krok 3.

Wyznaczamy wartość *h* ze wzoru *h=Rozstęp/K*

min	max	rozstęp	K	h
10,5	18,5	8	3	2,7

Krok 4.

W kolejnej części zadania wyznaczamy *Przedziały*. Pierwsza wartość równa jest wartości minimalnej wyliczonej w Kroku 1. Kolejne wartości wyliczamy kolejno: poprzedzająca wartość + *h*, formułę możemy „przeciągnąć" = powtórzyć dla pozostałych wartości, do momentu, gdy przedział przekroczy wartość maksymalną wyliczoną w punkcie 1.

Krok 5.

Wyznaczamy *Częstość,* korzystając z funkcji Częstość, kolejno wybierając tabelę z danymi z zadania, następnie tabelę z *Przedziałami* wyznaczonymi w punkcie 4. WAŻNE, by nie wybierać na tym etapie samego klawisza *Enter*, poprawne będzie wybranie kombinacji klawiszy *Ctrl* + *Shift* + *Enter*. Utworzona została tablica danych.

Przedziały	Częstość
10,5	1,00
13,2	5,00
15,8	7,00
18,5	7,00

Krok 6.

Sumujemy wartości z tabeli Częstość, żeby zweryfikować, czy dana ilość pokrywa się z ilością danych z treści zadania. Częstość możemy wyrazić za pomocą wartości procentowych.

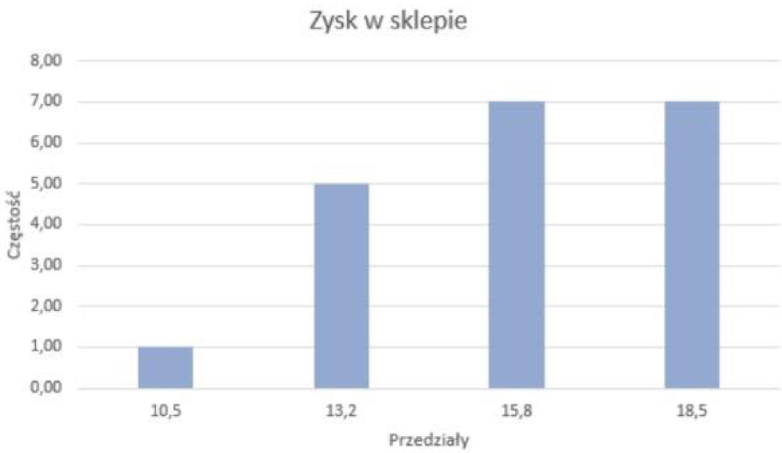
Krok 7.

Wyznaczamy środki klas poprzez wybranie kolejno dwóch przedziałów i podzielenie ich sumy przez 2.

Przedziały	Częstość	Środki klas	Częstość
10,5	1,00	12	5%
13,2	5,00	15	25%
15,8	7,00	17	35%
18,5	7,00	20	35%

Krok 8.

Wstawiamy wykres, który obrazuje częstość występowania konkretnych przedziałów kwot zysku w czasie 20 dni roboczych.



Zadania

Zadanie 1.

Zmierzono wzrost 80 studentek pewnego kierunku w Wyższej Szkole Ekonomii i Innowacji i otrzymano dane:

169, 153, 148, 154, 170, 157, 175, 159, 160, 165, 156, 160, 151, 174, 169, 186, 170, 159, 161, 173, 163, 164, 159, 158, 169, 155, 159, 148, 163, 161, 171, 171, 167, 180, 183, 162, 163, 177, 166, 159, 172, 162, 172, 161, 145, 162, 168, 155, 145, 154, 153, 167, 166, 153, 179, 162, 163, 162, 166, 168, 165, 163, 161, 177, 170, 149, 160, 164, 148, 154, 161, 172, 163, 149, 159, 155, 171, 156, 170, 154.

Określ zbiorowość statystyczną oraz badaną cechę. Zbuduj na podstawie uzyskanych danych szereg przedziałowy.

.....
.....
.....
.....
.....

Zadanie 2.

Liczba podręczników zakupionych podczas roku akademickiego przez studentów grupy liczącej 20 osób była następująca:
0, 1, 1, 6, 5, 3, 0, 2, 1, 3, 5, 3, 0, 4, 1, 6, 4, 3, 2, 5.

Określ zbiorowość statystyczną oraz badaną cechę. Zbuduj na podstawie uzyskanych danych szereg rozdzielczy punktowy.

.....
.....
.....
.....
.....

Studium przypadku 3.
Analiza struktury

Cel ćwiczenia

Poznanie metod i nabycie praktycznych umiejętności rozwiązywania problemów mieszczących się w zakresach wyznaczania podstawowych parametrów statystyki opisowej oraz interpretacji ich wyników wraz z graficzną prezentacją.

Wstęp teoretyczny

Analiza struktury polega na określeniu i zinterpretowaniu prawidłowości występujących w budowie zbiorowości z punktu widzenia określonej cechy statystycznej. Podstawowym miernikiem statystycznym wykorzystywanym w analizie struktury jest wskaźnik struktury.

Wskaźniki struktury możemy ogólnie podzielić na dwa typy:

parametryczne (klasyczne) – obliczane na podstawie wszystkich obserwacji (wszelkie odmiany średniej, odchylenie standardowe, skośność, kurtoza),

nieparametryczne – wyznaczane na podstawie ich miejsca w szeregu lub częstotliwości występowania (moda, mediana, kwartyle, kwantyle, centyle).

W zależności od rodzaju prezentacji danych w postaci szeregu (prostego, punktowego lub przedziałowego) stosujemy odpowiedni **wzór na obliczenie wskaźników struktury** – patrz tabela poniżej.

miara	rodzaje szeregów		
	szczegółowy prosty	**rozdzielczy punktowy**	**rozdzielczy przedziałowy**
Wskaźnik struktury	$\omega_i = \frac{n_i}{N}$ $\omega_i = \frac{n_{ik}}{N}$		
MIARY KLASYCZNE			
Klasyczne miary położenia			
średnia arytmetyczna	$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{k} x_i n_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{N} x_i n_i}{N}$
średnia harmoniczna	$\overline{X}_H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$	$\overline{X}_H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$	
średnia geometryczna	$\overline{X}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$		
Klasyczne miary zmienności (rozproszenia, dyspersji)			
wariancja	$s^2 = \frac{\sum_{i=1}^{N} \left(x_i - \overline{x}_i\right)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x}_i\right)^2 \cdot n_i}{N}$	$s^2 = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x}_i\right)^2 \cdot n_i}{N}$
odchylenie standardowe	$s = \sqrt{s^2}$		

obszar zmienności	$\bar{x} - s < x_{typ} < \bar{x} + s$		
odchylenie przeciętne	$d = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{N}$	$d = \frac{\sum_{i=1}^{k} \lvert x_i - \bar{x} \rvert \cdot n_i}{N}$	$d = \frac{\sum_{i=1}^{k} \lvert x_i - \bar{x} \rvert \cdot n_i}{N}$
współczynnik zmienności	$V_s = \frac{S}{\bar{x}}$		
Klasyczne miary asymetrii			
współczynnik asymetrii	$A_s = \frac{\bar{x} - D}{S}$		
MIARY POZYCYJNE			
Pozycyjne miary położenia			
dominanta	$D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D$		
mediana	$Me = \begin{cases} x_{\frac{n+1}{2}}, \text{ gdy } n \text{ jest nieparzyste} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) \text{ gdy } n \text{ jest parzyste} \end{cases}$	$Me = x_{Me} + \frac{h}{n_{Me}}\left(\frac{n}{2} - \sum_{i=1}^{k-1} n_i\right)$	
kwartyl 1		$Q_1 = X_{Q_1} + \frac{\frac{N}{4} - cum_{Q_1 - 1}}{n_{Q_1}} \cdot h_{Q_1}$	
kwartyl 3		$Q_3 = X_{Q_3} + \frac{\frac{3}{4}N - cum_{Q_3 - 1}}{n_{Q_3}} \cdot h_{Q_3}$	
Pozycyjne miary zmienności			
rozstęp	$R = X_{max} - X_{min}$		

odchylenie ćwiartkowe	$Q = \frac{(Q_3 - M) + (M - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$
typowy obszar zmienności	$M - Q < X_{typ} < M + Q$
współczynniki zmienności	$V_Q = \frac{Q}{M}$ $\quad V_{Q_{1Q_3}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$
Pozycyjne miary asymetrii	
współczynnik skośności	$A_Q = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} = \frac{Q_1 + Q_3 - 2M}{2Q}$ $\left(\frac{Q_3 - Q_1}{2}\right)$
równanie Pearsona	$2\overline{X} = 3M - D$

Wykorzystywane funkcje

Funkcja	Opis
INDEKS(tablica;nr_wiersza;[nr_kolumny])	Funkcja INDEKS zwraca wartość lub odwołanie do wartości, która znajduje się w tabeli lub zakresie. W formie tablicowej funkcji indeks występują następujące argumenty: tablica – argument wymagany. Zakres komórek lub stała tablicowa. nr_wiersza – wymagane, chyba że column_num jest obecny. Wskazuje wiersz tablicy, z którego zostanie zwrócona wartość. Jeśli row_num zostanie pominięty, wymagana jest column_num. nr_kolumny – argument opcjonalny. Wskazuje kolumnę tablicy, z której zostanie zwrócona wartość. Jeśli nr_kolumny zostanie pominięty, wymagana jest nr_wiersza.

Funkcja	Opis
JEŻELI(test_logiczny; wartość_jeżeli_prawda; [wartość_jeżeli_fałsz])	Funkcja umożliwia tworzenie logicznych porównań wartości z oczekiwanymi wynikami. W składni funkcji występują następujące wymagane argumenty: test_logiczny – warunek, który ma zostać sprawdzony wartość_jeżeli_prawda – wartość, która ma zostać zwrócona, jeśli wynik argumentu test_logiczny ma wartość PRAWDA. wartość_jeżeli_fałsz – wartość, która ma zostać zwrócona, jeśli wynik argumentu test_logiczny ma wartość FAŁSZ.
KWARTYL.PRZED. ZAMK(tablica;kwartyl)	Zwraca kwartyl zbioru danych na podstawie wartości percentylu z przedziału domkniętego od 0 do 1. W składni funkcji KWARTYL.PRZEDZ.ZAMK występują następujące argumenty: Tablica – argument wymagany. Tablica lub zakres komórek z wartościami liczbowymi, dla których ma zostać obliczona wartość kwartylu. Kwartyl – argument wymagany. Wskazuje, która wartość ma zostać zwrócona.
MAX(liczba1; [liczba2];...)))	Zwraca największą wartość w zbiorze wartości. W składni funkcji MAX występują następujące argumenty: Liczba1; liczba2;... Argument liczba1 jest wymagany, pozostałe są opcjonalne. Od 1 do 255 argumentów, dla których należy wyznaczyć wartość minimalną.
MIN(liczba1; [liczba2];...)))	Zwraca najmniejszą liczbę w zbiorze wartości. W składni funkcji MIN występują następujące argumenty: Liczba1; liczba2;... Argument liczba1 jest wymagany, pozostałe są opcjonalne. Od 1 do 255 argumentów, dla których należy wyznaczyć wartość minimalną.
MEDIANA(liczba1; [liczba2];...)	Zwraca wartość mediany dla podanych liczb. Mediana jest liczbą w środku zbioru liczb. W składni funkcji MEDIANA występują następujące argumenty: Liczba1; liczba2;... – argument liczba1 jest wymagany, pozostałe są opcjonalne. Od 1 do 255 argumentów, dla których należy wyznaczyć wartość mediany.

Funkcja	Opis
ODCH.STANDARD. PRÓBKI(liczba1; [liczba2];...)	Funkcja szacuje odchylenie standardowe próbki, ignorując wartości logiczne i tekst. Odchylenie standardowe jest miarą szerokości rozproszenia wartości od wartości średniej. W składni funkcji ODCH.STANDARD.PRÓBKI występują następujące argumenty: Liczba1 – argument wymagany. Pierwszy argument liczbowy odpowiadający próbce populacji. Zamiast argumentów rozdzielonych średnikami można użyć pojedynczej tablicy lub odwołania do tablicy. Liczba2;... – argument opcjonalny. Od 2 do 254 argumentów liczbowych odpowiadających próbce populacji. Zamiast argumentów rozdzielonych średnikami można użyć pojedynczej tablicy lub odwołania do tablicy.
PODAJ.POZYCJĘ (szukana_wartość; przeszukiwana_tablica)	W składni funkcji PODAJ.POZYCJĘ występują następujące argumenty: szukana_wartość – argument wymagany. Wartość, która ma być zgodna w przeszukiwana_tablica. Na przykład podczas wyszukiwania numeru innej osoby w książce telefonicznej jako wartości odnośnika jest używana nazwa osoby, ale numer telefonu jest odpowiednią wartością. Argument szukana_wartość może być wartością (liczbą, tekstem lub wartością logiczną) lub odwołaniem do komórki liczby, tekstu lub wartości logicznej. Przeszukiwana_tablica – argument wymagany. Zakres przeszukiwanych komórek.
SUMA(liczba1; [liczba2];...)	Funkcja Suma dodaje wartości. Możesz dodawać pojedyncze wartości, odwołania do komórek lub zakresów lub połączenie tych wszystkich trzech typów wyrażeń.
ŚREDNIA(wartość1; [wartość2];…)	Oblicza wartość średnią (średnią arytmetyczną) argumentów z listy. W składni funkcji ŚREDNIA.A występują następujące argumenty: Wartość1; wartość2;... Argument wartość1 jest wymagany, pozostałe są opcjonalne. Od 1 do 255 komórek, zakresów komórek lub wartości, dla których należy wyznaczyć średnią.

Funkcja	Opis
WYST.NAJCZĘŚCIEJ(liczba1; [liczba2];...)	W składni funkcji WYST.NAJCZĘŚCIEJ występują następujące argumenty: Liczba1 – argument wymagany. Pierwsza liczba zakresu, dla którego ma zostać obliczona dominanta. Liczba2;... – argumenty opcjonalne. Argumenty liczbowe od 2 do 255, dla których należy obliczyć dominantę. Zamiast argumentów rozdzielonych średnikami można użyć pojedynczej tablicy lub odwołania do tablicy.

Przypadek 1

Polecenie:

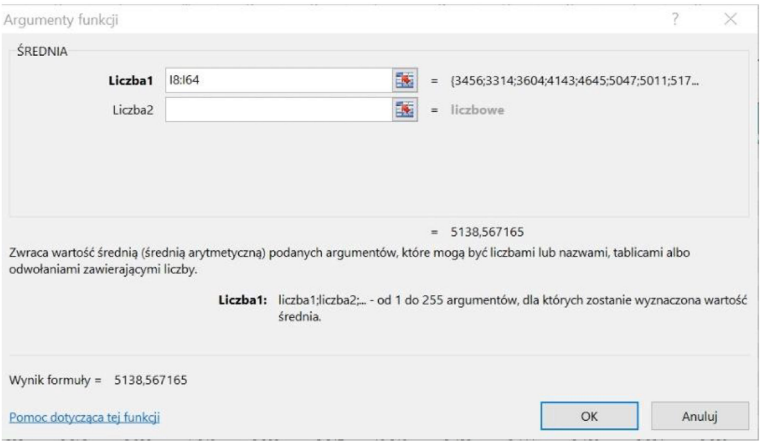
Załącznik3.xlsx zawiera dane dotyczące rynku nieruchomości w Polsce. Dane są publikowane kwartalnie na stronach internetowych Narodowego Banku Polskiego. Opis zawartych danych znajduje się w zakładce Opis danych. Oblicz dla miasta Lublin (ceny ofertowe) następujące charakterystyki:

- średnią arytmetyczną
- medianę
- rozstęp

Rozwiązanie:

Krok 1.

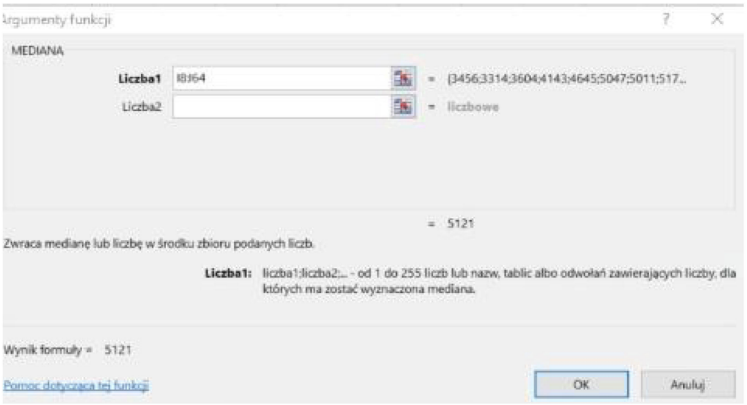
W komórce T4 użyj funkcji ŚREDNIA – okno dialogowe widoczne jest poniżej.



Uzyskana średnia wynosi 5 139 zł.

Krok 2.

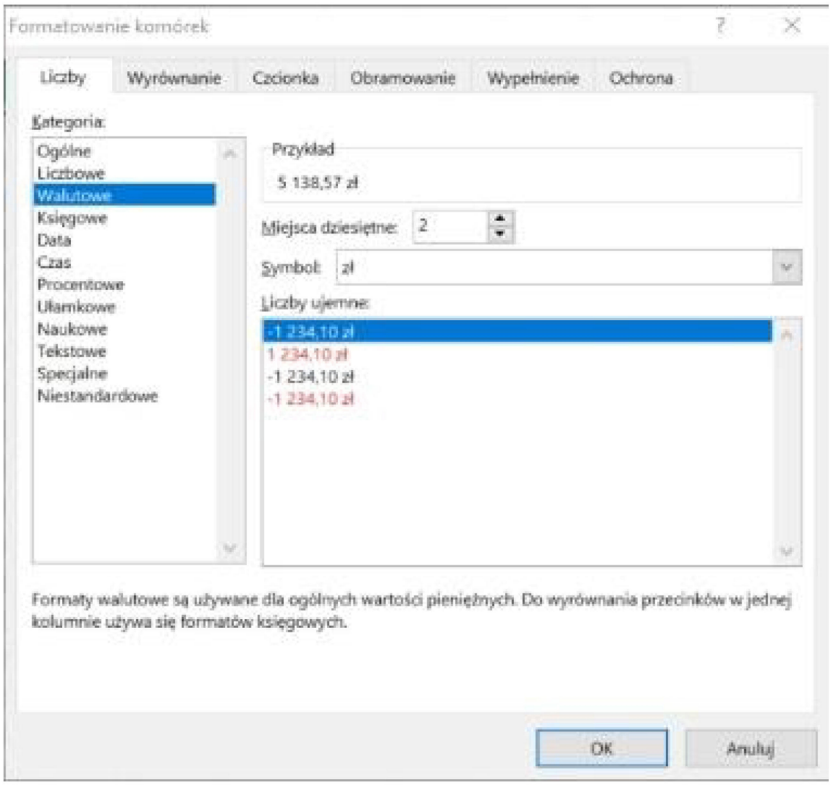
W komórce T5 użyj funkcji MEDIANA – okno dialogowe widoczne jest poniżej.



Uzyskana mediana wynosi 5 121 zł.

Krok 3.

Uzyskane wielkości warto sformatować. Zaznaczamy komórki T4 oraz T5 prawym przyciskiem myszy. Z menu kontekstowego wybieramy *Formatuj komórki*, okno dialogowe widoczne jest poniżej.



Po wciśnięciu przycisku *OK* efekt jest podobny jak na rysunku poniżej.

średnia	5 138,57 zł
mediana	5 121,00 zł

Krok 4.

Teraz obliczymy rozstęp danych. Excel nie ma funkcji *ROZSTĘP()*. W celu obliczenia rozstępu trzeba użyć poniższej formuły, zastępując podane argumenty odpowiednim adresem zakresu. W komórce T1 wpisujemy

=MAX(I8:I64)-MIN(I8:I64)

Krok 5.

Kolejno przejdziemy do obliczenia kwartylu pierwszego oraz trzeciego. W komórce T2 wpisujemy

=KWARTYL.PRZEDZ.ZAMK(I8:I64;1)

A następnie w komórce T3

=KWARTYL.PRZEDZ.ZAMK(I8:I64;3)

Krok 6.

Obliczając rozstęp kwartylowy, w komórce Q1 wpisujemy

=T3-T2

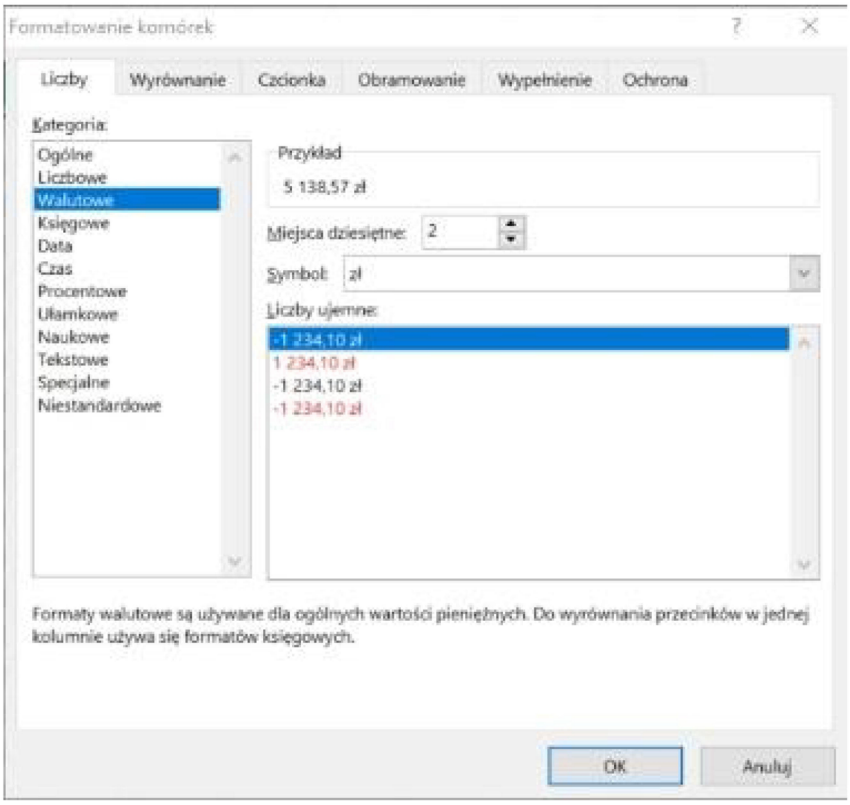
Krok 7.

Obliczając odchylenie standardowe w komórce Q2, wpisujemy

=ODCH.STANDARD.PRÓBKI(I8:I64)

Krok 8.

Uzyskane wielkości formatujemy. Zaznaczamy komórki T1, T2, T3, Q1, Q2 prawym przyciskiem myszy. Z menu kontekstowego wybieramy *Formatuj komórki*, okno dialogowe widoczne jest poniżej.



Po wciśnięciu przycisku *OK* efekt jest podobny jak na rysunku poniżej.

P	Q	R	S	T
Rozstęp kwartylowy	246,00 zł		Rozstęp	3 336,80 zł
Odchylenie standardowe	558,80 zł		Kwartyl 1	5 011,00 zł
			Kwartyl 3	5 257,00 zł

Przypadek 2

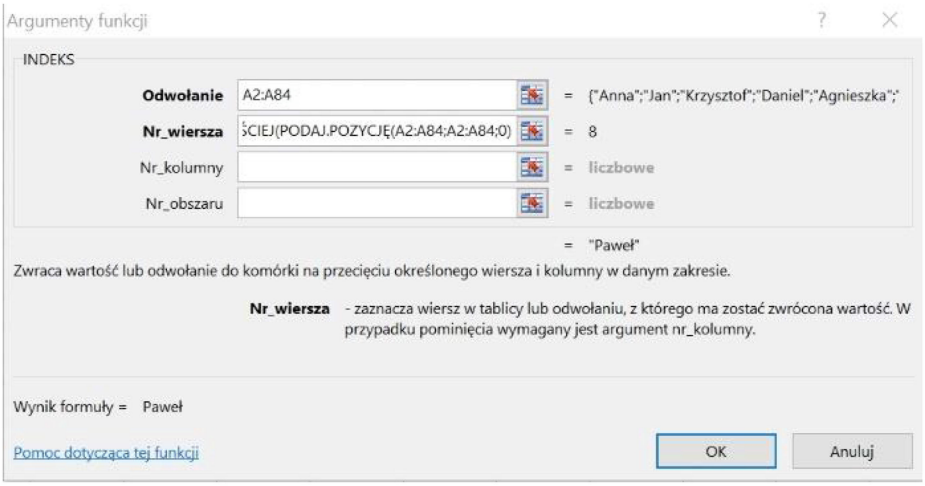
Polecenie:

Studenci WSEI zostali poproszeni o podanie swoich imion. Wyniki ankiety znajdują się w pliku Załącznik1.xlsx. Dla danych podanych w skali nominalnej (imiona) oblicz wartość modalną oraz dokonaj wizualizacji otrzymanych wyników.

Rozwiązanie:

Krok 1.
W polu E1 napisz formułę
=INDEKS(A2:A84;WYST.NAJCZĘŚCIEJ(PODAJ.POZYCJĘ(A2:A84;A2:A84;0)))

Okno dialogowe funkcji *INDEKS* widoczne jest poniżej



Krok 2.

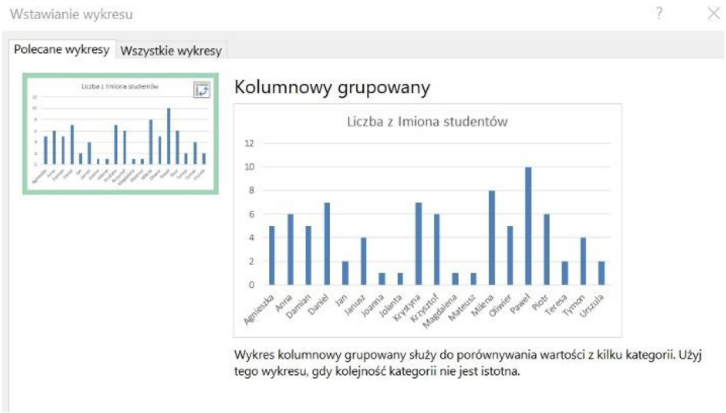
W polu E2 napisz formułę
=INDEKS(A2:A84;WYST.NAJCZĘŚCIEJ(PODAJ.POZYCJĘ(A2:A84;A2:A84;0)))

Krok 3.

W efekcie modą dla danego zbioru danych jest imię Paweł.

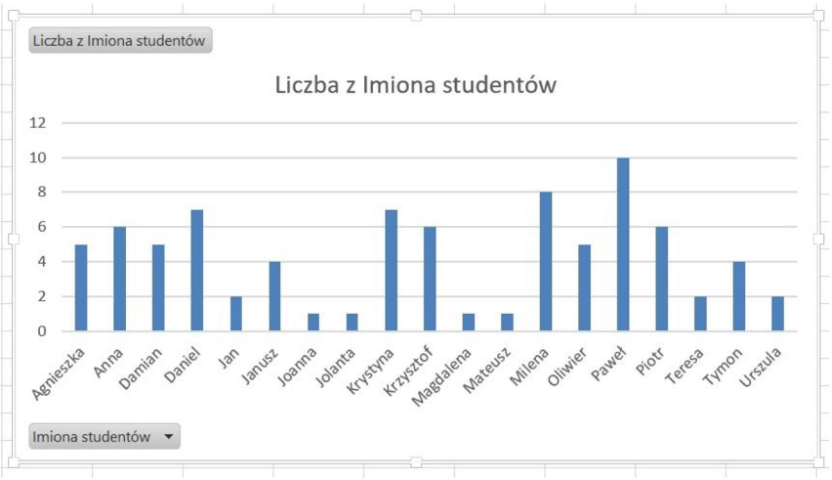
Krok 4.

W celu utworzenia wykresu przestawnego wykonaj następujące kroki. Zaznacz dane, z zakładki *Wstawianie* wybierz *Polecane wykresy* oraz wykres *Kolumnowy grupowany*.



W efekcie w nowym arkuszu otrzymasz tabelę przestawną oraz wykres przestawny – patrz Rysunek poniżej. Po wykonaniu tych czynności tabela przestawna i oparty na niej wykres przestawny są od razu aktualizowane.

	A	B
1	Imiona studentów	Liczba z Imiona studentów
2	Agnieszka	5
3	Anna	6
4	Damian	5
5	Daniel	7
6	Jan	2
7	Janusz	4
8	Joanna	1
9	Jolanta	1
10	Krystyna	7
11	Krzysztof	6
12	Magdalena	1
13	Mateusz	1
14	Milena	8
15	Oliwier	5
16	Paweł	10
17	Piotr	6
18	Teresa	2
19	Tymon	4
20	Urszula	2



Krok 5.

Aby obliczyć ile wartości modalnej (imię Paweł) występuje w zbiorze danych, w komórce E3 napisz następującą formułę
=SUMA(JEŻELI(A2:A84=E1;1;0))

wprowadź formułę za pomocą klawiszy *Ctrl+Shift+Enter* zamiast samego klawisza *Enter* używanego w przypadku zwykłej formuły Excela lub wartości.

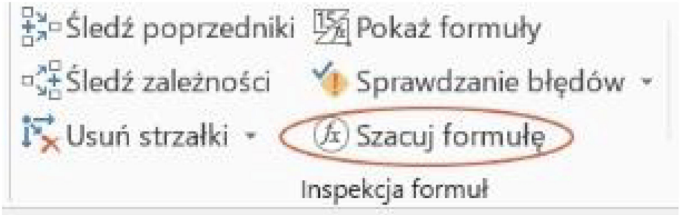
Krok 6.

W efekcie uzyskamy poniższe informacje

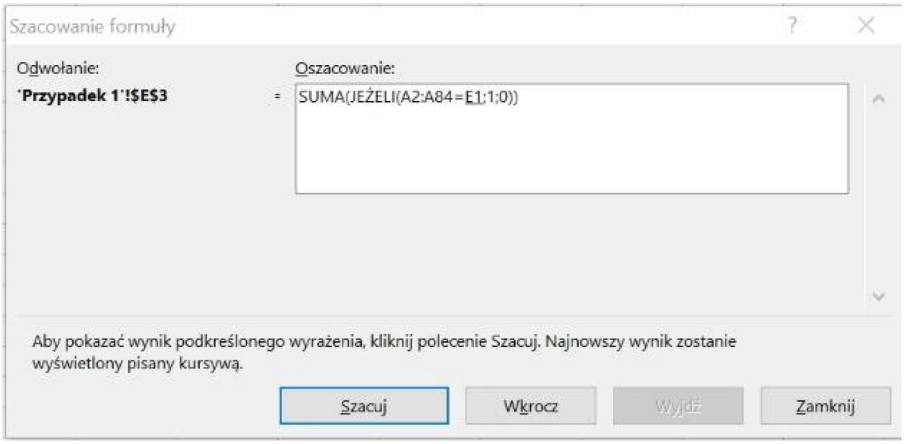
D	E
Moda	Paweł
Pozycja	8
Ile	10

Krok 7.

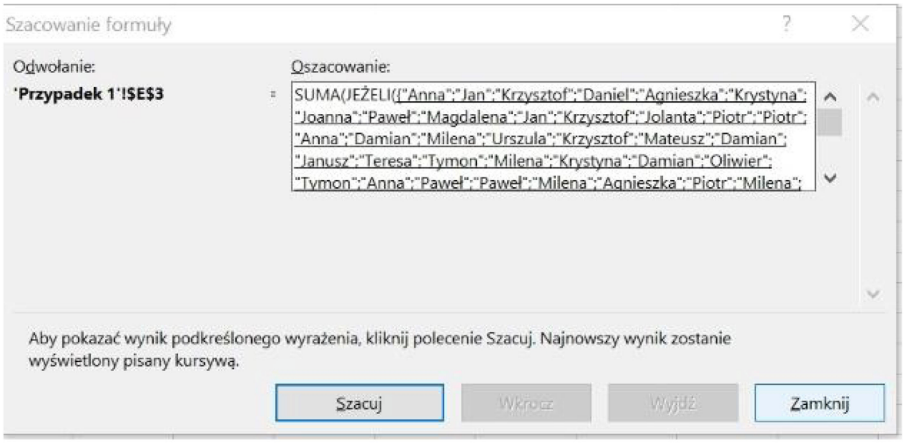
Skorzystajmy teraz z narzędzia szacowania formuły. Zaznacz ostatnio napisaną formułę, a następnie wybierz przycisk *Szacuj formułę* z karty *Formuły* w grupie *Inspekcja formuł.*



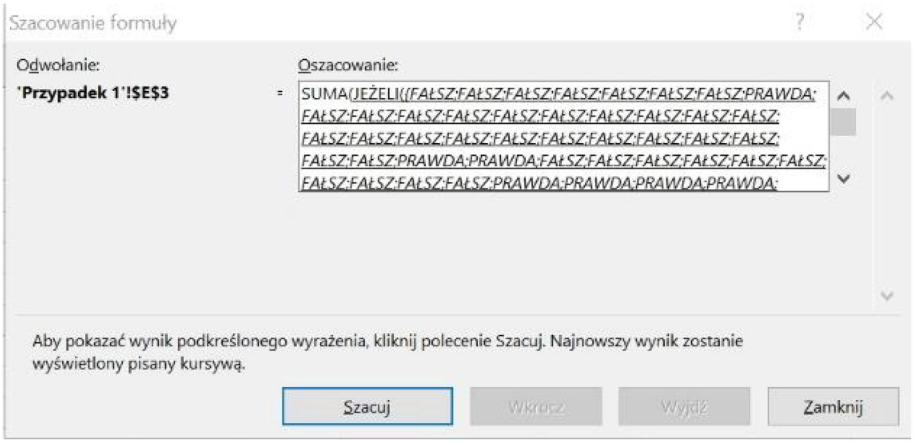
Pojawi się okno pokazane niżej.



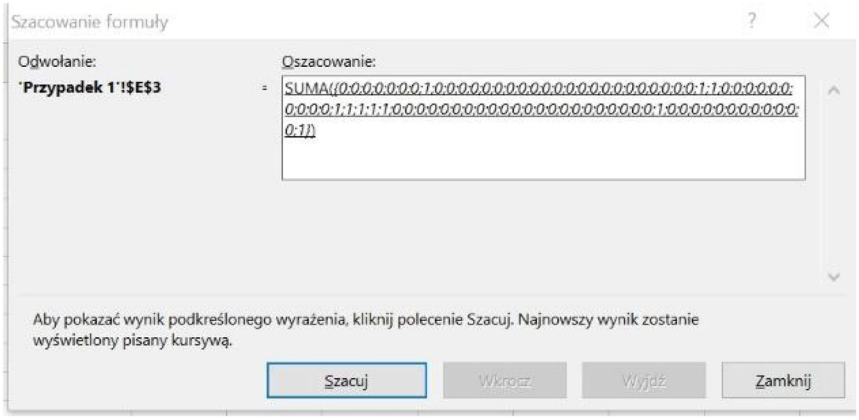
Jeżeli teraz klikniesz przycisk *Szacuj*, Excel zacznie obliczanie formuły od środka do zewnątrz i zmieni wyświetlane dane, jak zostało to zaprezentowane na rysunku poniżej.



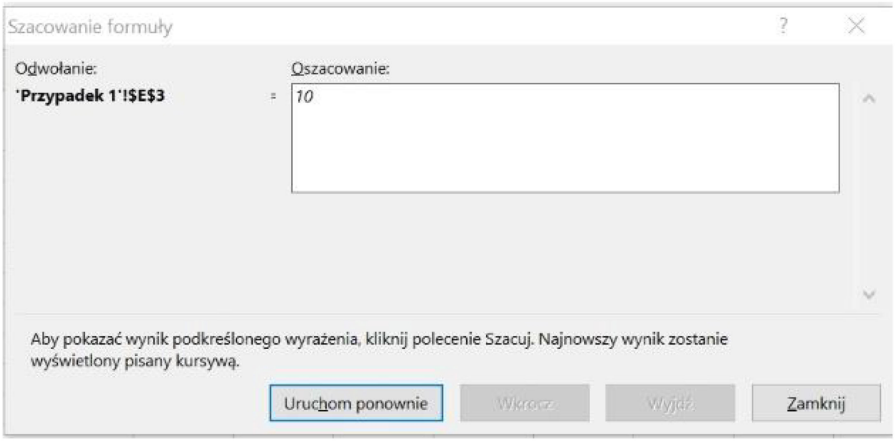
Kliknij ponownie przycisk *Szacuj*, a zobaczysz wyniki testu porównującego zawartość zaznaczonych komórek danych z imieniem Paweł, jak pokazano na rysunku poniżej.



Kliknij ponownie przycisk *Szacuj*, a okno pokaże wyniki funkcji JEŻELI(), które w tym przypadku zastąpią wartość PRAWDA liczbą 1, a FAŁSZ liczbą 0.



Ostatnie kliknięcie przycisku *Szacuj* pokazuje ostateczny wynik, w którym funkcja SUMA() dodaje wszystkie jedynki i zera, aby zwrócić liczbę wystąpień słowa Paweł w zaznaczonych komórkach danych.



Przypadek 3

Polecenie:

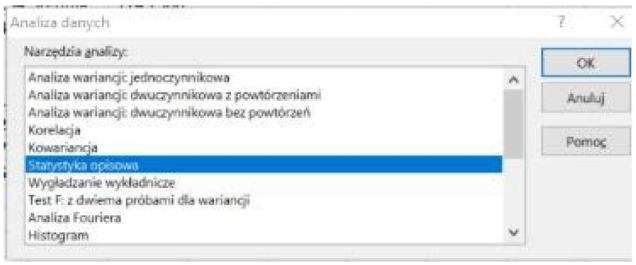
Dla danych opisanych w poniższej tabeli (patrz Załacznik5.xlsx) wyznacz statystyki opisowe.

Wiek	Waga	Wzrost
25	55	169
31	62	176
26	55	169
34	69	178
46	71	181
51	65	173
27	58	170
29	68	175
30	70	179
44	82	184
24	58	172
36	68	174
41	57	165

Rozwiązanie:

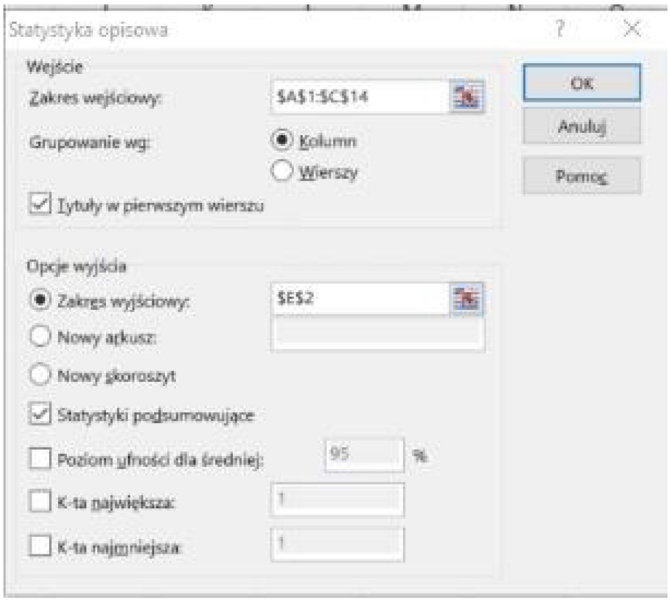
Krok 1.

Z karty *Dane* wybierz *Analiza danych,* a następnie *Statystyka opisowa*.



Krok 2.

Zaznaczamy *Zakres wejściowy* oraz wybieramy: *Tytuły w pierwszym wierszu* i *Statystyki podsumowujące*.



Krok 3.

W efekcie otrzymamy poniższe informacje w zakresie statystyk opisowych.

Wiek		*Waga*		*Wzrost*	
Średnia	34,15385	Średnia	64,46154	Średnia	174,2308
Błąd standardowy	2,441222	Błąd standardowy	2,194443	Błąd standardowy	1,48994
Mediana	31	Mediana	65	Mediana	174
Tryb	#N/D!	Tryb	55	Tryb	169
Odchylenie standardowe	8,801952	Odchylenie standardowe	7,912178	Odchylenie standardowe	5,372055
Wariancja próbki	77,47436	Wariancja próbki	62,60256	Wariancja próbki	28,85897
Kurtoza	-0,76574	Kurtoza	0,321708	Kurtoza	-0,44365
Skośność	0,687861	Skośność	0,668576	Skośność	0,158235
Zakres	27	Zakres	27	Zakres	19
Minimum	24	Minimum	55	Minimum	165
Maksimum	51	Maksimum	82	Maksimum	184
Suma	444	Suma	838	Suma	2265
Licznik	13	Licznik	13	Licznik	13

Przypadek 4

Polecenie:

Dwudziestoosobowa grupa studentów została poproszona o podanie liczby dzieci. Dane znajdują się w tabeli oraz pliku Załącznik6.xlsx.

Nr	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
Liczba dzieci	0	2	1	4	2	0	3	4	2	2	3	3	5	1	3	3	5	0	4	3

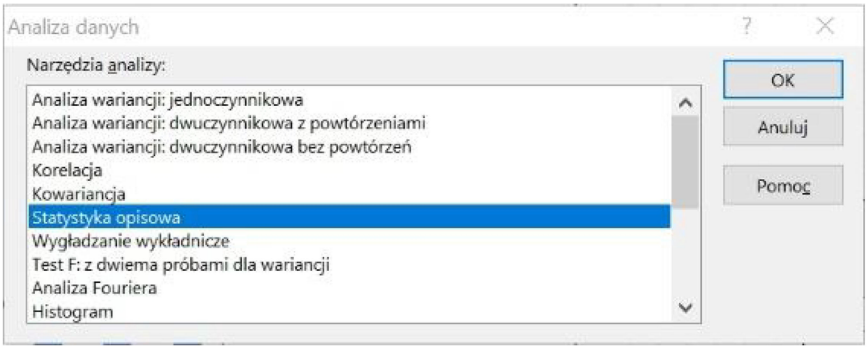
Wyznacz miary położenia, rozproszenia, asymetrii i koncentracji oraz podaj ich interpretację.

Rozwiązanie:

Poniżej sposób wykonania zadania w programie EXCEL, z wykorzystaniem dodatku *Analiza danych.*

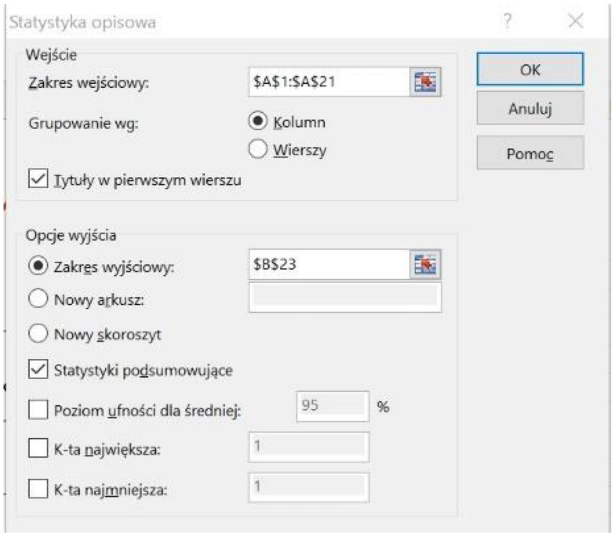
Krok 1.

Wczytaj dane z pliku Załącznik6.xlsx. Wybierz z menu *Dane/Analiza danych/Statystyka opisowa*



Krok 2.

W oknie dialogowym wskaż zakres wejściowy i zaznacz pole *Statystyki podsumowujące*.



Krok 3.

Program wyświetli poniższe statystyki.

Liczba dzieci	
Średnia	2,5
Błąd standardowy	0,34
Mediana	3
Tryb	3
Odchylenie standardowe	1,54
Wariancja próbki	2,37
Kurtoza	-0,69
Skośność	-0,19
Zakres	5
Minimum	0
Maksimum	5
Suma	50
Licznik	20

Krok 4.

Interpretacja:

- Średnia liczba dzieci wynosi
- Mediana wynosi 3, co oznacza, że 50% studentów ma troje lub mniej dzieci, a 50% - troje lub więcej.
- Tryb (czyli dominanta) wynosi 3, czyli najczęściej student ma troje dzieci.
- Rozkład jest symetryczny, ponieważ współczynnik skośności jest bliski zeru.
- Rozkład jest platykurtyczny, ponieważ kurtoza jest ujemna.

Przypadek 5

Polecenie:

Z egzaminu ze Statystyki można było dostać maksymalnie 16 punktów. Wyniki tego egzaminu przedstawia poniższa tabela oraz Załącznik9.xlsx.

Liczba punktów	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Liczba osób	3	10	6	18	26	37	49	36	27	29	28	20	12	20	8	1

1. Przedstaw graficznie dane.
2. Zbuduj szereg rozdzielczy przedziałowy.
3. Oblicz średnią arytmetyczną, wariancję, odchylenie standardowe oraz klasyczny współczynnik zmienności.
4. Oblicz klasyczny obszar zmienności, dominantę, medianę, kwartyl pierwszy, kwartyl trzeci, odchylenie ćwiartkowe, pozycyjny współczynnik zmienności.

Rozwiązanie:

Krok 1.

Z karty *Wstawianie/Wykresy* wybierz wykres *Kolumnowy grupowany.*

Krok 2.

Korzystając z procedury tworzenia szeregu rozdzielczego, otrzymujemy

D	E	F
Lewy koniec	Prawy koniec	Częstość ni
1	5	37
5	9	148
9	13	104
13	17	41

Szczegóły: n=16; k= $\sqrt{n} = \sqrt{16}$ =4; l= $\frac{x\max - x\min}{k} = \frac{16-1}{4} = \frac{15}{4} \approx 4$

Krok 3.

W celu obliczenia czterech pierwszych miar rozbudowujemy szereg.

D	E	F	G	H	I
Lewy koniec	Prawy koniec	Częstość ni	$\overline{xi}$	$\overline{xi}$ ni*	$(\overline{xi} - \overline{x})^2 * ni$
1	5	37	3	111	1247,28
5	9	148	7	1036	482,75
9	13	104	11	1144	500,59
13	17	41	15	615	1572,96
		330		2906	3803,59

Krok 4.

Średnia arytmetyczna

$$\overline{x} = \frac{1}{n} \sum ni * \overline{xi}$$

$$\overline{x} = \frac{1}{330} * 2906 = 8{,}81$$

Średnia liczba zdobytych punktów z egzaminu wyniosła 8,81.

Krok 5.

Wariancja

$$s^2 = \frac{1}{n} \sum (\overline{xi} - \overline{x})^2 * ni$$

$$s^2 = \frac{1}{330} * 3803{,}59 = 11{,}53$$

Odchylenie standardowe

$$s = \sqrt{11{,}53} = 3{,}39$$

Średnia odchylenia liczby zdobytych punktów od średniej arytmetycznej wynosi 3,39 punktu.

Krok 6.

Klasyczny współczynnik zmienności

$$V=\frac{s}{x}*100\%=\frac{3,39}{8,81}*100\%=38,5\%$$

Liczba zdobytych punktów wykazuje umiarkowane zróżnicowanie.

Średnia arytmetyczna	8,81
Wariancja	11,53
Odchylenie standardowe	3,39
Współczynnik zmienności	38,6%

Krok 7.

Klasyczny obszar zmienności

$$\bar{x}-S(x)<x_{typ}<\bar{x}+S(x)$$
$$8,81-3,39<x_{typ}<8,81+3,39$$
$$5,42<x_{typ}<12,2$$

Dominanta

$$D=_{o}x_D+\frac{n_D-n_{D-1}}{(n_D-n_{D-1})+(n_D-n_{D+1})}\cdot h_D$$
$$D=5+\frac{148-37}{(148-37)+(148-104)}\cdot 4=7,86$$

Najwięcej studentów dostało średnio 7,86 punktów.

Mediana

$$M=_{o}x_M+\frac{\frac{n}{2}-n_{sk-1}}{n_M}\cdot h_M=5+\frac{165-37}{148}\cdot 4=8,46$$

Połowa badanych studentów dostała średnio 8,46 razy i mniej, a druga połowa średnio 8,46 razy i więcej punktów.

Kwartyl pierwszy

$$Q_{1.4} = {}_{o}x_{Q_{3.4}} + \frac{\frac{n}{4} - n_{sk-1}}{n_{Q_{1.4}}} \cdot h_{Q_{1.4}} = 5 + \frac{82,5 - 37}{148} \cdot 4 = 6,22$$

25% badanych studentów dostała średnio 6,22 razy i mniej, a 75% badanych studentów dostała średnio 6,22 razy i więcej punktów z egzaminu.

Kwartyl trzeci

$$Q_{3.4} = {}_{o}x_{Q_{3.4}} + \frac{\frac{3n}{4} - n_{sk-1}}{n_{Q_{3.4}}} \cdot h_{Q_{3.4}} = 9 + \frac{247,5 - 185}{104} \cdot 4 = 11,4$$

75% badanych studentów dostało 11,4 razy i mniej, a 25% dostało 11,4 razy i więcej punktów z egzaminu.

Odchylenie ćwiartkowe

$$Q = \frac{Q_{3.4} - Q_{1.4}}{2} = \frac{11,4 - 6,22}{2} = 2,59$$

Średnie odchylenie liczby punktów od mediany wynosi 2,59 punktów.

Pozycyjny współczynnik zmienności

$$V_Q = \frac{Q}{M} \cdot 100\% = \frac{2,59}{8,46} \cdot 100\% = 30,61\%$$

Liczba zdobytych punktów z egzaminu w zawężonym obszarze wykazuje umiarkowane zróżnicowanie.

Pozycyjny obszar zmienności

$$M - Q < x_{typ} < M + Q$$
$$8,46 - 2,59 < x_{typ} < 8,46 + 2,59$$
$$5,87 < x_{typ} < 11,05$$

Zadania

Zadanie 1.

W pewnej uczelni przez 50 kolejnych dni notowano liczbę nieobecnych studentów na zajęciach ze Statystyki, otrzymując dane:

2, 1, 1, 4, 2, 3, 2, 0, 5, 0, 0, 0, 0, 1, 2, 0, 2, 1, 1, 2, 1, 2, 2, 1, 1, 3, 0, 0, 2, 2, 0, 3, 0, 1, 1, 1, 3, 2, 1, 4, 2, 0, 1, 1, 1, 0, 1, 0, 1, 0.

a) Zbuduj na podstawie powyższych danych szereg punktowy.
b) Oblicz średnią arytmetyczną, medianę i dominantę.
c) Oblicz kwartyle oraz decyle.
d) Oblicz odchylenie standardowe.
e) Wyznacz przedział typowych wartości cechy.
f) Oblicz klasyczny współczynnik zmienności, współczynnik skośności oraz klasyczny współczynnik asymetrii.

.....
.....
.....
.....
.....
.....
.....

Zadanie 2.

W województwie lubelskim zarejestrowanych jest 78 firm ubezpieczeniowych.

W pewnym miesiącu zebrano informacje o licznie zawartych przez nie umowach ubezpieczeniowych, które prezentuje poniższe zestawienie:

12 17 28 47 39 6 62 54 38 29 12 3 4 53 26 45 22 18 52 45 27 46 37 53 12 3 6 29 52 47 11 45 3 34 27 43 23 62 8 33 26 38 48 11 25 27 24 31 28 33 28 37 51 1 27 38 16 26 49 50 28 7 2 9 52 42 48 28 38 29 26 45 37 22 37 12 2 27

a) Zbuduj szereg rozdzielczy przedziałowy.
b) Jaka była łączna liczba zawartych umów w badanych firmach ubezpieczeniowych wyznaczona na podstawie danych indywidualnych oraz z danych przedstawionych w przedziałowym rozkładzie?
c) Z czego wynikają różnice w wynikach? Który z wyników wyraża faktyczną łączną liczbę zawartych umów?
d) Graficznie przedstaw rozkład liczebności zawartych umów względem badanych firm ubezpieczeniowych dla danych indywidualnych i pogrupowanych.

.....
.....
.....
.....
.....
.....
.....
.....

Zadanie 3.

Dla danych przedstawiających liczbę zjedzonych czekolad w miesiącu przez studentów pewnej lubelskiej uczelni, tj.:

2; 3; 4; 3; 6; 8; 3; 4; 1; 1; 1; 1; 4; 3; 5; 4; 6; 8; 2; 4; 5; 8; 6; 4; 2; 3; 6; 3; 5; 4.

a) Utwórz szereg rozdzielczy punktowy i przedziałowy o rozpiętości dwóch jednostek i przedstaw te szeregi na wykresie, zaznaczając wartość dominującą cechy.
b) Jaka była średnia liczba czekolad w badanej grupie studentów? Skorzystaj z danych pierwotnych.

.....
.....
.....
.....
.....
.....

Zadanie 4.

Przeprowadzono badanie sklepów branży meblarskiej ze względu na poziom rocznych zysków z działalności handlowej. Rozkład tych zysków ze względu na badaną cechę prezentuje tabelka:

Roczny zysk (w mln zł)	Liczba sklepów
0,5	10
1	17
1,5	12
2	8
2,5	3

Narysuj histogram, a następnie oblicz: średnią z próby, modę, medianę, wariancję, odchylenie standardowe oraz współczynnik zmienności. Podaj interpretację otrzymanych wyników.

.....

.....

.....

.....

.....

.....

Zadanie 5.

Wyznacz wskaźniki położenia oraz rozproszenia dla danych dotyczących czasu dojazdu pracowników z miejsca zamieszkania do pracy. Podaj interpretację wyznaczonych wskaźników.

Czas dojazdu	5-15	15-25	25-35	35-45	45-55	55-65	65-75
Liczba pracowników	4	7	15	24	11	6	2

.....

.....

.....

.....

.....

Zadanie 6.

Poniższa tabela przedstawia liczbę osób zatrudnionych w sklepie w następujących działach sprzedaży: mięsnym (A), spożywczym (B), papierniczym (C), obuwniczym (D), cukierniczym (E) i chemicznym (F) według płci (M-mężczyźni, K-kobiety) w pewnym pawilonie znajdującym się na terenie miasta Lublin:

Rodzaj działu	A	B	C	D	E	F
Kobiety	6	4	5	11	4	4
Mężczyźni	2	8	3	3	8	5

Wyznaczając odpowiednie miary statystyczne, odpowiedz na następujące pytania:

a) jaka była średnia liczba mężczyzn zatrudnionych w tym sklepie?
b) jak dalece przeciętnie różniło się zatrudnienie mężczyzn w tych działach w stosunku do przeciętnego zatrudnienia w tym sklepie?
c) która z badanych zbiorowości charakteryzowała się większym zróżnicowaniem liczby zatrudnionych w poszczególnych działach tego sklepu?

.....
.....
.....
.....
.....
.....

Zadanie 7.

Zbadano liczbę nieobecności na zajęciach ze statystyki w ciągu semestru i otrzymano następujące wyniki: 3,1,0,6,0,3,2,4,4,0,0,2,0,0,1.

Podaj wartość kwartyli I, II i III.

.....
.....
.....
.....
.....
.....

Zadanie 8.

Wśród 40 pracowników pewnego przedsiębiorstwa z województwa lubelskiego przeprowadzono badanie wynagrodzenia i otrzymano następujące wyniki w PLN:

1050, 1100, 1210, 1260, 1300, 1310, 3320, 1340, 1360, 1365, 1410, 2420, 1500, 1510, 1510, 1520, 1520, 2540, 1550, 1560, 1590, 1610, 2620, 4650, 17001520, 1520, 1540, 2550, 1560, 3520, 1610, 1620, 1650, 1900, 1720, 1980, 1830, 1920, 1980.

a) Zbuduj szereg rozdzielczy z przedziałami klasowymi.
b) Oblicz wskaźniki struktury.
c) Oblicz liczebności skumulowane.

.....
.....
.....
.....
.....
.....
.....
.....

Zadanie 9.

Zbadaj za pomocą odpowiedniego parametru stopień asymetrii płac pracowników firmy budowlanej, jeśli wiadomo, że:

mediana płac wynosi 2500 zł,
25% pracowników zarabia nie więcej niż 2000 zł,
75% pracowników zarabia nie więcej niż 3700 zł.

.....
.....
.....
.....
.....
.....
.....

Studium przypadku 4.
Analiza współzależności cech

Cel ćwiczenia

Poznanie metod i nabycie praktycznych umiejętności rozwiązywania problemów mieszczących się w zakresach korelacji oraz regresji oraz interpretacji ich wyników wraz z graficzną prezentacją.

Wstęp teoretyczny

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.

Jeżeli pomiędzy dwiema cechami zachodzi korelacja liniowa, to jej kierunek i siłę można zmierzyć za pomocą współczynnika korelacji liniowej Pearsona, który w skrócie nazwany jest współczynnikiem korelacji.

WAŻNE

Wzory

Współczynnik korelacji liniowej Pearsona mierzy związek między zmiennymi x i y, gdy są one mierzalne i podane w postaci szeregów szczegółowych.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{C(X, Y)}{S(X) \cdot S(Y)}$$

gdzie:
r_{xy} – współczynnik korelacji liniowej Pearsona
i=1,2,…, - numer jednostki statystycznej
x_i – wartość pierwszej cechy statystycznej u jednostki statystycznej o numerze i
y_i – wartość drugiej cechy statystycznej u jednostki statystycznej o numerze i
$\bar{x}$ – średnia arytmetyczna pierwszej cechy
$\bar{y}$ – średnia arytmetyczna drugiej cechy
S(X), S(Y) – odchylenia standardowe zmiennych X oraz Y

$\lvert r_{xy} \rvert$	Zależność liniowa
< 0.2	brak zależności
0.2 – 0.4	zależność niska
0.4 – 0.7	zależność uiarkowana
0.7 – 0.9	zależność znacząca
>0.9	zależność bardzo silna

Własności współczynnika korelacji liniowej Pearsona:

- Pozwala na określenie nie tylko siły, ale i kierunku zależności między zmiennymi.
- Dodatni znak współczynnika wskazuje na sytuację, gdy wzrostowi jednej cechy towarzyszy na ogół wzrost wartości drugiej cechy.
- Ujemny znak współczynnika wskazuje na sytuację, gdy wzrostowi jednej cechy towarzyszy na ogół spadek wartości drugiej cechy.

Przy wyznaczaniu współczynnika korelacji liniowej Pearsona interesowało nas zagadnienie siły i kierunku zależności między cechami X i Y. W wielu analizach warto scharakteryzować bardziej szczegółowo związek między nimi i wskazać na tę cechę, która jest przyczyną kształtowania się drugiej, czyli skutku. Mówimy wówczas o występowaniu związku przyczynowo-skutkowego. Do jego analizy posłużymy się metodą regresji prostej.

Celem regresji jest zbudowanie na podstawie dostępnych informacji modelu opisującego rzeczywistość. Taki model pełni funkcję poznawczą – dostarcza wiedzy na temat zjawiska, a także umożliwia prognozowanie (predykcję) nieznanej wartości analizowanej cechy.

WAŻNE

Wzory

I linia regresji

$$x = a_X \cdot y + b$$

$$a_X = \frac{C(X,Y)}{S_Y^{\,2}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

$$b = \overline{x} - a_X \cdot \overline{y}$$

gdzie:

x – zmienna objaśniana (zależna)

y – zmienna objaśniająca (niezależna)

a_X – parametr linii regresji (współczynnik kierunkowy prostej)

b – drugi parametr linii regresji.

Znaczenie poznawcze funkcji regresji jest tym większe, im silniejsza jest korelacja między badanymi zmiennymi. Silny związek oznacza bowiem, że zmienna niezależna determinuje w znacznym stopniu zmiany w poziomie zmiennej zależnej, niewielki natomiast jest wpływ indywidualnych odchyleń. Wskazuje na to kwadrat współczynnika korelacji liniowej Pearsona, czyli tzw. współczynnik determinacji

WAŻNE

Wzory

Współczynnik determinacji R^2 jest miarą unormowaną, $R^2 \in \langle 0; 1 \rangle$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

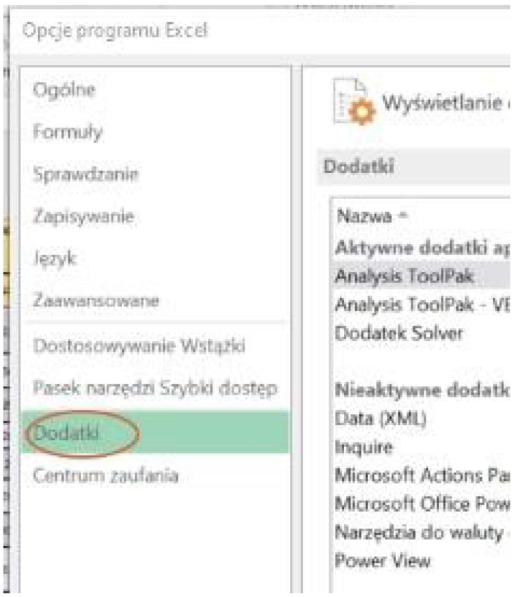
Współczynnik determinacji	Dopasowanie
0 – 0.5	niezadawalające
0.5 – 0.6	słabe
0.6 – 0.8	zadawalające
0.8 – 0.9	dobre
0.9 – 1.0	bardzo dobre

Wykorzystywane funkcje

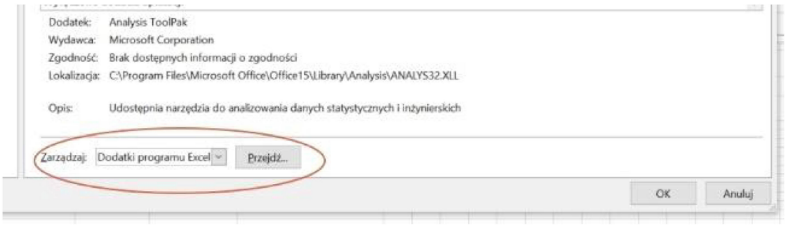
Podczas rozwiązywania zadań z tego rozdziału skorzystamy z dodatku *Analiza danych*. Jeśli nie masz zainstalowanego tego dodatku (jest widoczny na karcie *Dane*), to wykonaj poniższe kroki.

1. Kliknij kartę *Plik* i przycisk *Opcje* w jej pasku nawigacyjnym.

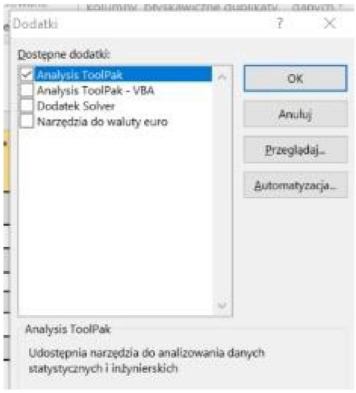
2. Otworzy się okno *Opcje programu Excel*. Kliknij zakładkę *Dodatki* w pasku nawigacyjnym.



3. W razie potrzeby wybierz pozycję *Dodatki programu Excel* z rozwijanej listy *Zarządzaj* i kliknij przycisk *Przejdź*.



4. Pojawi się okno dialogowe *Dodatki*. Jeżeli zobaczysz na liście dodatek *Analysis ToolPak*, upewnij się, że odpowiadające mu pole wyboru jest zaznaczone. (*Analysis ToolPak* to stare określenie tego dodatku). Kliknij przycisk *OK.* Przycisk *Analiza danych* powinien być teraz widoczny w grupie *Analiza* na karcie *Dane*.



Funkcja	Opis
PIERWIASTEK (liczba)	Zwraca dodatni pierwiastek kwadratowy liczby.
WSP.KORELACJI (tablica1;tablica2)	Funkcja zwraca współczynnik korelacji dwóch zakresów komórek. Współczynnik korelacji służy do określania relacji między dwiema własnościami. W składni funkcji WSP.KORELACJI występują następujące argumenty: tablica1 – argument wymagany. Zakres wartości komórek. tablica2 – argument wymagany. Drugi zakres wartości komórek.

Przypadek 1

Polecenie:

Średnie wartości podstawowych zmiennych makroekonomicznych w województwach Polski w latach 2008–2017 przedstawione są w poniższej tabeli, dodatkowe dane zawiera plik Załącznik4.xslx – zakładka Dane.

Województwo	Produkcja sprzedana przemysłu	Inwestycje	Wartość Brutto środków trwałych	Płace	Stopa bezrobocia rejestrow. %	Podmioty gospodarki
	Na mieszkańca w tys. zł		W tys. zł			
Dolnośląskie	35,22	4,63	140 713	3,83	10,67	30,27
Kujawsko-pomorskie	23,66	2,96	66 028	2,25	15,04	17,60
Lubelskie	13,99	1,94	47 076	3,40	12,24	15,24
Lubuskie	26,26	2,84	37 636	3,26	12,93	10,08
Łódzkie	26,20	3,52	95 203	3,43	11,16	21,39
Małopolskie	22,78	2,76	107 796	3,55	9,09	32,03
Mazowieckie	43,45	6,01	395 519	4,72	8,81	61,95
Opolskie	22,31	3,72	37 223	3,46	11,64	7,09

Województwo	Produkcja sprzedana przemysłu	Inwestycje	Wartość Brutto środków trwałych	Płace	Stopa bezrobocia rejestrow. %	Podmioty gospodarki
	Na mieszkańca w tys. zł		W tys. zł			
Podkarpackie	17,33	2,45	56 803	3,21	14,14	14,40
Podlaskie	17,01	2,10	29 497	3,36	12,37	8,86
Pomorskie	33,63	3,90	88 496	3,79	10,42	25,85
Śląskie	42,67	3,95	220 977	3,89	8,84	37,42
Świętokrzyskie	18,64	2,25	34 817	3,29	13,8	9,22
Warmińsko-mazurskie	18,57	2,17	34 327	3,20	18,14	11,56
Wielkopolskie	37,63	3,60	147 190	3,45	7,56	34,73
Zachodniopomorskie	19,54	2,82	58 055	3,48	15	19,03

Dla danych z tabeli należy obliczyć korelację między:

1) produkcją sprzedaną przemysłu na mieszkańca a wartością brutto środków trwałych,
2) stopami bezrobocia rejestrowanego a produkcją sprzedaną przemysłu na mieszkańca,
3) stopami bezrobocia rejestrowanego a nakładami inwestycyjnymi w przedsiębiorstwach na mieszkańca,
4) stopami bezrobocia rejestrowanego a wartością brutto środków trwałych w przedsiębiorstwach,
5) stopami bezrobocia rejestrowanego a przeciętnym miesięcznym wynagrodzeniem brutto,
6) stopami bezrobocia rejestrowanego a podmiotami gospodarki narodowej.

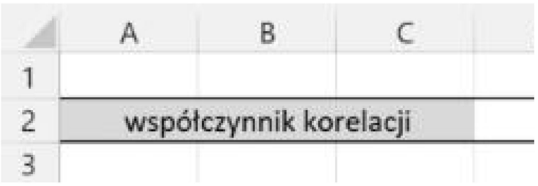
Rozwiązanie:

Obliczając kolejne podpunkty zadania, proponuję dodawać kolejne zakładki do pliku. Podpunkt pierwszy jest opisany szczegółowo, w kolejnych podpunktach należy wzorować się właśnie na tym rozwiązaniu. Zaprezentowane są bowiem odpowiedzi bez szczegółowych informacji dotyczących rozwiązania.

1. Dla danych z poniższej tabeli

Województwo	Produkcja sprzedana przemysłu	Wartość brutto środków trwałych
Dolnośląskie	35,22	140 713
Kujawsko-pomorskie	23,66	66 028
Lubelskie	13,99	47 076
Lubuskie	26,26	37 636
Łódzkie	26,2	95 203
Małopolskie	22,78	107 796
Mazowieckie	43,45	395 519
Opolskie	22,31	37 223
Podkarpackie	17,33	56 803
Podlaskie	17,01	29 497
Pomorskie	33,63	88 496
Śląskie	42,67	220 977
Świętokrzyskie	18,64	34 817
Warmińsko-mazurskie	18,57	34 327
Wielkopolskie	37,63	147 190
Zachodniopomorskie	19,54	58 055

W nowej zakładce, w komórce D2 wpisujemy formułę
=WSP.KORELACJI(Dane!B9:B24;Dane!D9:D24)
Otrzymujemy



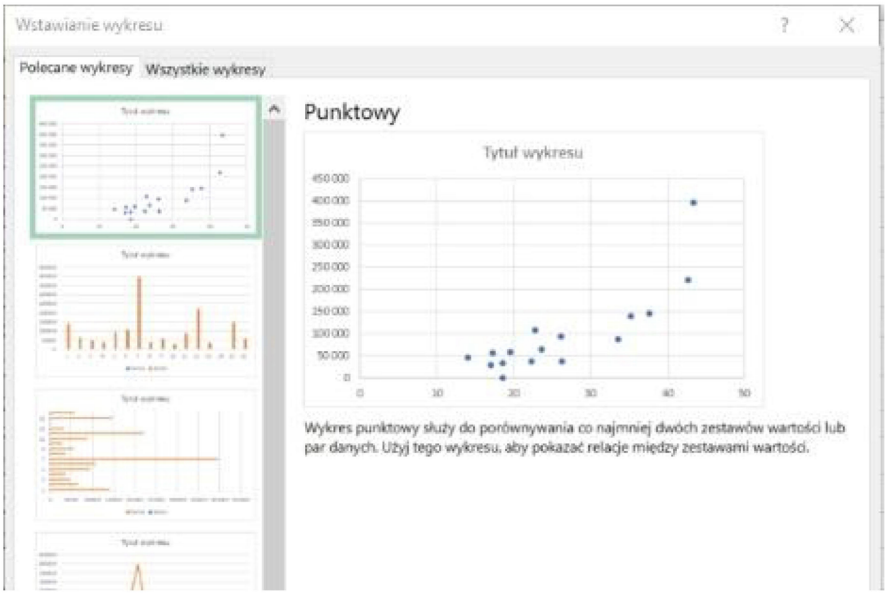
Wykonamy teraz wykres, ilustrujący nasze dane. W tym celu należy kolejno:

Krok 1.

Zaznaczyć dane źródłowe z tabeli (zakładka Dane, kolumna B oraz kolumna D).

Krok 2.

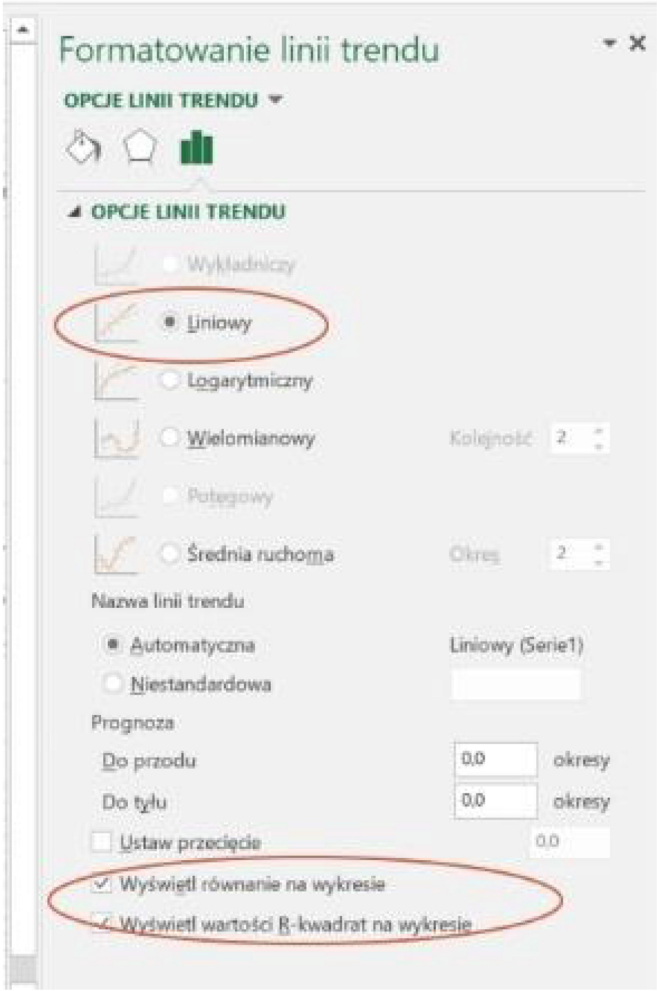
Z karty *Wstawianie* wybieramy *Wykresy polecane*, a następnie *Wykres punktowy*.



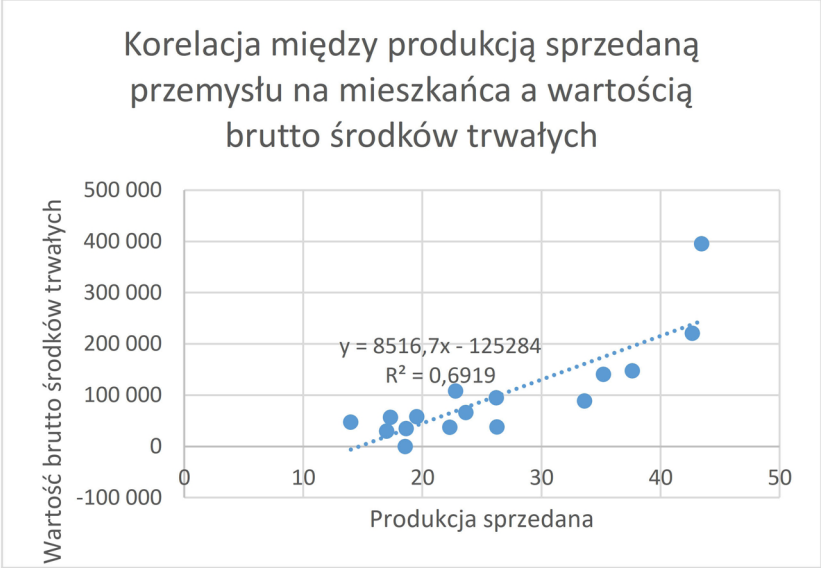
Krok 3.

Klikamy prawym przyciskiem myszy na dowolny punkt na wykre-

sie. Z menu kontekstowego wybieramy *Dodaj linię trendu*.



Wybieramy liniową linię trendu oraz zaznaczamy *Wyświetl równanie na wykresie* oraz *Wyświetl wartość R-kwadrat na wykresie*. Po sformatowaniu wykres wygląda podobnie jak na rysunku poniżej.



Interpretacja: Występuje silna korelacja między produkcją sprzedaną przemysłu na mieszkańca a wartością brutto środków trwałych. Im większa wartość środków trwałych, tym większa produkcja sprzedana.

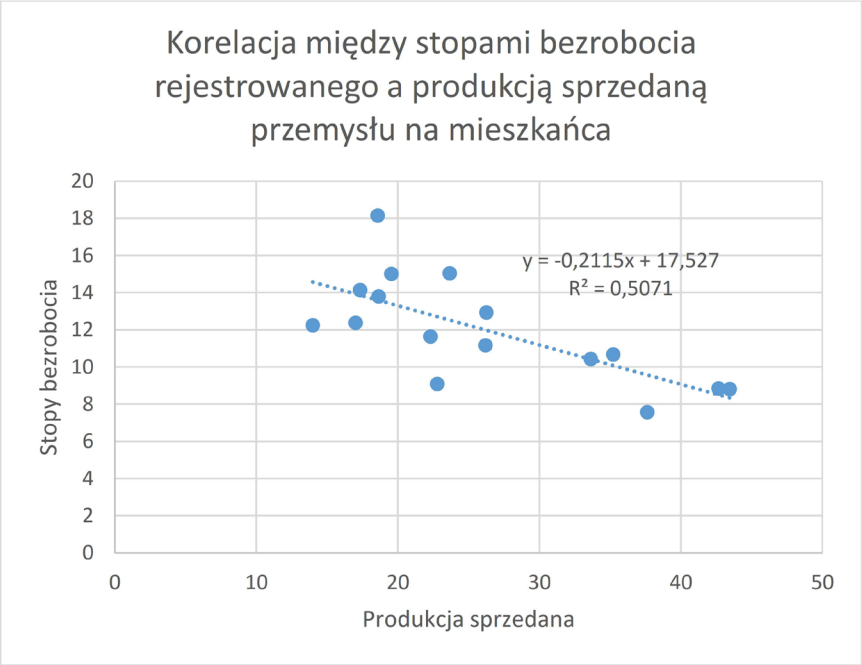
Zależność można opisać prostą o równaniu: Wartość brutto środków trwałych = 8516,7 Produkcja sprzedana przemysłu + 125284. Oznacza to, że zwiększenie Produkcja sprzedana przemysłu na 1 powoduje wzrost wartości brutto środków trwałych o 8516.7 tys. zł

W kolejnych podpunktach należy powtórzyć kolejne kroki z pierwszego podpunktu. Poniżej znajdują się wyniki wykonanych obliczeń.

2.

Województwo	Produkcja sprzedana przemysłu	Stopa bezrobocia rejestr. %
Dolnośląskie	35,22	10,67
Kujawsko-pomorskie	23,66	15,04
Lubelskie	13,99	12,24
Lubuskie	26,26	12,93

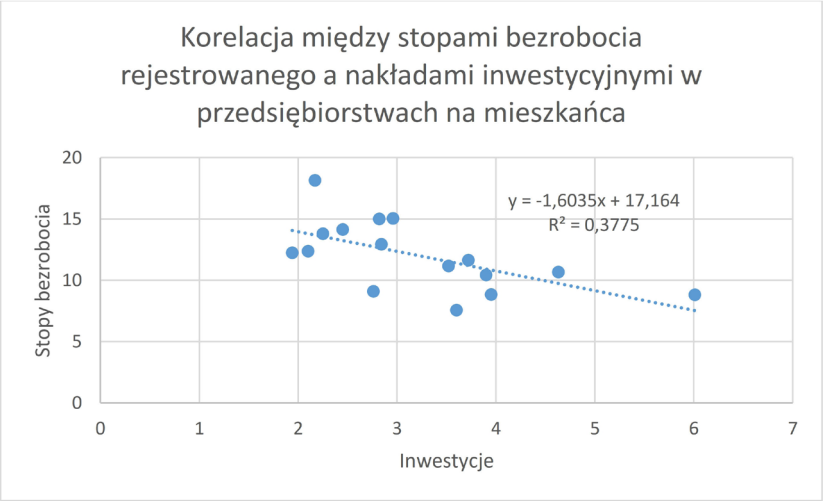
Województwo	Produkcja sprzedana przemysłu	Stopa bezrobocia rejestr. %
Łódzkie	26,2	11,16
Małopolskie	22,78	9,09
Mazowieckie	43,45	8,81
Opolskie	22,31	11,64
Podkarpackie	17,33	14,14
Podlaskie	17,01	12,37
Pomorskie	33,63	10,42
Śląskie	42,67	8,84
Świętokrzyskie	18,64	13,8
Warmińsko-mazurskie	18,57	18,14
Wielkopolskie	37,63	7,56
Zachodniopomorskie	19,54	15



współczynnik korelacji: -0,71213

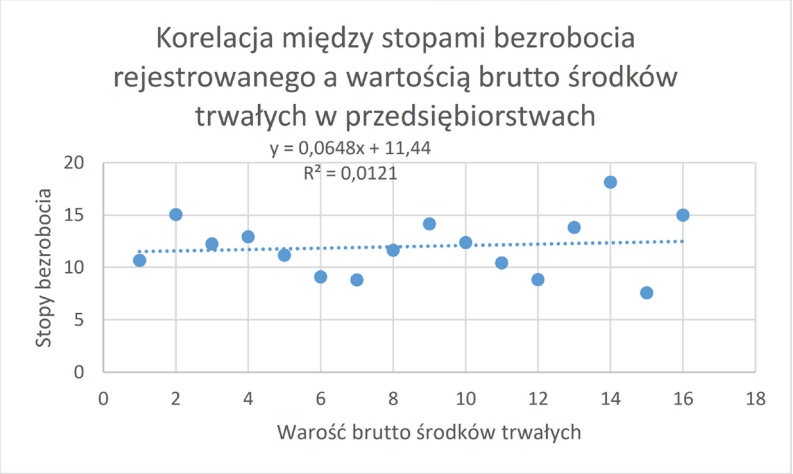
3.

Województwo	Inwestycje	Stopa bezrobocia rejestr. %
Dolnośląskie	4,63	10,67
Kujawsko-pomorskie	2,96	15,04
Lubelskie	1,94	12,24
Lubuskie	2,84	12,93
Łódzkie	3,52	11,16
Małopolskie	2,76	9,09
Mazowieckie	6,01	8,81
Opolskie	3,72	11,64
Podkarpackie	2,45	14,14
Podlaskie	2,1	12,37
Pomorskie	3,9	10,42
Śląskie	3,95	8,84
Świętokrzyskie	2,25	13,8
Warmińsko-mazurskie	2,17	18,14
Wielkopolskie	3,6	7,56
Zachodniopomorskie	2,82	15



4.

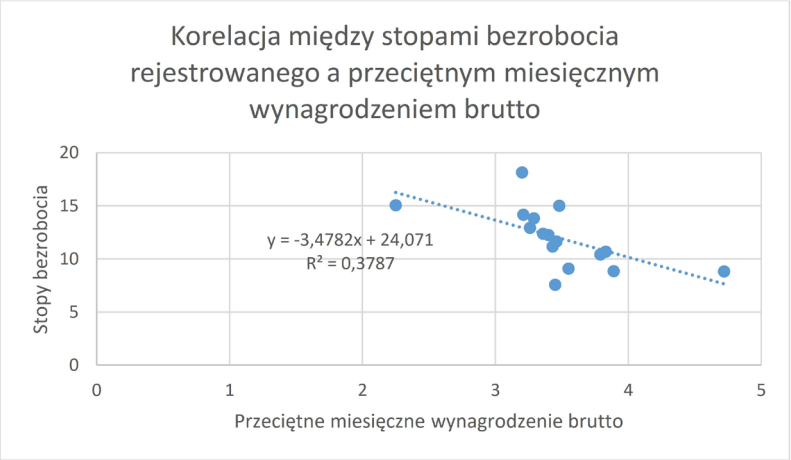
Województwo	Stopa bezrobocia rejestr. %	Wartość brutto środków trwałych
Dolnośląskie	10,67	140 713
Kujawsko-pomorskie	15,04	66 028
Lubelskie	12,24	47 076
Lubuskie	12,93	37 636
Łódzkie	11,16	95 203
Małopolskie	9,09	107 796
Mazowieckie	8,81	395 519
Opolskie	11,64	37 223
Podkarpackie	14,14	56 803
Podlaskie	12,37	29 497
Pomorskie	10,42	88 496
Śląskie	8,84	220 977
Świętokrzyskie	13,8	34 817
Warmińsko-mazurskie	18,14	34 327
Wielkopolskie	7,56	147 190
Zachodniopomorskie	15,00	58 055



współczynnik korelacji: -0,651

5.

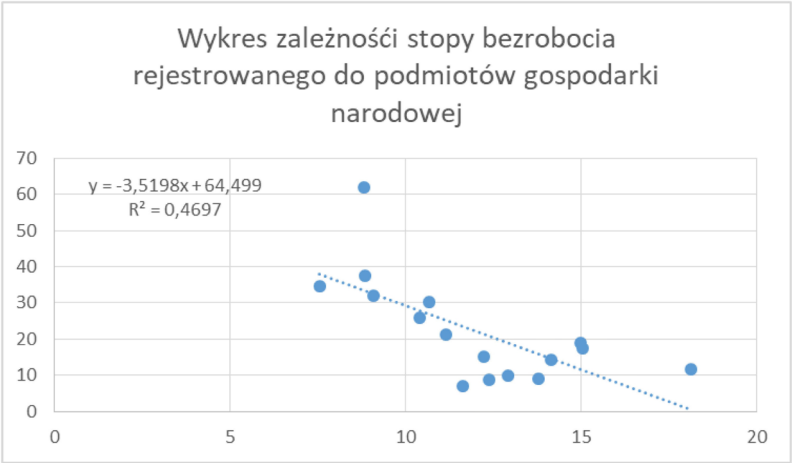
Województwo	Stopa bezrobocia rejestr. %	Wartość brutto środków trwałych
Dolnośląskie	10,67	3,83
Kujawsko-pomorskie	15,04	2,25
Lubelskie	12,24	3,4
Lubuskie	12,93	3,26
Łódzkie	11,16	3,43
Małopolskie	9,09	3,55
Mazowieckie	8,81	4,72
Opolskie	11,64	3,46
Podkarpackie	14,14	3,21
Podlaskie	12,37	3,36
Pomorskie	10,42	3,79
Śląskie	8,84	3,89
Świętokrzyskie	13,8	3,29
Warmińsko-mazurskie	18,14	3,2
Wielkopolskie	7,56	3,45
Zachodniopomorskie	15,00	3,48



współczynnik korelacji: -0,615

6.

Województwo	Stopa bezrobocia rejestr. %	Podmioty gospodarki narodowej
Dolnośląskie	10,67	30,27
Kujawsko-pomorskie	15,04	17,60
Lubelskie	12,24	15,24
Lubuskie	12,93	10,08
Łódzkie	11,16	21,39
Małopolskie	9,09	32,03
Mazowieckie	8,81	61,95
Opolskie	11,64	7,09
Podkarpackie	14,14	14,40
Podlaskie	12,37	8,86
Pomorskie	10,42	25,85
Śląskie	8,84	37,42
Świętokrzyskie	13,8	9,22
Warmińsko-mazurskie	18,14	11,56
Wielkopolskie	7,56	34,73
Zachodniopomorskie	15,00	19,03



współczynnik korelacji: -0,685

Przypadek 2

Polecenie:

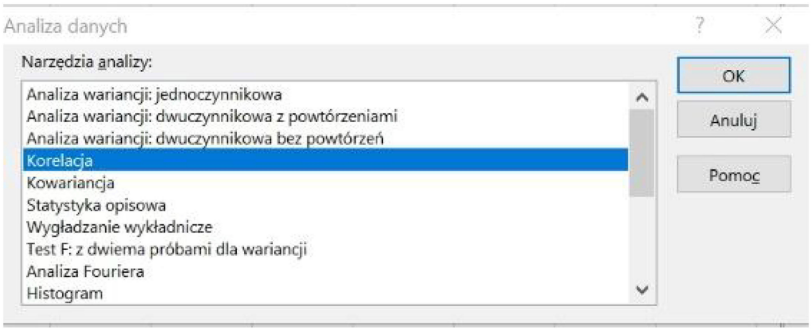
Dla danych opisanych w poniższej tabeli (patrz Załacznik5.xlsx) wyznacz współczynnik korelacji.

Wiek	Waga	Wzrost
25	55	169
31	62	176
26	55	169
34	69	178
46	71	181
51	65	173
27	58	170
29	68	175
30	70	179
44	82	184
24	58	172
36	68	174
41	57	165

Rozwiązanie:

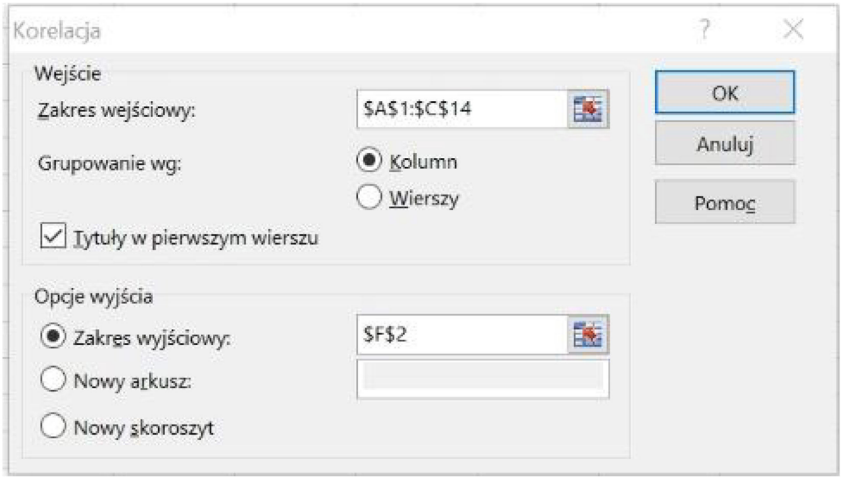
Krok 1.

Z karty *Dane* wybierz *Analiza danych,* a następnie *Korelację*.



Krok 2.

Kliknij przycisk *OK*, aby wyświetlić okno dialogowe *Korelacja* pokazane na rysunku poniżej. W dalszej kolejności uzupełnij pole *Zakres wejściowy* (tabela z danymi). Zaznacz opcję *Tytuły w pierwszym wierszu*. Kliknij pole edycji *Zakres wyjściowy*, a następnie komórkę arkusza, gdzie chcesz, aby zaczęło się wyświetlanie wyników. Kliknij przycisk *OK*, aby zacząć analizę.



Krok 3.

Wyniki korelacji są zaprezentowane poniżej. Na głównej przekątnej mamy wartości równe 1. Powyżej głównej przekątnej współczynniki korelacji nie są wyświetlane, ponieważ byłyby takie same jak te poniżej.

	Wiek	*Waga*	*Wzrost*
Wiek	1		
Waga	0,545735	1	
Wzrost	0,362237	0,912871	1

Przypadek 3

Polecenie:

Dane dotyczące sprzedaży (w kolejnych tygodniach) pewnego artykułu przedstawia poniższa tabela oraz Załącznik7.xlsx

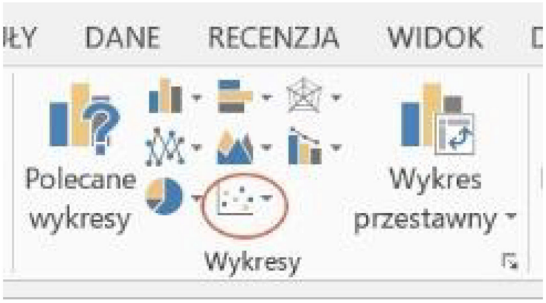
Tydzień	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
Sprzedaż w sztukach	23	26	25	26	29	30	32	35	34	38	40	39

a) Przedstaw graficznie dane.
b) Wyznacz równanie regresji.

Rozwiązanie:

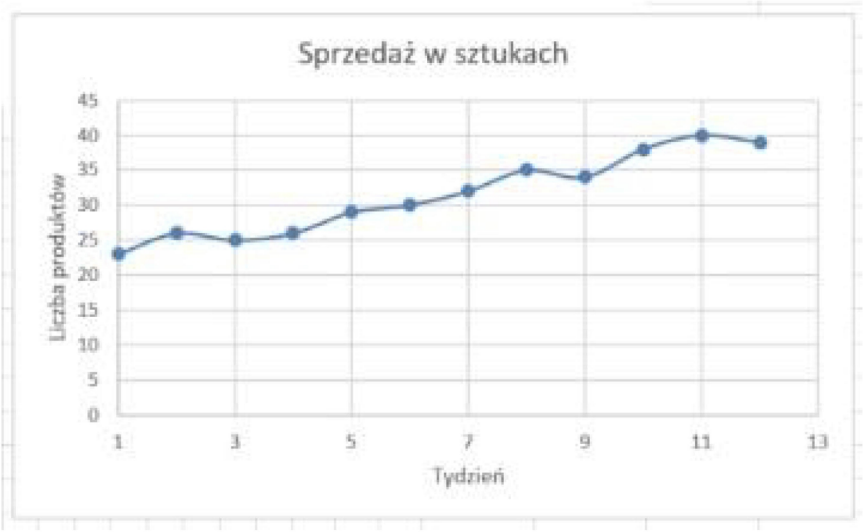
Krok 1.

Zaznaczamy dane. Następnie z zakładki *Wstawianie/Wykresy* wybieramy *Wykres punktowy z wygładzonymi liniami i znacznikami*.



Krok 2.

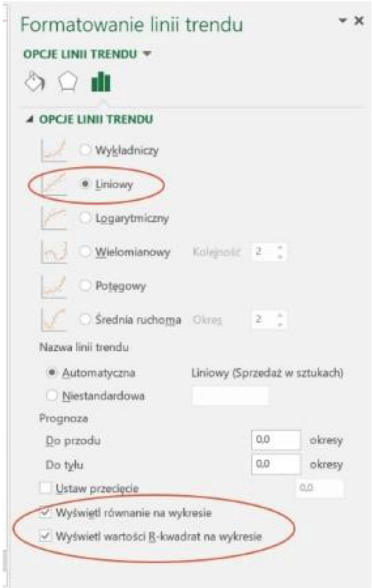
Otrzymamy graficzną ilustrację danych.

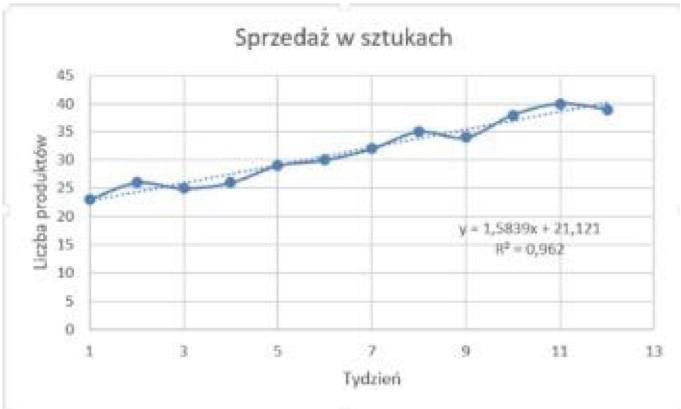


Na podstawie wykresu możemy stwierdzić, że sprzedaż wykazuje tendencję rosnącą (trend rosnący).

Krok 3.

Dodatkowo na wykresie możemy zamieścić linię trendu. Aby ją uzyskać, prawym klawiszem myszy klikamy w jeden z puktów wykresu i wybieramy *Dodaj linię trendu*. Otwiera się wtedy okno dialogowe *Formatowania linii trendu*. Po wybraniu poniższych opcji otrzymamy wykres z liniową funkcją trendu.





Na wykresie są zaznaczone linia regresji oraz wzór opisujący otrzymane równanie regresji.

Przypadek 4

Polecenie:

Przedsiębiorstwo produkujące damską odzież ma zakłady w ośmiu miastach Polski. Kierownictwo dostaje miesięczny raport o wartości średniej miesięcznej produkcji na osobę (zł) oraz o średniej miesięcznej wypłacie (zł). W ostatnim raporcie zamieszczono następujące dane:

Zakład:	1	2	3	4	5	6	7	8
Średnia miesięczna produkcja na os. (zł)	9850	11000	8250	12100	7200	9950	9500	11350
Średnia miesięczna wypłata (zł)	1050	1200	900	1300	800	1100	1000	1250

Zbadaj, czy między cechami istnieje zależność liniowa za pomocą współczynnika korelacji liniowej *r*.

Rozwiązanie:

Krok 1.

Obliczamy średnią arytmetyczną miesięcznej produkcji oraz wypłaty.

średnia produkcja	9900
średnia wypłata	1075

Krok 2.

Następnie do tabeli dodajemy nową kolumnę i wpisujemy do niej obliczoną różnicę pomiędzy miesięczną produkcją a średnią. Czynność powtarzamy dla miesięcznej wypłaty.

	A	B	C	D
	średnia miesięczna produkcja na osobę - x	średnia miesięczna wypłata - y	xi-x	yi-y
	9850	1050	-50	1050
	11000	1200	1100	125
	8250	900	-1650	-175
	12100	1300	2200	225
	7200	800	-2700	-275
	9950	1100	50	25
	9500	1000	-400	-75
	11350	1250	1450	175
0				
1				
2	średnia produkcja	9900		
3	średnia wypłata	1075		
4				

Krok 3.

Dodajemy kolejne kolumny do tabeli, wykonując odpowiednie obliczenia.

	A	B	C	D	E	F	G
1	średnia miesięczna produkcja na osobę - x	średnia miesięczna wypłata - y	xi-x	yi-y	(xi-x)*(yi-y)	(xi-x)^2	(yi-y)^2
2	9850	1050	-50	1050	-52500	2500	1102500
3	11000	1200	1100	125	137500	1210000	15625
4	8250	900	-1650	-175	288750	2722500	30625
5	12100	1300	2200	225	495000	4840000	50625
6	7200	800	-2700	-275	742500	7290000	75625
7	9950	1100	50	25	1250	2500	625
8	9500	1000	-400	-75	30000	160000	5625
9	11350	1250	1450	175	253750	2102500	30625
10				Suma	1896250	18330000	1311875
11							
12	średnia produkcja	9900					
13	średnia wypłata	1075					
14							

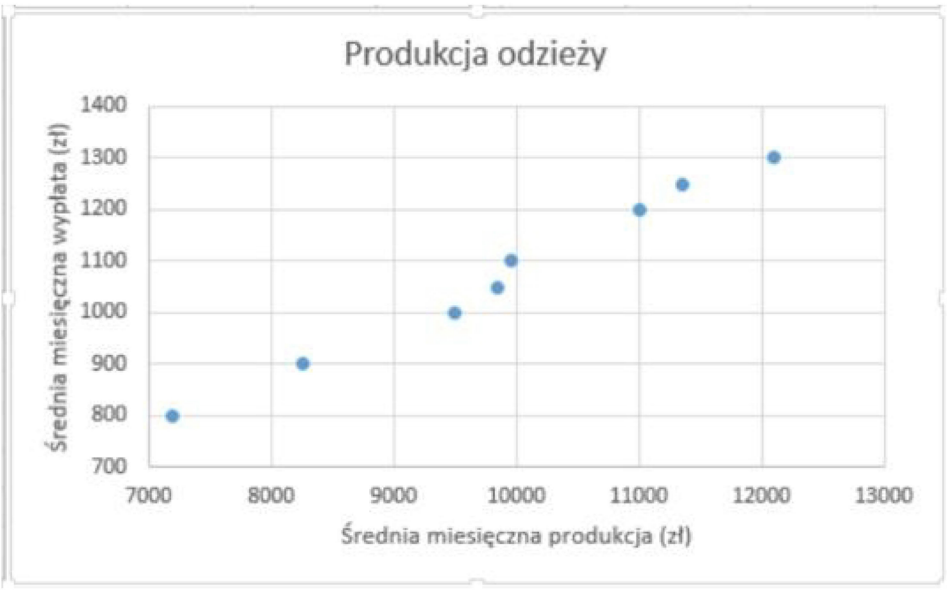
Krok 4.

Obliczamy korelację, wpisując w komórce B15 wyrażenie =E10/ PIERWIASTEK(F10*G10)

4		
5	koleracja	0,39
6		

Krok 5.

Zaznaczając dane, wybieramy wykres punktowy.



Zadania

Zadanie 1.

Przedsiębiorstwo produkujące męską odzież ma zakłady w ośmiu miastach Polski. Kierownictwo dostaje miesięczny raport o wartości średniej miesięcznej produkcji na osobę (zł) oraz o średniej miesięcznej wypłacie (zł). W ostatnim raporcie zamieszczono następujące dane:

Zakład:	1	2	3	4	5	6	7	8
Średnia miesięczna produkcja na os. (zł)	8850	11200	8350	12100	9200	9950	9500	11350
Średnia miesięczna wypłata (zł)	950	1205	1090	1300	890	1100	1000	1250

Zbadaj, czy między cechami istnieje zależność liniowa za pomocą współczynnika korelacji liniowej *r*.

.....

.....

.....

Zadanie 2.

Pięć zakładów produkujących ten sam wyrób zbadano ze względu na wielkość produkcji oraz koszty jednostkowe. Wyniki przedstawiono w tabeli:

X – produkcja w tys. sztuk	2	1	4	3	5
Y – koszty jednostkowe w zł	9	8	1	5	2

a) Oblicz współczynnik korelacji i zinterpretuj go.
b) Wyznacz prostą regresji kosztów jednostkowych względem produkcji.

c) Zinterpretuj parametr a prostej regresji kosztów jednostkowych względem produkcji.
d) Wyznacz współczynnik korelacji tylko na podstawie prostych regresji.
e) Oceń dopasowanie prostych regresji do punktów empirycznych.

.....
.....
.....
.....
.....

Zadanie 3.

W pewnej firmie produkcyjnej badano zależność zużycia głównego surowca od wielkości produkcji i otrzymano informacje zgodne z poniższą tabelą:

Zużycie surowca w tonach	1	2	3	4	5	6	7	8	9	10	11	12
Wielkość produkcji w tys. kg	8	10	13	18	20	21	22	24	25	27	30	32

a) Określ zmienną zależną i niezależną.
b) Wyznacz liniową funkcję regresji opisującą wpływ zużycia surowca na wielkość produkcji.
c) Wyznacz parametry dopasowania liniowej funkcji regresji do rzeczywistej wielkości produkcji.
d) Oszacuj wielkość produkcji, gdy zużycie surowca będzie na poziomie 15 ton.

.....
.....
.....
.....
.....
.....

Zadanie 4.

Staż w zawodzie i miesięczne zarobki (w tys. zł) w pewnej lubelskiej firmie są następujące

Staż (w latach)	1	2	3	4	5	6	7
Wynagrodzenie (w tys. zł)	2,5	2,7	2,8	3,2	3,2	3,6	4,3

a) Sporządź wykres rozrzutu punktów empirycznych i na jego
b) podstawie oceń, czy mamy liniową zależność między analizowanymi zmiennymi.
c) Oszacuj i zinterpretuj parametry równania.
d) Oceń jakość modelu, stosując odpowiedni miernik statystyczny.

.....

.....

.....

Zadanie 5.

Średni dzienny obrót wybranych cukierni lubelskich oraz powierzchnię ich sklepów przedstawia poniższa tabela.

Dzienny obrót (w tys. zł)	2	4	6	8	10	12
Powierzchnia cukierni (w m2)	30	45	50	55	62	70

a) Przedstaw dane na wykresie, na jego podstawie oceń wstępnie związek korelacyjny pomiędzy badanymi cechami.
b) Zweryfikuj przypuszczenia dotyczące korelacji metodą analityczną, stosując odpowiedni miernik.
c) Określ, która ze zmiennych jest zależna, a która niezależna oraz wyznacz funkcję regresji liniowej.
d) Zbadaj stopień dopasowania teoretycznej linii regresji do danych empirycznych.

.....

.....

.....

Zadanie 6.

Tabela poniżej przedstawia zależność ilości czasu przeznaczonego na naukę (w godz.) i ilości punktów zdobytych na egzaminie Statystyka. Na tej podstawie:

a) Określ korelację na podstawie wykresu.
b) Wyznacz i zinterpretuj równanie regresji liniowej.
c) Podaj prawdopodobną ilość uzyskanych punktów przez osobę, która uczyła się do egzaminu przez 8 godzin.

Ilość punktów	15	30	41	36	50	65
Czas nauki	2	4	5	6	7	9

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Studium przypadku 5.
Analiza dynamiki

Cel ćwiczenia

Poznanie metod i nabycie praktycznych umiejętności rozwiązywania problemów mieszczących się w zakresie **podstawowych metod badania dynamiki zjawisk rynkowych** oraz wypracowanie umiejętności analizowania i interpretowania rezultatów badań. W szczególności poruszane casy dotyczą prezentacji podstawowych metod analizy dynamiki zjawisk prezentowanych w postaci szeregów czasowych oraz kierunków i możliwości zastosowania tych metod w analizie dynamiki procesów rynkowych.

Wstęp teoretyczny

WAŻNE

Wzory

Przyrost bezwzględny o podstawie stałej jest różnicą między wielkością zjawiska w okresie badanym i podstawowym

$$\Delta y_s = y_n - y_t$$

gdzie:

Δy_s – przyrost absolutny o podstawie stałej

y_n – wielkość zjawiska w badanym okresie

y_t – wielkość zjawiska w okresie podstawowym

n – oznaczenie kolejnego okresu

t – oznaczenie okresu podstawowego

WAŻNE

Wzory

Przyrost bezwzględny o podstawie zmiennej jest różnicą między wielkością zjawiska w okresie badanym i poprzedzającym okres badany

$$\Delta y_{n/(n-1)} = y_n - y_{n-1}$$

gdzie:

$\Delta y_{n/(n-1)}$ – przyrost bezwzględny o podstawie stałej

y_n – wielkość zjawiska w badanym okresie

y_{n-1} – wielkość zjawiska w okresie poprzedzającym okres badany

n – oznaczenie kolejnego okresu

WAŻNE

Wzory

Szereg łańcuchowy przyrostów względnych tworzymy, dzieląc kolejne przyrosty absolutne łańcuchowe przez wartości cechy w okresie poprzednim

$$\Delta y_{n/(n-1)} = \frac{y_n - y_{n-1}}{y_{n-1}}$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_{n-1} – wielkość zjawiska w okresie poprzedzającym badany okres

n – oznaczenie kolejnego okresu

Wzory

Szereg jednopodstawowy przyrostów względnych tworzymy, dzieląc kolejne przyrosty absolutne łańcuchowe przez wartości cechy w okresie bazowym

$$\Delta y_{n/i} = \frac{y_n - y_i}{y_i}$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_i – wielkość zjawiska w okresie podstawowym

n – oznaczenie kolejnego okresu

i – oznaczenie okresu podstawowego

Wzory

Tempo wzrostu o podstawie zmiennej

$$T_z = \frac{y_n - y_{n-1}}{y_{n-1}} \cdot 100\%$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_{n-1} – wielkość zjawiska w okresie poprzedzającym badany okres

n – oznaczenie kolejnego okresu

Wzory

Tempo wzrostu o podstawie stałej

$$T_s = \frac{y_n - y_i}{y_i} \cdot 100\%$$

gdzie:

y_n – wielkość zjawiska w badanym okresie

y_i – wielkość zjawiska w okresie podstawowym

n – oznaczenie kolejnego okresu

i – oznaczenie okresu podstawowego

Definicja

Indeksem dynamiki nazywamy stosunek wielkości badanego zjawiska w danym okresie (momencie) do wielkości tego samego zjawiska w innym okresie (momencie), przyjętym za podstawę porównań:

$$i = \frac{y_n}{y_0} \quad \text{lub} \quad i = \frac{y_n}{y_0} \cdot 100\%$$

gdzie:

y_n – poziom zjawiska w pewnym okresie

y_0 – poziom zjawiska w okresie odniesienia

Definicja

Indeksy jednopodstawowe informują, jakie zmiany nastąpiły w poziomie zjawiska w kolejnych okresach w stosunku do okresu przyjętego jako podstawa (bazowego).

$$\frac{y_1}{y_0}, \frac{y_2}{y_0}, \frac{y_3}{y_0}, \ldots, \frac{y_n}{y_0}$$

Definicja

Indeksy łańcuchowe (o podstawie zmiennej) informują, jakie zmiany nastąpiły w poziomie zjawiska w kolejnym okresie w stosunku do okresu go poprzedzającego.

$$\frac{y_2}{y_1}, \frac{y_3}{y_2}, \frac{y_4}{y_3}, \ldots, \frac{y_n}{y_{n-1}}$$

Przypadek 1

Polecenie:

Liczbę studentów na Wydziale Ekonomicznym pewnej uczelni w latach 2014–2021 przedstawia następujący szereg:

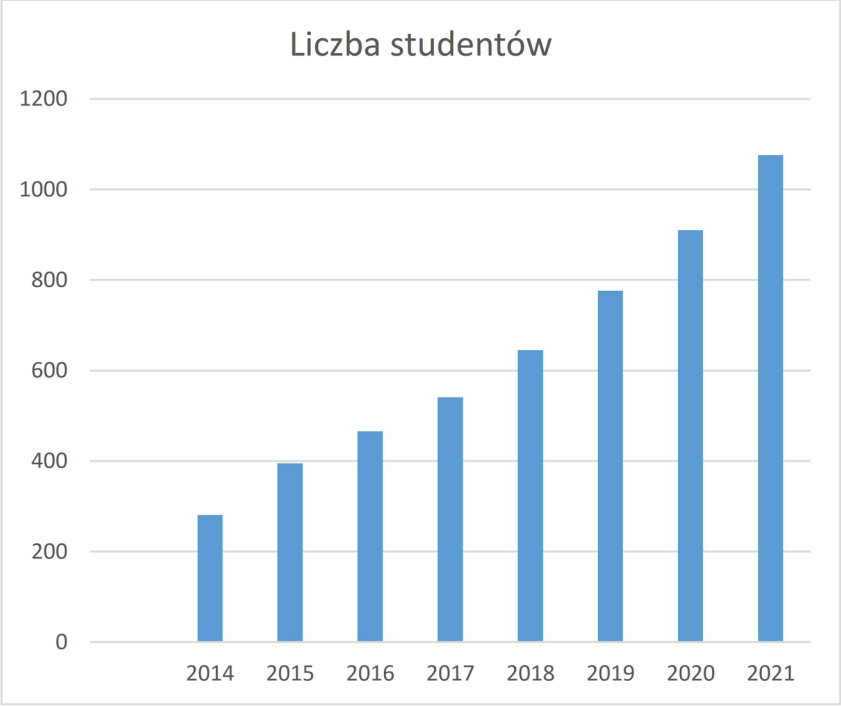
Lata	Liczba studentów
2014	281
2015	395
2016	466
2017	541
2018	645
2019	776
2020	910
2021	1076

1) Przedstaw graficznie dane.
2) Oblicz przyrosty absolutne łańcuchowe. Podaj interpretację dla 2019 roku.
3) Oblicz przyrosty jednopodstawowe, przyjmując poziom z roku 2014 za podstawę porównań. Podaj interpretację dla 2021 roku.
4) Oblicz indeksy łańcuchowe. Podaj interpretację dla 2021 roku.
5) Oblicz indeksy jednopodstawowe, przyjmując poziom z roku 2021 za podstawę porównań. Podaj interpretację dla 2021 roku.

Rozwiązanie:

Krok 1.

Zaznaczamy dane. Z zakładki *Wstawianie/Wykresy* wybieramy *Wykres kolumnowy grupowany*.



Krok 2.

W komórce C4 wpisujemy =B4-B3, a następnie kopiujemy na całą kolumnę C.

W komórce D3 wpisujemy =B3-B3, a następnie kopiujemy na całą kolumnę D.

W komórce E4 wpisujemy =B4/B3*100, a następnie kopiujemy na całą kolumnę E.

W komórce F3 wpisujemy =B3/B3*100, a następnie kopiujemy na całą kolumnę F.

Krok 3.

Wprowadzając odpowiednie formuły w Excelu, otrzymujemy wyniki jak w tabeli poniżej.

	A	B	C	D	E	F
1	Lata	Liczba studentów	Zmiana liczby studentów		Rok poprzedni = 100	Rok 2014 = 100
2			Do roku poprzedniego	Do roku 2014		
3	2014	281		0		100
4	2015	395	114	114	140,6	140,6
5	2016	466	71	185	118,0	165,8
6	2017	541	75	260	116,1	192,5
7	2018	645	104	364	119,2	229,5
8	2019	776	131	495	120,3	276,2
9	2020	910	134	629	117,3	323,8
10	2021	1076	166	795	118,2	382,9
11						

Interpretacja:

- w roku 2021 było o 166 studentów więcej niż w roku 2020;
- w roku 2021 było o 795 więcej studentów niż w roku 2014;
- 118,2 w roku 2021 było o 18,2% studentów więcej niż w roku 2020 (tempo wzrostu wynosiło 18,2%);
- 382,9 w roku 2021 było o 282,9% studentów więcej niż w roku 2014 (wzrost 3,8 raza).

Przypadek 2

Polecenie:

Dane dotyczące importu cukru w tys. ton w Polsce w latach 2016–2020 zebrano w tabeli:

Lata	Okres t	Import cukru (tys. ton) y_t
2016	1	23 957
2017	2	14 351
2018	3	18 813
2019	4	15 367
2020	5	19 273

Na podstawie tych danych oblicz i zinterpretuj:

a) indeksy dynamiki jednopodstawowe i łańcuchowe,
b) średniookresowe tempo zmian w czasie.

Rozwiązanie:

Krok 1.

Do obliczenia indeksów dynamiki jednopodstawowych wykorzystamy wzór:

$$\frac{y_1}{y_0}, \dots \frac{y_n}{y_0}$$

Do obliczenia indeksów dynamiki łańcuchowych wykorzystamy wzór:

$$\frac{y_2}{y_0 1}, \dots \frac{y_n}{y_{n-1}}$$

Aby obliczyć średniookresowy indeks łańcuchowy, wykorzystamy wzór:

$$G = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Aby obliczyć średniookresowe tempo zmian w czasie, wykorzystamy wzór:

$$T = G - 1$$

Krok 2.

Wprowadzamy dane do arkusza kalkulacyjnego. Wpisujemy następujące formuły:

W komórce D3 wyrażenie =C3/C$3*100, następnie kopiujemy je na całą kolumnę D.

W komórce E4 wyrażenie =C4/C3*100, następnie kopiujemy je na całą kolumnę E.

Otrzymujemy poniższą tabelę.

	A	B	C	D	E
1	Lata	Okres t	Import cukru (tys. ton) y_t	Indeksy dynamiki jednopodstawowe	Indeksy dynamiki łańcuchowe
2				(względem roku 2016)	(względem poprzedniego roku)
3	2016	1	23 957	100	
4	2017	2	14 351	60	60
5	2018	3	18 813	79	131
6	2019	4	15 367	64	82
7	2020	5	19 273	80	125

Krok 3.

Podajemy interpretację otrzymanych wartości.
W roku 2017 było o 40% mniej importu cukru niż w roku 2016.
W roku 2018 było o 21% mniej importu cukru niż w roku 2016.
W roku 2019 było o 36% mniej importu cukru niż w roku 2016.
W roku 2020 było o 20% mniej importu cukru niż w roku 2016.
W roku 2017 było o 40% mniej importu cukru niż w roku 2016.
W roku 2018 było o 31% więcej importu cukru niż w roku 2017.
W roku 2019 było o 18% mniej importu cukru niż w roku 2018.
W roku 2020 było o 25% więcej importu cukru niż w roku 2019.

Krok 4.

W komórce F9 wpisujemy wyrażenie =(C7/C3)^(1/4), otrzymując

średniookresowe tempo zmian w czasie.	0,95

Przypadek 3

Polecenie:

Dynamikę cen pewnego artykułu w latach podano w tabeli (w tabeli podano indeksy jednopodstawowe):

Lata	2016	2017	2018	2019	2020
Rok 2016 = 100%	106	101	92	87	102

a) Porównaj cenę artykułu w roku 2016 z ceną w roku 2020.
b) Jeśli cena w roku 2017 wynosiła 167 zł, to jaka była cena w 2020?

Rozwiązanie:

Krok 1.

Cena artykułu w 2020 roku w porównaniu do roku 2016 była o 2% wyższa, ponieważ indeks jednopodstawowy wynosi 102, czyli stosunek ceny w 2020 roku do ceny w 2016 roku to 1,02.

Krok 2.

Wprowadzamy dane do arkusza kalkulacyjnego i wpisujemy kolejno:

w komórce C2 wartość =C3*100/B3
w komórce C2 wartość 167
w komórce C2 wartość =B4*C$2/100 i kopiujemy na całą kolumnę C

	A	B	C
1	Lata	Indeksy jednopodstawowe	cena [zł]
2	2016	106	165,35
3	2017	101	167,00
4	2018	92	152,12
5	2019	87	143,85
6	2020	102	168,65

Przypadek 4

Polecenie:
Dynamikę cen pewnego produktu podaje tabelka:

Lata	2016	2017	2018	2019	2020
Rok poprzedni = 100 %	95	119	105	117	91

a) Oblicz cenę produktu w roku 2020, wiedząc, że cena produktu w roku 2016 wynosiła 29 zł.
b) Oblicz średniookresowe tempo zmian ceny w latach 2016–2020.

Rozwiązanie:

Krok 1.

Cena produktu w 2020 roku wynosiła ok. 38.58 zł.
Cenę produktów obliczamy ze wzoru

$$y_n = \frac{y_{n-1} * i_n}{100}$$

gdzie:
i_n – indeks łańcuchowy dla danego roku w %
y_n – cena w danym roku

Średniookresowe tempo zmiany cen w latach 2016–2020 liczymy ze wzoru:

$$T_z = \frac{y_n - y_{n-1}}{y_{n-1}} * 100\%$$

Po przekształceniu można obliczyć ze wzoru

$$T_z = \left(\frac{i_n}{100} - 1\right) * 100\%$$

i_n – indeks łańcuchowy dla danego roku w %

Krok 2.

Wprowadzamy dane do arkusza. Wpisujemy kolejno:

W komórce C2 wartość 29.
W komórce C3 wartość =C2*B3/100, kopiujemy ją na całą kolumnę C.
W komórce B2 wartość =(B2/100-1)*100, kopiujemy ją na całą kolumnę B.

	A	B	C	D
1	Lata	indeks łańcuchowy	cena [zł]	Średniookresowe tempo zmiany cen
2	2016	95	29,00	-5
3	2017	119	34,51	19
4	2018	105	36,24	5
5	2019	117	42,40	17
6	2020	91	38,58	-9

Zadania

Zadanie 1.

Dynamika eksportu zboża w latach 2010–2014 była następująca:

Lata	2010	2011	2012	2013	2014
Indeksy łańcuchowe	70	100	92	105	110

a) Oblicz indeks i zinterpretuj.
b) Jeśli eksport w roku 2014 wynosił 60 mln ton, to jaki był eksport w roku 2010?

.....
.....
.....

Zadanie 2.

Obroty sklepu w roku 2015 oraz w roku 2019 przedstawia poniższa tabelka.

Produkt	**Wielkość sprzedaży (w tys. sztuk)**		**Ceny jednostkowe (w zł)**	
	2015	**2019**	**2015**	**2019**
telewizor	1,5	1,2	1400	2100
zmywarka	5	3	1300	1200
lodówka	3,4	4	1500	1400

Odpowiedz na następujące pytania:
a) Jak zmieniła się wartość obrotów sklepu w roku 2019 w porównaniu z rokiem 2015?
b) Jaki wpływ na zmiany obrotów miały ceny, a w jakim stopniu wpłynęła na to zmiana ilości sprzedawanych wyrobów?

.....
.....
.....

Zadanie 3.

Na podstawie danych zawartych w tabeli poniżej oblicz:

a) Agregatowy indeks wartości produkcji dla 2019 w stosunku do 2017.
b) Agregatowy indeks produkcji dla 2019 w stosunku do 2017.
c) Agregatowe indeksy cen dla 2019 w stosunku do 2017.

Produkt	Wielkość sprzedaży (w tys. sztuk)		Ceny jednostkowe (w zł)	
	2017	2019	2017	2019
X	1	2,1	200	250
Y	2,1	5,3	130	152

.....
.....
.....
.....
.....

Zadanie 4.

Dynamikę produkcji przez pewną fabrykę w latach 2014–2020 opisuje następujący ciąg indeksów łańcuchowych: 1,6; 1,2; 1,3; 1,4; 0,8; 0,9.

a) Jak zmieniała się wielkość produkcji w poszczególnych latach w stosunku do roku 2014?
b) O ile średnio zmieniała się sprzedaż tego artykułu z roku na rok w badanym okresie?
c) Jeżeli sprzedaż tych produktów w 2014 r. wyniosła 60 tys. sztuk, to jakiej wielkości produkcji należy spodziewać się w 2021 roku?

.....
.....
.....
.....

Bibliografia

1. Domański, C. (Ed.). (2001). Metody statystyczne: teoria i zadania. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
2. Greń, J. (1982). Statystyka matematyczna: modele i zadania. Warszawa: Państwowe Wydawnictwo Naukowe.
3. Jóźwiak, J., Podgórski, J. (2006). Statystyka od podstaw. Warszawa: PWE.
4. Kassyk-Rokicka, H. (1999). Statystyka. Zbiór Zadań. Warszawa: PWE.
5. Kassyk-Rokicka, H. (1992). Statystyka nie jest trudna. Warszawa: PWE.
6. Kończak G., Trzpiot G. (2002). Analizy statystyczne z arkuszem kalkulacyjnym Microsoft Excel, Katowice: Wydawnictwo AE w Katowicach.
7. Kukuła, K. (2011). Elementy statystyki w zadaniach. Warszawa: Wydawnictwo Naukowe PWN.
8. Maksimowicz-Ajchel, A. (2019). Wstęp do statystyki: metody opisu statystycznego. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
9. Middleton, M.R. (2004). Microsoft Excel w analizie danych. Warszawa: Wydawnictwo RM.
10. Ostasiewicz, S., Rusnak, Z., Siedlecka, U. (2006). Statystyka: elementy teorii i zadania. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
11. Panek T. (red.). (2007). Statystyka społeczna. Warszawa: PWE.
12. Piłatowska M. (2007). Repetytorium ze statystyki. Warszawa: Wydawnictwo Naukowe PWN.
13. Pułaska-Turyna, B. (2008). Statystyka dla ekonomistów. Warszawa: Difin.
14. Regel W. (2003). Podstawy statystyki w Excelu. Warszawa: Mikom.
15. Rószkiewicz, M. (2002). Statystyka: kurs podstawowy. Warszawa: Efekt.

16. Wieczorkowska, G., Wierzbiński, J. (2007). Statystyka. Analiza badań społecznych, Warszawa: Wyd. Naukowe Scholar.
17. Zeliaś, A. (2000). Metody statystyczne. Warszawa: PWE.

Z recenzji:

Podręcznik upowszechnia metody statystyki opisowej oraz elementy statystyki matematycznej, służące wykrywaniu prawidłowości w zjawiskach gospodarczych i ekonomicznych. Autorka, oprócz omówienia podstawowych pojęć i metod badań statystycznych, przedstawia wyczerpujący opis cech zbiorowości statystycznej, oceny położenia rozkładu, klasyczne miary położenia obliczane na podstawie wszystkich obserwacji, pozycyjne miary położenia, czyli wartość miary, jakie wskazuje dana jednostka, a także oceny niezbędne do badania stopnia zróżnicowania wartości zmiennej, służące do badania asymetrii rozkładu i do analizy stopnia skupienia poszczególnych jednostek wokół średniej, określenia siły, kierunku oraz kształtu związku między badanymi cechami pewnej zbiorowości.

Prezentacja graficzna danych zdecydowanie wzbogaciła wszystkie objaśnione pojęcia i metody statystyczne. Poza tym opracowanie obejmuje zastosowania pakietu EXCEL do omówionych metod statystycznych, szczegółową klasyfikację danych, wyjaśnienie pojęć szeregu szczegółowego i szeregu rozdzielczego przedziałowego oraz niezbędne wzory do konstrukcji takich szeregów.

dr Anna Makarewicz

Wyższa Szkoła Ekonomii i Innowacji w Lublinie
20-209 Lublin, ul. Projektowa 4
tel.: +48 81 749 17 77
www.wsei.lublin.pl

WSEI LUBLIN
WYDAWNICTWO NAUKOWE
INNOVATIO PRESS

ISBN – wersja elektroniczna: 978-83-66159-83-9

KOMPETENCJE WIEDZA INNOWACJE Zintegrowany program rozwoju WSEI III ETAP
Fundusze Europejskie Program Regionalny
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Społeczny