

METODY ANALIZY DANYCH

DR MARTA DUDA

marta.dziechciarz@ue.wroc.pl

pok. 101 M

Literatura

- [1] Gatnar E., Walesiak M. [2004], *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE Wrocław, Wrocław.
- [2] Walesiak M. [1996], *Metody analizy danych marketingowych*, PWN, Wrocław.
- [3] Mynarski S. [2000], *Praktyczne metody analizy danych rynkowych i marketingowych*, Zakamycze, Kraków.
- [4] Aczel A. [2005], *Statystyka w zarządzaniu*, PWN, Warszawa.

Dane statystyczne

Dane statystyczne można podzielić z dwóch punktów widzenia.

- Skala jednostek:

- dane makroekonomiczne
- dane mikroekonomiczne

- Sposób ujęcia danych:

- dane przekrojowe
- dane w postaci szeregów czasowych
- dane przekrojowo-czasowe

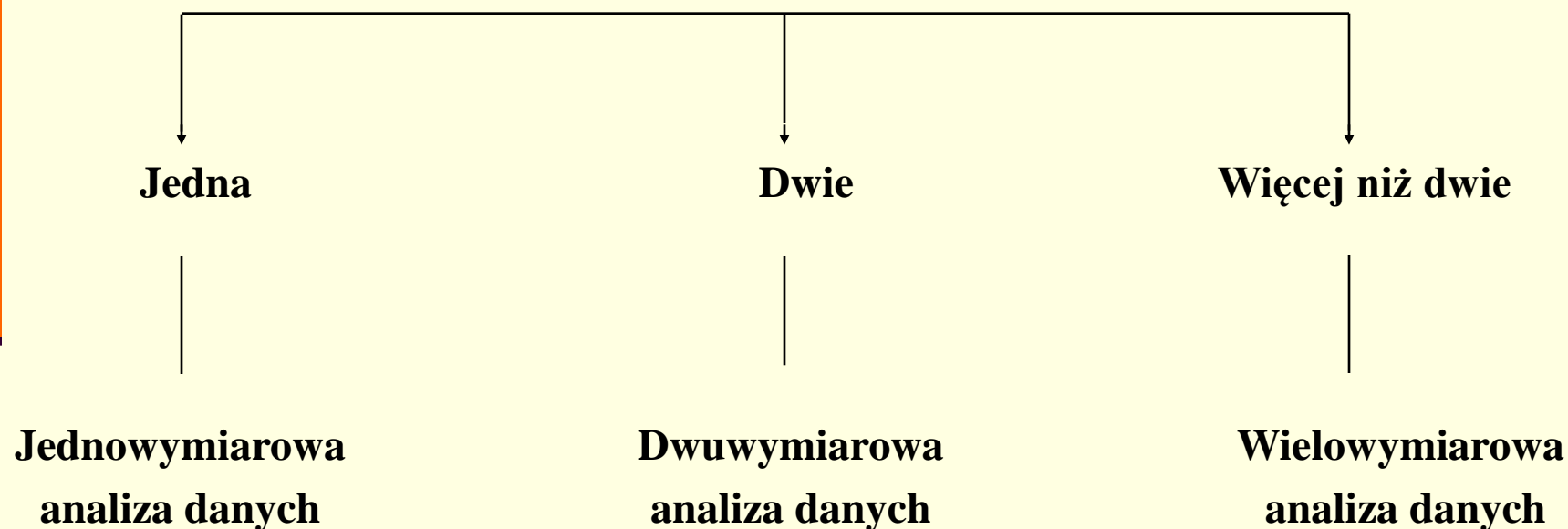
Analiza danych

Główne cele:

- poznawczy – realizowany przez opis zależności między zjawiskami społeczno-gospodarczymi przy pomocy metod statystycznych i ekonometrycznych
- predyktywny – realizowany przez metody prognozowania, czyli przewidywanie przebiegu i przyszłych stanów procesów społeczno-gospodarczych
- decyzyjny – realizowany przez wykorzystanie metod ekonometrycznych w symulacji i sterowaniu przebiegiem rozwoju zjawisk społeczno-gospodarczych

Analiza danych

**Liczba zmiennych będących przedmiotem
analizy w tym samym czasie**



Statystyczna analiza danych

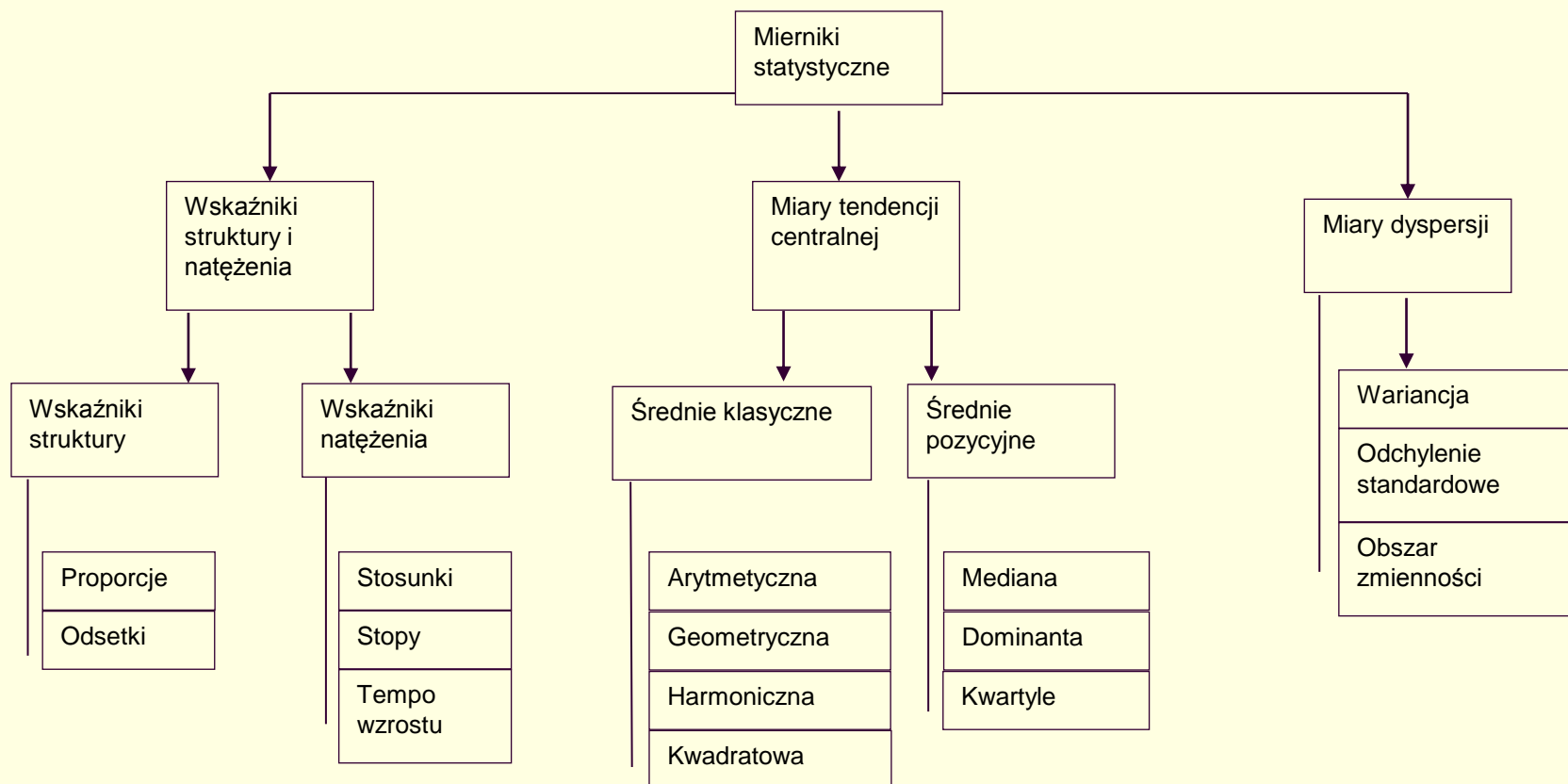
Jednowymiarowa

- Analiza rozkładu
 - Wartości przeciętne
 - Miary rozproszenia
 - Miary skośności
 - Miary koncentracji
- Ilustracja graficzna

Statystyczna analiza danych

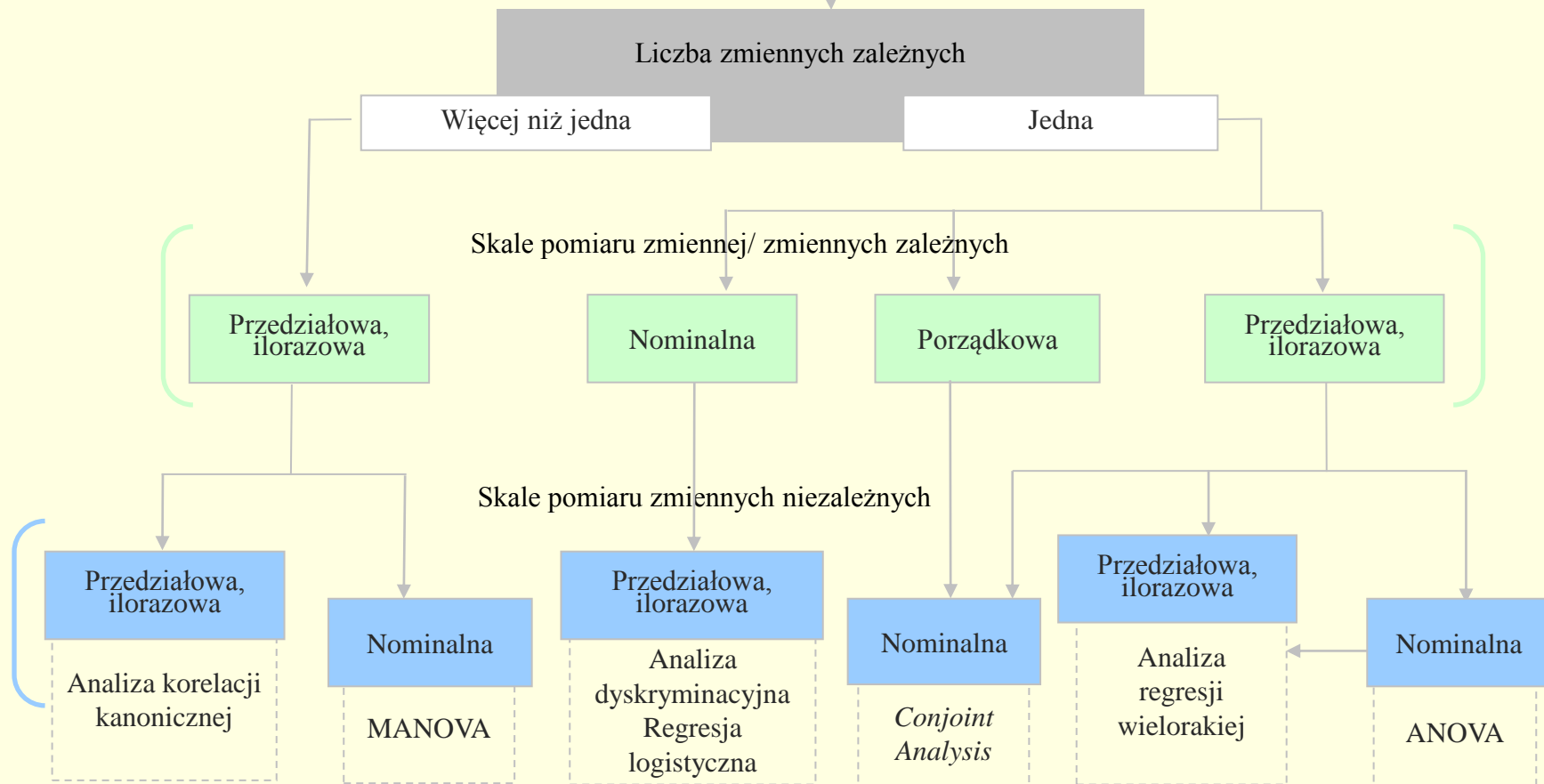
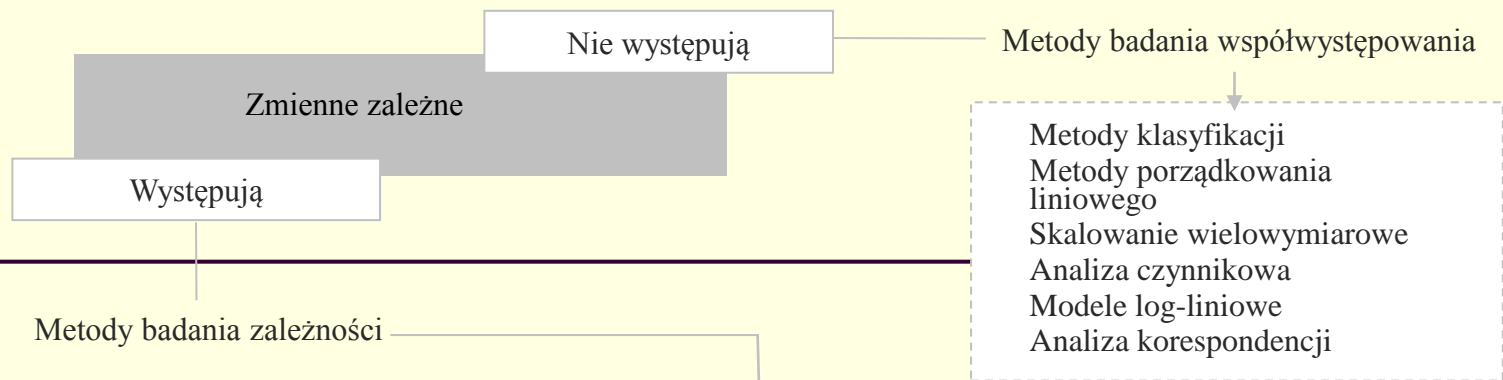
- Wartości przeciętne
 - Średnie
 - Arytmetyczna
 - Geometryczna
 - Harmoniczna
 - Przeciętne pozycyjne
 - Kwantyle
 - Mediana
 - Kwartyle
 - Kwintyle
 - Decyle
 - Percentyle

Przegląd mierników statystycznych



Statystyczna analiza danych

- Analiza wielowymiarowa
 - Analiza zależności
 - Korelacja
 - Regresja
 - Analiza współzależności
 - Klasyfikacja
 - Grupowanie
 - Dyskryminacja
 - Porządkowanie liniowe



CZYNNIKI DECYDUJĄCE O WYBORZE METOD ANALIZY DANYCH

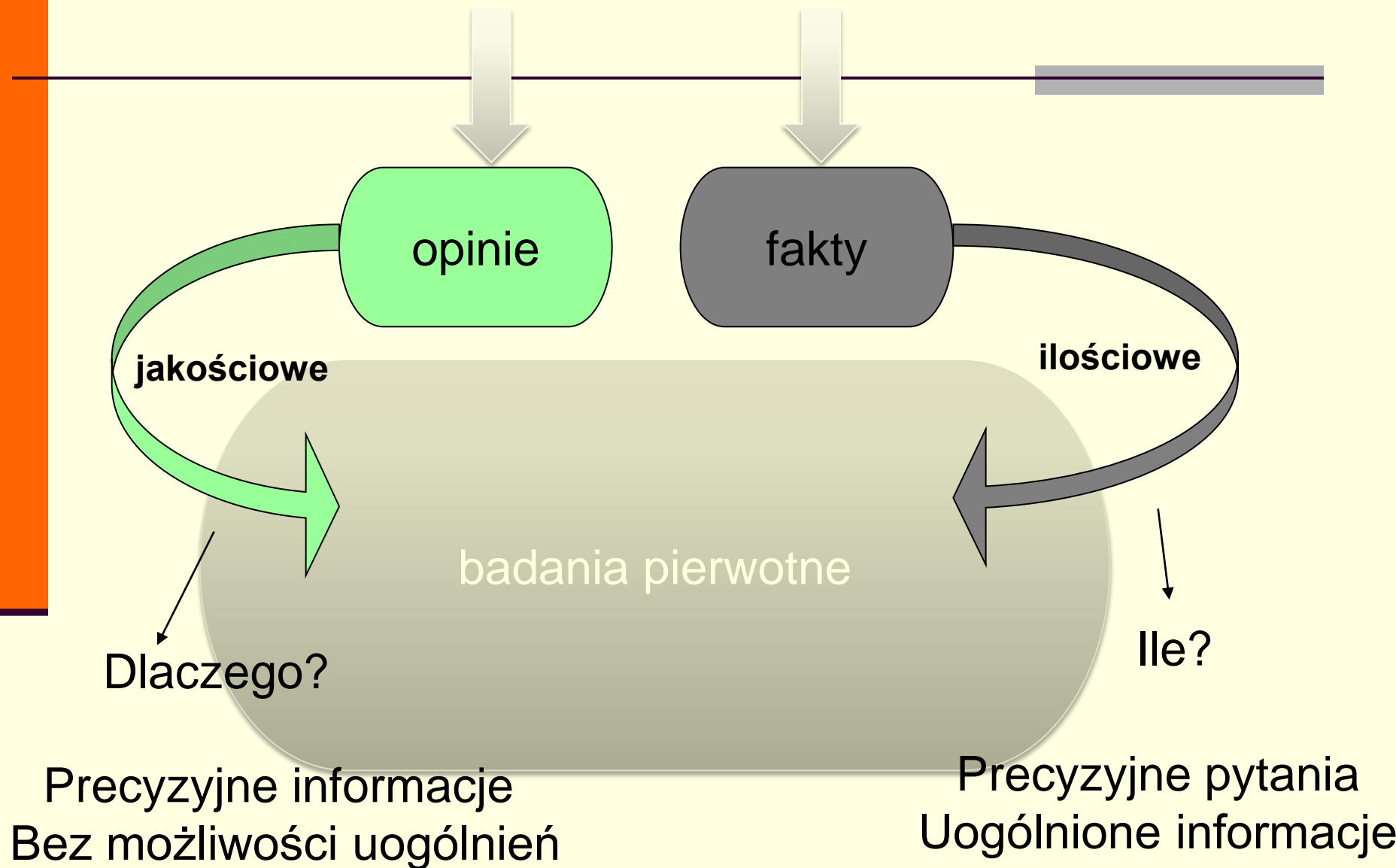
1. Cel badania.
2. Czy w zbiorze badanych zmiennych jesteśmy w stanie wyróżnić zmienną zależną (zmienne zależne) od innych zmiennych.
3. Skale pomiaru zmiennych.
Wynika to jednoznacznie z klasyfikacji jednowymiarowych i dwuwymiarowych metod analizy danych oraz klasyfikacji wielowymiarowych metod analizy danych.
4. Liczba analizowanych w badaniu zmiennych.
W zależności od liczby badanych zmiennych metody analizy danych dzieli się na: jednowymiarowe, dwuwymiarowe oraz wielowymiarowe.
5. Koszt badania (marketingowego, rynkowego).
6. Liczebność próby statystycznej.
Metody analizy wielowymiarowej wymagają na ogół liczniejszej próby (większa liczba zmiennych, przy danej liczbie obserwacji, oznacza mniejszą liczbę stopni swobody).

Proces badania statystycznego

- Przygotowanie badania
- Zebranie danych
- Opracowanie materiału statystycznego
- Statystyczna analiza wyników obserwacji



Cel badania



Badania jakościowe a badania ilościowe

jakościowe

- mała skala badania (10-40 osób)
- uczestnicy dobierani do badania w sposób subiektywny - celowy
- brak standaryzacji pomiaru
 - swoboda reakcji i wypowiedzi
 - kontekst badawczy
 - brak możliwości „uogólniania” wyników na całą populację

ilościowe

- duża skala badania (> 100 osób)
- losowy dobór respondentów
- standaryzacja pomiaru
 - kontrolowana reakcja i wypowiedzi
 - obiektywność pomiaru
 - analizy statystyczne uogólniające wyniki na całą populację

Zbieranie obserwacji do badania statystycznego. Pojęcia podstawowe

- Badana zbiorowość
 - generalna
 - próbna
- Reprezentatywność próby
- Źródła danych
- Skale pomiaru zmiennych

METODA REPREZENTACYJNA

- Do ograniczenia zakresu badań statystycznych do części populacji zmuszają zarówno względy organizacyjne, jak i finansowe
- Wśród badań częściowych na szczególną uwagę zasługują badania reprezentacyjne, a więc takie, dla których część populacji przewidziana do badania jest wybierana zgodnie z pewnym mechanizmem losowym
- Reprezentatywność jest warunkiem prawidłowych badań ilościowych

METODA REPREZENTACYJNA

- **Metoda reprezentacyjna** – metoda prowadzenia statystycznego badania częściowego, mającego charakter reprezentacyjny, to znaczy taki, że otrzymane wyniki można uogólnić na rozpatrywaną skończoną populację
- **Metoda reprezentacyjna** polega na tym, że:
 - W celu zbadania własności całej zbiorowości statystycznej wybiera się do badania tylko pewną liczbę jednostek statystycznych reprezentujących badaną zbiorowość.
 - Ta wybrana grupa jednostek nazywana jest próbą
- Podstawowym problemem w stosowaniu metody reprezentacyjnej jest odpowiedni dobór jednostek do próby
- Jednostki te powinny być tak wybierane, aby faktycznie reprezentowały całą zbiorowość, czyli tak, aby próba była miniaturowym odzwierciedleniem badanej zbiorowości statystycznej


METODA REPREZENTACYJNA

- Podstawowym pojęciem w metodzie reprezentacyjnej jest populacja generalna.
 - Przez **populację generalną** (w skrócie *populację*) rozumie się zbiór jednostek (elementów) będących przedmiotem zainteresowania statystyka
 - Elementy populacji nazywa się **jednostkami badania**
 - Jednostki badania wchodzące w skład populacji muszą mieć pewne cechy (własności) wspólne (pozwalają one rozstrzygnąć, czy określona jednostka należy do danej populacji) oraz muszą się różnić tymi cechami, które są przedmiotem zainteresowania badającego
 - Listę cech, które decydują o zaliczeniu jednostki do populacji określa statystyka
- Przykłady populacji:
 - Zbiór gospodarstw domowych w Polsce w dniu 1 stycznia 2010 roku.
 - Zbiór obywateli Rzeczypospolitej Polskiej uprawnionych do głosowania w ostatnich wyborach do Sejmu.
 - Zbiór klientów hipermarketu Tesco we Wrocławiu w ostatnią środę.


Zasady doboru próby

- Określenie populacji badanej
- Określenie operatu losowania (jeśli istnieje)
- Określenie jednostki wyboru
- Ustalenie wielkości próby


Wybór metody doboru jednostek:



Próba losowa: każda jednostka ma taką samą szansę trafienia do próby:
próba prosta, warstwowa, grupowa, wielostopniowa, systematyczna



Próba pseudo-losowa:
określona proporcja jednostek o wskazanych cechach, wybór jednostek losowy (kwotowa)



Próba celowa: w skład próby wchodzi wskazane jednostki z grupy docelowej

Proces doboru próby

1. **Zdefiniowanie populacji.** Zdefiniowanie zbiorowości jednostek różniących się pod względem badanej cechy. Populacja musi być zdefiniowana w terminach: elementów badania (cech), jednostek zbiorowości (obiektów), zakresu czasowego i zakresu przestrzennego
2. **Określenie operatu losowania.** Procedura, wg której poszczególne jednostki losowania dostają się do próby na podstawie listy, wykazu wszystkich jednostek reprezentujących populację. Np. wykaz miejscowości, rejestr, kartoteka, książka telefoniczna, lista płac.
 - **Operat doboru próby** – jest to wykaz, lista osób wylosowanych z jakiejś innej listy, która zawiera wykaz wszystkich osób wchodzącej w skład danej zbiorowości do badania. W przypadku wybierania próby osób najczęściej stosowanym operatem jest rejestr PESEL.
 - Obejmuje on wszystkich mieszkańców Polski i jest aktualizowany na bieżąco ze wzgl. na: urodzenia, zgony, zmiany miejsca zameldowania. W przypadku próby gospodarstw domowych jako operat doboru próby stosuje się najczęściej rejestr przygotowany przez GUS na podstawie spisów powszechnych.

Proces doboru próby

3. **Określenie jednostki losowania** jest podstawą tworzenia próby losowej: dzielnica, osiedle, ulica, rodzina, gospodarstwo domowe, firma, pracownik, osoba.
4. **Wybór metod doboru próby:**
 - metody doboru losowego (dają jednakową szansę dostania się jednostkom populacji do próby statystycznej. Można określić stopień reprezentatywności)
 - metody doboru nielosowego.
5. **Określenie liczebności próby.** Tylko dla metod doboru losowego. Zależy ono od rozmiaru populacji, zmienności badanej cechy oraz przyjętego poziomu dokładności.

SPOSOBY DOBORU PRÓBY

4. Wybór metod doboru próby

- Przyjmuje się dwa rodzaje doboru próby:
 - **dobór losowy**, przy którym tylko przypadek decyduje o tym, która jednostka zbiorowości generalnej będzie wylosowana. Cechą charakterystyczną tego doboru jest możliwość wnioskowania o charakterystykach populacji generalnej na podstawie zbadania losowo wybranej próby.
 - **dobór nielosowy**, w sytuacji tej nie ma możliwości uznania próby za reprezentatywną względem określonych cech statystycznych, jak również możliwości przeprowadzenia wnioskowania statystycznego o zbiorowości generalnej.

metody doboru próby

metody doboru losowego

■ **Dobór losowy prosty**

- **niezależny** – ze zwracaniem (každorazowe zwracanie wylosowanej jednostki z powrotem do populacji)
- **zależny** – bez zwracania (każda wylosowana jednostka już nie uczestniczy w dalszym losowaniu, a wybór pozostałych jednostek zależy od rezultatów wcześniejszego losowania).

■ **Sposoby:** Losowanie za pomocą urny, tablice liczb losowych

metody doboru próby

metody doboru losowego

- **Dobór losowy systematyczny** – jest odmianą doboru prostego, stosuje się go wówczas, gdy istnieje lista całej populacji generalnej. Należy do typu losowania bezpośredniego, ale ograniczonego. Dobór ten jest ograniczony do jednostek z pewnego przedziału liczbowego.
 - Polega na wyborze z uporządkowanego zbioru jednostek populacji odpowiedniej liczby jednostek w równych odstępach (interwałach).
 - Istnieją sytuacje, kiedy dobór losowy systematyczny może dawać istotne zniekształcenia (istnieje trend). Wynik będzie zależeć od punktu startu, od którego rozpoczyna się procedura losowania.

metody doboru próby

metody doboru losowego

- **Dobór losowy warstwowy** polega na podziale całej populacji generalnej na tzw. warstwy i dokonaniu bezpośredniego losowania niezależnych prób w obrębie każdej warstwy. Podział populacji na warstwy musi być przeprowadzony w taki sposób, aby każdy element wchodził tylko do jednej warstwy.
 - warstwy powinny być wewnętrznie jednorodne (ze względu na wyróżnioną cechę). Zróżnicowanie między warstwami - jak największe. W obrębie każdej warstwy dokonuje się oddzielnego, prostego doboru losowego.
 - Najczęściej stosowanymi kryteriami wyróżnienia warstw są znane cechy w populacji, które są silnie skorelowane z cechą badaną np. dochód, wielkość gospodarstwa domowego, pozycja społeczna, wiek, płeć, wielkość miejscowości.

- **Dobór losowy warstwowy proporcjonalny**
 - Polega na wyborze z każdej warstwy takiej liczby elementów, jaka pozostaje w proporcji do liczebności (udziału) tej warstwy w całej populacji

- **Dobór losowy warstwowy nieproporcjonalny**
 - Polega na wyróżnieniu warstw w populacji generalnej, które są niedoreprezentowane w próbie (ich udział w populacji jest niewielki).

■ **Dobór losowy warstwowy optymalny**

- Za optymalną alokację uznaje się taki dobór jednostek z poszczególnych warstw do próby, który jest proporcjonalny do iloczynu liczebności warstwy i wielkości odchylenia standardowego badanej cechy w danej warstwie, a odwrotnie proporcjonalny do pierwiastka kwadratowego z kosztu badania jednostki należącej do danej warstwy.

■ Dobór losowy zespołowy (grupowy)

- Losuje się nie jednostki badania, lecz zespoły, grupy tych jednostek, np. rodziny, klasy szkolne, dzielnice, itd. Losuje się do próby pewną liczbę grup i bada się wszystkie jednostki należące do tych grup.
- W tym przypadku należy podzielić populację generalną w ten sposób, aby zróżnicowanie wewnątrz grup było jak największe, a pomiędzy grupami minimalne. Dążymy by grupy były jak najbardziej zróżnicowane wewnętrznie i jak najbardziej jednorodne względem grupy (względem siebie podobne).

metody doboru próby

metody doboru losowego

- Dobór losowy zespołowy (grupowy): dobór zespołowy z jednakowymi i różnymi prawdopodobieństwami wyboru
- **Dobór zespołowy z jednakowymi prawdopodobieństwami wyboru:** ten sposób stosuje się wówczas, gdy grupy są tak samo liczebne. Sposób podobny do schematu doboru losowego prostego, z tym, że jednostkami losowania są nie poszczególne elementy, lecz ich zespoły.

metody doboru próby

metody doboru losowego

- Dobór losowy zespołowy (grupowy): dobór zespołowy z jednakowymi i różnymi prawdopodobieństwami wyboru
- **Dobór zespołowy z różnymi prawdopodobieństwami wyboru:** ten sposób stosuje się wówczas, gdy grupy różnią się pod względem liczebności. Polega na losowaniu grup z prawdopodobieństwami proporcjonalnymi do liczebności tych grup. Prawdopodobieństwa muszą sumować się do jedności i nie zmieniać się w trakcie losowania, co zabezpiecza losowanie ze zwracaniem.

- **Dobór losowy wielostopniowy:** dobór wielostopniowy z jednakowymi i różnymi prawdopodobieństwami wyboru
- Metoda kolejnego losowania zespołów jednostek coraz to niższego stopnia, powstających z podziału zespołów wyższego stopnia. W końcu dochodzi się do podstawowych jednostek badania.

metody doboru próby

metody doboru losowego

- Dobór losowy wielofazowy:
 - Jednostki poddawane badaniom są w kilku fazach zawsze w obrębie tych samych kategorii jednostek. Dobór jednostek tego samego stopnia.
 - 1 faza polega na wyborze dużej grupy, następnie wybiera się jedną lub kilka mniejszych grup jednostkowych.

- **Dobór kwotowy:** należy do najczęściej stosowanych metod nielosowego doboru próby w badaniach rynku. Opiera się on na znajomości struktury populacji generalnej według przyjętych cech (tzw. zmiennych kontrolnych) i narzuceniu tej struktury na skład próby. Stosowane cechy – kryteria to: wiek, płeć, wielkość rodziny, dochód, rodzaj grupy społecznej.
 - Liczebność grup (segmentów) w próbie ustala się na podstawie przemnożenia rozkładu procentowego wybranych cech w populacji generalnej przez ogólną liczebność próby.
 - Najczęściej poprzestaje się na 2 lub 3 cechach dających możliwość kwotowego określenia niewielkiej liczby segmentów. Przy większej liczbie cech jako zmiennych kontrolnych i większej liczbie segmentów trudniej skompletować odpowiedni skład zbiorowości próbnej.

■ Dobór jednostek typowych

- Polega na wyborze jednostek typowych (przeciętnych)
- Liczba jednostek nie musi być duża gdy chcemy określić przeciętne atrybuty, cechy; metoda ta nie daje właściwego zróżnicowania cech badanej zbiorowości, jest stosowana jako metoda pomocnicza, np. przy badaniach pilotażowych.

■ Dobór przez eliminację

- Poprzez eliminację jednostek nietypowych z grupy, nie daje ona odpowiedniego zróżnicowania populacji i jest metodą pomocniczą.

metody doboru próby

metody doboru nielosowego

- **Dobór celowy:** („wg wygody”) – najbardziej typowy przypadek doboru nielosowego. Polega na całkowicie subiektywnym wyborze jednostek badanych do próby w nadziei uzyskania najszerzej i najpełniejszej informacji. Metoda stosowana w wywiadach psychologicznych a zwłaszcza głębinowych, a także w badaniach eksperymentalnych. Elementami próby są te obiekty, które według badacza odpowiadają celom badań.
- Polega na subiektywnym doborze jednostek do badania, uzależniony jest od zakresu informacji jakimi ankieter dysponuje. Opinie tych celowych osób mogą jednak różnić się od krańcowych wyników. Dobór celowy jest świadomy – wynik może być krańcowo różny, nie może służyć jako wynik grupy populacji tylko do wąsko sprecyzowanej grupy.

metody doboru próby
metody doboru nielosowego

- **Dobór przypadkowy:** („przechwytywania po drodze”). Polega na przypadkowym (na „chybił – trafił”) doborze pewnych jednostek, które w danej (przypadkowej) sytuacji znalazły się w dogodnym zasięgu.
- Dobór przypadkowy – o wyborze jednostki do badania decyduje przypadek. Stosowane np. przy badaniach telefonicznych. Badanie niereprezentatywne.

metody doboru próby

metody doboru nielosowego

- **Dobór wygodny:** badani są znajomi, krewni, nasze otoczenie
- **Dobór sieciowy:** badani są klienci określonej sieci usługowej
- **Metoda kuli śniegowej:** respondenci wybrani przypadkowo, następnie dostarczani przez początkowych respondentów
 - dobór polegający na dotarciu do nielicznej grupy badanych, a następnie za ich pośrednictwem do kolejnych znanych im jednostek o podobnych cechach. W ten sposób następuje zwiększenie liczby jednostek w próbie aż do uzyskania zakładanej liczebności próby

Źródła danych

Najprostszym kryterium klasyfikacji metod analizy danych może być wykorzystywane narzędzie badania oraz źródła danych. Kierując się tym kryterium, w ramach ilościowych metod analizy danych wyróżnia się:

- Analiza źródeł wtórnych (dokumenty, opracowania, raporty)
- Analiza danych pierwotnych
 - Metody z wykorzystaniem kwestionariusza ankiety
 - Metody wykorzystujące eksperyment
 - Metody z wykorzystaniem obserwacji

Źródła danych

WTÓRNE



**Wykorzystanie istniejących
zasobów informacji lub danych**

PIERWOTNE



**Utworzenie zasobów danych
i przetworzenie ich w informację**

Dane wtórne



WEWNĘTRZNE

- statystyka sprzedaży
- statystyka zamówień
- sprawozdawczość kosztowa (koszty marketingu sprzedaży itp.)
- raporty komórek serwisowych



ZEWNĘTRZNE

- statystyka państwowa
- raporty organizacji konsumentów i/lub producentów
- publikacje naukowe
- periodyki ekonomiczne i handlowe
- komercyjne bazy danych ekonomicznych

Źródła danych pierwotnych

- **Metody uzyskania danych pierwotnych:**
 - **obserwacja rynku:** bierne rejestrowanie rzeczywistości
 - **eksperyment rynkowy:** świadoma ingerencja badacza w stan rzeczywistości i rejestracja efektów tej ingerencji
 - **wywiad i ankieta:**
 - wywiad - rozmowa bezpośrednia lub za pomocą wybranego środka komunikacji
 - ankieta - respondent samodzielnie wypełnia formularz ankiety

Utworzenie zasobów danych i przetworzenie ich w informację

Rejestrowanie danych pierwotnych

- **Sposoby rejestrowania danych pierwotnych:**
 - **Kwestionariusz, pomiar za pomocą specjalnych urządzeń**

- **Sposoby kontaktowania się w celu uzyskania danych pierwotnych:**
 - **kontakt osobisty**
 - **kontakt pocztowy**
 - **kontakt telefoniczny**

Metody zbierania danych w badaniach ilościowych

– ankieta (respondent samodzielnie wypełnia formularz)

- Ankieta pocztowa
- Ankieta audytoryjna
- Ankieta internetowa (CAWI – *Computer-Assisted Web Interviewing*)
 - wada: ze względu na ograniczony dostęp do Internetu badanie nie może być reprezentatywne
 - zaleta: bardzo szybkie w realizacji (kilkuset ankiet dziennie), niski koszt, odpowiednie dla specyficznych grup konsumentów (np. klientów banków internetowych)

Metody zbierania danych w badaniach ilościowych

– wywiad (w wypełnianiu formularza uczestniczy ankieter)

- PAPI (Paper & Pen Personal Interview) bezpośredni indywidualny wywiad kwestionariuszowy
- CAPI (Computer-Assisted Personal Interview) bezpośredni wywiad indywidualny ze wspomaganie komputerym
- CATI (Computer-Assisted Telephone Interview) indywidualny wywiad telefoniczny ze wspomaganie komputerym

Wywiad w badaniach ilościowych jest rodzajem zaaranżowanej, standaryzowanej rozmowy

Ogólna charakterystyka wybranych metod zbierania danych w badaniach ilościowych

Lp	Kryterium oceny	Metoda badawcza				
		PAPI	CAPI	CATI	CAWI	Ankieta pocztowa
1	Zakładana liczba pytań badawczych	+	+	■	—	—
2	Poziom trudności pytań badawczych	+	+	—	—	—
3	Stopień dostępności respondentów*	+	+	■	—	—
4	Stopień kontroli poprawności struktury próby badawczej	+	+	+	—	—
5	Możliwość wykorzystania zaawansowanych technik pomiaru	■	+	—	+	—
6	Możliwość wykorzystania materiałów multimedialnych	—	+	■	+	—
7	Stopień kontroli sytuacji badawczej	+	+	■	—	—
8	Stopień anonimowości respondenta	—	—	+	+	+
9	Czas realizacji projektu badawczego	—	■	+	+	—
10	Koszt realizacji badania	—	—	■	+	+

*Dla badań realizowanych na ogólnopolskich próbach reprezentatywnych

+ mocna strona metody **■** ocena pośrednia (zależna od sytuacji badawczej) **—** słaba strona metody

Skale pomiaru od najsłabszej do najsilniejszej

- Nominalna
- Porządkowa
- Przedziałowa (Interwałowa)
- Ilorazowa

Skale pomiaru – przykłady zmiennych

Typ skali	Nominalna	Porządkowa	Przedziałowa	Ilorazowa
Przykłady zmiennych	Płeć, Profil firmy, kolor, smak, zawód	Jakość produktu, nastawienie do produktu, stopień ryzyka	Temperatura, czas, wynik finansowy	Wielkość sprzedaży, cena, wydatki na reklamę, okres gwarancji, dochody konsumentów

Skale pomiaru – relacje i operacje arytmetyczne

Typ skali	Nominalna	Porządkowa	Przedziałowa	Ilorazowa
Dopuszczalne relacje	Równości, różności	Większości, mniejszości	Równość różnic i przedziałów	Równość ilorazów
Dopuszczalne operacje arytmetyczne	Zliczanie zdarzeń (liczba relacji równości, różności)	Zliczanie zdarzeń (liczba relacji równości, różności, większości, mniejszości)	Dodawanie, odejmowanie	Mnożenie, dzielenie

Skale pomiaru – metody i techniki

Typ skali	Nominalna	Porządkowa	Przedziałowa	Ilorazowa
Miary położenia	Modalna	Mediana	Średnia arytmetyczna	Średnia geometryczna, średnia harmoniczna
Miary rozrzutu	Miary informacji	Rozstęp ćwiartkowy	Wariancja, odchylenie standardowe	Współczynnik zmienności
Miary współzależności	Miara zależności statystycznej Pearsona	Współczynnik korelacji τ Kendalla	Współczynnik korelacji Pearsona	

POMIAR POSTAW NABYWCÓW

Przez pomiar rozumie się przyporządkowanie liczb obiektom zgodnie z określonymi regułami w taki sposób, aby liczby odzwierciedlały relacje zachodzące między tymi obiektami.

SKALE POMIARU POSTAW NABYWCÓW:

1. Skale podstawowe:

- a) skala nominalna
- b) skala pozycyjna (rating scale)
- c) skala rangowa
- d) skala stałych sum
- e) skala zamiarów zakupu
- f) skala porównywania parami

2. Skale specyficzne:

- a) skala różnicowania słownego (semantyczna)
- b) skala Stapela
- c) skala Likerta

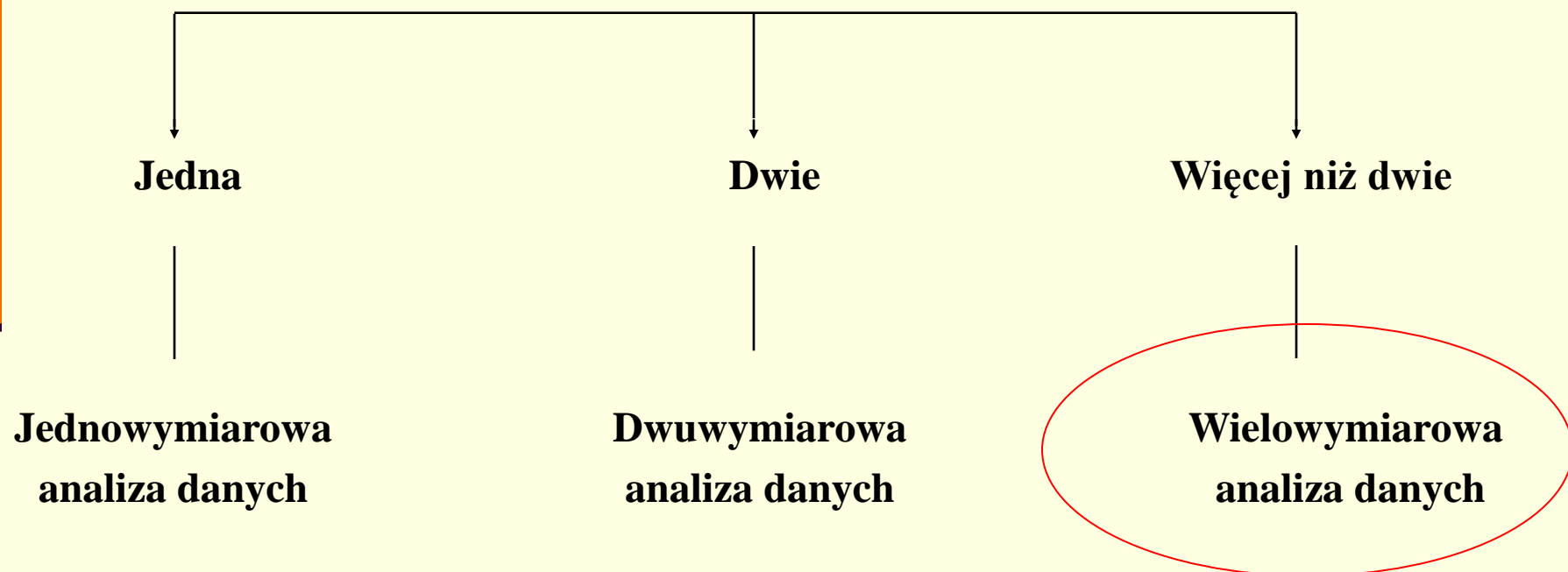
Problemy powstające przy konstrukcji skal pomiaru postaw

Przy konstrukcji skal pomiaru postaw powstaje wiele problemów. Wśród nich można wyróżnić takie, jak:

1. Liczba kategorii skali
2. Parzysta lub nieparzysta liczba kategorii
3. Skale wymuszające lub niewymuszające
4. Skala zrównoważona lub niezrównoważona
5. Skale z jednostką odniesienia i skale bez jednostki odniesienia
6. Forma prezentacji skali

Analiza danych

**Liczba zmiennych będących przedmiotem
analizy w tym samym czasie**



Metody analizy rozkładu wielu zmiennych

- Wielowymiarowa analiza porównawcza
 - Normalizacja
 - Macierz odległości
 - Metoda środków ciężkości
 - Taksonomia wrocławska
 - Metody porządkowania liniowego
 - Metoda wzorca rozwoju
 - Metoda sum standaryzowanych

Wielowymiarowa analiza porównawcza

■ Wielowymiarowa analiza porównawcza bada zjawiska złożone, czyli takie, które są opisane za pomocą co najmniej dwóch zmiennych.

■ Badany jest wektor zmiennych: $X = [X_1, X_2, \dots, X_m]$

■ gdzie m – liczba zmiennych ($m \geq 2$)

■ Dane surowe prezentowane są w postaci macierzy obserwacji:

■ gdzie x_{ij} – wartość j -tej zmiennej dla i -tego obiektu

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

Wielowymiarowa analiza porównawcza

- Za pomocą metod wielowymiarowej analizy porównawczej rozwiązuje się zwykle dwa następujące zagadnienia:
 - Badanie podobieństw obserwacji pod względem zjawiska złożonego z wykorzystaniem metod klasyfikacji
 - Uszeregowanie obserwacji pod względem poziomu zjawiska złożonego z wykorzystaniem metod porządkowania liniowego zbioru

Wielowymiarowa analiza porównawcza

- W badaniach praktycznych zdarza się z reguły, że zmienne wchodzące w skład wektora zmiennych wyrażone są w różnych jednostkach miary.
- Z formalnego punktu widzenia większość metod wielowymiarowej analizy porównawczej można stosować tylko wtedy, gdy wszystkie zmienne są wyrażone w tych samych jednostkach miary o zbliżonych rzędach wielkości.
- Aby to uzyskać stosuje się normalizację zmiennych

Wielowymiarowa analiza porównawcza

Standaryzacja

- Standaryzację przeprowadza się według wzoru:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1, 2, \dots, m)$$

gdzie: x_j – średnia arytmetyczna zmiennej X_j

s_j – odchylenie standardowe zmiennej X_j

z_{ij} – standaryzowana wartość zmiennej X_j dla i -tego obiektu

Wielowymiarowa analiza porównawcza

Unitaryzacja

- Innym sposobem normalizacji zmiennych jest unitaryzacja

$$z_{ij} = \frac{x_{ij} - \min_i x_{ij}}{O_j}$$

Gdzie O_j – rozstęp zmiennej X_j ; czyli
 $O_j = \max(X_j) - \min(X_j)$

Wielowymiarowa analiza porównawcza

Macierz standaryzowanych obserwacji

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1j} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2j} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{i1} & z_{i2} & \dots & z_{ij} & \dots & z_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nj} & \dots & z_{nm} \end{bmatrix}$$

Gdzie:

z_{ij} – znormalizowana wartość zmiennej X_j w i -tym obiekcie

Wielowymiarowa analiza porównawcza

Macierz standaryzowanych obserwacji

- Pojęciem umożliwiającym wzajemne porównywanie obiektów pod względem zjawiska złożonego jest podobieństwo. Podobieństwo obiektów w WAP najczęściej jest mierzone za pomocą ODLEGŁOŚCI.
- Zwykle stosuje się tzw. odległość euklidesową

$$d_{il} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{lj})^2} \quad (i, l = 1, 2, \dots, n)$$

gdzie: d_{il} – odległość między obiektem i-tym a l-tym

z_{ij} – wartość znormalizowanej zmiennej X_j w obiekcie i

z_{lj} – wartość znormalizowanej zmiennej X_j w obiekcie l

Macierz odległości

- Po obliczeniu odległości między każdą parą obiektów otrzymuje się macierz odległości:

$$D = \begin{bmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & 0 \end{bmatrix}$$

D – macierz kwadratowa o wymiarze $n \times n$; gdzie n – liczba obiektów

Metody klasyfikacji zbioru obserwacji

- Metody klasyfikacji umożliwiają podział zbioru n obserwacji na K rozłącznych i niepustych podzbiorów, zwanych klasami, tak aby:
 - obserwacje należące do tych samych klas były jak najbardziej podobne
 - obserwacje należące do różnych klas były jak najmniej podobne

Metody klasyfikacji zbioru obserwacji

Taksonomia wrocławska

- Taksonomia wrocławska służy do nieliniowego porządkowania zbioru obiektów opisywanych przez zespół cech oraz do podziału tego zbioru na grupy obiektów podobnych ze względu na przyjęte cechy.

Porządkowanie liniowe

- Pozwala uszeregować obiekty pod względem poziomu zjawiska złożonego. Sposób normalizacji zmiennych uzależniony jest m.in. od charakteru analizowanych zmiennych.
- Metody porządkowania liniowego stosowane są:
 - do oceny obiektów (produktu, przedsiębiorstwa, kraju), pozwalając określić pozycję badanego obiektu na tle obiektów konkurencyjnych
 - np. do oceny możliwości istniejących (nowych) produktów; pozwalają określić pozycję produktu na rynku na tle produktów konkurencyjnych

Porządkowanie liniowe

- Typy zmiennych (cech):
 - Stymulanty – zmienne, których wzrost wartości jest pożądaný
 - Destymulanty – zmienne, których spadek wartości jest pożądaný
 - Nominanty – zmienne, dla których zalecana jest pewna wartość nominalna lub przedział wartości

Porządkowanie liniowe

Destymulanta

- Destymulanta może być łatwo sprowadzona do postaci stymulanty np. Poprzez:
 - pomnożenie wszystkich jej wartości przez -1 (dla danych standaryzowanych), albo przez
 - odjęcie od jedności wszystkich jej wartości (dla zmiennych unitaryzowanych)

Porządkowanie liniowe

- Celem stosowania metod porządkowania liniowego jest uporządkowanie obserwacji w kolejności od „najlepszej” do „najgorszej”, przy czym kryterium uporządkowania jest właśnie na poziomie zjawiska złożonego
- U podstaw tych metod tkwi założenie, że istnieje pewne nadrzędne kryterium, które pozwala na liniowe uporządkowanie obserwacji

Normalizacja uwzględniająca charakter zmiennych

■ Stymulanta

$$Z_{it} = \frac{x_{it}}{\max_t x_{it}}$$

■ Nominanta z zalecanym przedziałem wartości

$[x_{i,\min}, x_{i,\max}]$

$$Z_{it} = \begin{cases} 1 & \text{dla } x_{i,\min} \leq x_{it} \leq x_{i,\max} \\ \frac{x_{it}}{x_{i,\min}} & \text{dla } x_{it} < x_{i,\min} \\ \frac{x_{i,\max}}{x_{it}} & \text{dla } x_{it} > x_{i,\max} \end{cases}$$

■ Destymulanta

$$Z_{it} = \frac{\min_t x_{it}}{x_{it}}$$