

1. Statystyka jako nauka

• Podstawowe cele rozdziału •

- Wyjaśnić, czym zajmuje się nauka nazywana statystyką.
- Pokazać obszary zastosowań statystyki.
- Wyjaśnić podstawowe pojęcia – zbiorowość statystyczna, cecha statystyczna.
- Omówić etapy badania statystycznego.

1.1. Co to jest statystyka?

Statystyka to nauka o metodach badań liczbowo wyrażalnych własności zbiorowości¹.

Ta definicja jest jedną z wielu, które można spotkać w literaturze przedmiotu. Jej zaletą jest to, że w zwięzły sposób pokazuje istotę nauki, jaką jest statystyka. Po pierwsze, statystyka bada nie jednostki, ale zbiorowości statystyczne, jest to podmiot jej badania. Na przykład, aby odpowiedzieć na pytanie, ile zarabiają Polacy, nie wystarczy spytać o to jedną osobę, dopiero obserwacja zarobków wielu osób pozwoli na odpowiedź. Po drugie, celem statystyki jest poznanie własności tych zbiorowości, wyrażenie ich w sposób wymierny, za pomocą liczb. Na przykład, odpowiadając na pytanie dotyczące zarobków Polaków, nie zadowoli nas odpowiedź „mało” lub „dużo”, będziemy chcieli wiedzieć, jaka jest średnia zarobków, czy są bardzo zróżnicowane, jak silnie zależą od wykształcenia, jak zmieniają się w czasie itd. Po trzecie, jak precyzuje przytoczona na początku definicja, przedmiotem badania statystyki jako nauki są metody, które pozwolą na realizację wspomnianego celu. Statystyka odkrywa, rozwija i oferuje nam metody, które pozwalają w sposób wymierny określić na przykład, jaka jest struktura zarobków w Polsce, czym uwarunkowana jest ich wysokość, jaka jest ich dynamika. Zatem zawsze tam, gdzie odpowiedź na pytanie badawcze wymaga empirycznego rozstrzygnięcia, wykorzystamy metody statystyczne. Warto wspomnieć, że w opinii wielu naukowców tylko odpowiedź potwierdzona empirycznie może nosić miano naukowej.

Sto lat temu H.G. Wells, pisarz angielski, przewidywał, że pewnego dnia myślenie statystyczne stanie się tak niezbędne, jak umiejętność czytania i pisanie. Rzeczywiście, w świecie współczesnym jesteśmy zasypywani danymi liczbowymi opisującymi różne sfery otaczającej nas rzeczywistości. Musimy nauczyć się je rozumieć i wykorzystywać do

¹ B. Szulc, *Statystyka dla ekonomistów. Opis statystyczny*, PWE, Warszawa 1976, s. 13.

swoich celów. Znajomość statystyki ułatwia nam funkcjonowanie we współczesnym świecie, a z drugiej strony „ułatwia światu” lepsze funkcjonowanie. Zatem, nawet jeśli nie rozstrzygamy naukowych problemów, gdzie jak wspomniano wykorzystanie metod statystycznych jest niezbędne, statystykę warto znać.

Więcej argumentów potwierdzających potrzebę znajomości metod statystycznych daje wiedza o tym, jak duże są możliwości ich zastosowania – jakie problemy można rozwiązać, wykorzystując metody statystyczne, oraz w jakich dziedzinach wiedzy i działalności społecznej użyteczne są metody statystyczne?

Jeśli chodzi o pierwszy obszar zastosowania, metody statystyczne ułatwiają, a czasem wręcz umożliwiają:

- klasyfikację zjawisk i procesów (pozwalają uporządkować informacje na ich temat),
- zrozumienie zjawisk i procesów,
- ocenę zjawisk i procesów obecnie i w przeszłości,
- przewidywanie przyszłych zjawisk i procesów,
- podejmowanie decyzji.

Wspomaganie procesu podejmowania decyzji uważa się za najważniejszy cel stosowania metod statystycznych. Podjęcie decyzji jest na ogół (a przynajmniej powinno być) zakończeniem procesu przebiegającego od klasyfikacji zjawisk i procesów poprzez ich zrozumienie, ocenę i prognozowanie. W każdym z etapów wykorzystuje się metody statystyczne. Biolog na przykład, wykorzystując m.in. metody statystyczne, podejmie decyzję o ochronie danego gatunku roślin, lekarz zdecyduje, czy zalecić pacjentom określoną szczepionkę, polityk zdecyduje o uczestnictwie w kolejnych wyborach, kredytobiorca o wyborze banku, menedżer w przedsiębiorstwie zdecyduje np. o wycofaniu lub wprowadzeniu produktu, o ograniczeniu kosztów, podjęciu akcji promocyjnej, zmianie metody produkcji, ogłoszeniu bankructwa, podjęciu nowej inwestycji.

W odniesieniu do drugiego obszaru zastosowań metod statystycznych należy powiedzieć, że niemal wszystkie dyscypliny naukowe wykorzystują metody statystyczne. Użyteczne są wszędzie tam, gdzie przedmiotem badania są zjawiska masowe. Należą do bardzo ważnych metod badawczych takich dyscyplin, jak: psychologia, socjologia, demografia, historia, medycyna, biologia, politologia, a przede wszystkim nauki ekonomiczne. Wśród dyscyplin związanych z biznesem metody statystyczne są niezbędne w marketingu (np. badania produktu, badania rynku, wizerunek firmy itp.), finansach (np. wybór inwestycji, analiza trendów wskaźników finansowych), rachunkowości (np. wybór próby przy badaniu sprawozdania finansowego), zarządzaniu (np. doskonalenie jakości produktu, procesu produkcji, serwisu), ubezpieczeniach (np. ustalanie wysokości składki ubezpieczenia, analiza ryzyka).

Na zakończenie kilka słów na temat historii statystyki. **Termin statystyka**² pochodzi od łacińskiego słowa *status*, czyli stan rzeczy, państwo. Po raz pierwszy termin ten zo-

² W języku potocznym termin statystyka jest też używany dla określania: zbioru informacji, na ogół liczbowych (np. statystyka zgonów, statystyka rolnictwa), procesu gromadzenia danych (prowadzić statystyki, tzn. gromadzić i opracowywać dane), instytucji związanych z gromadzeniem i opracowywaniem danych (np. działy statystyki w szpitalach, przedsiębiorstwach), miary opisującej zbiorowość (np. średnia wartość, dominująca wartość).

stał użyty w XVIII w., dla określenia nauki zajmującej się opisem stanu państwa na podstawie gromadzonych danych liczbowych (tzw. państwowznawstwo). Jednakże działalność polegająca na gromadzeniu danych niezbędnych do kierowania państwami znana była dużo wcześniej, na pewno od czasów, kiedy istnieją historyczne źródła pisane. W okresie starożytnych cywilizacji, np. egipskiej, greckiej i rzymskiej, przeprowadzane były spisy ludności i jej majątku. Rejestrowano też między innymi dane dotyczące woj-ska, podatków, wpływów z wojen. W okresie średniowiecza instytucje kościelne zajmowały się rejestracją ruchu naturalnego w parafiach (urodzenia, zgony, małżeństwa) oraz gromadziły dane dotyczące dóbr królewskich czy klasztornych. Można zatem powiedzieć, że statystyka ma długą prehistorię. Jednak jej historia jest całkiem krótka. Gromadzenie danych w sposób usystematyzowany (według określonych zasad, z określoną częstotliwością) i zinstytucjonalizowany (przez centralną, państwową instytucję) rozpoczęło się dopiero w XIX w. Pierwszy Centralny Urząd Statystyczny na świecie powstał we Francji w 1880 roku³. Natomiast rozwój metod, które dały podwaliny nowoczesnej statystyce, nastąpił pod koniec XVI w. i w XVII w.

Rozwój nowoczesnej statystyki przebiega w trzech nurtach:

- rozwój sposobów gromadzenia danych statystycznych,
- rozwój metod analizy danych statystycznych,
- rozwój technik rejestracji i przetwarzania danych.

Jest wiele czynników, które sprzyjają rozwojowi nowoczesnej statystyki. Przede wszystkim należy wymienić: rosnące zapotrzebowanie rządów, instytucji, podmiotów gospodarczych na dane statystyczne, rozwój matematyki, a w szczególności teorii prawdopodobieństwa, oraz rozwój technik komputerowych.

1.2. Podstawowe pojęcia statystyki – zbiorowość, cecha statystyczna

Mimo, że pojęcia badania statystycznego i zbiorowości są intuicyjnie znane i pojawiły się już w definicji statystyki, wymagają sprecyzowania.

Załóżmy, że w pewnej firmie system wynagradzania pracowników budzi niezadowol-enie. Jesteśmy odpowiedzialni za wyjaśnienie przyczyn niezadowol-enia i zapropon-owanie ewentualnych zmian. Najlepszym rozwiązaniem będzie przeprowadzenie ba-dania. Będzie ono polegało na zebraniu odpowiednich danych (np. wysokość płacy, płeć, staż pracy, zajmowane stanowisko), opracowaniu i analizie zebranych danych i przedstawieniu w zwięzłej formie wyników. Takie badanie nazywa się **badaniem sta-tystycznym**. Obejmiemy nim wszystkich pracowników firmy. Pracownicy tworzą zbiór nazywany w statystyce zbiorowością statystyczną.

³ W Polsce Główny Urząd Statystyczny powstał w 1918 r. Więcej informacji na temat historii badań statystycznych w Polsce można znaleźć w: *Rozwój myśli i instytucji statystycznych na ziemiach polskich*, Biblioteka Wiadomości Statystycznych, tom 42, Warszawa 1994.

Zbiorowość statystyczna, nazywana też **populacją statystyczną**, jest to zbiór elementów wyodrębnionych z punktu widzenia celu badania statystycznego.

Na przykład, jeżeli chcemy poznać, jak kształtują się płace w firmie – badamy zbiorowość pracowników. Chcemy dowiedzieć się, jak kształtują się wydatki na żywność gospodarstw domowych – badamy zbiorowość gospodarstw domowych. Chcemy dowiedzieć się, jakie obroty osiągały firmy farmaceutyczne w ostatnim miesiącu – badamy zbiorowość firm farmaceutycznych itd.

Element zbiorowości statystycznej (populacji) jest nazywany **jednostką statystyczną**. Na przykład w zbiorowości pracowników jednostką statystyczną jest jeden pracownik. Liczba jednostek statystycznych, czyli elementów zbiorowości, jest nazywana **liczebnością zbiorowości**.

Jednostka statystyczna może być obiektem materialnym (np. pracownik, firma) bądź zjawiskiem (np. sesja giełdy, urodzenie). Jednostka statystyczna badania, powinna być ściśle określona pod względem rzeczowym (co? lub kto?), przestrzennym (gdzie?) i czasowym (kiedy?). Wskazujemy w ten sposób, co łączy wszystkie jednostki i pozwala zaliczyć je do jednej zbiorowości statystycznej. Na przykład w badaniu płac pracowników, jednostki statystyczne to pracownicy zatrudnieni w tej samej firmie (co? – pracownik firmy X) w tym samym czasie (kiedy? – czerwiec 2006) i miejscu (gdzie? – Warszawa, Polska).

Aby lepiej poznać zbiorowość statystyczną, można podzielić ją na części nazywane **podzbiorowościami (podpopulacjami, subpopulacjami)** i badać każdą z nich osobno oraz porównywać. Na przykład pracowników firmy można podzielić według płci, by sprawdzić, czy płace w tych podzbiorowościach są podobne, czy też wyraźnie różnią się.

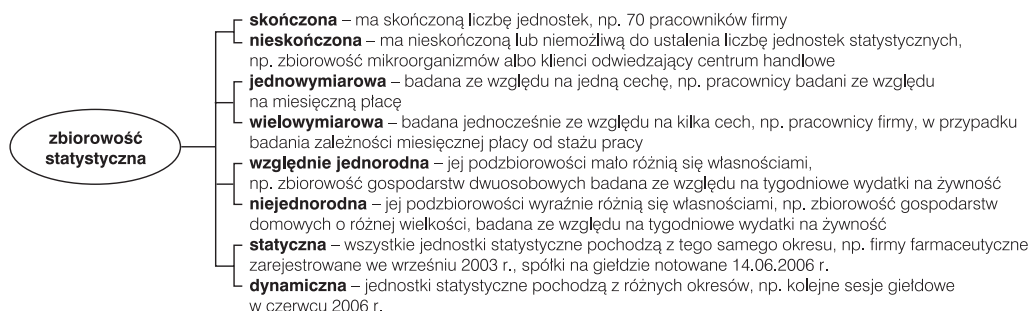
Zbiorowość statystyczną badamy ze względu na pewną jej cechę, nazywaną cechą statystyczną.

Cecha statystyczna (nazywana też zmienną) to właściwość elementów zbiorowości statystycznej będąca przedmiotem badania statystycznego.

Na przykład, we wspomnianym badaniu dotyczącym niezadowolenia pracowników firmy z wynagrodzeń, cechy statystyczne to: wysokość płacy, płeć, staż pracy, zajmowane stanowisko. Z kolei zbiorowość firm farmaceutycznych możemy badać ze względu na takie cechy, jak: miesięczne obroty, liczba przedstawicieli handlowych, średnia płaca pracowników, liczba pracowników, kwartalne koszty, roczny zysk itd.

Cecha statystyczna, która jest przedmiotem badania statystycznego pozwala rozróżnić jednostki zbiorowości między sobą. Na przykład, cecha wysokość płacy różni badanych pracowników. Pierwszy pracownik zarabia 2 tys. zł, drugi 1,5 tys. zł, następny 3 tys. zł ... itd.

Zbiorowość statystyczna i cecha statystyczna to dwa podstawowe pojęcia dyscypliny naukowej, jaką jest statystyka. Niżej przedstawiamy ich klasyfikacje. Różne sposoby klasyfikowania zbiorowości w zależności od przyjętego kryterium przedstawia rys. 1.1. Natomiast rys. 1.2. zawiera propozycję klasyfikacji cech statystycznych.



Rys. 1.1. Różne klasyfikacje zbiorowości statystycznych (ze względu na cztery nierozłączne kryteria)

Przy klasyfikacjach zbiorowości wykorzystano cztery nierozłączne kryteria, które prawdopodobnie nie wyczerpują tematu. Ważniejszym celem przedstawionych podziałów jest wprowadzenie pojęć odnoszących się do zbiorowości, którymi posługuje się statystyka i które używane są w tym podręczniku.



Rys. 1.2. Klasyfikacja cech statystycznych

Cecha mierzalna to cecha, której odmiany (warianty, kategorie) są wyrażone za pomocą liczb (są wartościami liczbowymi).

W tym miejscu należy podkreślić, że nie jest cechą mierzalną na przykład płeć albo wykształcenie, chociaż w trakcie kodowania odmianom tych cech przypisano liczby (kobieta – 1, mężczyzna – 2, wykształcenie podstawowe – 1, średnie – 2, wyższe – 3). W tym przypadku liczby spełniają taką samą rolę jak nazwy, nie są wartościami określającymi nasilenie cechy u badanych jednostek.

Cecha mierzalna skokowa może przyjmować skończoną lub przeliczalną liczbę wartości, w pewnym przedziale zmienności może przyjąć tylko niektóre wartości. Na przykład, liczba osób w gospodarstwie domowym może przyjąć wartości 1, 2, 3, 4, 5, 6, nigdy natomiast nie przyjmie wartości 3,2 lub 4,5. Przykładami cech skokowych są też: liczba samochodów wyprodukowanych w ciągu jednego dnia, liczba towarów uszkodzonych w danej dostawie, dzienna liczba zleceń firmy itp.

Obszar zmienności cechy skokowej można przedstawić następująco: • • • • • • • •

Cecha mierzalna ciągła może przyjmować nieprzeliczalnie wiele wartości. Mówimy, że cecha jest ciągła w pewnym przedziale zmienności, jeżeli może przyjąć każdą wartość z tego przedziału. Na przykład, cecha statystyczna, jaką jest wiek, może przyjąć każdą wartość z przedziału od 0 do 110 lat, a zatem również wartość 7,8654789. Inne przykłady cech ciągłych to: czas dostarczenia przesyłki, wzrost, wielkość produkcji mleka, powierzchnia mieszkania.

Obszar zmienności cechy ciągłej można przedstawić następująco: _____

Określając typ cechy mierzalnej (ciągła czy skokowa), zastanawiamy się, jaka jest natura cechy, czy może ona przyjąć teoretycznie każdą wartość z pewnego przedziału zmienności – wtedy jest ciągła, czy też może przyjąć tylko niektóre wartości (liczba tych wartości jest skończona lub przeliczalna) – wtedy jest skokowa.

Czasem wyróżnia się jeszcze jedną grupę cech mierzalnych, tzw. cechy quasi-ciągłe (prawie ciągłe).

Cecha mierzalna quasi-ciągła to cecha skokowa, która ze względu na dużą liczbę wartości traktowana jest jak ciągła, np. płaca pracownika, cena towaru, miesięczny dochód gospodarstwa domowego, kapitał własny przedsiębiorstwa, zysk przedsiębiorstwa oraz wszystkie inne cechy wyrażone w złotych lub jakiegokolwiek innej walucie. Cechy mierzalne quasi-ciągłe będziemy zaliczać do jednej grupy razem z ciągłymi.

Cecha niemierzalna to cecha, której odmiany (warianty, kategorie) są wyrażone za pomocą opisu słownego.

Przykładami cech niemierzalnych są: płeć (mężczyźni, kobiety), preferowana marka samochodu (Toyota, Fiat, Nissan,), rodzaj lub profil firmy, wykształcenie.

Warto zauważyć, że podstawą najbardziej ogólnego podziału cech na mierzalne i niemierzalne jest sposób ich pomiaru⁴. Dla wielu cech jest możliwe zaklasyfikowanie do jednej bądź drugiej grupy w zależności od tego, jaki sposób pomiaru zastosowano.

Skala pomiaru	Własności	Przykłady
Nominalna	Pozwala jedynie podzielić elementy zbiorowości na rozłączne podzbiorowości (rozróżnić jednostki)	Płeć – kobiety, mężczyźni Rodzaj wykształcenia – ekonomiczne, techniczne itd.
Porządkowa	Dodatkowo pozwala uporządkować te podzbiorowości (lub jednostki) według intensywności występowania cechy, nie ustalając odległości między nimi	Ocena produktu – bardzo zła, zła, średnia, dobra, bardzo dobra Wykształcenie – podstawowe, zawodowe, średnie, wyższe
Przedziałowa (interwałowa)	Dodatkowo pozwala ustalić różnice (odległości) intensywności występowania cechy między podzbiorowościami (lub jednostkami). Początek skali to tzw. zero względne (umowne)	Ocena produktu – 1, 2, 3, 4, 5 Temperatura w °C Rok urodzenia Godzina rozpoczęcia egzaminu
Ilorazowa (stosunkowa)	Dodatkowo pozwala ustalić iloraz intensywności występowania cechy między podzbiorowościami (lub jednostkami). Posiada zero bezwzględne	Dochód, cena, zysk, wiek, liczba zleceń, czas dojazdu, wydatki

⁴ Pomiar – przyporządkowanie określonych symboli (liczby, nazwy, znaki) własnościom mierzonych obiektów.

Na przykład wiek osób tworzących zbiorowość mierzony w latach jest cechą mierzalną. Jeśli natomiast wiek badanych osób będziemy określać jako młody, średni lub starszy, będziemy mieli do czynienia z cechą niemierzalną.

Typ skali pomiarowej jest podstawą jeszcze innej, niż zaprezentowano wyżej, klasyfikacji cech. Rodzaj cechy jest utożsamiany z rodzajem skali zastosowanej do jej pomiaru. Jakie są podstawowe własności różnych skal pomiarowych, wyjaśnia tabela na s. 16. Skale uporządkowano od najsłabszej do najsilniejszej, to znaczy każda następna posiada dodatkowe własności poza własnościami poprzedniej skali.

Podział cech na nominalne, porządkowe, przedziałowe, ilorazowe nie jest konkurencyjny do podziału na mierzalne i niemierzalne. Pomiar w skali nominalnej i porządkowej odnosi się do cech niemierzalnych (uzyskujemy opisowe kategorie cechy), natomiast pomiar w skali przedziałowej i ilorazowej do cech mierzalnych (uzyskujemy wartości liczbowe). Cecha, dla której możliwy jest pomiar w skali silniejszej, może być też mierzona we wszystkich słabszych skalach. Na przykład wysokość wkładów oszczędnościowych indywidualnych klientów banku można zmierzyć w skali ilorazowej, ale w zależności od celu badania pomiar może być dokonany w słabszej skali. Wyjaśnia to poniższy przykład.

Przykład. Określanie typu zbiorowości i cechy statystycznej

Celem badania jest poznanie wysokości wkładów oszczędnościowych (w tys. zł) indywidualnych klientów PKO BP. Chcemy dowiedzieć się, jaka jest struktura klientów ze względu na wysokość wkładu, ile wynosi przeciętny wkład, jak bardzo są one zróżnicowane. Badanie przeprowadzono w dniu 31.12.2006.

Zbiorowość statystyczna – indywidualni klienci PKO BP.

Jednostka statystyczna – jeden klient (kto? – klient indywidualny PKO BP, kiedy? – 31.12.06 r., gdzie? – Polska).

Typ zbiorowości – skończona, jednowymiarowa, względnie jednorodna, statyczna.

Cecha statystyczna – wysokość wkładu oszczędnościowego w tys. zł.

Typ cechy statystycznej – mierzalna, quasi-ciągła, pomiar w skali ilorazowej.

Jeśli jednak cel naszego badania będzie inny niż podano wyżej, dla cechy – wysokość wkładu oszczędnościowego, możliwy jest słabszy pomiar.

– Jeśli celem badania jest wskazanie klientów, którym bank proponuje Złote konto, wystarczy **pomiar nominalny**. Kategorie cechy: wysokość wkładu wystarcza do otwarcia złotego konta, wysokość wkładu nie wystarcza do otwarcia złotego konta. Przy tej skali pomiaru będziemy mogli wyłącznie odróżnić klientów spełniających od niespełniających warunki Złotego konta, podzielić ich na dwie subpopulacje.

– Jeśli bank zamierza wprowadzić dodatkowe bezpłatne usługi dla grup klientów o różnej wysokości wkładów (więcej dodatkowych usług im większa wysokość wkładu), zastosujemy **pomiar porządkowy**. Kategorie cechy np.: 1 (niski wkład, poniżej 20 tys. zł), 2 (średni wkład, 20–80 tys. zł), 3 (wysoki, powyżej 80 tys. zł). Przy tej skali pomiaru będziemy mogli nie tylko odróżnić klientów o niskiej wysokości wkładu od tych posiadających średnie czy wysokie wkłady (podzielić na trzy subpopulacje), ale także upo-

rządkować. Klienci zaliczeni do drugiej grupy mają wyższy dochód od tych zaliczonych do pierwszej. Klienci zaliczeni do trzeciej grupy mają wyższy dochód od tych zaliczonych do drugiej i pierwszej. Nie wiemy jednak, jakie są różnice wysokości wkładu pomiędzy klientami, którym przypisano wariant cechy 1, 2 lub 3.

Dokładne ustalenie różnic, a także stosunku wysokości wkładu jest możliwe dopiero przy **miarze ilorazowym**, np. wkład w wysokości 60 tys. zł jest o 40 tys. zł większy oraz trzy razy większy od wkładu o wysokości 20 tys. zł.

Nawiązując do podanego przykładu, wyjaśnimy jeszcze różnice między pomiarem przedziałowym i ilorazowym. Załóżmy, że celem badania jest poznanie ocen klientów banku odnośnie jakości obsługi. Dla cechy – ocena jakości obsługi klienta banku, możemy zastosować pomiar w **skali przedziałowej**. Na przykład, poprosimy o podanie oceny w skali od 1 do 10, gdzie 1 oznacza najniższą ocenę, a 10 najwyższą. Przy tej skali pomiaru poza porządkowaniem możliwe jest też ustalenie różnic (umownych) oceny jakości obsługi między klientami lub grupami klientów. Na przykład klienci, którzy wskazali wartość 6, oceniają jakość obsługi o 3 umowne punkty więcej niż ci, którzy wskazali wartość 3. Nie oznacza to jednak, że oceniają jakość obsługi klientów banku 2 razy lepiej.

1.3. Etapy badania statystycznego

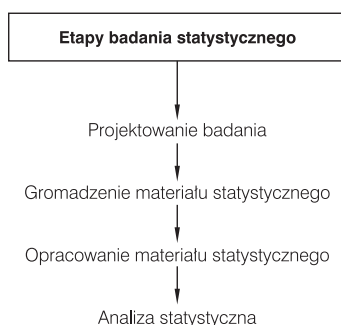
Badanie statystyczne to ogół prac mających na celu poznanie własności zbiorowości statystycznej.

W zależności od zakresu, badania statystyczne dzieli się na pełne i częściowe.

Badanie statystyczne pełne obejmuje wszystkie jednostki zbiorowości.

Badanie statystyczne częściowe obejmuje tylko niektóre jednostki statystyczne. Jednostki te są wybrane w określony sposób ze zbiorowości i nazywane **próbą statystyczną**.

Niezależnie od rodzaju badania, bez względu na to jaką zbiorowość analizujemy, wspólne wszystkim badaniom statystycznym są cztery etapy (rys. 1.3).

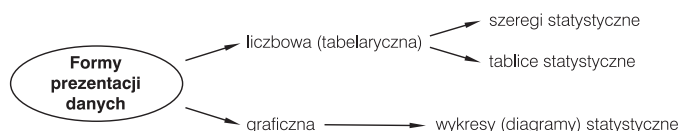


Rys. 1.3. Etapy badania statystycznego

Projektowanie badania. W tym etapie przygotowujemy badanie od strony merytorycznej i technicznej. Najważniejsze jest sprecyzowanie celu badania, zarówno diagnostycznego (co i dlaczego chcemy badać), jak i praktycznego (komu i czemu badanie ma służyć). Określamy zbiorowość statystyczną i cechy statystyczne podlegające badaniu oraz podejmujemy decyzję o rodzaju badania (pełne, częściowe). Wybieramy rodzaj danych statystycznych, określamy sposób ich pozyskania i przetwarzania. W przypadku konieczności uzyskania danych wcześniej nieistniejących niezbędne jest skonstruowanie instrumentu pomiarowego, jakim najczęściej jest kwestionariusz wywiadu⁵. Etap projektowania obejmuje też przygotowanie badania od strony technicznej, na przykład: wylosowanie próby, przygotowanie ankiet, przeszkolenie ankierów itp. Konieczne jest też ustalenie harmonogramu prac.

Gromadzenie materiału statystycznego. Wszystkie jednostki zbiorowości bądź próby statystycznej poddajemy obserwacji. Gromadzimy informacje o wartościach cech statystycznych ilościowych lub odmianach cech jakościowych. Docieramy również do danych, które zostały zgromadzone wcześniej, np. przez Główny Urząd Statystyczny (GUS). Materiał statystyczny poddajemy kontroli formalnej i merytorycznej w celu wyeliminowania ewentualnych błędów powstałych w trakcie jego gromadzenia. Jak dobrać dane odpowiednie dla danego badania statystycznego wyjaśniamy w następnym rozdziale.

Opracowanie materiału statystycznego polega na porządkowaniu, grupowaniu i prezentacji danych statystycznych. Sposoby opracowania danych, a w szczególności formy prezentacji, różnią się w zależności od ich rodzaju. Omówimy je szczegółowo w dalszych częściach podręcznika. W tym miejscu zwracamy uwagę, że można wyróżnić dwie formy prezentacji wspólne dla wszystkich typów danych (por. rys. 1.4).



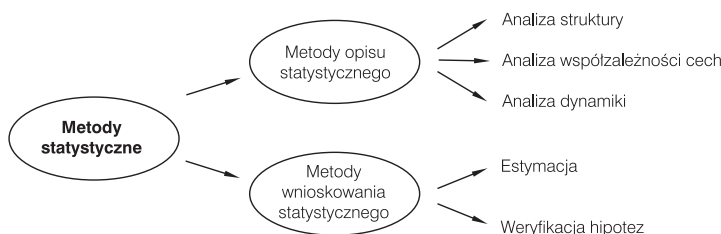
Rys. 1.4. Podstawowe formy prezentacji danych statystycznych

Szereg statystyczny to ciąg danych statystycznych (wartości cechy ilościowej bądź odmian cechy jakościowej) uporządkowany według określonego kryterium.

Tablica statystyczna to zestawienie dwu lub więcej szeregów statystycznych.

Analiza statystyczna polega na ocenie własności zbiorowości statystycznych, przy wykorzystaniu właściwych metod statystycznych. Ich klasyfikację prezentuje rys. 1.5.

⁵ W opinii wielu autorów projektowanie kwestionariuszy jest nadal sztuką, a nie nauką (por. G.A. Churchill, *Badania marketingowe. Podstawy metodologiczne*, Wydawnictwo Naukowe PWN, Warszawa 2002). Niemniej jednak wiedza na ten temat jest dosyć duża. Zainteresowanych odsyłam do pozycji C. Frankfort-Nachmias, D. Nachmias, *Metody badawcze w naukach społecznych*, Zys i S-ka Wydawnictwo, Poznań 2001, rozdz. 11.



Rys. 1.5. Klasyfikacja metod statystycznych

Metody opisu statystycznego pozwalają opisać własności całej zbiorowości (w przypadku badania pełnego) lub próby statystycznej (w przypadku badania częściowego) badanej ze względu na wybraną cechę lub cechy statystyczne. Opis statystyczny może dotyczyć:

- własności jednowymiarowej zbiorowości statystycznej (analiza struktury),
- własności wielowymiarowej zbiorowości statystycznej (analiza współzależności cech),
- własności zbiorowości dynamicznej (analiza dynamiki zjawisk).

Metody opisu statystycznego są przedmiotem tego podręcznika. Szczegółowo omawiamy je w następnych rozdziałach.

Metody wnioskowania statystycznego pozwalają poznać (z określoną precyzją) własności nieznannej zbiorowości statystycznej na podstawie próby. Jeśli chcemy, np. dowiedzieć się, ile wynoszą przeciętne wydatki na telefon wśród studentów UW, gdy w próbie losowej 100 studentów średnia ta jest równa 60 zł, zastosujemy metody szacowania, inaczej estymacji. Jeśli natomiast chcemy sprawdzić przypuszczenie, że przeciętne wydatki na telefon studentów UW przekraczają 50 zł, wykorzystamy metody weryfikacji hipotez. Estymacja i weryfikacja hipotez może dotyczyć zarówno struktury nieznannej zbiorowości, współzależności cech, jak i dynamiki zjawisk.



1.4. Pytania sprawdzające

1. Czym zajmuje się statystyka?
2. Jakie są główne obszary zastosowań statystyki? Spróbuj podać własne przykłady.
3. Co to jest zbiorowość statystyczna, podzbiorowość, próba, jednostka? Podaj przykłady.
4. Podaj przykład zbiorowości i określ pod względem rzeczowym, przestrzennym i czasowym jednostkę statystyczną.
5. Co to jest cecha statystyczna? Podaj, ze względu na jakie cechy statystyczne mógłbyś zbadać wybraną przez siebie zbiorowość?
6. Omów, podając przykłady, klasyfikację zbiorowości i cechy statystycznej.
7. Wymień etapy badania statystycznego i najważniejsze działania podejmowane w każdym z etapów.
8. Omów znaną ci klasyfikację metod statystycznych.