

Iwona Jażdżewska

**STATYSTYKA
DLA GEOGRAFÓW**

Wydanie II



WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

STATYSTYKA DLA GEOGRAFÓW



40 LAT

WYDAWNICTWA

UNIWERSYTETU
ŁÓDZKIEGO

Iwona Jażdżewska

**STATYSTYKA
DLA GEOGRAFÓW**

Wydanie II



WYDAWNICTWO
UNIwersytetu
ŁÓDZKIEGO

ŁÓDŹ 2013

Iwona Jażdżewska – Instytut Geografii Miast i Turyzmu
Wydział Nauk Geograficznych, Uniwersytet Łódzki, 90-323 Łódź, ul. Kopcińskiego 31

RECENZENT

Jerzy Runge

REDAKTOR WYDAWNICTWA UŁ

Danuta Bąk

SKŁAD I ŁAMANIE

AGENT PR

PROJEKT OKŁADKI

Ewa Laśkiewicz

© Copyright by Uniwersytet Łódzki, Łódź 2013

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie II. W.03575.13.1.S

ISBN (wersja drukowana) 978-83-7525-984-1
ISBN (ebook) 978-837969-334-4

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63, faks (42) 665 58 62

SPIS TREŚCI

WSTĘP	7
1. WPROWADZENIE	9
1.1. Podstawowe pojęcia matematyczne	10
1.2. Podstawowe pojęcia statystyczne	13
1.3. Metoda reprezentacyjna	16
1.4. Zadania	22
2. PREZENTACJA DANYCH STATYSTYCZNYCH	31
2.1. Szeregi statystyczne	32
2.2. Tablice statystyczne	41
2.3. Graficzna prezentacja danych statystycznych	43
2.4. Zadania	57
3. ROZKŁADY ZMIENNYCH LOSOWYCH I ICH WŁASNOŚCI	67
4. ANALIZA JEDNEJ ZMIENNEJ	72
4.1. Miary średnie	73
4.1.1. Zadania	94
4.2. Miary rozproszenia	105
4.2.1. Zadania	119
4.3. Miary asymetrii i koncentracji	127
4.3.1. Zadania	146
5. ANALIZA WSPÓLZALEŻNOŚCI	152
5.1. Zadania	172
6. ANALIZA DYNAMIKI	183
6.1. Wskaźniki dynamiki	183
6.2. Wyznaczanie tendencji rozwojowych	187
6.3. Zadania	193
TABLICA STATYSTYCZNA. ROZKŁAD t -Studenta	204
LICZBY LOSOWE	205
INDEKS TERMINÓW	206
LITERATURA	212

WSTĘP

Niniejszy podręcznik przeznaczony jest dla studentów wyższych uczelni kierunków geograficznych, którzy uczęszczają na wykłady i ćwiczenia ze statystyki. Głównym jego zadaniem jest pomoc w rozwiązywaniu zadań i ich interpretacji.

Prezentowane zagadnienia pogrupowane zostały w poszczególne działy o podobnej strukturze. Każdy z nich rozpoczynają wiadomości teoretyczne wykorzystywane w danej części, następnie rozwiązywane są przykładowe zadania, w których umieszczono algorytm obliczeń oraz właściwą interpretację wyników, dalej zaprezentowane są zadania do rozwiązania samodzielnego. Rozwiązania części z nich podane zostały na końcu rozdziałów. Zadania mają różny stopień trudności, a także ich wykonanie wymaga różnego czasu. Wiele z nich można rozwiązać, przy użyciu dostępnych komputerowych pakietów statystycznych lub arkuszy kalkulacyjnych. Specyfika studiów geograficznych wymaga badania danych w przestrzeni, stąd w każdym z rozdziałów znajduje się kilka przykładów prezentujących wyniki na mapie. Więcej przykładów przestrzennej prezentacji wyników analiz statystycznych należy szukać w opracowaniach kartograficznych lub w Systemach Informacji Geograficznej (GIS).

Większość zadań powstała w trakcie ćwiczeń ze statystyki prowadzonych ze studentami studiującymi na Wydziale Nauk Geograficznych Uniwersytetu Łódzkiego. Studenci w trakcie zajęć nie tylko rozwiązywali zadania, ale również sami je układali. Przedstawiali także własny problem badawczy, poszukiwali do jego rozwiązania odpowiednich metod statystycznych i prezentowali ich interpretację. Chciałabym wyrazić podziękowanie moim byłym studentom za ich trud i intelektualną przygodę, jaką niektórzy z nich przedstawili w swoich pracach. Nie zawsze i nie wszystkim studentom studiowanie statystyki przychodziło z łatwością, wielu z nich miało z nią problemy. Swoim młodszym koleżankom i kolegom przekazują oni następujące wskazówki:

1. Przygotuj się do ćwiczeń z zagadnień, które były omawiane na wykładach lub zostały zadane do samodzielnego przygotowania na poprzednich ćwiczeniach, nie zostawiaj na ostatnią chwilę przyswojenia teorii.

2. Jeśli jesteś przygotowany do zajęć, możesz zadać wykładowcy pytania dotyczące zagadnień, których nie rozumiesz; nie bój się pytać; jeśli jesteś nieprzygotowany, nie masz możliwości wzięcia udziału w dyskusji.

3. Jeśli nie rozumiesz tematu, nie przejmuj się, tylko zrób przerwę i przeczytaj go jeszcze raz.

4. Rób notatki na wykładach i ćwiczeniach, wielu komentarzy i przykładów przedstawianych przez wykładowców na zajęciach nie znajdziesz w podręczniku; notatki przydadzą Ci się przed kolokwium i egzaminem.

5. Rozwiązuj zadania zamieszczone w podręczniku – często na kolokwium są przykłady podobne do tych w podręczniku.

6. Przygotowuj się do egzaminu z innymi koleżankami i kolegami ze studiów, w grupie łatwiej się uczyć. Wymyślajcie własne zadania podobne do tych, które rozwiązywaliście wcześniej.

7. Egzamin lub kolokwium w pierwszym terminie są często najłatwiejsze.

8. Jeśli Ci zależy na dobrej ocenie i zrozumieniu tematu, przeglądaj tematykę wykładów i ćwiczeń z wyprzedzeniem. Wykładowca powinien podać kolejność prezentacji zagadnień, warto zapoznać się z nią i wcześniej ją przestudiować, łatwiej Ci będzie zrozumieć prezentowany temat.

9. Proste obliczenia wykonuj samodzielnie lub tylko przy pomocy kalkulatora.

10. Wiele obliczeń możesz wykonać przy użyciu narzędzi informatycznych. Zwróć uwagę na interpretację wyników oraz założenia, jakie muszą być spełnione, aby te obliczenia miały sens.

11. Szukaj danych do przykładów w wiarygodnych źródłach, np. GUS lub Eurostatu.

12. Jeśli ten podręcznik Ci nie odpowiada, poszukaj innego, może będzie dla Ciebie bardziej przystępny.

13. Nie bój się statystyki, przyda Ci się podczas pisania pracy licencjackiej i magisterskiej.

Jeśli masz własne uwagi, które powinny być uwzględnione w następnych wydaniach podręcznika, napisz do autorki iwona.jazdzewska@uni.lodz.pl

Iwona Jażdżewska

1. WPROWADZENIE

Statystyka jest działem matematyki, stąd też pewnie bierze się niechęć do niej niektórych osób studiujących geografię, które wybierając ten kierunek studiów, miały nadzieję, iż pożegnały się z matematyką na zawsze. Z mojego doświadczenia wynika, iż wielu z Was, zwłaszcza Ci, którzy mieli w szkole kłopoty z matematyką, uważa statystykę za przedmiot trudny i podchodzi do niej z obawą¹. Spróbujemy przezwyciężyć te trudności, a poznane metody statystyczne wykorzystać do rozwiązywania własnych problemów geograficznych.

W badaniach geograficznych występuje wiele informacji, które zazwyczaj (choć nie zawsze) przyjmują formę liczb. Niekiedy jest ich kilka i można je natychmiast poddać analizie, z czasem przybywa ich, może ich być kilkaset czy nawet kilka tysięcy i wówczas, aby je zinterpretować czy wykryć pewne prawidłowości, trzeba się posłużyć metodami badawczymi oferowanymi przez statystykę.

PRZYKŁAD 1

Przypuśćmy, że ktoś z Was bada turystów, którzy wybrali się do Muzeum na Wawelu. Interesujące jest, kiedy i kto przychodzi do tej placówki. Możemy dowiedzieć się, ile osób odwiedziło to muzeum w ciągu roku, miesiąca, tygodnia, dnia, a nawet godziny. Dane można przedstawić w formie tabelarycznej. Wyobraźmy sobie zestawienie roczne, w którym umieszczony będzie każdy dzień z informacją o zwiedzających w poszczególnych przedziałach czasowych, przykładowo co 2 godz. Po analizie jednego dnia łatwo możemy wysnuć wnioski, np. że największe natężenie wizyt turystów było w godz. 12–14 lub że połowa z nich przyszła do godz. 12. Jednak, gdy sporządzimy zestawienie miesięczne czy roczne, ogrom cyfr wymusza na nas obliczenia.

Aby zbadać zjawiska masowe, korzysta się z informacji zebranych osobiście (np. przez ankiety), jak również z danych oferowanych przez placówki powołane do ich zbierania, np. GUS. Przetwarzanie tak wielkiej ilości informacji wymaga znajomości metod, a także odpowiedniego oprogramowania komputerowego, które eliminuje czasochłonne obliczenia. Nie chcąc być „niewolnikiem” komputera, można sobie poradzić w sytuacjach, gdy go nie będzie pod ręką i trzeba będzie wykonać proste obliczenia. Procedury tych obliczeń są prezentowane w przykładach.

¹ Osoby, które chciałyby przypomnieć sobie więcej wiadomości z matematyki, powinny skorzystać z prac Huk (2001) lub Jokiel, Kostrubiec (1981), pisanych przez matematyków dla geografów.

By rozwiązać zadania ze statystyki, potrzebna jest umiejętność posługiwania się podstawowymi operacjami, symbolami i oznaczeniami matematycznymi, takimi jak np.: $+$, $-$, \times , \div , $\sqrt{\quad}$, \leq , \geq , \neq , \in , \notin , \sum , \prod , a także znajomość kilku liter greckich, wykorzystywanych w matematyce: α , β , γ , ν , μ , δ , Σ , σ , π , Π , ε , ξ , χ , φ .

1.1. PODSTAWOWE POJĘCIA MATEMATYCZNE

Większość prezentowanych w zbiorze pojęć matematycznych występuje w programie matematyki na poziomie szkoły średniej, jednakże sens niektórych z nich zostanie krótko wyjaśniony.

Zbiory oznaczamy będziemy dużymi literami (np. X , Z), a ich elementy odpowiednio małymi literami. Fakt przynależności elementu x do zbioru X zapisywać będziemy $x \in X$, natomiast $y \notin X$ oznacza, że element y nie należy do zbioru X .

Jeżeli zbiór X ma skończoną liczbę elementów, symbol x_i oznacza i -ty element tego zbioru, przy czym i może przyjmować wartości całkowite od 1 do n włącznie, co zapisujemy $i = 1, \dots, n$.

Zbiory X , Y nazywamy **rozłącznymi**, jeżeli nie mają wspólnego elementu, tzn. jeżeli $X \cap Y = \emptyset$.

Jeżeli każdy element ze zbioru X należy do zbioru Y , to mówimy, że zbiór X **zawiera się** w zbiorze Y i piszemy $X \subset Y$ lub $Y \supset X$.

Jeżeli $X \subset Y$, to X nazywamy **podzbiorem** zbioru Y .

Przedział otwarty $(a; b)$ określamy następująco:

$$x \in (a; b) \Leftrightarrow a < x < b$$

W przedziale tym nie ma liczby najmniejszej i nie ma liczby największej. Liczby a i b są **kresami**: dolnym i górnym zbioru $(a; b)$.

Przedział zamknięty (domknięty) $\langle a; b \rangle$ określamy następująco:

$$x \in \langle a; b \rangle \Leftrightarrow a \leq x \leq b$$

W przedziale tym istnieje liczba najmniejsza a i największa b .

Przedział lewostronnie domknięty określamy:

$$x \in \langle a; b \rangle \Leftrightarrow a \leq x < b$$

Przedział prawostronnie domknięty określamy:

$$x \in (a; b] \Leftrightarrow a < x \leq b$$

Wartość bezwzględna liczby x oznaczamy symbolem $|x|$ i określamy:

$$|x| = \begin{cases} x, & \text{gdy } x \geq 0 \\ -x, & \text{gdy } x < 0 \end{cases}$$

Sumę n elementów $x_1 + x_2 + x_3 + \dots + x_n$ możemy zapisać w postaci symbolu $\sum_{i=1}^n x_i$. Wskaźnik i nie musi przyjmować wartości począwszy od 1, np. $\sum_{i=4}^6 x_i$ oznacza sumę 3 elementów $x_4 + x_5 + x_6$.

Iloczyn n elementów $x_1 \times x_2 \times x_3 \times \dots \times x_n$ możemy zapisać w postaci: $\prod_{i=1}^n x_i$.

Symbol $n!$ (czyt. ***n* silnia**) oznacza iloczyn $n! = 1 \times 2 \times 3 \times \dots \times n$, gdzie $n \in N$ oraz $0! = 1! = 1$. Liczba ***k*-elementowych kombinacji bez powtórzeń *n* elementów**, którą oznaczamy symbolami C_n^k lub $\binom{n}{k}$ (czyt. ***n* nad *k***), jest równa

$$\frac{n!}{k!(n-k)!}$$

Macierzą $M \times N$ nazywamy układ $m \times n$ liczb ustawionych w formie tablicy prostokątnej mającej m wierszy i n kolumn. Liczby tworzące tablicę nazywamy elementami macierzy. Zapisujemy ją:

$$A_{m \times n} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix}$$

Pierwszy wskaźnik jest numerem wiersza, a drugi kolumny, np. macierz 3×2 jest macierzą

$$A_{m \times n} = \begin{vmatrix} 2 & 1 \\ 3 & 5 \\ 4 & 6 \end{vmatrix}$$

Tablica zwana **macierzą geograficzną** B. J. L. Berry'ego to taka tablica, w której elementy pewnego zbioru (mogą to być np. województwa czy inne jednostki przestrzenne) będą opisywane zbiorem pewnych cech. Macierz Berry'ego ma tyle wierszy, ile jest badanych jednostek i tyle kolumn, ile jest cech w tych jednostkach. Każdy z wierszy odpowiada jednej jednostce, czyli wiersz i -ty zawiera wartości wszystkich cech opisujących tę i -tą jednostkę. Symbol x_{ik} oznacza wartość k -tej cechy w i -tej jednostce. Każda kolumna odnosi się do jednej cechy dla wszystkich jednostek zbioru.

W statystyce występują liczby **bezwzględne (absolutne)** i **względne**. Liczby absolutne są to wielkości, które otrzymujemy w wyniku mierzenia lub sumowania jednostek zbiorowości. Każda z nich, wyrażająca rozmiar badanego zjawiska, jest mianowana. Liczby względne powstają przez porównanie ze sobą dwóch liczb. Odgrywają ważną rolę przy porównywaniu zjawisk. Są to przykładowo wielkości procentowe lub wskaźniki natężenia opisujące relacje między różnymi zbiorowościami, np.: gęstość zaludnienia, spożycie cukru na jednego mieszkańca, plony z 1 ha.

Zdarzeniami losowymi nazywamy takie wyniki uzyskiwane przez realizację danego doświadczenia (procesu), które mogą w określonym zespole warunków wystąpić lub nie wystąpić.

W doświadczeniach można wyróżnić **zdarzenia złożone** i **zdarzenia elementarne**. Zdarzenie złożone składa się ze zdarzeń elementarnych. Ściśle określone zdarzenie złożone składa się z danej liczby zdarzeń elementarnych.

PRZYKŁAD 1.1.1

Zdarzeniem złożonym będzie wyrzucenie (kostką do gry) parzystej liczby oczek. Składa się ono z trzech zdarzeń elementarnych:

- uzyskania dwóch oczek,
- uzyskania czterech oczek,
- uzyskania sześciu oczek.

Jeśli każdorazowa realizacja określonego doświadczenia daje w wyniku to samo zdarzenie A , to zdarzenie to nazywamy **zdarzeniem pewnym**. Zdarzeniu pewnemu przyporządkowany jest zbiór wszystkich zdarzeń elementarnych. Jeśli każdorazowa realizacja określonego doświadczenia nie daje w wyniku zdarzenia A , to zdarzenie A nazywamy **zdarzeniem niemożliwym**. Zdarzeniu niemożliwemu odpowiada pusty zbiór zdarzeń elementarnych. Jeśli realizacja określonego doświadczenia niekiedy prowadzi do zdarzenia A , a niekiedy do zdarzenia \bar{A} nie prowadzi, to zdarzenie A jest **zdarzeniem losowym** (przypadkowym).

PRZYKŁAD 1.1.2

Jeżeli w urnie znajdują się wyłącznie kule białe, to zdarzenie polegające na wyciągnięciu kuli białej będzie zdarzeniem pewnym. Natomiast zdarzenie polegające na wyciągnięciu kuli czerwonej będzie wówczas zdarzeniem niemożliwym.

Jeśli w urnie będą znajdowały się zarówno kule białe, jak i czerwone, to wylosowanie kuli białej jest zdarzeniem losowym.

Własności zdarzeń elementarnych wchodzi w zakres materiału ze szkoły średniej. W tym miejscu przypomnimy jedynie niektóre z nich:

a) każdemu zdarzeniu należącemu do danego zbioru zdarzeń przyporządkowana jest pewna liczba $P(A)$, która spełnia warunek:

$$0 \leq P(A) \leq 1$$

Liczba ta jest prawdopodobieństwem zdarzenia A ;

b) prawdopodobieństwo zdarzenia pewnego jest równe jedności;

c) prawdopodobieństwo zdarzenia niemożliwego równa się zeru;

d) prawdopodobieństwo sumy zdarzeń jest równe sumie prawdopodobieństw poszczególnych zdarzeń;

e) suma prawdopodobieństw danego zdarzenia i zdarzenia do niego przeciwnego jest równa jedności;

f) znając prawdopodobieństwo zdarzeń elementarnych, można obliczyć prawdopodobieństwo dowolnego zdarzenia losowego.

Kiedy zajmujemy się statystyką, wcześniej lub później musimy się zetknąć z liczbą e i wiedzieć, co ona oznacza i do czego służy (np. rozkład prawdopodobieństwa Poissona, normalny itp.).

Wartość stałej e jest równa sumie wyrazów nieskończonego, malejącego szeregu:

$$e = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} + \dots$$

Szereg ten jest zbieżny². Wartość $e = 2,7183$ po uwzględnieniu czterech miejsc po przecinku.

1.2. PODSTAWOWE POJĘCIA STATYSTYCZNE

Zbiorowością statystyczną (populacją statystyczną³) nazywamy ogół elementów (np. przedmiotów, jednostek administracyjnych, osób, zdarzeń) poddanych badaniu statystycznemu. Przedmiotem badań geograficznych są struktury i procesy obserwowane w jednostkach przestrzennych. Stąd, **geograficzna populacja generalna** to zbiór obiektów mających wspólne cechy geograficzne, czyli precyzyjnie określone położenie geograficzne (zlewiska, dorzecza, kontynenty, gminy). W zależności od potrzeb, skali badań i naszych umiejętności można posługiwać się różnymi metodami podawania położenia geograficznego obiektu. Są to metody: nominalna (podajemy nazwę kraju lub obiektu, np.: Kraków, Katowice,

² Wyrażenie $a_1 + a_2 + \dots + a_n + \dots$ nazywamy szeregiem nieskończonym o składnikach a_n i oznaczamy $\sum_{i=1}^{\infty} a_n$. Ciąg (S_n) taki, że $s_n = a_1 + a_2 + \dots + a_n$ dla $n \geq 1$ nazywamy ciągiem sum częściowych szeregu. Jeżeli ciąg (S_n) jest zbieżny, to szereg nazywamy zbieżnym (Dziubiński, Świątkowski 1980).

³ Obydwa określenia są równorzędne, lecz niekiedy przyjmuje się, że populacja generalna (od łac. *populatio* – 'ludność') dotyczy zbioru ludności.

Wrocław), współrzędne lokalne (określamy obiekt i odległość, np. w promieniu 500 m od leśniczówki), porządkowa (numeracja domów, np. wzdłuż ulicy Piotrkowskiej), współrzędne kartograficzne (długość i szerokość geograficzna), topologiczna (czyli sąsiedztwa obiektów, np. graniczy z Morzem Bałtyckim). Populacja może być jednocechowa (jednowymiarowa) i wielocechowa.

PRZYKŁAD 1.2.1

Można mówić o zbiorowości województw, wierzchołków górskich, miast, jezior, mieszkańców miast, przedsiębiorstw, dróg. Każda z tych populacji ma pewne cechy i każdemu z jej elementów zostały przyporządkowane wartości tych cech.

Jednostkami statystycznymi nazywamy elementy zbiorowości statystycznej, powiązane ze sobą logicznie tak, aby można je było przyporządkować danej populacji. Każdej z jednostek można przyporządkować pewne cechy i ich wartości.

PRZYKŁAD 1.2.2

Jeśli jednostką statystyczną jest kino, to cechą wspólną łączącą ją z innymi placówkami kulturalnymi jest funkcja, jaką ma do spełnienia, a cechami różniącymi wybrane kina od innych są takie wielkości, jak liczba widzów, liczba zatrudnionych, liczba seansów.

Jednostki statystyczne powinny być precyzyjnie określone pod względem **rzeczowym** (co lub kogo badamy), **przestrzennym** (gdzie odbywają się badania) oraz **czasowym** (w jakim okresie lub w jakiej chwili się one odbywają). Własności, jakimi charakteryzują się jednostki statystyczne, nazywamy **cechami statystycznymi**. Zbiór cech dzielimy na mierzalne i niemierzalne.

Cechy mierzalne, zwane wymiernymi lub ilościowymi, są to te własności jednostek statystycznych, które zostały zmierzone lub zważone i dają się wyrazić za pomocą liczb. Podane są one w jednostkach miary (np. m, kg, zł). Cechy mierzalne mogą być skokowe lub ciągłe. **Cecha mierzalna ciągła** to cecha, której wartości są liczbami z pewnego przedziału liczbowego i mogą przyjmować wszystkie wartości z tego przedziału, np. temperatura powietrza, wysokość nad poziomem morza, powierzchnia działki. **Cecha mierzalna skokowa (dyskretna)** to cecha, której wartości mogą przyjmować jedynie skończoną lub przeliczaną liczbę wartości, np. liczba dzieci w rodzinie, liczba teatrów w mieście.

Cechy niemierzalne, zwane niewymiernymi, jakościowymi, są to te własności jednostek statystycznych, których nie można zmierzyć. Przykładem cech niemierzalnych jest płeć, narodowość. Wśród cech niemierzalnych mogą wystąpić cechy stopniowalne lub dwudzielne. **Cechy stopniowalne** mają określoną liczbę wariantów, np. niski, wyższy i najwyższy, natomiast **cechy dwudzielne** określają, czy dane jednostki posiadają daną cechę czy nie (innych możliwości nie ma), np. odpowiedź na pytanie: Tak/Nie.

PRZYKŁAD 1.2.3

Na podstawie danych zawartych w tabeli określ: zbiorowość statystyczną, jednostkę statystyczną, liczebność.

Tabela 1.2.1

Szkolnictwo w Polsce w roku szkolnym 1980/1981

Szkoły	Liczba szkół
Podstawowe	13 524
Szkolne punkty filialne	6 468
Artystyczne I stopnia	199
Niepełne średnie – zawodowe	3 507
Średnie	7 207
ogólnokształcące	1 230
zawodowe	5 977
Policealne	1 181
Wyższe	91

Źródło: „Rocznik Statystyczny” [GUS], 1981, s. 515.

Jakie cechy statystyczne można wyróżnić w tej zbiorowości?

Zbiorowością statystyczną będą wszystkie szkoły działające w Polsce w roku szkolnym 1980/1981. Jednostką statystyczną będzie jedna szkoła. Według „Rocznika Statystycznego” liczebność charakteryzowanej zbiorowości wynosiła 32 177 szkół.

W zależności od celu przeprowadzanego badania można brać pod uwagę następujące cechy statystyczne:

- liczbę absolwentów, liczbę nauczycieli – cecha mierzalna skokowa,
- powierzchnię działki, którą zajmuje szkoła – cecha mierzalna ciągła,
- źródło finansowania szkoły – samorządowa, społeczna, prywatna,
- rodzaj szkoły (podstawowa, zawodowa itd.) – cecha niemierzalna stopniowalna (według tej cechy podzielono zbiorowość w tab. 1.2.1),
- działalność drużyny ZHP (tak lub nie) – cecha niemierzalna dwudzielna.

Nie zawsze jednak będziemy mogli zbadać całą zbiorowość statystyczną. W takim przypadku badania prowadzimy w części zbiorowości. **Próba** nazywamy część populacji statystycznej wybraną za pomocą określonego sposobu losowania w celu zbadania własności całej populacji. Aby informacje pochodzące z próby były obiektywne i wiarygodne, musi być ona losowana według specjalnych zasad, nie może być pobierana w sposób tendencyjny. Oznacza to, iż fakt zaliczenia obiektu do próby nie może zależeć od wielkości cechy przypisanej obiektowi.