

**Anna Krajewska**

# **STATYSTYKA**

**dla pedagogów**

**Anna Krajewska**

# **STATYSTYKA**

**dla pedagogów**

*wybrane zagadnienia*

**Trans Humana**

**Białystok 2001**

Recenzja naukowa:    prof. dr hab. Jerzy Niemiec  
                                 dr hab. Anna Kwiatkowska

Projekt okładki:        Mieczysław Rabczko  
Skład:                     Roman Sakowski

© Copyright by        Trans Humana Wydawnictwo Uniwersyteckie  
                                 15-328 Białystok, ul. Świerkowa 20  
                                 tel./fax (085) 745-74-23  
                                 <http://pip.uwb.edu.pl/transhumana>  
                                 e-mail: [transhum@uwb.edu.pl](mailto:transhum@uwb.edu.pl).

Białystok 2001  
Wydanie III

**ISBN 83-86696-85-0**

Druk i oprawa: „AG-BART”, Białystok

# SPIS TREŚCI

|                                                           |    |
|-----------------------------------------------------------|----|
| <b>WSTĘP</b> .....                                        | 7  |
| ROZDZIAŁ I                                                |    |
| <b>WPROWADZENIE W PROBLEMATYKĘ</b> .....                  | 9  |
| 1. PODSTAWOWE POJĘCIA .....                               | 10 |
| 2. POMIAR I SKALE POMIAROWE .....                         | 14 |
| 3. ETAPY I RODZAJE BADANIA STATYSTYCZNEGO .....           | 16 |
| 4. METODY DOBORU PRÓBY REPREZENTACYJNEJ .....             | 20 |
| ROZDZIAŁ II                                               |    |
| <b>GRUPOWANIE I PREZENTACJA ZEBRANEGO MATERIAŁU</b> ..... | 23 |
| 1. GRUPOWANIE STATYSTYCZNE .....                          | 23 |
| 2. BUDOWA SZEREGU ROZDZIELCZEGO .....                     | 26 |
| 3. TABLICA STATYSTYCZNA .....                             | 31 |
| 3.1. Budowa tablicy .....                                 | 31 |
| 3.2. Rodzaje tablic .....                                 | 35 |
| 4. WYKRESY .....                                          | 38 |
| PYTANIA I ZADANIA .....                                   | 51 |

### ROZDZIAŁ III

## **CHARAKTERYSTYKA ZBIOROWOŚCI STATYSTYCZNEJ** ..... 53

|                                                          |    |
|----------------------------------------------------------|----|
| 1. OPIS ROZKŁADU CECHY JAKOŚCIOWEJ .....                 | 53 |
| 1.1. Wskaźniki struktury i natężenia .....               | 54 |
| 1.2. Układ i analiza wskaźników struktury w tablicy..... | 57 |
| 2. OPIS ROZKŁADU CECHY ILOŚCIOWEJ.....                   | 64 |
| 2.1. Miary przeciętne.....                               | 66 |
| 2.1.1. Średnia arytmetyczna zwykła i ważona.....         | 67 |
| 2.1.2. Średnia harmoniczna .....                         | 70 |
| 2.1.3. Mediana.....                                      | 71 |
| 2.1.4. Dominanta .....                                   | 74 |
| 2.2. Miary rozproszenia .....                            | 76 |
| 2.2.1. Odchylenie przeciętne .....                       | 77 |
| 2.2.2. Wariancja i odchylenie standardowe .....          | 79 |
| 2.3. Miary asymetrii .....                               | 83 |
| 2.4. Obliczanie i interpretacja miar zmiennej .....      | 86 |
| PYTANIA I ZADANIA .....                                  | 89 |

### ROZDZIAŁ IV

## **WSPÓŁZALEŻNOŚĆ ZJAWISK W BADANIACH PEDAGOGICZNYCH** ..... 93

|                                                      |     |
|------------------------------------------------------|-----|
| 1. ZWIĄZEK CECH JAKOŚCIOWYCH.....                    | 95  |
| 2. ZWIĄZEK CECH ILOŚCIOWYCH.....                     | 96  |
| 2.1. Współczynnik korelacji Pearsona.....            | 97  |
| 2.2. Współczynnik korelacji rangowej Spearmana ..... | 102 |
| PYTANIA I ZADANIA .....                              |     |

### ROZDZIAŁ V

## **PODSTAWY WNIOSEKOWANIA STATYSTYCZNEGO** ..... 108

|                                      |     |
|--------------------------------------|-----|
| 1. ZMIENNA LOSOWA I JEJ RODZAJE..... | 109 |
| 2. ROZKŁAD NORMALNY .....            | 110 |
| 3. ESTYMACJA PARAMETRÓW .....        | 112 |

# W S T Ę P

Metody statystyczne mają zastosowanie nie tylko w naukach matematyczno-przyrodniczych, ale także społecznych, w tym w pedagogice. Wykrycie i opis prawidłowości zachodzących w rzeczywistości pedagogicznej wymaga analizy rzeczy i zjawisk nie tylko od strony jakościowej, ale także ilościowej.

W opracowaniu i analizie wyników badań pamiętać należy o konieczności wyważenia potrzeby zakresu stosowania metod statystycznych. Nie można bezkrytycznie dokonywać różnorodnych obliczeń statystycznych. Zanim badacz zdecyduje się na stosowanie określonej metody statystycznej, powinien mieć pewność co do zasadności i racjonalności takiego postępowania. Metody statystyczne powinny badaczowi pomagać w opisie, analizie i przewidywaniu badanych zjawisk, a nie mogą być celem samym w sobie.

Pomimo że statystyka często jest niezbędna w analizie współczesnych badań pedagogicznych, to jednak pełni w nich rolę pomocniczą i powinna być uzupełnieniem jakościowej analizy wyników badań pedagogicznych.

Celem prezentowanego skryptu jest wprowadzenie studentów pedagogiki (szczególnie studiujących systemem zaocznym) w najbardziej elementarne zagadnienia ze statystyki dla pedagogów.

W związku z tym, stosownie do programu przedmiotu – Metodologia badań pedagogicznych z elementami statystyki, dokonano odpowiedniego wyboru zagadnień. W trosce o komunikatywny sposób ich prezentacji pominięto niektóre wzory, dowody i symbole matematyczne, a także dokonano pewnych uproszczeń. Przedstawiono jedynie wybrane, podstawowe schematy postępowania w dokonywaniu opisu i analizy statystycznej zgromadzonego materiału empirycznego, bardziej zainteresowanych Czytelników odsyłając do zamieszczonej bibliografii.

Skrypt składa się z dwóch części. Część pierwszą rozpoczynają zagadnienia wstępne, wprowadzające w problematykę, w dalszej kolejności przedstawiono zagadnienia dotyczące gromadzenia materiału statystycznego, jego grupowania i prezentacji. Omówiono również wybrane, najczęściej stosowane charakterystyki zbiorowości statystycznej, a także przedstawiono współczynniki korelacji.

Część druga skryptu zawiera elementarne zagadnienia z zakresu wnioskowania statystycznego, a także weryfikacji hipotez statystycznych. Przedstawiono tu stosowanie wybranych testów statystycznych. Całość kończy podsumowanie zawierające rozważania na temat możliwości wykorzystania komputerów w przetwarzaniu wyników badań w pedagogice.

Skrypt zawiera pytania i zadania do samodzielnego opracowania. W celu usprawnienia posługiwania się skryptem, w końcowej jego części zamieszczono aneks zawierający niezbędne tablice statystyczne, ułatwiające korzystanie ze skryptu bez konieczności poszukiwania ich w podręcznikach.

Autorka serdecznie dziękuje P.P. Doktor habilitowanej Annie Kwiatkowskiej i Profesorowi Jerzemu Niemcowi za cenne sugestie i uwagi, które były bardzo pomocne w przygotowaniu ostatecznej postaci skryptu.

## Rozdział I

# WPROWADZENIE W PROBLEMATYKĘ

Statystyka jako nauka zajmuje się badaniem zjawisk masowych przy pomocy metod ilościowych. Zjawiska będące obiektem badań statystyki są obserwowalne i można je wyrazić przy pomocy różnych jednostek miar, są też masowe, występując bowiem w większej ilości wykazują pewne prawidłowości niemożliwe do zaobserwowania w pojedynczych przypadkach. Są pod wpływem dwojakiego rodzaju przyczyn: głównych (podstawowych), mających charakter „wewnętrzny”, związanych z istotą zjawiska, działających na wszystkie jednostki jednakowo, w określonym kierunku, i ubocznych (przypadkowych), mających charakter „zewnętrzny”, nie związanych z istotą zjawiska<sup>1</sup>.

Wyodrębnia się dwa działy statystyki: statystykę opisową i statystykę matematyczną (indukcyjną). Statystyka opisowa zajmuje się przede wszystkim metodami gromadzenia (pomiar, spis, rejestr) i prezentacji danych (tablice i wykresy) oraz ich opisem, czyli obliczaniem podstawowych miar statystycznych – odsetek, średnich i miar rozproszenia. Statystyka matematyczna (indukcyjna) natomiast jest nauką o budowie reguł wnioskowania statystycznego, w którym na podstawie danych dotyczących części zbiorowości, czyli tzw. próby reprezentacyjnej wnioskuje się z pewnym prawdopodobieństwem błędu o własnościach całej zbiorowości.

Metody statystyczne to metody stosowane podczas zbierania danych, przy ich opracowywaniu, analizie i interpretacji. Podporządkowane są one kolejnym etapom dochodzenia do praw naukowych (obserwacja zjawisk, tworzenie

---

<sup>1</sup> Z. Rogoziński, *Wstęp do statystyki społecznej*, Warszawa 1969, s. 71.

hipotezy, wnioskowanie, weryfikacja) i pozwalają badaną rzeczywistość opisywać, analizować i przewidywać.<sup>2</sup> Obejmują sposoby postępowania i techniki badawcze występujące w statystyce opisowej i wnioskowaniu statystycznym.

Statystyka ma duże znaczenie w badaniach i dostarcza szeregu korzyści: umożliwia ścisły i jednoznaczny opis poprzez zestawienie wyników w treściwej i dogodnej formie, umożliwia dostrzeżenie ogólnego obrazu w dużej liczbie wyników, sformułowanie wniosków ogólnych, przy czym proces ich formułowania powinien być dokonywany zgodnie z przyjętymi zasadami. Ponadto dostarcza możliwości przewidywania badanych zjawisk, a także wyodrębnienia pewnych przyczyn ze złożonych zjawisk, których innymi metodami nie można by ustalić.

## 1. Podstawowe pojęcia

Jak już wskazano, statystyka bada zjawiska masowe. Oznacza to, że zajmuje się badaniem zbiorowości składających się z dużej liczby jednostek.

**Zbiorowość statystyczna** to ogół osób, rzeczy, przedmiotów, zjawisk, zdarzeń poddanych badaniom statystycznym, jednoznacznie określony i wyodrębniony od pozostałych elementów. Poszczególne jednostki wchodzące w skład badanej zbiorowości statystycznej charakteryzują się pewną cechą wspólną, a jednocześnie różnią się innymi cechami, które podlegają badaniu. Badana zbiorowość statystyczna powinna być szczegółowo określona pod względem rzeczowym – co i kogo badamy, terytorialnym – na jakim terenie przeprowadzane jest badanie i czasowym – moment, termin lub okres badania. Jeżeli, np.: badamy formy spędzania czasu wolnego przez nauczycieli nauczania początkowego na terenie miasta Białegostoku w roku szkolnym 1995/96, to cechą wspólną nauczycieli tworzących zbiorowość statystyczną jest to, że są w badanym okresie nauczycielami nauczania początkowego na terenie miasta Białegostoku. Różnią się natomiast takimi cechami, jak forma spędzania czasu wolnego, rodzaj zainteresowań, staż pracy w zawodzie, kwalifikacje, stan cywilny.

---

<sup>2</sup> Cz. Nowaczyk, *Podstawy metod statystycznych dla pedagogów*, Warszawa 1985, s. 13.

**Jednostki statystyczne** to poszczególne elementy (osoby, rzeczy, przedmioty, zjawiska, zdarzenia) wchodzące w skład badanej zbiorowości statystycznej. Bezpośredniej obserwacji i opisowi w badaniach podlegają jednostki wchodzące w skład zbiorowości. W podanym wyżej przykładzie jednostkami statystycznymi będą poszczególni nauczyciele nauczania początkowego.

Ogólna liczba jednostek w zbiorowości to jej **liczebność**. Zebrane w badaniach informacje gromadzone oddzielnie dla wszystkich jednostek tworzą pełny materiał statystyczny. 4

Badaniom statystycznym podlegają określone zbiorowości. Cała zbiorowość to zbiorowość generalna lub krócej **populacja**. Najczęściej jednak, gdy badanie całej populacji jest niemożliwe lub niecelowe, badaniom poddaje się próbę. **Próba** jest to część populacji dobrana w określony sposób w celu przeprowadzenia badań. Próba powinna być reprezentacyjna dla populacji, z której ją pobrano, czyli pod względem występowania charakteryzujących ją istotnych dla badania cech powinna odzwierciedlać całą populację, np.: pod względem płci, pochodzenia społecznego badanych; próba powinna mieć strukturę zbliżoną do populacji.

**Cechy statystyczne** to właściwości charakteryzujące jednostki badanej zbiorowości, służące do ich różnicowania i identyfikacji. Stanowią kryterium podziału badanej zbiorowości lub jego klasyfikacji, w wyniku czego można wykryć pewne prawidłowości. Cechy stałe to takie, które są wspólne dla całej badanej zbiorowości, służą do jej określenia i nie podlegają badaniu.

Badaniom podlegają tylko cechy różniące poszczególne jednostki statystyczne, czyli takie, które są zmienne. **Zmienna** jest właściwością – cechą, która przybiera różne wartości, np.: płeć, poziom wykształcenia, staż pracy. Każda zmienna (cecha) może występować w rozmaitej formie lub różnych wielkościach.

W zależności od wielkości zbioru, z którego dana zmienna przyjmuje wartości, a także od liczby możliwych do wystąpienia jej wariantów, wyróżnia się zmienne (cechy) dwudzielne, dychotomiczne, np.: płeć – kobieta, mężczyzna i wielodzielne, np.: poziom aspiracji edukacyjnych – wysoki, średni, niski.

Ze względu na rodzaj cech (zmiennych) wyróżnia się: jakościowe, ilościowe, a czasem również porządkujące<sup>3</sup>.

---

<sup>3</sup> Cechy porządkujące to takie cechy, które wyraża się opisowo, np.: „duży”, „mały”, a nie liczbowo, lecz opis ten określa w pewien sposób natężenie tej cechy. [za:] J. Brzeziński. *Metodologia badań psychologicznych*, Warszawa 1997, s. 188.

**Cechy jakościowe** (niemierzalne, opisowe) to właściwości ludzi, przedmiotów lub zjawisk wchodzących w skład badanej zbiorowości, których określanie polega na zakwalifikowaniu badanego obiektu do dokładnie jednej z dopuszczalnych kategorii, np.: płeć, zawód, rodzaj

**Cechy ilościowe** (mieralne) to właściwości ludzi, przedmiotów lub zjawisk wchodzących w skład badanej zbiorowości, których wielkość lub nasilenie mierzy się w pewnych jednostkach, przy pomocy liczb. Wartości liczbowe cechy ilościowej mogą występować w postaci zmienności ciągłej lub w postaci zmienności skokowej.

**Zmienność ciągła** występuje wtedy gdy wartość cechy może przybierać dowolne wartości w danych granicach. Oznacza to, że między dwiema wartościami liczbowymi cechy, mogą wystąpić inne wartości pośrednie, np.: wiek – do grupy dzieci w wieku 7 i 8 lat należą dzieci mające 7 lat i 2 miesiące, dzieci będące w wieku 7 lat, 2 miesiące i 2 tygodni itp. Innym przykładem zmienności ciągłej jest wzrost: może być wyrażony w metrach, centymetrach, milimetrach w zależności od dokładności.

**Zmienność skokowa** (dyskretna) występuje wtedy gdy badaną cechę można wyrażać tylko w liczbach zmieniających się skokami, bez wartości pośrednich między nimi, np.: liczba dzieci w rodzinie, liczba członków rodziny. Zmienność skokowa o bardzo dużej liczebności wariantów traktowana jest jako prawie ciągła (quasi ciągła).

W wielu praktycznych sytuacjach cechy ilościowe wygodnie jest traktować jako cechy jakościowe. Na przykład w dowodzie osobistym wpisuje się w rubryce wzrost: wysoki, średni, niski, chociaż wzrost jest cechą ilościową. Zamianę taką określa się mianem nominalizacji<sup>4</sup>. Jej najprostszą postacią jest wyodrębnienie z cechy ilościowej cechy jakościowej dychotomicznej (dwudzielnej) lub cechy wielodzielnej.

Chcąc dokonać zamiany cechy ilościowej na cechę jakościową dwudzielną, należy wybrać z zakresu wartości cechy ilościowej pewną liczbę, oznaczając ją  $X$ . Pierwszą kategorię utworzą wszystkie pomiary, które są mniejsze lub równe  $X$ , a do drugiej kategorii można zaliczyć te pomiary, które są większe od  $X$ , np.: jeśli wyniki przeprowadzonego testu osiągnięć szkolnych przyjęły wartości od 0 do 60 punktów, to można przyjąć, że badani, którzy uzyskali powyżej 30 punktów, będą stanowili kategorię zaliczających test, a ci, którzy uzyskali 30 punktów lub mniej – nie zaliczających test. W ten sposób dokonano zamiany cechy ilościowej na cechę jakościową dwudzielną.

---

<sup>4</sup> A. Góralski (red.), *Metody badań pedagogicznych w zarysie*. WSPS, Warszawa 1994, s. 101.

Podobnie, np.: przedstawiając sytuację materialną badanych, określoną poprzez średni dochód na osobę w rodzinie, dopuszczalne jest przedstawienie jej poprzez ustalenie odpowiednio kategorii: bardzo dobra, dobra, wystarczająca, zła, przy czym wiadomo, że poszczególne kategorie, będące wariantami cechy jakościowej, odpowiadają zróżnicowanemu średniemu dochodowi na osobę w rodzinie wyrażonemu w postaci liczbowej.

Ze względu na rodzaj relacji zachodzących między zmiennymi (cechami) wyróżnia się: zmienne zależne, zmienne niezależne, pośredniczące, a także inne.

**Zmienna zależna** to taka, która zmienia się pod wpływem oddziaływania innej zmiennej, której związki z innymi zmiennymi będą w badaniach wyjaśniane (określane).

**Zmienna niezależna** to taka, która powoduje zmiany w innej zmiennej, zależnej.

**Zmienna pośrednicząca** to zmienna wpływająca pośrednio na zmienną zależną, mogąca modyfikować związek między zmienną niezależną i zależną.

W badaniach pedagogicznych istnieją pewne typowe układy zmiennych. Jako zmienne niezależne często przyjmuje się: środowisko rodzinne, kulturę pedagogiczną rodziców, sytuację ekonomiczną rodziny, metody nauczania, metody wychowania itp. Typowe zmienne zależne to osiągnięcia szkolne, przyrost wiadomości, niepowodzenia dydaktyczne, zmiany w postawach, zachowaniach, aspiracje edukacyjne, plany życiowe itp. Jako zmienne pośredniczące w badaniach pedagogicznych często przyjmuje się: płeć, wiek, miejsce zamieszkania itp.

Podkreślić jednak należy, że przyjęcie określonych zmiennych jako zmiennej niezależnej, zależnej i pośredniczącej uzależnione jest od celu badań. Ta sama zmienna czasem może być przyjęta jako zmienna niezależna, czasem jako zależna, a czasem jako pośrednicząca.

W zależności od przyjmowanych kryteriów podziału można wyodrębnić również inne rodzaje zmiennych. Ze względu na stopień ogólności zmiennych wyróżnia się **zmienne globalne** i **szczegółowe**. W badaniach eksperymentalnych wskazuje się zmienne kontrolowane i niekontrolowane<sup>5</sup>.

**Zmienna kontrolowana** to taka, na którą badacz może mieć wpływ, powodując jej wystąpienie lub regulując natężenie, np.: ruchy, czynności, procesy psychiczne wywołane przez stworzenie odpowiedniej sytuacji eksperymentalnej.

---

<sup>5</sup> A. Skorny, *Prace magisterskie z psychologii i pedagogiki*, WSiP, Warszawa 1984, s. 48.

**Zmienna niekontrolowana** to taka, na którą badacz nie ma wpływu, nie może jej wywołać, ani też zmieniać natężeń, np.: niektóre fakty występujące spontanicznie w zachowaniu.

## 2. Pomiar i skale pomiarowe

Pomiar to przyporządkowywanie liczb przedmiotom, osobom, zjawiskom, zdarzeniom w celu przedstawienia ich cechy w taki sposób, aby przyporządkowane liczby odzwierciedlały relacje zachodzące między nimi.

Wyróżnione w badaniach cechy charakteryzujące badaną zbiorowość – cechy jakościowe i ilościowe – wymagają odpowiednich narzędzi – skal do ich pomiaru. Ze względu na stopień dokładności pomiaru wyróżnia się cztery podstawowe rodzaje skal pomiarowych:

1. skalą nominalną,
2. skalą porządkową (rangową),
3. skalą przedziałową (interwałową),
4. skalą ilorazową (stosunkową).

**Skala nominalna** jest najprostszym rodzajem pomiaru. Polega na przyporządkowaniu liczb przedmiotom, zjawiskom, zdarzeniom, osobom w taki sposób, aby oznaczały one ich przynależność do rozłącznych kategorii. Liczby pełnią tu rolę umownych symboli i nie można dokonywać na nich żadnych operacji matematycznych. Możliwe jest tylko liczenie jednostek w poszczególnych kategoriach i na tej podstawie obliczanie np.: proporcji czy odsetek. W praktyce dla oznaczenia kategorii zamiast liczb używa się oznaczeń słownych i na ich podstawie dokonuje się klasyfikacji przedmiotów, zjawisk, zdarzeń, osób, np.: płeć: kobieta, mężczyzna – skala nominalna dwudzielna; pochodzenie społeczne – robotnicze, chłopskie, inteligenckie, rzemieślnicze – skala nominalna wielodzielna. Podstawą zaliczenia do danej kategorii jest wyłącznie fakt posiadania określonego wariantu cechy. O poszczególnych wariantach cech można tylko powiedzieć, że są jednakowe lub różne.

**Skala porządkowa** umożliwia nie tylko rozróżnianie przedmiotów, zdarzeń, osób, ale także ich porównywanie i porządkowanie. W przypadku tej skali porównuje się ze sobą przedmioty i ocenia je ze względu na to, czy

przejaw danej cechy w tych przedmiotach jest większy, mniejszy czy równy. W skali tej przyporządkowuje się liczby przedmiotom w taki sposób, że odzwierciedlają one ich uporządkowanie ze względu na nasilenie mierzonej cechy. Uczniów można uporządkować według poziomu aktywności podczas lekcji, wysokości samooceny, czyli można dokonać rangowania przypisując rangę najwyższą uczniowi o największym stopniu nasilenia danej cechy. Można także przeprowadzić rangowanie w kierunku odwrotnym. Wadą skali rangowej jest brak możliwości mierzenia różnicy między rangami.

**Skala przedziałowa** posiada wszystkie własności skali porządkowej, a ponadto umożliwia określenie odległości, dystansu między jej punktami. Podstawą tej skali są przedziały o równej wielkości, a punkt zerowy przyjęty jest umownie. Przykładem może być pomiar temperatury na skali Celsjusza, zawierającej umowne zero. Można stwierdzić, że np.: woda w jednym naczyniu ma  $30^{\circ}\text{C}$ , w drugim  $20^{\circ}\text{C}$ , a w trzecim  $10^{\circ}\text{C}$ , co oznacza, że różnica między temperaturą wody w poszczególnych naczyniach wynosi  $10^{\circ}\text{C}$ . Błędem natomiast byłoby twierdzić, że woda w pierwszym naczyniu jest trzykrotnie cieplejsza od tej w trzecim, ponieważ nie wynika to wcale z tego rodzaju skali. Na skali przedziałowej dopuszczalne jest dokonywanie większości operacji matematycznych. W praktyce pedagogicznej, przy pewnych zastrzeżeniach, skalę przedziałową budować można na podstawie wyników badań testowych.

**Skala ilorazowa** zachowuje własności trzech poprzednich skal. Zbudowana jest podobnie jak skala interwałowa z tym, że posiada absolutne zero, co umożliwia dokonywanie na niej wszystkich operacji matematycznych. Zero jest przyporządkowane przedmiotowi mającemu cechę mierzoną w stopniu zerowym. Jeżeli porównujemy ze sobą dwie osoby pod względem wieku, to można powiedzieć nie tylko, że jedna osoba jest starsza od drugiej, że jest starsza o konkretną liczbę lat, ale ponadto, że jest starsza, np.: trzy razy. Jest to słuszne, ponieważ zmienną „wiek” można przedstawić na skali ilorazowej.

Wymienione skale charakteryzują się narastającym stopniem dokładności pomiaru. Każda kolejna jest dokładniejsza, mocniejsza, zawiera wszystkie własności poprzedniej, a także nowe.

W psychologii występuje określanie zmiennych (cech) w zależności od postaci skali pomiarowej, na której występuje zbiór wartości zmiennej.

W związku z tym wyróżnia się zmienną nominalną, zmienną porządkową, zmienną interwałową i zmienną ilorazową<sup>6</sup>.

Podejmując określone badania, badacz powinien uświadamiać sobie, jakie zmienne są dla niego istotne z punktu widzenia celu badań, a także przy pomocy jakich skal można wyrażać te zmienne. Jak już wskazano, zmienną „wiek” można wyrazić na skali ilorazowej, ale także można ją przedstawić na skali nominalnej dwudzielnej, wyróżniając dwie kategorie osób: np.: poniżej 18 lat i powyżej 18 lat, ale także może być ona (przy pewnych zastrzeżeniach) wyrażona na skali przedziałowej.

### 3. Etapy i rodzaje badania statystycznego

Każde badanie statystyczne składa się z kilku etapów. Są nimi:

- a/ przygotowanie (zaplanowanie) badań statystycznych,
- b/ przeprowadzenie badań (obserwacja statystyczna),
- c/ opis zbiorowości statystycznej (opis statystyczny),
- d/ analiza wyników obserwacji (analiza statystyczna).

**Przygotowanie i zaplanowanie badań statystycznych** obejmuje m.in. następujące działania:

- określenie celu badań, problemów i hipotez badawczych,
- określenie badanej zbiorowości i jednostek, ustalenie zmiennych (cech) i związków między nimi,
- ustalenie rodzaju badania (wyczerpujące, reprezentacyjne),
- wskazanie lub opracowanie metod, technik i narzędzi badawczych,
- przygotowanie badań od strony organizacyjnej.

Precyzyjne określenie celu badań jest zazwyczaj kilkuczęściowe i powinno uświadamiać badaczowi, do czego zmierza. W celu poznawczym najczęściej jest zawarta odpowiedź na pytanie: co badacz chce osiągnąć? Cel praktyczny badań wskazuje na sposób wykorzystania wyników i powinien zawierać odpowiedź na pytanie: komu i czemu ma to służyć. Mogą być też formułowane i inne cele badań, np.: diagnostyczny, prognostyczny, osobisty.

---

<sup>6</sup> J. Brzeziński. *Metodologia* ..., op. cit., s. 188-189.

Precyzyjne ustalenie przedmiotu badań, określenie badanej zbiorowości statystycznej pod względem rzeczowym, przestrzennym i czasowym, to kolejne czynności. Ustalając cele badań i określając zbiorowość należy wyrazić określić jednostkę statystyczną. Jednostką statystyczną może być każdy badany uczeń, ale także klasa szkolna, szkoła itp. Zależy to od celu badań, jeśli badacz chce np.: ustalić, jakie są aspiracje edukacyjne absolwentów klas VIII na terenie miejscowości X, w roku szkolnym 1995/96, to zbiorowością statystyczną będą absolwenci kl. VIII na terenie tej miejscowości, a jednostką statystyczną będzie każdy uczeń.

Sprecyzowanie celu badań wyznacza problematykę badawczą. Sformułowanie problemu badawczego, sprowadzenie go do szeregu pytań rozstrzygnięcia, sformułowanie hipotez, to kolejne czynności. Następnie można dokonać wskazania zmiennych (cech), które poddane zostaną badaniom. Każda badana zbiorowość posiada wiele cech, ale wybiera się tylko te, które są istotne z punktu widzenia celu podjętych badań. Dla przyjętych do badań cech ustala się zbiór ich wartości, ich warianty, które występują u poszczególnych jednostek. W podanym wyżej przykładzie cechami istotnymi z punktu widzenia celu badań będą np.: płeć ucznia, a jej warianty to: dziewczynka, chłopiec, oraz rodzaj wybranej szkoły ponadpodstawowej: zasadnicza szkoła zawodowa, technikum, liceum zawodowe, liceum ogólnokształcące.

Wybrane do badań cechy (zmienne) powinny wykazywać związek między sobą, a rodzaj tego związku badacz powinien określić. Pozwala to na pełniejszą analizę badanych zjawisk. Ważne jest więc wskazanie, które ze zmiennych przyjmuje się w badaniach jako zmienne niezależne, zależne, pośredniczące lub inne, a także określenie skal pomiarowych, na których zmienne przyjmują wartości.

W podanym wyżej przykładzie zmienna płeć może być traktowana jako zmienna niezależna, przyjmująca warianty (wartości) na skali nominalnej, dwudzielnej: dziewczynka, chłopiec, a zmienna – rodzaj wybranej szkoły ponadpodstawowej może być przyjęta jako zmienna zależna, a jej warianty wyrażone są na skali nominalnej wielodzielczej: zasadnicza szkoła zawodowa, technikum, liceum zawodowe, liceum ogólnokształcące.

Przygotowując badanie badacz powinien rozstrzygnąć, jaki rodzaj badania chce przeprowadzić: wyczerpujące czy częściowe (reprezentacyjne).

**Badania wyczerpujące** to takie, w których obejmuje się badaniami wszystkie jednostki statystyczne wchodzące w skład zbiorowości generalnej

(populacji). Badanie aspiracji edukacyjnych absolwentów kl. VIII z Białegostoku będzie badaniem wyczerpującym, jeśli badaniem objęci zostaną wszyscy absolwenci kl. VIII z tego miasta w danym roku. W praktyce pedagogicznej takie badania przeprowadza się **bardzo rzadko**, ponieważ niewiele jest populacji występujących w małej liczebności. Często przeprowadzanie takich badań jest niemożliwe ze względów **technicznych**, lub też niecelowe jest badanie całej np.: populacji uczniów w szkole, mieście czy województwie. W takich przypadkach przeprowadza się badania częściowe.

**Badania częściowe** (reprezentacyjne) to takie, w których badaniami obejmuje się tylko część zbiorowości statystycznej, niektóre jednostki czy próbę wybraną w taki sposób, aby jej właściwości w sposób wiarygodny reprezentowały właściwości całej zbiorowości (populacji). Winna to być zatem próba reprezentacyjna.

Wybór rodzaju badań ma odpowiednie konsekwencje dla sposobu opracowywania, jak i sposobu interpretacji wyników. W badaniach wyczerpujących opracowanie wyników polega na opisie statystycznym, a w badaniach reprezentacyjnych opracowanie wyników obejmuje opis statystyczny, ale także wnioskowanie statystyczne, w którym na podstawie poznanych właściwości próby reprezentacyjnej formułuje się wnioski o właściwościach populacji.

Wyniki badań wyczerpujących są **stwierdzeniami kategorycznymi**, dotyczącymi wprost badanej populacji, natomiast wyniki badań reprezentacyjnych są zawsze **stwierdzeniami hipotetycznymi**, o określonych granicach niepewności wniosku.

W pedagogice większość prowadzonych badań to badania reprezentacyjne, wymagające odpowiedniego doboru próby<sup>7</sup>.

Przygotowanie badań wymaga również wskazania i uzasadnienia, a czasem również opracowania metod, technik i narzędzi badawczych, które będą wykorzystane w celu gromadzenia materiału.

Uzyskanie zgody odpowiednich władz na przeprowadzenie badań (jeśli taka jest wymagana) często kończy etap przygotowania i zaplanowania badań.

**Przeprowadzenie badań** (obserwacja statystyczna). Przeprowadzenie badań właściwych (zasadniczych) czasem poprzedzają **badania wstępne** (pilotażowe). Badania wstępne przeprowadza się na mniejszej próbie niż badania właściwe. Umożliwiają uzyskanie orientacji w rodzajach badanych zjawisk, ich

---

<sup>7</sup> Metody doboru próby reprezentacyjnej przedstawiono w kolejnym podrozdziale.

przebiegu, rodzaju występujących związków, kompletności zmiennych. Mogą prowadzić do zmodyfikowania problemów i hipotez. W rezultacie badań wstępnych następuje również kontrola zastosowanych narzędzi badawczych, istotna szczególnie wtedy, gdy są samodzielnie opracowywane przez badacza. Sprawdzeniu podlega również teren, a także organizacja badań.

**Badania właściwe** (zasadnicze) dostarczają materiałów empirycznych umożliwiających znalezienie odpowiedzi na sformułowane problemy badawcze poprzez weryfikację hipotez.

Rezultatem przeprowadzonych badań są nieuporządkowane materiały empiryczne, określane jako materiały surowe. Zgromadzony materiał powinien zostać poddany **kontroli merytorycznej** ze względu na jego wiarygodność oraz kontroli formalnej ze względu na jego kompletność. Należy odrzucić np.: nieczytelnie wypełnione formularze, ogólnikowo i powierzchownie przedstawione obserwacje, a także wypowiedzi uzyskane podczas wywiadu budzące wątpliwości dotyczące prawdomówności rozmówców. W ramach **kontroli formalnej** należy sprawdzić, czy są wszystkie materiały oraz wyeliminować te, które nie zawierają kompletnych informacji.

Podczas kontroli szczególną uwagę należy zwracać na najczęściej występujące błędy tendencyjne i przypadkowe. Błędy tendencyjne często wynikają z celowej dezinformacji badającego, mającej miejsce np.: wówczas, gdy osoba badana nie chce stawiać się w niekorzystnym świetle (np.: odpowiedzi na pytania dotyczące stosowania kar w wychowaniu). Błędy mogą być również rezultatem źle skonstruowanych narzędzi badawczych, a także mogą powstać w wyniku wadliwego pogrupowania wariantów badanych zmiennych i mogą znacznie zniekształcać rezultaty badań.

Błędy przypadkowe są najczęściej drobnymi pomyłkami wynikającymi z niewiedzy i nie są groźne.

Kolejne etapy badania statystycznego – **opis zbiorowości statystycznej** (opis statystyczny) oraz etap **analizy wyników obserwacji** (analiza statystyczna) zostały przedstawione w kolejnych rozdziałach skryptu.

## 4. Metody doboru próby reprezentacyjnej

Jak już wspomniano, w praktyce pedagogicznych badań empirycznych bardzo rzadko badaniom podlegają populacje. Najczęściej podejmuje się badania części populacji, dobranej w specjalny sposób, nazywanej próbą reprezentacyjną (reprezentatywną). Uzyskane wyniki zostają uogólnione poprzez wnioskowanie na całą populację. Warunkiem trafności tych uogólnień jest reprezentatywność próby dla populacji. Oznacza to, że struktura i zróżnicowanie właściwości w badanej próbie reprezentacyjnej powinny odzwierciedlać proporcjonalnie strukturę i właściwości całej populacji, z której tę próbę pobrano.

Nie zawsze jednak tak jest, gdyż próba może być reprezentatywna ze względu na kilka cech, ale jednocześnie może taką nie być z punktu widzenia innych cech. Ponadto reprezentatywność próby nie zawsze zależy od jej wielkości, ponieważ istnieje możliwość pobrania względnie dużej próby (duża liczebność), a próba może być mało reprezentatywna, co może być rezultatem tendencyjności w jej doborze. Próba będzie tym bardziej reprezentatywna, im mniej błędów zostanie popełnionych podczas jej doboru.

Wielkość próby reprezentacyjnej w dużym stopniu zależy od<sup>8</sup>:

- stopnia różnorodności cech (zmiennych) badanej zbiorowości; im bardziej różnorodna zbiorowość, tym większa powinna być próba,
- ilości kategorii statystycznych, na jakie ma być podzielona badana zbiorowość; im więcej kategorii badacz zamierza wyróżnić i analizować, tym większa powinna być próba,
- stopnia dokładności i pewności, jakie badacz zamierza osiągnąć w badaniach; im większa próba reprezentacyjna, tym większa dokładność wniosków o populacji, którą ta próba reprezentuje.

Nie można wskazać żadnych reguł ogólnie obowiązujących, dotyczących wielkości próby. Używanie określenia próba „mała” lub „duża” zależy od potrzeb badacza i jego możliwości, np.: dla socjologa badającego poglądy próba licząca 100 jednostek jest małą, podczas gdy dla lekarza badającego rzadko występującą chorobę zgromadzenie np.: 20 przypadków będzie próbą dużą.

W badaniach pedagogicznych przyjmuje się najczęściej, że próba licząca poniżej 30 jednostek jest próbą małą, a próba o liczbie jednostek powyżej 30 jest próbą dużą<sup>9</sup>.

---

<sup>8</sup> J. Sztumski, *Wstęp do metod i technik badań społecznych*, Katowice 1995, s. 105-106.

<sup>9</sup> G. Clauss, H. Ebner, *Podstawy statystyki dla psychologów, pedagogów i socjologów*. PZWS, Warszawa 1972, s. 188.

Na ogół stosuje się dwie metody doboru próby:

- **dobór losowy** (probabilistyczny),
- **celowy**.

Przy doborze losowym badacz nie może mieć wpływu na dobór jednostek z populacji generalnej, a każda jednostka musi mieć jednakowe szanse znalezienia się w próbie losowej. Wyróżnia się dwa rodzaje losowania: losowanie niezależne i zależne.

**Losowanie niezależne** ma miejsce wówczas gdy element (jednostka) raz wylosowany ponownie wraca i może być powtórnie wylosowany. W tym rodzaju losowania każdy element ma jednakowe szanse dostania się do próby.

**Losowanie zależne** (bez powtórzeń) to takie, gdy elementu raz wylosowanego nie włącza się do ponownego losowania. W takim przypadku następne losowania zwiększają szanse pozostałych elementów. W badaniach głównie stosuje się losowanie zależne i takie będzie przedmiotem rozważań.

**Dobór losowy prosty** polega na tym, że każdej jednostce w populacji generalnej przyporządkowuje się określony numer kolejny od 1 w kierunku rosnącym, a następnie z tak przygotowanej listy dokonuje się losowania potrzebnej ilości elementów. Przy małej populacji generalnej można korzystać z urny, ale najczęściej losowania takiego dokonuje się przy wykorzystaniu tzw. tablic liczb losowych. Na tablicy ustala się dowolny punkt, od którego w dowolnym kierunku wybiera się kolejno taką ilość liczb, która odpowiada liczebności próby.

**Dobór systematyczny** jest podobny do wcześniej omówionego i wystąpi wówczas gdy do próby badawczej włączy się, np.: co piątą jednostkę (1, 5, 10, 15, ... itd.) lub cyfry podzielne przez np.: osiem (8, 16, 24, 32, ... itd.).

**Dobór warstwowy** stosowany jest w przypadku, gdy populacja nie jest jednorodna. Wówczas dzieli się populację na jednorodne warstwy i dokonuje się losowania oddzielnie z każdej warstwy. Dobór jednostek w ramach poszczególnych warstw może być różny. Najczęściej wyróżnia się dwa rodzaje doboru warstwowego: proporcjonalny (z poszczególnych warstw do próby dobiera się jednostki w ilościach proporcjonalnych do liczebności warstw) i nieproporcjonalny (dobór elementów do próby nie jest uzależniony od wielkości warstw, a ilość jednostek wchodzących do próby z każdej z warstw uzasadnia się kryteriami praktycznymi lub merytorycznymi, np.: szczególne zainteresowanie się nią badacza).

**Dobór zespołowy** jest wykorzystywany wówczas gdy badaczowi zależy na losowym doborze nie pojedynczych jednostek, ale całych zespołów, np.: klas, szkół, gmin. Próbę badawczą stanowią wszystkie jednostki wylosowanych zespołów. W zależności od różnicowania liczebności zespołów dokonuje się dla zespołów jednakowo licznych – doboru zespołowego z jednakowym prawdopodobieństwem wyboru (szanse wylosowania każdego zespołu są jednakowe), a dla zespołów różniących się pod względem liczebności – doboru zespołowego z różnym prawdopodobieństwem wyboru (szanse wylosowania poszczególnych zespołów są nierówne, w zależności od ich liczebności).

Odmianą doboru zespołowego jest **dobór dwustopniowy**. Obejmuje on dwa etapy: w pierwszym następuje losowanie pewnej liczby zespołów, a w drugim wylosowanie z każdego dobranego do próby zespołu pewnej tylko liczby elementów do zbadania. Stosowanie takiego doboru w nieco rozszerzonej postaci jest doborem trzy- i wielostopniowym.

Wykorzystanie przez badacza konkretnego doboru próby uzależnione jest od celu i warunków badań, posiadanych informacji o populacji, a także możliwości technicznych i organizacyjnych badacza.

**Celowy dobór próby** to metoda, w której decyzja dotycząca tego, które jednostki będą stanowiły próbę, uzależniona jest od informacji dotyczących zbiorowości posiadanych przez badacza. Dobór celowy świadomie może powodować włączenie do próby jednostek nietypowych lub o szczególnie charakterystycznych cechach.

Dobór celowy proporcjonalny wymaga dokładnego określenia populacji pod względem cech interesujących badacza, np.: płeć, zawody. Jednostki do próby dobierane są w taki sposób i w takich proporcjach, aby próba posiadała, pod względem badanych cech, strukturę bardzo zbliżoną do populacji.

W praktyce czasem dzieje się tak, że student–badacz z różnych powodów nie ma możliwości poprawnego doboru próby i prowadzi badania na próbie, której reprezentatywność trudno ustalić. J.P. Guilford takie próby nazywa **incydentalnymi** dlatego, że są najłatwiejsze do uzyskania<sup>10</sup>. W takiej sytuacji badacz powinien pamiętać o ograniczonym zasięgu wniosków końcowych z badań.

---

<sup>10</sup> J.P. Guilford, *Podstawowe metody statystyczne w psychologii i pedagogice*, Warszawa 1964, s. 171.

## Rozdział II

# GRUPOWANIE I PREZENTACJA ZEBRANEGO MATERIAŁU

### 1. Grupowanie statystyczne

Podczas przeprowadzenia badań materiał empiryczny gromadzony jest przy pomocy różnych metod, technik i narzędzi badawczych dobranych w zależności od celu badań. Do najczęściej stosowanych należą: kwestionariusz, ankieta, obserwacja, analiza dokumentów, testy, a także różnorodne narzędzia badawcze konstruowane specjalnie przez badaczy w celu dokładniejszego poznania badanych przez nich zjawisk. Dane empiryczne zgromadzone podczas badań stanowią materiał surowy, wymagający odpowiedniego grupowania.

**Grupowanie statystyczne** polega na podziale zebranego materiału, stanowiącego zbiór dużej ilości szczegółowych informacji na jednorodne grupy (kategorie). Podstawą ich wyodrębnienia są cechy, czyli zmienne.

Poprawne dokonanie grupowania zebranego materiału wymaga przyjęcia jako kryteria grupowania takich cech, które są istotne dla badanej zbiorowości z punktu widzenia celu badań, a także uwzględnienia badanych zjawisk we wzajemnym powiązaniu. Istotne jest również ustalenie dla każdej cechy (zmiennej), według której dokonuje się grupowania liczby kategorii, a także ich wielkości w taki sposób, aby w rezultacie grupowania możliwe było przedstawienie specyficznych stron badanej zbiorowości i różnic występujących między jednostkami tej zbiorowości.

Grupowanie statystyczne ma podstawowe znaczenie dla wyników badań, od ustalenia prawidłowych zasad podziału całej zbiorowości na poszczególne

grupy (kategorie) zależy bowiem poprawność wniosków dotyczących tej zbiorowości. Prawidłowo przeprowadzone grupowanie umożliwia przedstawienie struktury badanej zbiorowości, ustalenie jej jakościowych i ilościowych charakterystyk, a także ustalenie związków między zmiennymi.

Najczęściej wyróżnia się następujące rodzaje grupowania: typologiczne, wariancyjne, czasem również kombinowane.

**Grupowanie typologiczne** polega na wydzieleniu grup względnie jednorodnych z niejednorodnej zbiorowości i odnosi się głównie do cech jakościowych. Przejawem takiego grupowania jest podział badanej zbiorowości na grupy według takich cech, jak np.: płeć, pochodzenie społeczne, wykształcenie. Dokonując grupowania typologicznego można uwzględnić jedną cechę (Tablica 1) lub dwie (Tablica 2).

Tablica 1

*Kandydaci na kierunek studiów X według typu ukończonej szkoły średniej*

| Typ ukończonej szkoły średniej | Liczba kandydatów |
|--------------------------------|-------------------|
| Liceum ogólnokształcące        | 96                |
| Liceum zawodowe                | 27                |
| Technikum                      | 12                |
| Ogółem                         | 135               |

Źródło: dane umowne

Tablica 2

*Wybór szkoły ponadpodstawowej przez absolwentów kl. VIII według wykształcenia rodziców*

| Wykształcenie rodziców | Wybór szkoły ponadpodstawowej |           |    |    |        |
|------------------------|-------------------------------|-----------|----|----|--------|
|                        | ZSZ                           | Technikum | LZ | LO | Ogółem |
| Wyższe                 | –                             | –         | 2  | 12 | 14     |
| Średnie                | 4                             | 10        | 4  | 10 | 28     |
| Zawodowe               | 7                             | 18        | 8  | 16 | 49     |
| Podstawowe             | –                             | 2         | 3  | 5  | 10     |
| Ogółem                 | 11                            | 30        | 17 | 43 | 101    |

Źródło: dane umowne

W zależności od celu badań czasem jest potrzebny podział na grupy mniej lub bardziej szczegółowe, np.: w przypadku takiej cechy, jaką jest wykształcenie, może być: podstawowe, zawodowe, średnie, wyższe. Nazwy grup, na które dzieli

się zbiorowość, wymienione w odpowiedniej kolejności tworzą tzw. wykaz klasyfikacyjny. Powinien być tak skonstruowany, aby możliwe było zaszeregowanie każdej jednostki do jednej z grup klasyfikacyjnych wykazu. Kolejność grup w wykazie powinna być logiczna i uzasadniona. Można utworzyć grupę o nazwie „inne” lub „pozostałe” dla tych odmian cechy, które nie mają istotnego znaczenia w badaniach lub występują bardzo rzadko (Tablica 3). Jednak utworzenie takiej grupy jest zasadne tylko wówczas, jeżeli jej liczebność nie jest zbyt duża w stosunku do wszystkich jednostek objętych badaniami (nie powinna ujmować więcej niż 15% jednostek)<sup>11</sup>. Ponadto przy ustalaniu wykazu należy zwracać uwagę na to, aby grupy wyszczególnione na wykazie były rozłączne i nie pokrywały się nawzajem, np.: robotnik, pracownik umysłowy, urzędnik. Urzędnika można zakwalifikować do pracowników umysłowych.

Tablica 3

*Miejsce zamieszkania studentów kierunku X*

| Miejsce zamieszkania | Liczba studentów |
|----------------------|------------------|
| 1. Akademik          | 23               |
| 2. Stancja           | 12               |
| 3. U rodziców        | 18               |
| 4. Inne              | 6                |
| Ogółem               | 59               |

Źródło: dane umowne

**Grupowanie wariacyjne** polega na łączeniu poszczególnych jednostek statystycznych według wielkości interesującej badacza cechy i dotyczy cech ilościowych. Można dokonać grupowania wariacyjnego uwzględniając jedną cechę (Tablica 4) lub dwie i więcej (Tablica 5).

Tablica 4

*Wyniki testu z matematyki uczniów kl. V*

| Wyniki testu | Liczba uczniów |
|--------------|----------------|
| 9 – 13       | 11             |
| 14 – 18      | 26             |
| 19 – 23      | 23             |
| 24 – 28      | 15             |
| Ogółem       | 75             |

Źródło: dane umowne

<sup>11</sup> Z. Peuker, *Statystyka*, WSiP, Warszawa 1987, s. 51.

Wyniki testu z matematyki uczniów kl. V według odległości do szkoły

| Wyniki testu | Odległość do szkoły (w km) |            |            |            |        |
|--------------|----------------------------|------------|------------|------------|--------|
|              | $0 \div 1$                 | $2 \div 3$ | $4 \div 5$ | $6 \div 7$ | Ogółem |
| 9 – 13       | 1                          | 1          | 2          | 7          | 11     |
| 14 – 18      | 10                         | 8          | 5          | 3          | 26     |
| 19 – 23      | 12                         | 7          | 3          | 1          | 23     |
| 24 – 28      | 11                         | 2          | 2          | —          | 15     |
| Ogółem       | 34                         | 18         | 12         | 11         | 75     |

Źródło: dane umowne

**Grupowanie kombinowane** (analityczne) to takie, w którym uwzględnia się dwie lub więcej cech powiązanych ze sobą z punktu widzenia celu badań, co prowadzi do uchwycenia związków między cechami. Grupowanie to odnosi się do cech jakościowych i ilościowych w różnych układach: jakościowa-jakościowa, ilościowa-ilościowa, jakościowa-ilościowa.

Podział badanej zbiorowości na grupy według wyników testu z matematyki i odległości do szkoły (Tablica 5) pozwala na dokładniejsze jej poznanie, a także uwidacznia różnice w rozkładzie wyników.

## 2. Budowa szeregu rozdzielczego

Rezultatem grupowania i zliczania danych są **szeregi statystyczne**, będące zbiorem informacji uporządkowanych według pewnej cechy. Najprostszy szereg składa się z dwóch wierszy lub kolumn, z których jedna podaje warianty cechy, a druga liczbę jednostek w danym wariantcie.

Przyjmując jako kryterium sposób budowy szeregu wyróżnia się: szeregi szczegółowe, szeregi rozdzielcze, szeregi kumulacyjne, natomiast z punktu widzenia celu badań – szeregi dynamiczne (chronologiczne), terytorialne i inne.

**Szereg szczegółowy** (indywidualny) uwzględnia wszystkie wartości cechy występujące w zbiorowości. Nie są one pogrupowane, tylko uporządkowane według pewnego kryterium, np.: odległości z domu do szkoły (w km) grupy uczniów, uporządkowanej od najmniejszej do największej: 0,5; 0,8; 1,1; 1,2... itd.

Przydatność takiego szeregu jest niewielka i często, nawet gdy badana zbiorowość jest nieliczna, przekształcany jest w tzw. szereg rozdzielczy.

**Szereg rozdzielczy** powstaje w rezultacie podziału zbiorowości z punktu widzenia pewnej cechy na grupy (klasy) i przyporządkowania im liczebności (częstotliwości) występowania poszczególnych wariantów cechy.

Zabieg ten jest prosty w przypadku cechy ilościowej nieciągłej, skokowej, np.: jeśli chcemy przedstawić badane rodziny według liczby dzieci, to otrzymamy szereg rozdzielczy taki jak w Tablicy 6, jeśli nie wystąpiły rodziny o liczbie dzieci większej niż 5.

W przedstawionym przykładzie zbiorowość to liczba rodzin,  $N = 82$ , a cechą, według której dokonano grupowania, jest liczba dzieci w rodzinie. Wyodróżniona cecha – liczba dzieci w rodzinie – przyjęła tu następujące wartości: 1, 2, 3, 4 i 5. Obok każdej wartości podano jej liczebność ( $f$ ).

Tablica 6

*Liczba dzieci w rodzinach uczniów kl. III*

| Liczba dzieci<br>$X$ | Liczba rodzin<br>$f$ |
|----------------------|----------------------|
| 1                    | 10                   |
| 2                    | 37                   |
| 3                    | 22                   |
| 4                    | 9                    |
| 5                    | 4                    |
| Ogółem               | $N = 82$             |

$N$  – liczebność badanej zbiorowości

Źródło: dane umowne

W przypadku, gdy w szeregu warianty cechy przyjmują bardzo wiele wartości liczbowych, wskazane jest łączenie w jedną klasę kilku zbliżonych do siebie wartości. **Klasa** to zbiór wszystkich wartości liczbowych leżących wewnątrz przedziału oznaczonego ustalonymi granicami. Najniższa wartość liczbową klasy to jej dolna granica, a najwyższa jej wartość liczbową stanowi górną granicę.

**Środek przedziału** to średnia arytmetyczna jego dolnej i górnej granicy, np. w przedziale  $10 \div 14$  dolną granicą jest 10, górną – 14, środek przedziału wynosi 12, ponieważ:  $(10+14):2 = 12$ .

**Rozpiętość (wielkość) klasy**, przy zmiennych skokowych jest to ilość wartości liczbowych zawartych w danej klasie. Jest to różnica dwóch średnich arytmetycznych klas następujących bezpośrednio po sobie.

Przedział jest zamknięty, jeśli jego granice określone są liczbowo, np.:  $10 \div 14$ . Przedział otwarty występuje wtedy gdy wartości liczbowe leżące powyżej lub poniżej granicy dolnej albo górnej są ujęte razem, np.: do 6 lat, powyżej 40 lat itp.

Obliczenia wielkości przedziału klasowego,  $i = \frac{X_{max} - X_{min}}{r}$  dokonuje się

w ten sposób, że znajduje się różnicę między wartościami największą ( $X_{max}$ ) i najmniejszą ( $X_{min}$ ), a następnie dzieli ją przez pożądaną liczbę klas ( $r$ ).

Dokonując budowy szeregu rozdzielczego należy na wstępie ustalić liczbę klas, ich wielkość, a także punkt wyjściowy klasy początkowej.

W rezultacie przyjęcia dużej wielkości klasy uzyskuje się małą liczbę klas, a przyjęcie małej wielkości klasy powoduje dużą liczbę klas. Należy pamiętać o tym, że zbyt duża liczba klas zbytnio uszczegóławia rozkład badanej cechy, nie przedstawiając jej ogólnej struktury, natomiast zbyt mała liczba klas prowadzi w konsekwencji do zacierania się miejsc nasilania się lub braku występowania badanej cechy, co również jest niepożądane w poprawnej analizie.

Podjmując decyzję o liczbie klas, badacz powinien poddać wnikliwej analizie rozkład cechy (zmiennej) w szeregu statystycznym i przyjąć ich taką liczbę, która pozwoli na uwzględnienie w przedziałach klasowych najbardziej istotnych miejsc nasilania się i braku występowania badanej cechy.

Jako wielkości dla przedziałów klasowych zalecane są następujące liczby: 1, 2, 3, 5 lub 10, 20, 30, 50. Najbardziej praktycznym wydaje się rozpoczęcie ustalania dolnej granicy pierwszego przedziału klasowego od najniższej zarejestrowanej w badaniu wartości, licząc w górę, lub od najwyższej, licząc w dół. Jeżeli najmniejsza wartość uzyskana z pomiaru nie jest zerem, istnieje możliwość ustalenia dolnej granicy pierwszego przedziału klasowego od dowolnej wartości, nawet takiej, która nie wystąpiła w badaniu, jednak w utworzonym przedziale klasowym powinna znaleźć miejsce chociażby jedna wartość uzyskana w badaniu.

Niżej zamieszczono przykład budowy szeregu rozdzielczego (Tablica 7) przedstawiającego rozkład wyników testu osiągnięć szkolnych  $N = 70$  uczniów

kl. V, w którym największy wynik wynosi  $X_{max}=28$ , najniższy  $X_{min}=9$ . Przedstawiono rozkład wyników dla 10, 7, 5 i 3 klas.

Tablica 7

Rozkład liczebności przy różnej wielkości klasy

| $i = 1$                                                                                                     | $f$ | $i = 2$<br>$r = 10$                    | $f$ | $i = 3$<br>$r = 7$                    | $f$ | $i = 4$<br>$r = 5$                    | $f$ | $i = 7$<br>$r = 3$                    | $f$ |
|-------------------------------------------------------------------------------------------------------------|-----|----------------------------------------|-----|---------------------------------------|-----|---------------------------------------|-----|---------------------------------------|-----|
| 9                                                                                                           | 1   | 9–10                                   | 1   | 9–11                                  | 3   | 9–12                                  | 4   | 9–15                                  | 12  |
| 10                                                                                                          | –   | 11–12                                  | 3   | 12–14                                 | 5   | 13–16                                 | 14  | 16–22                                 | 49  |
| 11                                                                                                          | 2   | 13–14                                  | 4   | 15–17                                 | 20  | 17–20                                 | 37  | 23–29                                 | 9   |
| 12                                                                                                          | 1   | 15–16                                  | 10  | 18–20                                 | 27  | 21–24                                 | 10  | $N = 70$<br>$i = (28-9):3$<br>$i = 7$ |     |
| 13                                                                                                          | 2   | 17–18                                  | 21  | 21–23                                 | 8   | 25–28                                 | 5   |                                       |     |
| 14                                                                                                          | 2   | 19–20                                  | 16  | 24–26                                 | 6   | $N = 70$<br>$i = (28-9):5$<br>$i = 4$ |     |                                       |     |
| 15                                                                                                          | 4   | 21–22                                  | 6   | 27–29                                 | 1   |                                       |     |                                       |     |
| 16                                                                                                          | 6   | 23–24                                  | 4   | $N = 70$<br>$i = (28-9):7$<br>$i = 3$ |     |                                       |     |                                       |     |
| 17                                                                                                          | 10  | 25–26                                  | 4   |                                       |     |                                       |     |                                       |     |
| 18                                                                                                          | 11  | 27–28                                  | 1   |                                       |     |                                       |     |                                       |     |
| 19                                                                                                          | 9   | $N = 70$<br>$i = (28-9):10$<br>$i = 2$ |     |                                       |     |                                       |     |                                       |     |
| 20                                                                                                          | 7   |                                        |     |                                       |     |                                       |     |                                       |     |
| 21                                                                                                          | 4   |                                        |     |                                       |     |                                       |     |                                       |     |
| 22                                                                                                          | 2   |                                        |     |                                       |     |                                       |     |                                       |     |
| 23                                                                                                          | 2   |                                        |     |                                       |     |                                       |     |                                       |     |
| 24                                                                                                          | 2   |                                        |     |                                       |     |                                       |     |                                       |     |
| 25                                                                                                          | 3   |                                        |     |                                       |     |                                       |     |                                       |     |
| 26                                                                                                          | 1   |                                        |     |                                       |     |                                       |     |                                       |     |
| 27                                                                                                          | –   |                                        |     |                                       |     |                                       |     |                                       |     |
| 28                                                                                                          | 1   |                                        |     |                                       |     |                                       |     |                                       |     |
| $i$ – wielkość przedziału klasowego,<br>$r$ – liczba klas,<br>$f$ – częstość występowania (liczebność klas) |     |                                        |     |                                       |     |                                       |     |                                       |     |

Przy wielkości klasy  $i = 2$  i liczbie klas  $r = 10$  otrzymano rozkład bardzo szczegółowy, nie przedstawiający ogólnej struktury badanej cechy, natomiast przy wielkości klasy  $i = 7$  i liczbie klas  $r = 3$  uzyskano rozkład zbyt ogólny, zacierający specyfikę badanej cechy.

Dla innego przykładu szeregu rozdzielczego, zawierającego wyniki pewnego testu, zamieszczonego w Tablicy 8, przedstawiono środki przedziałów klasowych ( $x_i$ ), a także utworzono szereg kumulacyjny ( $f_c$ ).

**Szereg kumulacyjny** powstaje najczęściej z szeregu rozdzielczego w rezultacie kolejnego dodawania (kumulowania) przedziałów klasowych wraz z odpowiadającymi im liczebnościami. Zabieg taki jest pożądany wówczas gdy badacz zainteresowany jest nie tym, jakie liczebności przyjmują poszczególne klasy, ale tym, jakie wartości znajdują się poniżej lub powyżej pewnej wartości zmiennej (np.: może to być określona norma).

Środek przedziału ( $x_i$ ) stanowi średnia arytmetyczna jego górnej i dolnej granicy. Tak ustalone granice przedziałów klasowych nie wskazują jednak dokładnie, gdzie przedział zaczyna się i kończy na linii pomiarowej. Pamiętać jednak należy, że wynik np.: 7 oznacza faktycznie od 6,5 do 7,5, wynik 8 – od 7,5 do 8,5. Oznacza to, że przedział, np.: od 7 do 8 rozciąga się właściwie od 6,5 do 8,5.

Tablica 8

**Środki przedziałów ( $x_i$ ) i kumulacyjny rozkład liczebności ( $f_c$ )**

| Wyniki testu<br>$X$ | Środek przedziału<br>$x_i$ | Dokładne granice<br>klasy | $f$ | $f_c$ |
|---------------------|----------------------------|---------------------------|-----|-------|
| 7 – 8               | 7,5                        | 6,5 – 8,5                 | 1   | 1     |
| 9 – 10              | 9,5                        | 8,5 – 10,5                | 2   | 3     |
| 11 – 12             | 11,5                       | 10,5 – 12,5               | 3   | 6     |
| 13 – 14             | 13,5                       | 12,5 – 14,5               | 9   | 15    |
| 15 – 16             | 15,5                       | 14,5 – 16,5               | 13  | 28    |
| 17 – 18             | 17,5                       | 16,5 – 18,5               | 17  | 45    |
| 19 – 20             | 19,5                       | 18,5 – 20,5               | 8   | 53    |
| 21 – 22             | 21,5                       | 20,5 – 22,5               | 5   | 58    |
| 23 – 24             | 23,5                       | 22,5 – 24,5               | 4   | 62    |
| 25 – 26             | 25,5                       | 24,5 – 26,5               | 3   | 65    |

Źródło: dane umowne

Wartości w kolumnie  $f_c$  uzyskano w rezultacie sumowania liczebności, rozpoczynając od liczebności najniższej, kolejno sumując dalsze. Liczby zamieszczone w kolumnie  $f_c$  należy odczytywać w sposób następujący: jedna osoba uzyskała z testu mniej niż 8,5 pkt., 3 osoby uzyskały mniej niż 10,5 pkt., 6 osób mniej niż 12,5 pkt. itd. Kumulowanie przeprowadzić można również w odwrotnym kierunku, wówczas sumowanie należy rozpoczynać od liczebności odpowiadającej najwyższej wartości zmiennej.

### 3. Tablica statystyczna

#### 3.1. Budowa tablicy

Zebrany w badaniach materiał empiryczny po pogrupowaniu umieszczany jest w tablicach w tekście opracowania, a czasem również w aneksie. Prawidłowo zbudowana tablica pozwala na szybką orientację w prezentowanym materiale statystycznym, umożliwia porównywanie wyników badań i analizę struktury zbiorowości, ułatwia wykrywanie prawidłowości, a także umożliwia obliczanie miar statystycznych. Sposób prezentowania wyników uzależniony jest od celu i funkcji, jaką ma spełniać tablica.

Każda tablica składa się z następujących części:

- tytułu tablicy,
- tablicy właściwej,
- informacji dotyczących źródła danych zawartych w tablicy.

Tytuł tablicy umieszcza się nad tablicą. Nad tytułem, po prawej lub lewej stronie umieszcza się napis tablica z cyfrą lub literą oznaczającą kolejność tablicy w tekście opracowania, przy czym nie wskazane jest umieszczanie przed cyfrą skrótu „nr”, np.: Tablica 3, Tablica B itp. Tytuł powinien określać treść tablicy w zakresie rzeczowym (co i kogo badamy?), przestrzennym (gdzie przeprowadzono badanie?), czasowym (kiedy przeprowadzono badanie?) i cechy według której dokonano klasyfikacji, np.:

| Treść tytułu                                              | Zakres       |
|-----------------------------------------------------------|--------------|
| liczba książek przeczytanych przez uczniów kl. V SP nr 34 | reczowy      |
| w Białymstoku                                             | przestrzenny |
| w roku szkolnym 1995/96                                   | czasowy      |
| według ocen z języka polskiego                            | cechy        |

Z praktyki wynika, że jeżeli badacz przygotowując relację z badań, dokonuje w odpowiednim miejscu szczegółowej charakterystyki badanych osób, a także czasu i miejsca przeprowadzonych badań, to wówczas w tytułach tablic znajdujących się w dalszej części opracowania nie zamieszcza się informacji dotyczących osób badanych, czasu i miejsca badań i wystarczająco w wyżej zamieszczo-

nym przykładzie byłby następujący tytuł tablicy: Liczba książek przeczytanych przez uczniów kl. V według ocen z języka polskiego.

Jeżeli w tablicy przedstawia się wartości liczbowe lub procentowe, albo jedno i drugie jednocześnie, to wówczas można zaznaczyć tę informację w tytule tablicy, na końcu w nawiasie albo umieszczając w tablicy odpowiednie kolumny lub wiersze z napisem: Liczba, %. W tytule tablicy nie należy używać sformułowań typu „Zestawienie procentowe liczby nauczycieli ...”.

Tablica właściwa zbudowana jest z pewnej liczby poziomych wierszy i pionowych kolumn. Poziomych wierszy nie oddziela się zazwyczaj liniami poziomymi. Górna część tablicy nosi nazwę główki tablicy, a boczna – boczka tablicy, pozostała jej część to korpus. W główce i boczku tablicy umieszcza się nazwy cech, pod względem których zebrany materiał jest uporządkowany lub pogrupowany. Przykład poprawnie wykonanej makiety tablicy przedstawiono w Tablicy 9.

*Tablica 9*

*Makieta tablicy*

| Tytuł boczku<br>(podmiotu) | Tytuł główki (orzeczenie) |  |  |  |  |
|----------------------------|---------------------------|--|--|--|--|
|                            |                           |  |  |  |  |
|                            |                           |  |  |  |  |

*Legenda:*

*Źródło:*

W tablicach stosuje się następujące znaki umowne:

- kreska (–) – oznacza brak występowania danego zjawiska,
- zero (0) – oznacza, że zjawisko występuje w ilościach mniejszych od rzędu liczb w tablicy,
- kropka (.) – oznacza brak informacji,
- krzyżyk (x) – oznacza, że tablica w tym miejscu nie może być wypełniona ze względu na układ tablicy.

Umieszczanie większej liczby cech i ich odmian w główce powoduje w konsekwencji konieczność budowania główki „wielopiętrowej”, co czyni ją trudno czytelną, natomiast rozbudowywanie boczka jest znacznie łatwiejsze

i w rezultacie otrzymuje się tablicę bardziej przejrzystą. Należy dążyć do mniej więcej równomiernego rozdzielenia cech między główką i boczkiem, gdyż w przeciwnym razie otrzymujemy tablice niezgrabne i nieestetyczne, tzn. zbyt małą liczbę kolumn w stosunku do zbyt dużej liczby wierszy lub tablicę ze zbyt dużą liczbą kolumn w stosunku do niewielkiej liczby wierszy.

Umieszczenie nazwy poszczególnych cech w główce w zasadzie nie ma istotnego znaczenia i zależy od badacza. Jednak z praktyki wynika, iż w przypadku badań weryfikacyjnych (przyczynowo-skutkowych), tzn. wówczas gdy w podejmowanych badaniach badacz określa zmienne niezależne (globalną i szczegółowe) oraz zmienne zależne (globalną i szczegółowe) wskazane jest, aby w tytule boczku tablicy umieszczać zmienną niezależną szczegółową (przyczyna) i jej warianty, natomiast w główce tablicy – zmienną zależną szczegółową (skutek) i jej warianty. Taki sposób budowy tablicy wydaje się być bardziej logiczny i umożliwia poprawniejszą analizę zebranego materiału.

Tytuły kolumn i wierszy (nazwy cech – zmiennych) powinny być formułowane możliwie krótko, precyzyjnie i jasno. Należy unikać przenoszenia do tytułów kolumn lub wierszy długich zdań pytających przeniesionych z kwestionariusza lub ankiety, np.: pytanie „Czym się kierowałaś (eś) w wyborze kierunku studiów?” można ująć w tablicy jako „Motywy wyboru kierunku studiów”.

W przypadku, gdy warianty dla którejkolwiek ze zmiennych są wyrażone przy pomocy dłuższych sformułowań, czy zdań wskazane jest, aby w celu uzyskania przejrzystej tablicy nazwy poszczególnych wariantów zastąpić, np.: literami a, b, c, d wpisując je w odpowiednie miejsca w tablicy, a pod tablicą wyjaśnić znaczenie poszczególnych liter (legenda). Tego rodzaju sytuacja może mieć miejsce np.: podczas przedstawiania w tablicach odpowiedzi na pytania z kwestionariusza.

Sumy „Ogółem” lub „Razem” w tablicy umieszcza się albo u dołu i z prawej strony (tzn. w ostatnim wierszu i w ostatniej kolumnie) albo u góry i z lewej strony (tzn. w pierwszym wierszu i w pierwszej kolumnie). W tablicy wskazane jest konsekwentne używanie określenia „Ogółem” albo „Razem”. Nie należy natomiast stosować w tej samej tablicy, np.: w wierszu – „Ogółem”, a w kolumnie – „Razem”. Z praktyki wynika, iż podawanie sum u góry i z lewej strony jest wygodne ze względu na to, że na ogół przy analizie danych statystycznych wychodzimy od wielkości ogólnych, charakteryzujących całą badaną zbiorowość statystyczną, a następnie przechodzimy do szczegółów dotyczących poszczególnych części zbiorowości. Jeśli natomiast prowadzimy analizę od szczegółu do ogółu, zasadne jest podawanie sum u dołu i z prawej strony.

W zależności od stopnia złożoności i zawartości tablicy, umieszczając odpowiednie odsyłacze przy liczbach lub procentach, pod tablicą można podawać informacje dodatkowe, mogące ułatwić odczytywanie danych zawartych w tablicy. W przypadku opracowywania materiału empirycznego uzyskanego przy pomocy ankiety lub kwestionariusza zdarza się (w pytaniach wielokrotnego wyboru), iż w tablicy przedstawiającej wyniki dotyczące któregoś z pytań liczba wyborów dokonanych przez respondentów nie sumuje się z liczbą osób badanych. W takiej sytuacji warto zaznaczyć to przez umieszczenie gwiazdki (\*) na końcu tytułu tablicy, pod tablicą umieszczając gwiazdkę (\*) z informacją, np.: liczba dokonanych wyborów nie sumuje się z liczbą osób badanych, ponieważ badani mogli wybrać więcej niż jedną odpowiedź. Podobnie, wyjaśnienia wymaga sytuacja, gdy liczba informacji czy wyborów jest mniejsza niż liczebność próby badanej.

Zamieszczanie informacji dotyczącej źródła danych zawartych w tablicy jest niezbędne, ponieważ stwarza możliwość sprawdzenia przedstawionych danych, a także przedstawione dane czyni bardziej wiarygodnymi.

Tablica 10, zamieszczona jako przykład, spełnia określone wyżej warunki poprawności.

*Tablica 10*

*Liczba książek przeczytanych przez uczniów kl. III SP nr 10 w Białymstoku  
w roku szkolnym 1995/96 według ocen z języka polskiego*

| Liczba przeczytanych<br>książek | Oceny z języka polskiego |    |    |   |   |        |
|---------------------------------|--------------------------|----|----|---|---|--------|
|                                 | 5                        | 4  | 3  | 2 | 1 | Ogółem |
| 6 – 10                          | —                        | 1  | 8  | 2 | 2 | 13     |
| 11 – 15                         | 3                        | 4  | 7  | 1 | — | 15     |
| 16 – 20                         | 11                       | 6  | 5  | 1 | — | 23     |
| 21 – 25                         | 16                       | 7  | 2  | — | — | 25     |
| Ogółem                          | 30                       | 18 | 22 | 4 | 2 | 76     |

*Źródło: Obliczenia własne*

Tablice powinny spełniać zasadę samoobjaśnialności, czyli mieć taką budowę i opis, że zarówno ich tytuł jak i treść nagłówek winny dostarczać takich informacji, aby na ich podstawie można było w pełni zrozumieć zawartość tablicy, bez potrzeby poszukiwania dodatkowych objaśnień w tekście opracowania.

### 3.2. Rodzaje tablic

W zależności od celu i zakresu badań, a także stopnia złożoności analizowanych zjawisk uporządkowany i pogrupowany materiał badawczy może być przedstawiony w postaci różnorodnych tablic statystycznych. Z różnego punktu widzenia tablice statystyczne można podzielić na:

- tablice robocze i wynikowe,
- tablice proste i złożone,
- tablice analityczne i syntetyczne.

**Tablice robocze i wynikowe.** Tablice robocze wykorzystywane są podczas zliczania zebranego materiału empirycznego (Tablica 11), a wynikowe zamieszczane są w tekście opracowania.

Tablica 11

Tablica robocza

| Płeć       | Rodzaj ukończonej szkoły średniej |           |        |
|------------|-----------------------------------|-----------|--------|
|            | LO                                | Technikum | Ogółem |
| Dziewczęta | ☐☐☐☐☐L 27                         | ☐☐☐ 15    | 42     |
| Chłopcy    | ☐☐☐☐ 18                           | ☐☐ 10     | 28     |
| Ogółem     | 45                                | 25        | 70     |

**Tablice proste i złożone.** Tablice proste charakteryzują badaną zbiorowość pod względem jednej cechy i zawierają jeden szereg statystyczny, a tablice złożone przedstawiają badaną zbiorowość według dwóch lub więcej cech jednocześnie i mogą zawierać kilka szeregów statystycznych.

Tablice proste stosowane są głównie w celu charakterystyki badanej próby, także w przypadku podstawowej prezentacji uzyskanych wyników i stanowią przejrzysty obraz struktury badanej zbiorowości, lecz dostarczają stosunkowo mało informacji o niej. Natomiast tablice złożone stosowane są najczęściej podczas szczegółowej analizy wyników badań, gdy poszczególne cechy są analizowane w zestawieniu z innymi cechami, z którymi mogą mieć związek. Taki tok postępowania jest racjonalny, jednak przy dużej liczbie cech można otrzymać tablice nadmiernie skomplikowane i nieczytelne. W praktyce zaleca się, aby w tablicy nie umieszczać więcej niż cztery cechy.

**Tablice analityczne i syntetyczne.** W tablicach analitycznych, zamieszczanych najczęściej w aneksie, prezentuje się bardzo szczegółowe wyniki badań w odniesieniu np.: do poszczególnych grup wiekowych, roczników, klas, lat studiów itp. Tego rodzaju wyniki mogą być przedstawione w postaci zbiorczej tablicy analitycznej (jedna duża tablica zawierająca uzyskane wyniki dla np.: badanych klas, grup wiekowych itp.) lub też może to być kilka oddzielnych tablic, w których przedstawiono wyniki poszczególnych badanych klas, grup wiekowych itp. (Tablica 12, 13, 14).

*Tablica 12*

*Liczba dzieci w rodzinie uczniów kl. III „a” SP 3 w miejscowości X w roku szkolnym 1995/96, a ich wyniki w nauce*

| Liczba dzieci | Wyniki w nauce |         |        |        |
|---------------|----------------|---------|--------|--------|
|               | wysokie        | średnie | niskie | Ogółem |
| 1             | 5              | 3       | –      | 8      |
| 2             | 4              | 6       | 2      | 12     |
| 3 i więcej    | 1              | 3       | 3      | 7      |
| Ogółem        | 10             | 12      | 5      | 27     |

*Źródło: dane umowne*

*Tablica 13*

*Liczba dzieci w rodzinie uczniów kl. III „b” SP 3 w miejscowości X w roku szkolnym 1995/96, a ich wyniki w nauce*

| Liczba dzieci | Wyniki w nauce |         |        |        |
|---------------|----------------|---------|--------|--------|
|               | wysokie        | średnie | niskie | Ogółem |
| 1             | 3              | 3       | –      | 6      |
| 2             | 2              | 7       | 2      | 11     |
| 3 i więcej    | 1              | 3       | 4      | 8      |
| Ogółem        | 6              | 13      | 6      | 25     |

*Źródło: dane umowne*

Tablica 14

*Liczba dzieci w rodzinie uczniów kl. III „c” SP 3 w miejscowości X  
w roku szkolnym 1995/96, a ich wyniki w nauce*

| Liczba dzieci | Wyniki w nauce |         |        |        |
|---------------|----------------|---------|--------|--------|
|               | wysokie        | średnie | niskie | Ogółem |
| 1             | 4              | –       | 3      | 7      |
| 2             | 3              | 8       | 3      | 14     |
| 3 i więcej    | 2              | 1       | 4      | 7      |
| Ogółem        | 9              | 9       | 10     | 28     |

Źródło: dane umowne

W tablicach syntetycznych, zamieszczanych w tekście zasadniczym opracowania, przedstawia się syntezę, ogólne ujęcie zawartości tablic analitycznych, wartości przeciętne i miary rozproszenia charakteryzujące wyodrębnione grupy badanych. Jedna tablica syntetyczna może być opracowana na podstawie kilku tablic analitycznych. Tablica 15 jest tablicą syntetyczną opracowaną na podstawie danych zawartych w Tablicy 12, 13, 14.

Tablica 15

*Liczba dzieci w rodzinie uczniów kl. III SP 3 w miejscowości X  
w roku szkolnym 1995/96, a ich wyniki w nauce*

| Liczba dzieci | Wyniki w nauce |         |        |        |
|---------------|----------------|---------|--------|--------|
|               | wysokie        | średnie | niskie | Ogółem |
| 1             | 12             | 6       | 3      | 21     |
| 2             | 9              | 21      | 7      | 37     |
| 3 i więcej    | 4              | 7       | 11     | 22     |
| Ogółem        | 25             | 34      | 21     | 80     |

Źródło: dane umowne

Dokonując analizy danych z tablicy syntetycznej, badacz powinien odwoływać się do danych zawartych w tablicach analitycznych (z aneksu), które to były podstawą do budowy tablic syntetycznych.

Podział tablic wynikowych na analityczne i syntetyczne jest pożądany szczególnie wówczas gdy zebrany materiał jest bardzo obszerny, a umieszczenie w tekście opracowania dużej liczby tablic ze szczegółowymi wynikami zakłóciłoby przejrzystość analiz.

## 4. Wykresy

Równorzędnym sposobem przedstawienia zebranego materiału, obok tablic są wykresy, których odczytanie często jest łatwiejsze i bardziej zrozumiałe niż nawet najlepiej skonstruowane tablice.

Podstawową funkcją wykresów jest pokazywanie w sposób obrazowy kształtu rozkładu uzyskanych wyników badań. Istnieją różne typy wykresów. Zastosowanie w praktyce określonego rodzaju wykresu uzależnione jest od tego, jakie cechy (zmienne) zawiera tablica, dla których sporządzany jest wykres. Innego rodzaju wykres wymagany jest dla przedstawienia graficznego cech(zmiennych) o charakterze ilościowym, a inny dla cech jakościowych.

W przypadku cech ilościowych do najczęściej stosowanych form wykresów zaliczamy: **wykresy liniowe, histogramy i wieloboki liczebności** (diagramy). Przy graficznym przedstawianiu cech jakościowych najczęściej zastosowanie znajdują **wykresy słupkowe, koła, kule, obrazki i kartogramy**<sup>12</sup>.

**Wykresy liniowe** sporządza się najczęściej na I ćwiartce (dodatniej) układu współrzędnych Kartezjusza. W przypadku przedstawiania zbiorowości według jednej cechy zaobserwowane wartości zmiennej odkłada się na osi odciętych (poziomej), a częstości jej występowania (wyrażone liczbowo lub procentowo) na osi rzędnych (pionowej).

Podczas wykonywania wykresów liniowych najczęściej wykorzystywana jest podziałka zwykła lub półlogarytmiczna. Podziałkę należy tak dobierać, aby wykres najwłaściwiej oddawał charakterystykę danej zbiorowości i poszczególne elementy obrazu graficznego przedstawiały przejrzystość wartości badanej zmiennej, aby powierzchnia przeznaczona na wykres pozwalała zarówno na odzwierciedlenie najmniejszych, jak i największych częstości występowania (wyrażonych liczbowo lub procentowo).

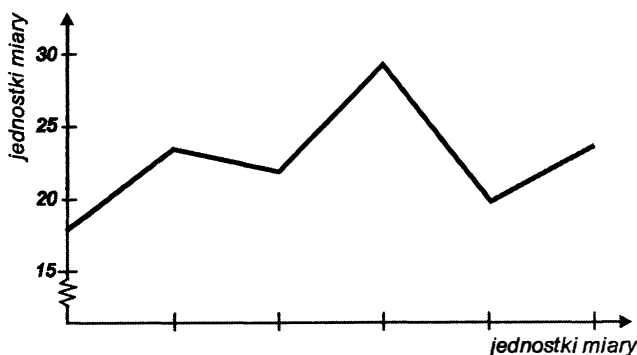
Od podziałki zaznaczonej na obu osiach zależy prezentacja graficzna obrazu. Zaleca się dobór podziałki w taki sposób, by wysokość graficznego obrazu, do maksymalnej szerokości tego obrazu pozostawała w stosunku 3 : 5<sup>13</sup>. Skala jednostek odkładanych zarówno na osi pionowej jak i poziomej może być różna.

---

<sup>12</sup> K. Podoski, *Metody statystyczne. Podstawowe pojęcia i zastosowanie w naukach społecznych*. WSP, Gdańsk 1967, s. 48-67; Z. Rogoziński, *Statystyka opisowa i indukcyjna*, PWN, Warszawa 1971, s. 43-68.

<sup>13</sup> G. Clauss, H. Ebner, *Podstawy .... op.cit.*, s. 56.

Wykres jest budowany najczęściej na podstawie wyników zawartych w tablicy i jest odpowiednikiem jej całości lub wybranych fragmentów. Składa się z podobnych elementów jak tablica – części tekstowej oraz samego wykresu graficznego. Część tekstowa powinna zawierać odpowiedni napis, np. wykres, rysunek, rycina z numerem, a także tytuł określający w sposób zwięzły to, co wykres zawiera w zakresie rzeczowym, przestrzennym i czasowym. W opracowaniach naukowych zaleca się podpisywanie wykresów na dole, pod wykresem. Pod wykresem również wskazuje się na źródło materiałów, a także podaje się legendę zawierającą stosowne objaśnienia (Wykres 1).



*Legenda (objaśnienie):*

*Wykres 1. Tytuł wykresu*

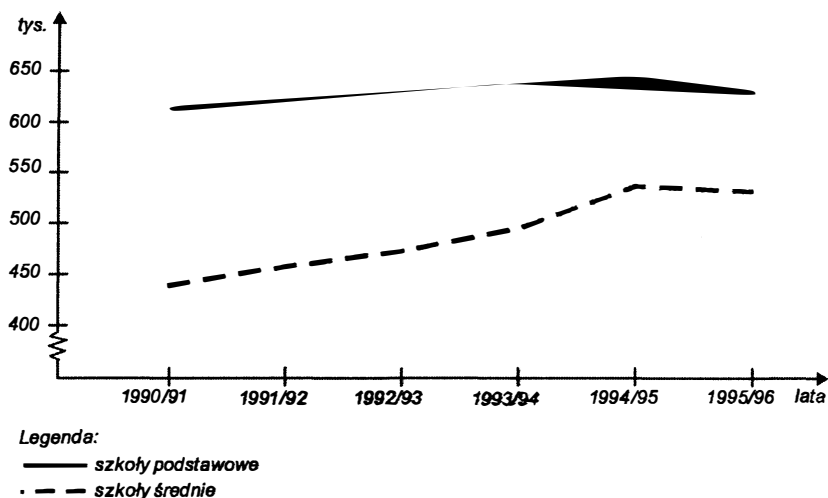
*Źródło:*

W przypadku przedstawiania cechy o bardzo dużej częstości występowania, w celu zmniejszenia wykresu stosuje się przerywanie osi linią zygzakowatą lub przerywaną, a ponad tym miejscem oznacza się odpowiednią wielkość (Wykres 1).

Jeżeli, w sytuacji prezentowania wyników badań istnieje potrzeba wykonania kilku lub kilkunastu graficznych obrazów prezentujących rozkład interesującej badacza cechy np. w różnych częściach zbiorowości, należy dla celów porównawczych stosować takie same podziałki i oznaczenia na wszystkich wykresach.

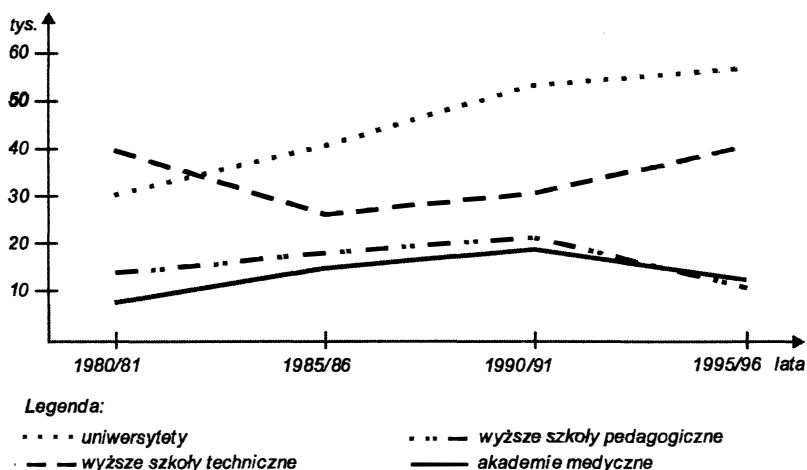
Na osi pionowej nie zaznacza się wartości liczbowych lub procentowych które uzyskano w badaniu, tylko zaznacza się odpowiednią, przyjętą podziałkę, np. jeśli w badaniu uzyskano następujące wartości liczbowe (częstości występowania): 3, 7, 10, 18 i 23 to nie zaznacza się ich na osi pionowej, lecz można tu zastosować podziałkę następującą: 0, 5, 10, 20, 25. Punkt zerowy, przypadający w miejscu przecięcia obu osi nie musi być podawany.

Wykresy liniowe umożliwiają graficzne przedstawienie zjawisk przebiegających w ciągu określonego czasu, ujętych w szereg dynamiczny, t.j. zawierający lata, kwartały, miesiące, dni (Wykres 2 i 3).



**Wykres 2. Absolwenci szkół podstawowych i średnich w Polsce w latach 1991-1996 (w tys.)**

Źródło : Rocznik Statystyczny 1997, GUS, Warszawa 1997, s.226.

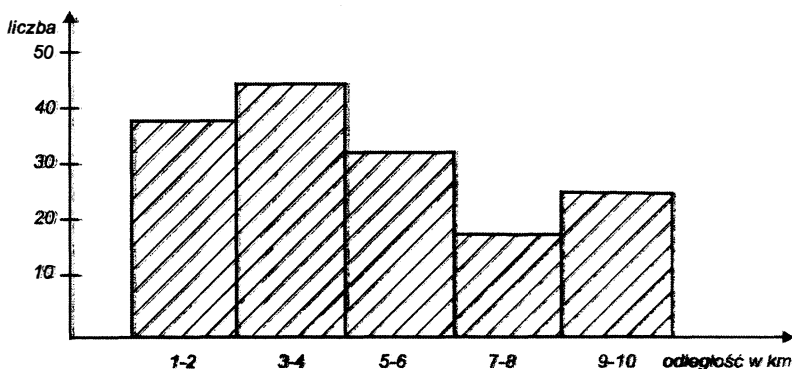


**Wykres 3. Studenci szkół wyższych w Polsce otrzymujący stypendia w latach 1981-1996 (w tys.)**

Źródło : Rocznik Statystyczny, 1997, GUS, Warszawa 1997, s.233.

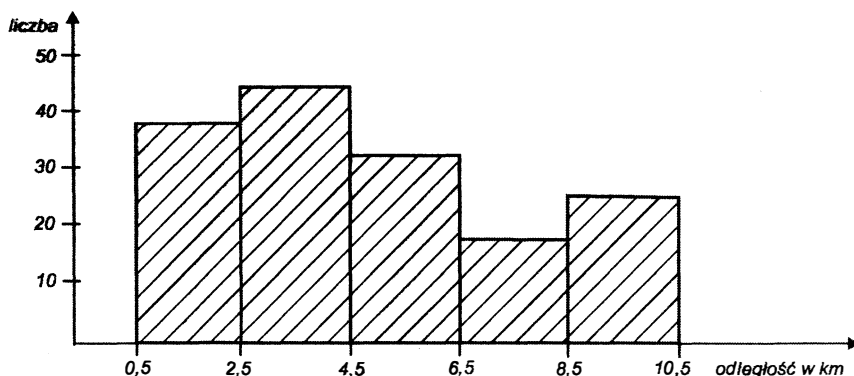
W celu zobrazowania graficznego cechy ilościowej wykorzystuje się także histogramy i wieloboki liczebności (diagramy).

**Histogram** składa się z szeregu przylegających figur prostokątnych, których wysokość odpowiada częstości występowania w wyróżnionych przedziałach klasowych (w liczbach lub w procentach), a podstawy są równe długości (rozpiętości) przedziałów. Wykonując histogram jako podstawy prostokątów na osi poziomej można zaznaczyć przedziały klasowe (Wykres 4) lub granice przedziałów klasowych (Wykres 5)<sup>14</sup>.



**Wykres 4.** Odległość do szkoły (w km) dojeżdżających uczniów kl. V-VIII z miejscowości X w roku szkolnym 1997/98

Źródło: dane umowne

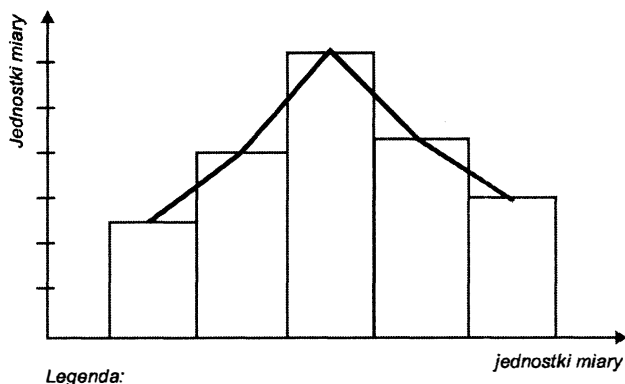


**Wykres 5.** Odległość do szkoły (w km) dojeżdżających uczniów kl. V-VIII z miejscowości X w roku szkolnym 1997/98

Źródło: dane umowne

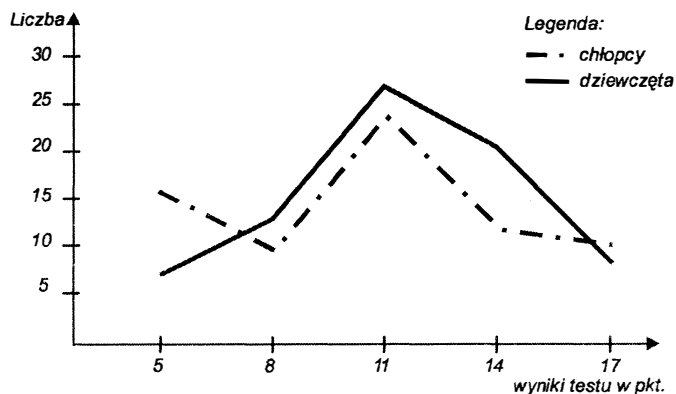
<sup>14</sup> G. Clauss, H. Ebner, *Podstawy ...., op.cit.*, s. 57.

**Wielobok liczebności (diagram)** budowany jest w dwojaki sposób: albo na podstawie histogramu, albo bezpośrednio na podstawie szeregu rozdzielczego, w którym oznaczono środki przedziałów klasowych. Łącząc środki przedziałów klasowych z odpowiadającymi im częstościami występowania otrzymuje się linię łamaną (Wykres 6 i 7). Przy pomocy wieloboku liczebności (zaokrąglając go) można określić przybliżony kształt krzywej ciągłej, nazywanej krzywą liczebności.



Wykres 6. Przykładowy histogram i wielobok liczebności

Źródło:



Wykres 7. Wyniki testu z przedmiotu X uczniów kl. VIII z miejscowości Y w roku szkolnym 1997/98 w zależności od płci

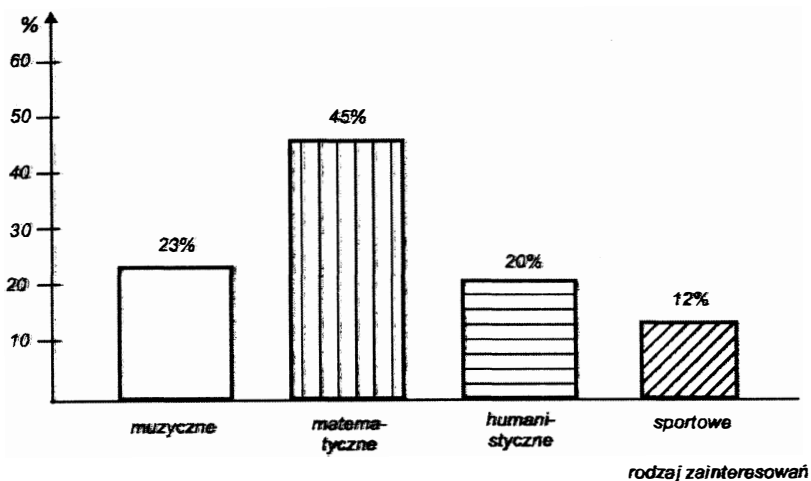
Źródło: dane umowne

Celem omawianych wyżej histogramów i wieloboków liczebności jest przedstawienie graficzne danych o charakterze ilościowym. W związku z tym, jak już sygnalizowano, przy ich rysowaniu uwzględniać należy dwie podziałki: na osi poziomej umieszcza się przedziały klasowe dla uzyskanych wartości zmiennej (Wykres 4) lub granice przedziałów klasowych (Wykres 5) albo środki przedziałów klasowych (Wykres 7), natomiast na osi pionowej zaznacza się częstości występowania pomiarów (w liczbach lub procentach).

W celu prezentacji graficznej wyników badań będących cechami jakościowymi najczęściej stosuje się wykresy słupkowe, wykresy kołowe i kartogramy.

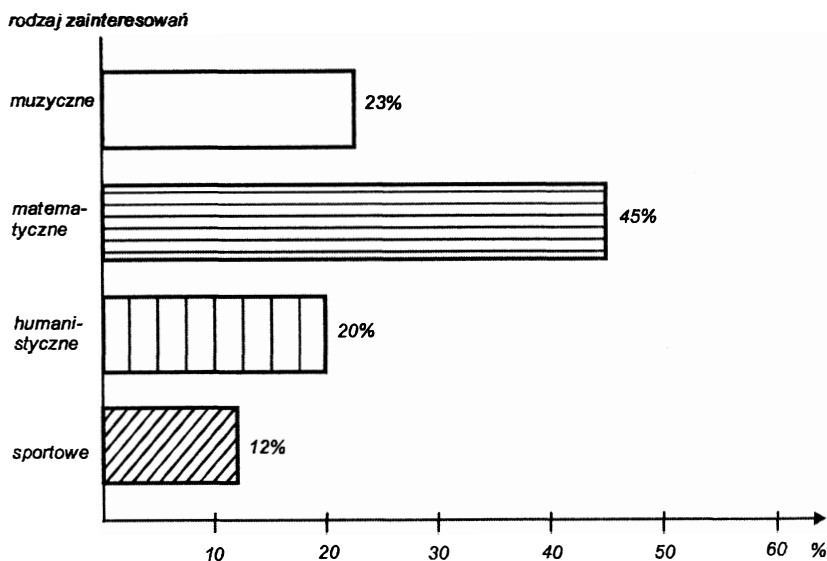
**Wykres słupkowy** składa się z prostokątów, których długość lub wysokość jest proporcjonalna do częstości występowania cechy jakościowej. Cechy jakościowe w statystyce wyrażane są tylko przy pomocy częstości ich występowania (liczba lub procent) i w związku z tym w celu przedstawienia takich danych wykorzystuje się podziałkę umieszczoną wzdłuż osi pionowej (Wykres 8) lub poziomej (Wykres 9).

W przypadku, gdy pewną zbiorowość klasyfikuje się według jednej cechy, stosownie do wariantów jej występowania, np. zainteresowania według ich rodzaju: matematyczne, humanistyczne, muzyczne, sportowe itp., to każdy rodzaj zainteresowań przedstawia się przy pomocy oddzielnego prostokąta (słupka) odpowiedniej wielkości. Słupki mogą występować po sobie w różnej kolejności, należy zachować równą ich szerokość, pomiędzy nimi pozostawia się wolną przestrzeń (np. równą połowie szerokości słupka), aby zaznaczyć skokowy charakter cechy jakościowej. Wykorzystując wykres słupkowy można przedstawić rozkład analizowanej cechy jakościowej w całej badanej zbiorowości (Wykres 8 i 9).



**Wykres 8. Rodzaj zainteresowań uczniów kl. V SP z miejscowości X w roku szkolnym 1997/98**

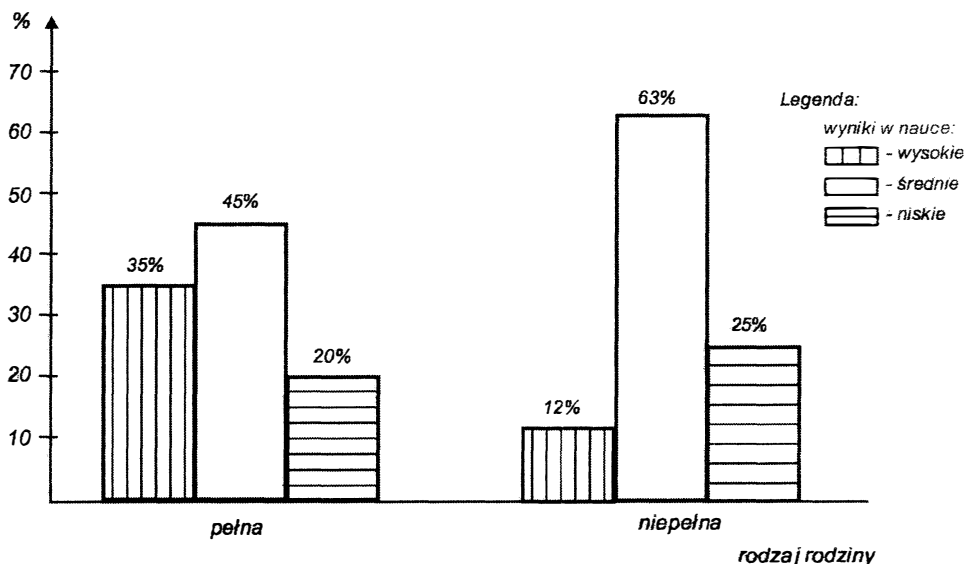
Źródło: dane umowne



**Wykres 9. Rodzaj zainteresowań uczniów kl. V SP z miejscowości X w roku szkolnym 1997/98**

Źródło: dane umowne

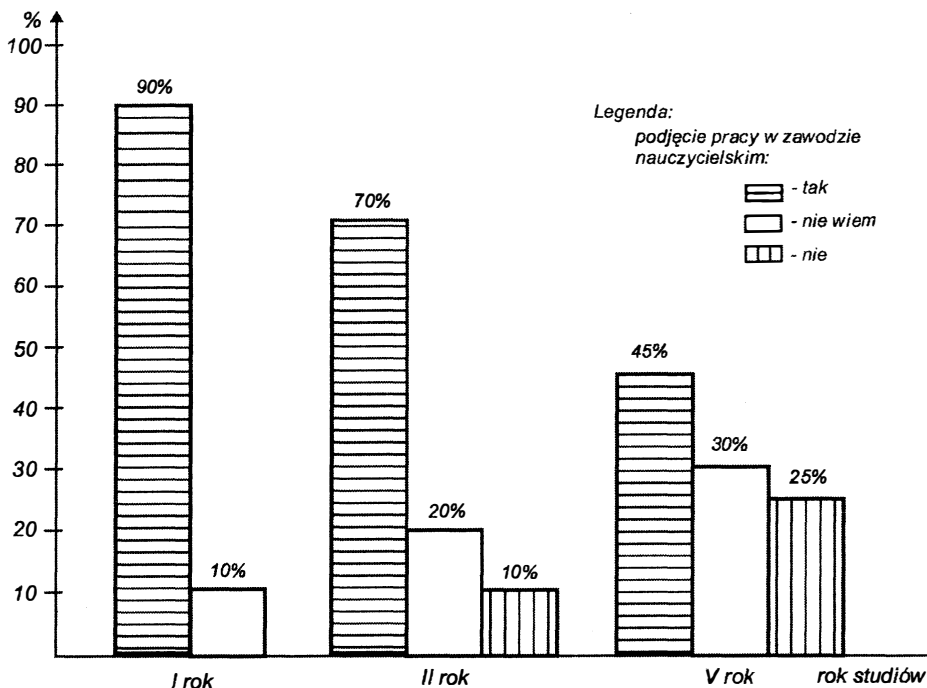
Wykresy słupkowe umożliwiają w rezultacie zastosowanie dla poszczególnych słupków (prostokątów) różnych barw lub zacięniowań przedstawienie rozkładu jednej interesującej badacza cechy (np. na Wykresie 10 – wyniki w nauce: wysokie, średnie, niskie), w różnych wariantach występowania innej cechy (np. na wykresie 10 – rodzaj rodziny: pełna, niepełna), istotnej z punktu widzenia celu podjętych badań (Wykres 10 i 11).



Wykres 10. Wyniki w nauce uczniów kl. III SP z miejscowości X w roku szkolnym 1997/98 w zależności od rodzaju rodziny

Źródło: dane umowne

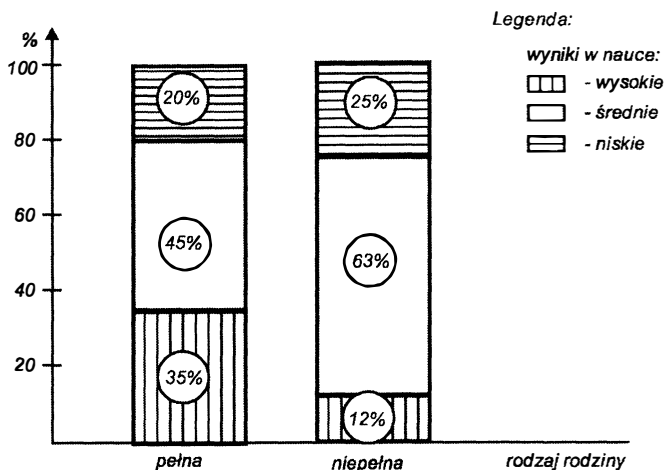
Prezentacja graficzna przedstawiona na Wykresie 10 i 11 obrazuje rozkład wyników badań z uwzględnieniem dwóch cech charakteryzujących badaną zbiorowość. Analizując kształt poszczególnych słupków (prostokątów, powierzchni) nie łatwo jest w sposób bezpośredni dostrzec wzajemne ustosunkowanie części zbiorowości wyodrębnionych dla poszczególnych wariantów występowania analizowanej cechy jakościowej.



Wykres 11. Deklaracja podjęcia pracy w zawodzie nauczycielskim przez studentów I, III i V roku kierunku X w uczelni Y w roku akademickim 1997/98

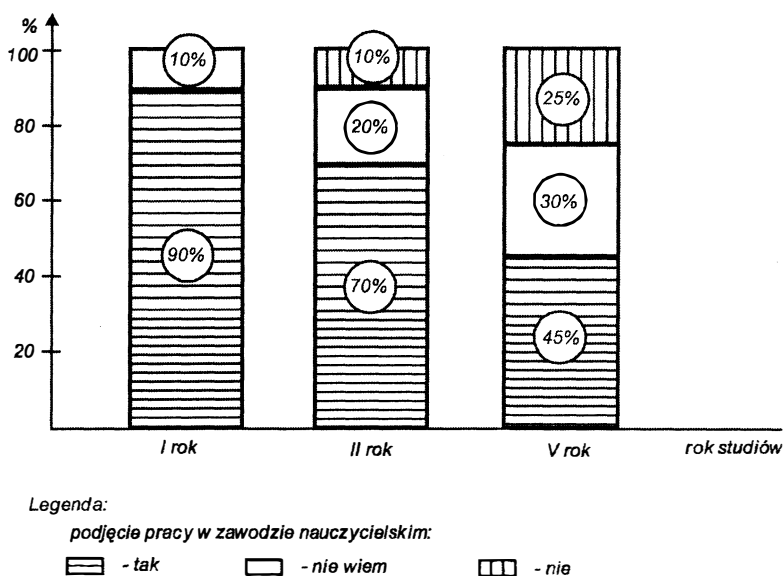
Źródło: dane umowne

W związku z tym korzystniejsze jest przedstawienie złożoności struktury badanej zbiorowości poprzez narysowanie kilku słupków (prostokątów) jednakowej wysokości, obrazujących powierzchnię 100% każdy. Wewnątrz każdego słupka można zaznaczyć, np. kolorami lub poprzez zacieniowanie odpowiednich powierzchni, rozkład interesującej badacza cechy (Wykres 12 i 13). Takie postępowanie umożliwia porównanie rozkładu danej cechy wewnątrz poszczególnych, wyróżnionych części zbiorowości.



Wykres 12. Wyniki w nauce uczniów kl. III SP z miejscowości X w roku szkolnym 1997/98 w zależności od rodzaju rodziny

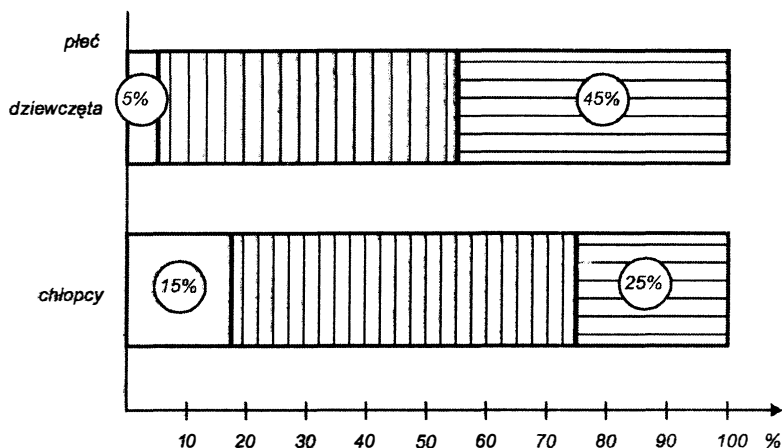
Źródło: dane umowne



Wykres 13. Deklaracja podjęcia pracy w zawodzie nauczycielskim przez studentów I, III i V roku kierunku X w uczelni Y w roku akademickim 1997/98

Źródło: dane umowne

Podczas wykonywania Wykresu 12 i 13 wykorzystano **podziałkę** umieszczoną wzdłuż osi pionowej, ale można tu zastosować również **podziałkę** usytuowaną wzdłuż osi poziomej (Wykres 14).



Legenda:

rodzaj wybranej szkoły ponadpodstawowej:

□ - ZSZ

▤ - Technikum

▥ - Liceum

Wykres 14. Rodzaj wybranej szkoły ponadpodstawowej przez uczniów kl. VIII z miejscowości X w roku szkolnym 1997/98 w zależności od płci

Źródło: dane umowne

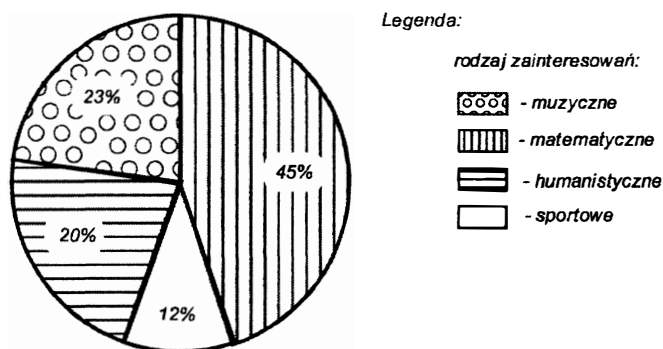
Przy pobieżnej obserwacji wydaje się, że wykres słupkowy jest bardzo podobny do histogramu. Jednak obie formy prezentacji różnią się zasadniczo:

1. oś, na której rysuje się słupki nie zawiera żadnej podziałki lecz określenie cechy jakościowej, np. rodzaj zainteresowań: muzyczne, matematyczne, humanistyczne, sportowe itp.;
2. położenie osi przedstawiającej liczbę lub procent badanych na wykresie słupkowym może być dowolne (poziome lub pionowe);
3. słupki mogą mieć szerokość dowolną, ale powinna być ona jednakoowa dla wszystkich i powinny być one od siebie oddzielone (cecha jakościowa jest cechą zmieniającą się skokowo)<sup>15</sup>.

<sup>15</sup> G. Claus, H. Ebner, *Podstawy...*, op. cit., s. 62.

**Wykres kołowy** stanowi bardzo często stosowany sposób prezentacji graficznej zbiorowości o cesze jakościowej. Wymaga przedstawienia występujących liczebności w przeliczeniu procentowym. Cała powierzchnia koła odpowiada 100%.  $360^{\circ}$  stanowi 100%, w związku z tym  $3,6^{\circ}$  stanowi 1%, czyli dla oznaczenia, np. 10% udziału w całości należy odmierzyć kątomierzem w kole kąt środkowy  $36^{\circ}$ . Poszczególne odmierzone powierzchnie można odpowiednio zakreskować lub zacieniować, aby uzyskać bardziej wyrazisty obraz.

Wyniki badań przedstawione na wykresach słupkowych – Wykres 8 i 9 – można zaprezentować również na wykresie kołowym – Wykres 15.

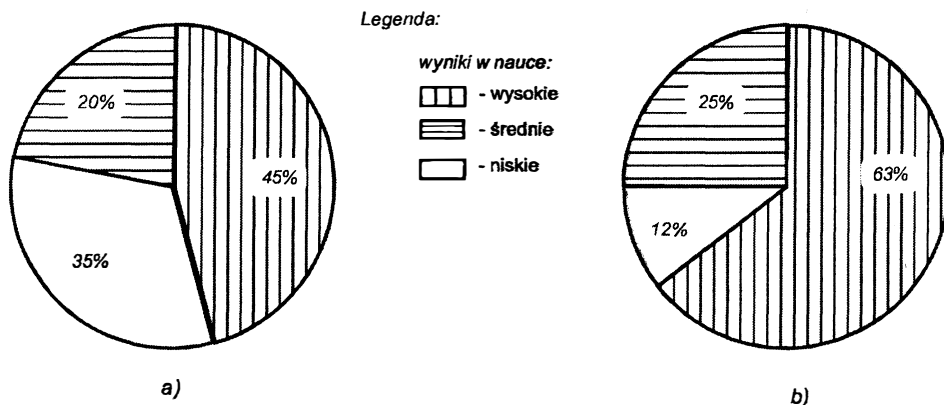


**Wykres 15. Rodzaj zainteresowań uczniów kl. V SP z miejscowości X w roku szkolnym 1997/98**

*Źródło: dane umowne*

Wykorzystanie wykresu kołowego pozwala nie tylko na przedstawienie rozkładu analizowanej cechy jakościowej w całej badanej zbiorowości (Wykres 15), ale także umożliwia prezentację graficzną jednej cechy w różnych wariantach występowania innej, interesującej badacza cechy – Wykres 16.

Wyniki badań przedstawione na wykresie kołowym – Wykres 16 – prezentowano już na wykresie słupkowym (Wykres 12). Z analizy kształtu obu wykresów wynika, że w tym przypadku jednak wykres słupkowy w sposób bardziej jednoznaczny przedstawia różnice w rozkładzie analizowanej cechy w obu wyróżnionych częściach zbiorowości.



Wykres 16. Wyniki w nauce uczniów kl. V pochodzących z rodzin: a) pełnych, b) niepełnych z miejscowości X w roku szkolnym. 1997/98

Źródło: dane umowne

Prezentacja graficzna wyników badań na dwóch lub więcej wykresach kołowych bardziej zasadna wydaje się jednak wówczas gdy rozkład analizowanej cechy w wyróżnionych częściach zbiorowości jest zasadniczo różny, wówczas te różnice są również bardzo widoczne na wykresach kołowych.

**Kartogram** to zazwyczaj mapa geograficzna, na której zaznacza się dane uzyskane na podstawie badań danego problemu w różnych częściach kraju lub województwa, gminy itp. Służy do zobrazowania rozmieszczenia zbiorowości w przestrzeni w takich zagadnieniach jak np. struktura ludności, przestępczość, alkoholizm, analfabetyzm itp. Ważne jest tu ściśle określenie jednostki terytorialnej.

Wszystkie wyżej omówione sposoby prezentowania graficznego danych empirycznych o charakterze ilościowym i jakościowym, aby należycie spełniały swą funkcję, powinny być wykonane zgodnie z określonymi zasadami:

1. powinny być przejrzyste, estetyczne, starannie i dokładnie wykonane;
2. nie powinny być przeładowane i zawierać skomplikowanego opisu;
3. w sposób prosty, jednoznaczny i czytelny powinny obrazować to, co badacz chce przedstawić.

## Pytania i zadania

1. Na czym polega grupowanie statystyczne?
2. Na wybranych przykładach dokonaj grupowania typologicznego i wariacyjnego, wskaż różnice.
3. Jakich ustaleń wymaga budowa szeregu rozdzielczego?
4. Niżej podane wyniki testu osiągnięć szkolnych z fizyki ( $X$ ) uczniów kl. VIII przedstaw w postaci szeregu rozdzielczego zawierającego:

a/ 10 klas,

b/ 7 klas,

c/ 5 klas.

$X$ : 10, 13, 14, 14, 14, 15, 16, 16, 16, 17, 17, 17, 17, 17, 18, 18, 18, 18, 19, 19, 19, 19, 19, 19, 19, 20, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 21, 21, 21, 22, 22, 22, 22, 22, 22, 22, 22, 22, 23, 23, 23, 23, 23, 23, 23, 23, 24, 24, 24, 24, 24, 24, 24, 25, 25, 25, 25, 26, 26, 26, 26, 27, 27, 27, 28, 28.

5. Niżej podane wyniki testu osiągnięć szkolnych z matematyki ( $X$ ) uczniów kl. V przedstaw w postaci szeregu rozdzielczego:

a/  $X$ : 22, 45, 54, 44, 22, 11, 31, 40, 38, 35, 25, 23, 15, 12, 17, 20, 33, 56, 55, 67, 68, 58, 38, 29, 28, 24, 19, 29, 50, 60, 40, 41, 60, 52, 68, 66, 31, 31, 35, 64, 64, 20, 53, 52, 47, 29, 27, 24.

b/  $X$ : 1, 45, 3, 45, 5, 7, 43, 7, 7, 40, 7, 43, 40, 7, 7, 10, 40, 7, 40, 16, 10, 40, 20, 38, 25, 37, 35, 37, 35, 20.

c/  $X$ : 42, 88, 37, 75, 98, 93, 73, 62, 96, 80, 52, 76, 66, 54, 73, 69, 83, 62, 53, 79, 69, 56, 81, 75, 52, 65, 49, 80, 67, 59, 88, 80, 44, 71, 72, 87, 91, 82, 89, 79.

6. Przedstaw propozycję tablicy (wraz z tytułem) zawierającą następujące dane:
- a/ miejsce zamieszkania (wieś, gmina, miasto), typ ukończonej szkoły średniej (LO, LZ, T), wybór szkoły wyższej (uniwersytet, wyższa szkoła pedagogiczna, politechnika, akademia medyczna, inne);
  - b/ wykształcenie matki (podstawowe, zawodowe, średnie, pomaturalne, wyższe) oraz odpowiedź na następujące pytanie z kwestionariusza: Jak często uczestniczy Pani w spotkaniach rodziców z wychowawcą klasy? a/ zawsze, b/ często, c/ nigdy.
7. Przedstaw propozycję tablicy (wraz z tytułem) zawierającej:
- a/ dwie powiązane ze sobą cechy statystyczne jakościowe,
  - b/ dwie powiązane ze sobą cechy statystyczne ilościowe,
  - c/ dwie powiązane ze sobą cechy statystyczne – ilościową i jakościową.
8. W tablicy, z zadania 7b:
- a/ wpisz umowne dane liczbowe,
  - b/ dla wybranych danych z tablicy, zaproponuj ich graficzny obraz.
9. Przedstaw propozycję trzech tablic analitycznych, dla których:
- a/ wpisz umowne dane liczbowe,
  - b/ przedstaw dane liczbowe w tablicy syntetycznej,
  - c/ zaproponuj odpowiedni wykres dla danych zawartych w tablicy syntetycznej.

## Rozdział III

# CHARAKTERYSTYKA ZBIOROWOŚCI STATYSTYCZNEJ

Grupowanie zebranego materiału, budowa szeregów statystycznych, tablic, wykresów, to czynności wykonywane po przeprowadzeniu badań. Głównym jednak celem każdego badania jest analiza uzyskanych wyników, sformułowanie wniosków dotyczących badanego zjawiska, a także wykrycie i uchwycenie występujących prawidłowości i związków.

W toku analizy zebranego materiału dokonuje się obliczeń różnych miar statystycznych, będących charakterystykami liczbowymi szeregu statystycznego. Jak już wcześniej wskazano, szereg statystyczny może przedstawiać warianty oraz liczbę jednostek cechy ilościowej lub jakościowej w każdym wariancie.

### 1. Opis rozkładu cechy jakościowej

Umieszczone w szeregu dane dotyczące cechy jakościowej przedstawiają strukturę badanej zbiorowości z punktu widzenia tej właśnie cechy, czyli przedstawiają jej rozkład. Aby dokonać analizy zbiorowości przedstawionej w takim szeregu, należy posłużyć się pewnymi charakterystykami liczbowymi, które umożliwiają określenie właściwości rozkładu.

**Cechy jakościowe** są niemierzalne i wyrażane na skali nominalnej. Dlatego liczbowa ich charakterystyka jest możliwa tylko poprzez przedstawienie liczebności (liczb) bezwzględnych i względnych.

**Liczby bezwzględne** (absolutne) uzyskuje się w wyniku zliczania poszczególnych jednostek lub w rezultacie sumowania cech mierzalnych, są one zwykle mianowane, wyrażane w określonych jednostkach miary: waga, długość, powierzchnia i inne (np. powierzchnia szkoły wynosi  $1500 \text{ m}^2$ , do klasy pierwszej przyjęto 90 uczniów, zasoby biblioteki szkolnej liczą 3000 woluminów). Posługiwanie się w analizie liczbami bezwzględnymi często nie jest wystarczające, ponieważ nie pozwala na porównywanie niejednorodnych zbiorowości.

W związku z tym oprócz liczb bezwzględnych wykorzystuje się w analizie liczby względne (stosunkowe), wyrażające stosunek dwóch wielkości w postaci ułamków, procentów, promili. Powszechnie stosowane są wskaźniki struktury i natężenia, rzadziej natomiast wskaźniki dynamiki.

## 1.1. Wskaźniki struktury i natężenia

**Wskaźniki struktury** przedstawiają stosunek wielkości poszczególnych części zbiorowości do wielkości całej zbiorowości. Wskaźnik struktury jest obliczany w rezultacie podzielenia liczebności danej części przez liczebność całej zbiorowości.

Wskaźniki struktury można obliczać jako<sup>16</sup>:

- wskaźniki ułamkowe (suma wskaźników dotyczących poszczególnych części zbiorowości jest równa 1),
- wskaźniki procentowe (suma wskaźników dotyczących poszczególnych części zbiorowości równa się 100% ),
- wskaźniki promilowe (suma wskaźników dotyczących poszczególnych części zbiorowości równa się 1000).

Jeśli przyjmiemy, że  $N_1, N_2, N_3 \dots N_x$ , to są liczbowe części zbiorowości, np. badana zbiorowość liczy  $N = 126$  rodziców, w tym:  $N_1 = 28$  legitymuje się wykształceniem wyższym,  $N_2 = 56$  – średnim,  $N_3 = 27$  – zawodowym,  $N_4 = 15$  – podstawowym, to wówczas otrzymamy następujące wzory na obliczanie wskaźników struktury:

---

<sup>16</sup> Z. Peuker, *Statystyka ...*, op. cit., s. 85-86.

### 1. Wskaźniki ułamkowe:

- dla części zbiorowości:  $\frac{N_1}{N}, \frac{N_2}{N}, \dots, \frac{N_x}{N}$
- dla całości zbiorowości:  $\frac{N_1}{N} + \frac{N_2}{N} + \dots + \frac{N_x}{N} = 1$

### 2. Wskaźniki procentowe:

- dla części zbiorowości:  $\frac{N_1}{N} \cdot 100, \frac{N_2}{N} \cdot 100, \dots, \frac{N_x}{N} \cdot 100$
- dla całości zbiorowości:  $\frac{N_1}{N} \cdot 100 + \frac{N_2}{N} \cdot 100 + \dots + \frac{N_x}{N} \cdot 100 = 100\%$

### 3. Wskaźniki promilowe:

- dla części zbiorowości:  $\frac{N_1}{N} \cdot 1000, \frac{N_2}{N} \cdot 1000, \dots, \frac{N_x}{N} \cdot 1000$
- dla całości zbiorowości:  $\frac{N_1}{N} \cdot 1000 + \frac{N_2}{N} \cdot 1000 + \dots + \frac{N_x}{N} \cdot 1000 = 1000$

Najczęściej stosowane są wskaźniki procentowe, jako najwygodniejsze. Rzadziej wykorzystywane są wskaźniki promilowe, głównie wówczas gdy inne wskaźniki, np. procentowe byłyby wyrażone bardzo małymi liczbami ułamkowymi, z wieloma cyframi po przecinku.

Wskaźniki struktury mają bardzo duże znaczenie praktyczne, ponieważ w sposób bardzo przejrzysty przedstawiają całą zbiorowość, wskazując na udział jej poszczególnych części w całości, co można przedstawić w tablicy, ale także na wykresie, np. kołowym lub słupkowym. Posługując się wskaźnikami struktury można dokonać porównania kilku zbiorowości przyjmując liczebności każdej zbiorowości za 100% (przy obliczaniu wskaźników procentowych), niezależnie od ich liczebności wyrażonych liczbami bezwzględными (Tablica 16).

Na podstawie liczb bezwzględnych zamieszczonych w Tablicy 16 trudno jest jednoznacznie stwierdzić, w której ze szkół struktura poziomu wykształcenia jest korzystniejsza. Z wartości liczb bezwzględnych wynika, że w szkole A zatrudnionych jest więcej niż w szkole B nauczycieli z ukończonymi studiami podyplomowymi, więcej jest także nauczycieli z wykształceniem wyższym. Natomiast z odpowiednich wskaźników procentowych wynika, że w obu szkołach procent nauczycieli posiadających ukończone studia podyplomowe jest identyczny, a w przypadku nauczycieli z wykształceniem wyższym w szkole B posiada go nawet większy procent (odsetek) nauczycieli niż w szkole A.

*Nauczyciele szkoły A i B według wykształcenia*

| Poziom wykształcenia nauczycieli | Nauczyciele w szkole |    |       |       |
|----------------------------------|----------------------|----|-------|-------|
|                                  | A                    | B  | A     | B     |
|                                  | Liczba               |    | %     |       |
| Studium Podyplomowe              | 8                    | 5  | 10,0  | 10,0  |
| Wyższe                           | 36                   | 24 | 45,0  | 48,0  |
| WSN                              | 22                   | 13 | 27,5  | 26,0  |
| SN                               | 10                   | 3  | 12,5  | 6,0   |
| Liceum Pedagogiczne              | 4                    | 5  | 5,0   | 10,0  |
| Ogółem                           | 80                   | 50 | 100,0 | 100,0 |

*Źródło: dane umowne*

Porównując wskaźniki struktury, np. procentowe obliczone dla podobnych kategorii w kilku zbiorowościach można wskazać, w której zbiorowości określona grupa jest stosunkowo liczniejsza, a ponadto można stwierdzić, że w którejś zbiorowości określona grupa jest np. kilkakrotnie mniejsza lub większa, w porównaniu z takimi grupami z innych zbiorowości, np. w szkole A pracuje dwukrotnie więcej nauczycieli legitymujących się wykształceniem na poziomie SN niż w szkole B.

**Wskaźniki natężenia** określają stosunek dwóch zbiorowości pozostających ze sobą w związku logicznym, wyrażają stopień natężenia:

$$W_n = \frac{N_1}{N_2},$$

gdzie:

- $N_1$  – liczebność jednej zbiorowości, np. liczba studentów określonego kierunku studiów,
- $N_2$  – liczebność drugiej zbiorowości, np. liczba nauczycieli akademickich zatrudnionych na tym kierunku studiów.

W badaniach pedagogicznych stosowane są takie wskaźniki natężenia, jak np. liczba uczniów przypadających na 1 m<sup>2</sup> powierzchni klasy szkolnej, na jednego nauczyciela, liczba woluminów w bibliotece szkolnej przypadająca na jednego ucznia, liczba studentów przypadająca na jednego profesora itp. Znaczenie wskaźników natężenia polega na tym, że pozwalają na porównywanie wielkości, które ujmowane w liczbach bezwzględnych nie pozwalają na sformułowanie pełniejszych wniosków.

## 1.2. Układ i analiza wskaźników struktury w tablicy

Spośród przedstawionych wskaźników struktury w badaniach pedagogicznych najczęściej stosowane są wskaźniki procentowe. Z praktyki wynika, że istnieją różne sposoby umieszczania w tablicach liczb bezwzględnych i odpowiadających im wskaźników procentowych. Zróżnicowanie to wynika głównie z celu i typu badań, a także z zamiarów badacza. Celem przedstawienia w tablicy liczb bezwzględnych i wskaźników procentowych może być podanie informacji o strukturze badanej zbiorowości, a także przedstawienie związków między badanymi zmiennymi i zbiorowościami.

Przykłady zamieszczania w tablicach liczb bezwzględnych i wskaźników procentowych przedstawiono w Tablicach 17, 18, 19, 20, 21.

*Tablica 17*

*Uczniowie kl. III w miejscowości X w roku szkolnym 1995/96  
według pochodzenia społecznego*

| Pochodzenie społeczne | Liczba | %     |
|-----------------------|--------|-------|
| Robotnicze            | 42     | 36,8  |
| Chłopskie             | 37     | 32,5  |
| Inteligenckie         | 28     | 24,6  |
| Inne                  | 7      | 6,1   |
| Ogółem                | 114    | 100,0 |

*Źródło: dane umowne*

*Tablica 18*

*Rodzaj rodziny uczniów kl. III w miejscowości X  
w roku szkolnym 1995/96, a ich wyniki w nauce*

| Rodzaj rodziny   | Wyniki w nauce |         |        |        |         |         |        |        |
|------------------|----------------|---------|--------|--------|---------|---------|--------|--------|
|                  | Liczba         |         |        |        | %       |         |        |        |
|                  | Wysokie        | Średnie | Niskie | Ogółem | Wysokie | Średnie | Niskie | Ogółem |
| Rodzina pełna    | 51             | 55      | 8      | 114    | 44,7    | 48,3    | 7,0    | 100,0  |
| Rodzina niepełna | 2              | 21      | 6      | 29     | 6,9     | 72,4    | 20,7   | 100,0  |
| Ogółem           | 53             | 76      | 14     | 143    | 37,1    | 53,1    | 9,8    | 100,0  |

*Źródło: dane umowne*

Tablica 19

*Rodzaj rodziny uczniów kl. III w miejscowości X w roku szkolnym 1995/96,  
a ich wyniki w nauce*

| Rodzaj rodziny   | Wyniki w nauce |      |         |      |        |      |        |       |
|------------------|----------------|------|---------|------|--------|------|--------|-------|
|                  | Wysokie        |      | Średnie |      | Niskie |      | Ogółem |       |
|                  | L              | %    | L       | %    | L      | %    | L      | %     |
| Rodzina pełna    | 51             | 44,7 | 55      | 48,3 | 8      | 7,0  | 114    | 100,0 |
| Rodzina niepełna | 2              | 6,9  | 21      | 72,4 | 6      | 20,7 | 29     | 100,0 |
| Ogółem           | 53             | 37,1 | 76      | 53,1 | 14     | 9,8  | 143    | 100,0 |

Źródło: dane umowne

Tablica 20

*Rodzaj rodziny uczniów kl. III w miejscowości X w roku szkolnym 1995/96,  
a ich wyniki w nauce*

| Rodzaj rodziny   |   | Wyniki w nauce |         |        |        |
|------------------|---|----------------|---------|--------|--------|
|                  |   | Wysokie        | Średnie | Niskie | Ogółem |
| Rodzina pełna    | L | 51             | 55      | 8      | 114    |
|                  | % | 96,2           | 72,4    | 57,1   | 79,7   |
| Rodzina niepełna | L | 2              | 21      | 6      | 29     |
|                  | % | 3,8            | 27,6    | 42,9   | 20,3   |
| Ogółem           | L | 53             | 76      | 14     | 143    |
|                  | % | 100,0          | 100,0   | 100,0  | 100,0  |

Źródło: dane umowne

Tablica 21

*Rodzaj rodziny uczniów kl. III w miejscowości X w roku szkolnym 1995/96,  
a ich wyniki w nauce*

| Rodzaj rodziny   | Wyniki w nauce |         |        |        |
|------------------|----------------|---------|--------|--------|
|                  | Wysokie        | Średnie | Niskie | Ogółem |
| Liczba           |                |         |        |        |
| Rodzina pełna    | 51             | 55      | 8      | 114    |
| Rodzina niepełna | 2              | 21      | 6      | 29     |
| Ogółem           | 53             | 76      | 14     | 143    |
| %                |                |         |        |        |
| Rodzina pełna    | 96,2           | 72,4    | 57,1   | 79,7   |
| Rodzina niepełna | 3,8            | 27,6    | 42,9   | 20,3   |
| Ogółem           | 100,0          | 100,0   | 100,0  | 100,0  |

Źródło: dane umowne

Wskazane jest, jeśli to możliwe, aby przyjęty sposób umieszczania w tablicy wynikowej wskaźników procentowych był konsekwentnie stosowany we wszystkich tablicach, w całym tekście opracowania.

Analiza danych uzyskanych w badaniach jest kolejną czynnością w postępowaniu badawczym. Polega na rozpatrywaniu danych uzyskanych w rezultacie badań z różnych punktów widzenia, w taki sposób, aby przedstawić badany problem lub zjawisko w sposób możliwie dokładny i wyczerpujący.

Zebrany materiał przedstawiony w tablicach wynikowych powinien być poddany na wstępie analizie jakościowej, a w dalszej kolejności może być poddany analizie ilościowej.

**Analiza jakościowa** polega na poszukiwaniu i uwydatnianiu prawidłowości, związków i zależności występujących w tablicy zawierającej dane empiryczne. Dokonując analizy danych z tablicy zawierającej dane liczbowe, jak i wskaźniki procentowe, wskazane jest, aby posługiwać się wskaźnikami procentowymi. Należy dokonać porównania między sobą poszczególnych wyników, jednocześnie poszukując pewnych prawidłowości, np.  $X$  jest dwukrotnie większy od  $Y$ , ale jednocześnie o połowę mniejszy niż  $Z$ ; określona cecha występuje trzykrotnie częściej wśród jednej kategorii badanych niż wśród drugiej; około 20% lub inaczej – tylko co piąty badany charakteryzuje się cechą  $X$ ; w określonej grupie badanych, np. 63,5% charakteryzującej się pewną cechą, występuje bardzo często inna cecha – prawdopodobnie zachodzi tu związek itp.

Ponadto dokonując analizy danych w tablicy w celu zaznaczenia występujących prawidłowości można używać zamiennie zamiast wskaźników procentowych np. określeń prawie co trzeci, czy co piąty badany,  $1/3$ ,  $2/3$ ,  $3/4$  badanych, ale jednocześnie nie jest pożądane stosowanie określeń – co jedenasty, co dwudziesty badany czy też  $5/6$  lub  $7/9$  badanych, ponieważ tego rodzaju sformułowania powodują, że dokonywana analiza zebranego materiału traci komunikatywność i przejrzystość.

Do stosunkowo często popełnianych błędów w analizie jakościowej zawartości tablicy należy dokonywanie jej opisu, polegającego na odtworzeniu w postaci zdań tego, co już w tablicy jest, np. spośród badanych należących do kategorii  $A$  – 23,5% charakteryzuje się cechą  $a_1$ , 48,4% cechą  $a_2$ , a 28,1% cechą  $a_3$ . Takie informacje już są w tablicy i nie należy ich powtarzać, natomiast pożądane jest przeprowadzenie ich analizy. W takim przypadku można by napisać, że spośród badanych należących do kategorii  $A$ , prawie połowa

charakteryzuje się jednocześnie cechą  $a_2$ , a tylko u co czwartego badanego występuje cecha  $a_1$  lub  $a_3$ .

Z praktyki wynika, iż w przypadku analizowania danych przedstawionych w stosunkowo dużej tablicy wynikowej, np. 4 lub więcej wierszy i 4 lub więcej kolumn, by nie zagubić istoty problemu w masie danych, można bardziej szczegółowo przeanalizować dane znajdujące się w kategoriach skrajnych, a tylko w uzasadnionych przypadkach, jeśli byłyby szczególnie znaczące, odwołać się do wyników uzyskanych w kategoriach pośrednich.

W celu pełniejszego przedstawienia wyników badań, podczas dokonywania analizy jakościowej danych z tablicy wskazane jest zamieszczenie wykresu obrazującego w postaci graficznej niektóre, bardziej interesujące wyniki badań zawarte w tablicy.

Po zakończeniu analizy jakościowej danych z tablicy należy zamieścić jedno lub kilka zdań będących ogólniejszym wnioskiem z przeprowadzonej analizy.

W celu zweryfikowania wniosku sformułowanego na podstawie analizy jakościowej można przystąpić do **analizy ilościowej** (statystycznej), polegającej na interpretacji wartości uzyskanych w obliczeniach. Wyniki obliczeń statystycznych przeprowadzonych dla danych zawartych w tablicy zamieszcza się w dodatkowym (ostatnim) wierszu w tablicy.

W Tablicy 22 zamieszczono dane liczbowe i wskaźniki procentowe, a niżej dokonano analizy jakościowej danych zawartych w tej tablicy.

Tablica 22

*Rodzaj rodziny uczniów kl. III w miejscowości X w roku szkolnym 1995/96,  
a ich wyniki w nauce*

| Rodzaj rodziny   | Wyniki w nauce |         |        |        |         |         |        |        |
|------------------|----------------|---------|--------|--------|---------|---------|--------|--------|
|                  | Liczba         |         |        |        | %       |         |        |        |
|                  | Wysokie        | Średnie | Niskie | Ogółem | Wysokie | Średnie | Niskie | Ogółem |
| Rodzina pełna    | 51             | 55      | 8      | 114    | 44,7    | 48,3    | 7,0    | 100,0  |
| Rodzina niepełna | 2              | 21      | 6      | 29     | 6,9     | 72,4    | 20,7   | 100,0  |
| Ogółem           | 53             | 76      | 14     | 143    | 37,1    | 53,1    | 9,8    | 100,0  |

Źródło: dane umowne

Z analizy danych zawartych w Tablicy 22 wynika, że spośród dzieci pochodzących z rodzin pełnych prawie co drugie uzyskało w nauce wyniki wysokie lub średnie, a tylko 7% spośród nich rezultaty niskie. Natomiast spośród dzieci pochodzących z rodzin niepełnych prawie 3/4 uzyskało wyniki średnie, 20,7% – niskie i tylko 6,9% osiągnęło wysokie rezultaty.

Z porównania osiągnięć w nauce uzyskanych przez obie wyróżnione kategorie badanych wynika, że dzieci pochodzące z rodzin pełnych uzyskały sześciokrotnie więcej rezultatów wysokich i jednocześnie trzykrotnie mniej wyników niskich niż dzieci pochodzące z rodzin niepełnych.

W rezultacie przeprowadzonej analizy sformułowano następujący wniosek: rodzaj rodziny, w której wychowują się badani uczniowie, ma związek z wysokością ich wyników w nauce, przy czym rodzina pełna stwarza zdecydowanie korzystniejsze warunki do uzyskiwania wyższych rezultatów.

Nie zawsze jednak układ wskaźników procentowych jest tak jednoznaczny jak w Tablicy 22. Z praktyki wynika, że czasem wskaźniki procentowe przedstawiające rozkład jednej interesującej badacza cechy układają się w bardzo podobny sposób we wszystkich wyróżnionych wariantach drugiej cechy (Tablica 23). W takim przypadku, aby uniknąć monotonnej analizy można po wstępnym przeanalizowaniu ogólnych rezultatów badań przystąpić do kompleksowej, zbiorczej analizy wskaźników procentowych podkreślając, że procentowy rozkład jednej cechy w poszczególnych wariantach innej cechy jest podobny i można tu posłużyć się odpowiednimi przykładami. Analizę należy zakończyć formułując jednoznaczny wniosek. Niżej zamieszczono przykładowy sposób dokonania takiej analizy dla danych z Tablicy 23.

Tablica 23

*Sytuacja materialna rodziny a rodzaj wybranej szkoły ponadpodstawowej  
przez uczniów kl. VIII w miejscowości X w roku szkolnym 1997/98*

| Sytuacja materialna rodziny | Rodzaj wybranej szkoły ponadpodstawowej |              |    |        |      |              |      |        |
|-----------------------------|-----------------------------------------|--------------|----|--------|------|--------------|------|--------|
|                             | Liczba                                  |              |    |        | %    |              |      |        |
|                             | ZSZ                                     | Technikum LZ | LO | Ogółem | ZSZ  | Technikum LZ | LO   | Ogółem |
| bardzo dobra                | 10                                      | 25           | 26 | 61     | 16,4 | 41,0         | 42,6 | 100,0  |
| dobra                       | 6                                       | 9            | 10 | 25     | 24,0 | 36,0         | 40,0 | 100,0  |
| zła                         | 5                                       | 9            | 8  | 22     | 22,7 | 40,9         | 36,4 | 100,0  |
| Ogółem                      | 21                                      | 43           | 44 | 108    | 19,5 | 39,8         | 40,7 | 100,0  |

Źródło: dane umowne

Z analizy rezultatów badań przedstawionych w Tablicy 23 wynika, że około 80% badanych uczniów jako dalsze drogi kształcenia wybiera pełne szkoły średnie. Połowa spośród nich to kandydaci do liceum ogólnokształcącego lub liceów zawodowych albo techników. Jednocześnie co piąty badany zamierza rozpocząć dalszą naukę w zasadniczej szkole zawodowej.

Interesujących spostrzeżeń dostarcza analiza planów kształceniowych badanej młodzieży z punktu widzenia sytuacji materialnej ich rodzin. Okazuje się, że ta zmienna nie jest czynnikiem różnicującym rodzaj wybranej szkoły ponadpodstawowej przez badanych. Odsetki uczniów wskazujących liceum ogólnokształcące jako dalszą drogę kształcenia są zbliżone w grupach uczniów pochodzących z rodzin o bardzo dobrej (42,6%), dobrej (40,0%), i złej sytuacji materialnej (36,4%). Podobnie, w przypadku młodzieży planującej rozpoczęcie dalszej nauki w technikach lub liceach zawodowych. Niewielkie różnice obserwuje się tylko w grupie badanych zamierzających podjąć naukę w zasadniczych szkołach zawodowych: stosunkowo najmniejszy odsetek badanych podejmujących taką decyzję pochodzi z rodzin o bardzo dobrej sytuacji materialnej – 16,4%, podczas gdy z rodzin o dobrej i złej sytuacji ekonomicznej szkołę zawodową wybiera odpowiednio 24% i 22,7% badanych.

W związku z tym z przeprowadzonej analizy danych zawartych w Tablicy 23 wynika, że nie zachodzi związek między sytuacją materialną rodzin badanych uczniów kl. VIII i rodzajem wybranej przez nich szkoły ponadpodstawowej.

Nieco odmiennego sposobu analizy jakościowej wymagają wyniki badań przedstawione w tablicy zawierającej trzy cechy (Tablica 24). Ze względu na stosunkowo dużą liczbę informacji zawartych w takiej tablicy, analizy wyników badań można dokonać rozpoczynając od porównania częstości występowania (wyrażonych w %) poszczególnych przyczyn lęku przed nauczycielami wymienianych przez uczniów ze szkół miejskich i wiejskich, w dalszej kolejności poddając analizie i porównując powody lęku przed nauczycielami wskazywane przez dziewczęta i chłopców, kończąc sformułowaniem wniosku ogólnego i dokonaniem interpretacji wyników.

Tablica 24

*Przyczyny lęku przed nauczycielami w opinii uczniów kl. V  
ze szkół miejskich i wiejskich (w %)*

| Przyczyny lęku przed nauczycielami | Miasto     |         | Wieś       |         |
|------------------------------------|------------|---------|------------|---------|
|                                    | Dziewczęta | Chłopcy | Dziewczęta | Chłopcy |
| Brak wyrozumiałości                | 25,0       | 29,0    | 30,0       | 25,0    |
| Bezwzględność, surowość            | 20,0       | 21,0    | 20,0       | 23,0    |
| Niesprawiedliwe ocenianie          | 17,0       | 15,0    | 16,0       | 12,0    |
| Zastraszanie                       | 15,0       | 10,0    | 12,0       | 6,0     |
| Zmienne wymagania                  | 13,0       | 12,0    | 12,0       | 16,0    |
| Ośmieszanie                        | 10,0       | 13,0    | 10,0       | 18,0    |
| Ogółem                             | 100,0      | 100,0   | 100,0      | 100,0   |

Źródło: dane umowne

Z przeprowadzonych badań wynika (Tablica 24), że uczniowie kl. V odczuwający lęk przed nauczycielami jako przyczynę najczęściej wymieniają brak wyrozumiałości, na co wskazuje ok. 30% badanych ze szkół miejskich i wiejskich. Bezwzględność i surowość przejawiana w zachowaniach nauczycieli, to kolejny powód lęku występujący u co piątego ucznia, przy czym nieco częściej wskazywany jest przez badanych ze szkół usytuowanych w środowisku wiejskim. Dla około 15% uczniów (nieco częściej ze szkół miejskich) przyczyną lęku przed nauczycielami jest niesprawiedliwe ocenianie. Zastraszenie jako przyczynę lęku przed nauczycielami wymienia ponad 10% badanych, przy czym prawie dwukrotnie częściej dotyczy to uczniów ze szkół miejskich. Ponad 10% uczniów wyraża opinię, że zmienność wymagań stawianych przez nauczycieli oraz ośmieszanie uczniów to najczęstsze powody lęku przed nauczycielami.

Z analizy danych zawartych w Tablicy 24 wynika, że wśród przyczyn lęku przed nauczycielami wymienianymi przez chłopców (niezależnie od usytuowania szkoły) dominuje brak wyrozumiałości, bezwzględność, surowość oraz ośmieszanie przed klasą. Podobne zachowania nauczycieli powodujące lęk wskazują dziewczęta, przy czym częściej niż chłopcy wymieniają niesprawiedliwe ocenianie i zastraszanie.

W rezultacie dokonanej analizy należy wskazać, że rodzaj środowiska, w którym usytuowana jest szkoła oraz płeć badanych uczniów kl. V nie różnicuje w stopniu znaczącym ich opinii dotyczących przyczyn lęku przed nauczycielami.

Po przeprowadzeniu analizy jakościowej i ilościowej danych z tablicy, badacz powinien podjąć próbę wyjaśnienia uzyskanych wyników, związków i prawidłowości, czyli powinien podjąć próbę racjonalnego i wyczerpującego objaśnienia, dlaczego uzyskano właśnie takie wyniki.

W celu zwiększenia rzetelności i kompletności przeprowadzonych analiz wnioski ogólniejsze wynikające z analizy danych w tablicy (potwierdzone już analizą statystyczną) należy porównać (jeśli to możliwe) z wynikami podobnych badań, przeprowadzonymi przez innych badaczy, na innych próbach, na innym terenie. Odwołanie się do literatury przedmiotu badań jest zadaniem bardzo ważnym, ponieważ badacz nie tylko ma możliwość przedstawienia swoich wyników badań na tle innych, ale także prezentuje znajomość literatury problemu.

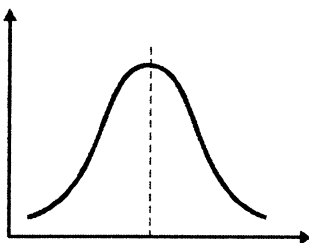
Przedstawione wyżej propozycje prezentowania danych liczbowych i procentowych w tablicy, dokonywania ich analizy jakościowej i ilościowej są w opinii autora skryptu jednymi z możliwych i nie pretendują do uznawania ich za powszechnie sprawdzające się i zasadne.

## 2. Opis rozkładu cechy ilościowej

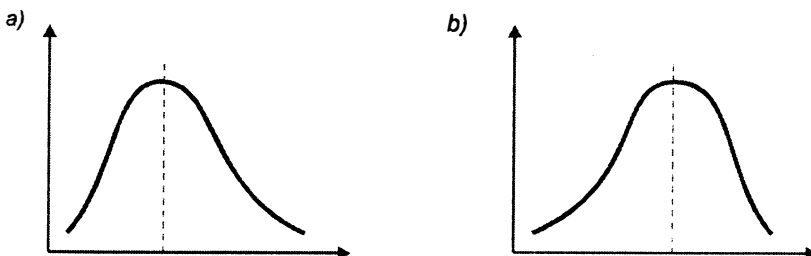
Umieszczone w szeregu statystycznym wartości, będące przejawem występowania cechy ilościowej, przedstawiają badaną zbiorowość z punktu widzenia tej cechy, czyli przedstawiają jej rozkład.

**Cechy ilościowe** są mierzalne, można dokonywać ich pomiaru na skalach dających większe możliwości działań matematycznych. W związku z tym wykorzystuje się wiele miar opisujących ich rozkład.

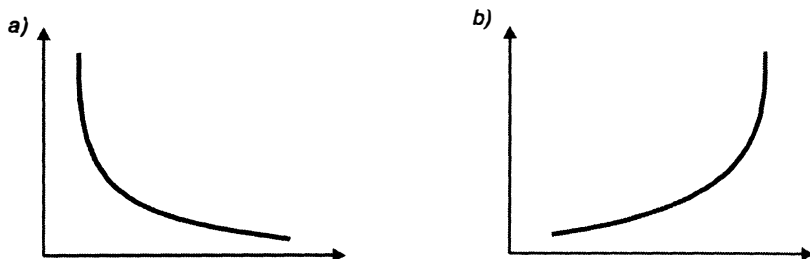
Obliczanie poszczególnych miar uzależnione jest od formy uzyskanego rozkładu, czego najbardziej obrazowym przykładem jest wykres i dlatego na wstępie przedstawiono typowe formy rozkładów. Do najczęściej występujących form rozkładu należą: symetryczny, asymetryczny, jedno- lub wielowierzchołkowe, a także spłaszczone i wypukłe.



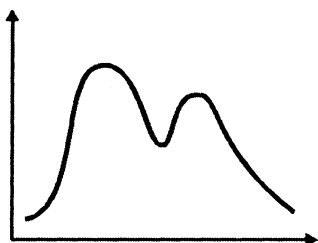
Rys. 1. Rozkład symetryczny



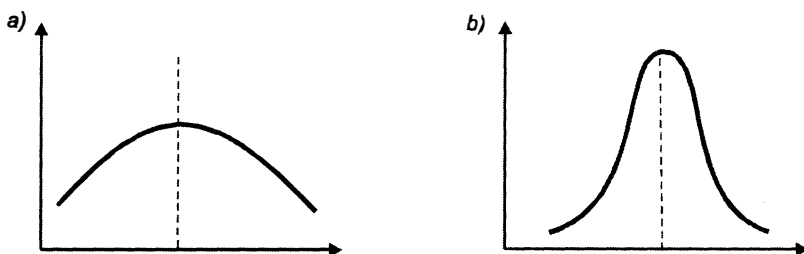
Rys. 2. Rozkłady asymetryczne: a) asymetria prawostronna, b) asymetria lewostronna



Rys. 3. Rozkłady skrajnie asymetryczne



Rys. 4. Rozkład dwuwierchołkowy



Rys. 5. Rozkłady: a) spłaszczony, b) wypukły

Rozkład symetryczny (Rys. 1) charakteryzuje się tym, że największe i najmniejsze wartości występują rzadko, częściej występują wartości średnie, wierzchołek znajduje się w środku rozkładu. Jego wykres charakteryzuje się tym, że jest on symetryczny względem osi dzielącej go na dwie części.

W rozkładzie asymetrycznym następuje „koncentracja” wartości jednego lub drugiego końca rozkładu. Wykres może przybierać dwa różne kształty: prawe ramię wykresu wydłużone – asymetria prawostronna (Rys. 2a) lub lewe ramię wykresu wydłużone – asymetria lewostronna (Rys. 2b).

Szczególnym przypadkiem rozkładu asymetrycznego jest rozkład, tzw. skrajnie asymetryczny (Rys. 3a i 3b), w którym wartości pomiarowe tylko jednego końca rozkładu występują najczęściej.

Rozkład jedno-, dwu- lub wielowierzchołkowy (Rys. 4) charakteryzuje się tym, że niektóre wartości, czasem jedna z nich lub wartości, skrajne występują podobną ilość razy, co powoduje, że wykres ma jeden, dwa lub kilka maksimów (na wykresie są to wierzchołki).

Rozkład spłaszczony jest rezultatem mniejszego skupienia wartości pomiarowych, czyli większego rozproszenia wyników (Rys. 5a). Przeciwnieństwem jest rozkład wypukły, o intensywnym skupieniu zbliżonych do siebie wartości (Rys. 5b).

Znając formę uzyskanego rozkładu można przystąpić do bardziej dokładnego przedstawienia badanej cechy ilościowej przez obliczenie odpowiednich miar charakteryzujących jej rozkład. Należą do nich:

- miary przeciętne (miary tendencji centralnej, miary położenia, średnie),
- miary rozproszenia,
- miary asymetrii.

Wymienione miary nazywane są parametrami opisowymi, a stosowanie konkretnych parametrów uzależnione jest od typu rozkładu zmiennej.

## 2.1. Miary przeciętne

Dokładna charakterystyka cechy ilościowej wymaga przedstawienia jej przeciętnego poziomu, np. przeciętny wzrost, przeciętna waga itp. Wśród miar przeciętnych wyróżnia się:

- **średnie klasyczne**, przy obliczaniu których uwzględnia się wszystkie wartości szeregu: średnia arytmetyczna, harmoniczna, geometryczna, kwadratowa;
- **średnie pozycyjne**, będące konkretnymi wartościami zajmującymi w danym szeregu określone miejsce (pozycję): mediana, kwartyle, dominanta.

Średnie klasyczne i pozycyjne są obliczane głównie dla szeregów o rozkładzie jednomodalnym (jednowierzchołkowym), o jednorodnej strukturze. Spośród średnich klasycznych do najczęściej stosowanych w praktyce należą średnia arytmetyczna i harmoniczna, i te miary zostaną bliżej scharakteryzowane.

### 2.1.1. Średnia arytmetyczna zwykła i ważona

**Średnia arytmetyczna**, nazywana wartością przeciętną, jest sumą wszystkich wartości zmiennej, podzieloną przez liczebność zbiorowości. Można to wyrazić ogólnym wzorem:<sup>17</sup>

$$\bar{x} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N} = \frac{\sum X}{N},$$

gdzie:

$X_1, X_2, X_3$  – poszczególne wartości zmiennej,

$\sum$  (sigma) – duża litera alfabetu greckiego, jest znakiem sumowania,

$N$  – liczebność zbiorowości.

Wzór ten jest właściwy dla średniej arytmetycznej zwykłej (prostej), takiej, którą oblicza się dla wartości przedstawionych w szeregu szczegółowym, tzn. wówczas gdy wartości zmiennej występują pojedynczo. Wartość obliczonej w ten sposób średniej nie musi być równa którejkolwiek wartości z szeregu. Wynika z tego, że średnia arytmetyczna jest wartością abstrakcyjną i zaciera występujące różnice pomiędzy poszczególnymi wartościami zmiennej.

**Średnia arytmetyczna ważona** nie różni się pod względem treści od średniej arytmetycznej zwykłej, ale obliczana jest wówczas, gdy wartości zmiennej ujęte są w szeregu rozdzielczym, np. ze zmienną skokową, lub w szeregu rozdzielczym zawierającym przedziały klasowe. Nazwa „średnia arytmetyczna ważona” jest rezultatem tego, że przy jej obliczaniu należy uwzględnić tzw. wagę, czyli liczebność każdej grupy (klasy) szeregu.

**Obliczanie średniej arytmetycznej ważonej z szeregu rozdzielczego zawierającą zmienną skokową** wymaga zastosowania następującego wzoru:

$$\bar{x} = \frac{X_1 f_1 + X_2 f_2 + \dots + X_n f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum Xf}{N},$$

gdzie:

$X_1, X_2, \dots, X_n$  – kolejne wartości zmiennej,

$f_1, f_2, \dots, f_n$  – liczebności kolejnych wartości zmiennych,

$N$  – liczebność zbiorowości.

---

<sup>17</sup> Zastosowano tu uproszczoną postać wzoru, bez symboli nad, pod i obok znaku sumy (sigma). W celu uproszczenia symbole te będą odtąd również pomijane.

Przykład obliczania średniej arytmetycznej ważonej z szeregu rozdzielczego ze zmienną skokową zamieszczono w Tablicy 25 zawierającej szereg przedstawiający liczbę książek zebranych przez uczniów kl. V.

Tablica 25

*Uczniowie kl. V według liczby zebranych książek*

| Liczba książek<br>$X$ | Liczba uczniów<br>$f$ | $Xf$              |
|-----------------------|-----------------------|-------------------|
| 1 ( $X_1$ )           | 5 ( $f_1$ )           | 5 ( $X_1 f_1$ )   |
| 2 ( $X_2$ )           | 20 ( $f_2$ )          | 40 ( $X_2 f_2$ )  |
| 3 ( $X_3$ )           | 37 ( $f_3$ )          | 111 ( $X_3 f_3$ ) |
| 4 ( $X_4$ )           | 9 ( $f_4$ )           | 36 ( $X_4 f_4$ )  |
| 5 ( $X_5$ )           | 8 ( $f_5$ )           | 40 ( $X_5 f_5$ )  |
|                       | $N=79$                | $\Sigma Xf=232$   |

Źródło: dane umowne

Po podstawieniu do wzoru uzyskano:

$$\bar{x} = \frac{\Sigma Xf}{N} = \frac{232}{79} = 2,94$$

Z przedstawionych obliczeń wynika, że jeden uczeń zebrał podczas zbiórki przeciętnie, średnio 3 egzemplarze książek.

**Obliczanie średniej arytmetycznej ważonej z szeregu rozdzielczego zawierającego przedziały klasowe** wymaga zastosowania następującego wzoru:

$$\bar{x} = \frac{x_{i1}f_1 + x_{i2}f_2 + \dots + x_{in}f_n}{f_1 + f_2 + \dots + f_n} = \frac{\Sigma x_i f}{N},$$

gdzie:

- $x_{i1}, x_{i2}$  – środki kolejnych przedziałów klasowych,
- $f_1, f_2$  – liczebność kolejnych przedziałów klasowych,
- $N$  – liczebność zbiorowości.

Przykład obliczania średniej arytmetycznej ważonej z szeregu rozdzielczego zawierającego przedziały klasowe zamieszczono w Tablicy 26 zawierającej szereg przedstawiający wiek studentów uczestniczących w zajęciach sportowych.

Studenty – uczestnicy zajęć sportowych według wieku

| Wiek studentów<br>$X$ | Liczebność<br>$f$ | Środki przedziałów<br>$x_i$ | $x_i f$               |
|-----------------------|-------------------|-----------------------------|-----------------------|
| 18 - 20               | 30 ( $f_1$ )      | 19 ( $x_{i1}$ )             | 570                   |
| 21 - 23               | 19 ( $f_2$ )      | 22 ( $x_{i2}$ )             | 418                   |
| 24 - 26               | 15 ( $f_3$ )      | 25 ( $x_{i3}$ )             | 375                   |
| 27 - 29               | 10 ( $f_4$ )      | 28 ( $x_{i4}$ )             | 280                   |
|                       | $N=74$            |                             | $\Sigma x_i f = 1643$ |

Źródło: dane umowne

Po podstawieniu do wzoru uzyskano:

$$\bar{x} = \frac{\sum x_i f}{N} = \frac{1643}{74} = 22,2$$

Z przedstawionych obliczeń wynika, że średni wiek studenta uczestniczącego w zajęciach sportowych wynosił 22,2 lat.

**Średnia arytmetyczna** jest jedną z najwygodniejszych i najczęściej stosowanych miar, jednakże przy jej obliczaniu pamiętać należy o **jej następujących właściwościach**:

- obliczania średniej arytmetycznej dokonuje się uwzględniając wszystkie wartości zmiennej znajdujące się w szeregu,
- średnia arytmetyczna jest wartością mianowaną, tzn. wyrażaną w konkretnych jednostkach miar (np. lata, sztuki, kilogramy),
- średnia arytmetyczna nie może być wartością mniejszą niż najmniejsza wartość zmiennej, a także nie może być wartością większą niż największa wartość zmiennej,
- duży wpływ na wartość średniej arytmetycznej mają wartości skrajne i im większa rozpiętość wyrazów skrajnych, tym średnia jest mniej dokładna i czasem jej obliczanie jest bezsensowne,
- suma odchyleń poszczególnych wartości od średniej arytmetycznej wynosi zawsze zero.

### Nie należy obliczać średniej arytmetycznej:

- w szeregach rozdzielczych o otwartych przedziałach klasowych, np. poniżej 5 lat, powyżej 20 m<sup>2</sup> i więcej,
- gdy rozkład wyników jest dwu- lub wielowierzchołkowy,
- w szeregach zawierających wartości niejednorodne, np. obliczanie przeciętnej płacy miesięcznej dla nauczycieli akademickich, pracowników administracyjnych, sprzątaczek itp.,
- gdy rozkład jest wyraźnie asymetryczny, a próba bardzo mała.

### 2.1.2. Średnia harmoniczna

**Średnia harmoniczna** jest odwrotnością średniej arytmetycznej i stosowana jest wówczas gdy poszczególne wartości charakteryzujące zbiorowość podane są w formie odwrotności, kiedy należy wyliczyć średnią z wielkości logicznie nie nadających się do dodawania. Z powodzeniem jest stosowana w takich rozkładach, w których poszczególne wartości wyrażają miary sprawności w wykonywaniu pewnego zadania, przy nie limitowanym czasie jego wykonania. Na przykład w eksperymentach psychologicznych z zakresu psychologii uczenia się, kiedy badane osoby otrzymują polecenie wyuczenia się na pamięć określonego tekstu przez wielokrotne jego powtarzanie, określenie przeciętnego czasu uczenia się lub przeciętnej liczby powtórzeń wymaga zastosowania średniej harmonicznej<sup>18</sup>.

**Dla szeregu szczegółowego** średnia harmoniczna wyrażona jest następującym wzorem:

$$H = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}},$$

gdzie:

- $X_1, X_2$  – np. czas wykonania jakiegoś zadania, liczba powtórzeń przy wykonywaniu jakiejś czynności,  
 $N$  – liczebność zbiorowości.

Jeśli w grupie 5 uczniów ( $N=5$ ) dokonano pomiaru czasu wykonania określonego zadania z matematyki i uzyskano następujące wyniki w minutach: 10, 14, 11, 20, 15, to średni czas wyniesie:

---

<sup>18</sup> G. Clauss, H. Ebner, *Podstawy ...*, op. cit., s. 89.

$$H = \frac{5}{\frac{1}{10} + \frac{1}{14} + \frac{1}{11} + \frac{1}{20} + \frac{1}{15}} = \frac{5}{0,38} = 13,2$$

Oznacza to, że średni czas wykonania tego zadania dla badanej grupy uczniów wyniósł 13,2 minuty.

### 2.1.3. Mediana

**Mediana** jest to wartość środkowa dzieląca zbiorowość na dwie równe części: część dolną obejmującą wartości mniejsze lub równe medianie i część górną zawierającą wartości większe lub równe medianie. Wstępnym warunkiem dla wyznaczania mediany jest uporządkowanie szeregu według wartości wzrastających lub malejących.

**W szeregu o nieparzystej liczbie wartości** zmiennej, medianę stanowi będzie wartość zajmująca centralne (środkowe) miejsce w szeregu, czyli jest to wartość, której numer kolejny jest połową liczby obserwacji powiększoną o 1:

$$M_e \text{ (pozycja mediany)} = X_{\frac{N+1}{2}}$$

W szeregu liczącym np.  $N = 9$  obserwacji – wzrost uczniów w cm: 112, 114, 115, 117, 120, 122, 123, 123, 129 medianą jest wartość zajmująca piąte miejsce, czyli wartość 120.

**W szeregu o parzystej liczbie wartości** medianą jest średnia arytmetyczna dwóch wartości leżących w środku:

$$M_e \text{ (pozycja mediany)} = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2}$$

W szeregu liczącym np.  $N = 10$  obserwacji (waga uczniów w kg): 30, 32, 32, 33, 35, 36, 36, 37, 39, 39 medianą jest średnia dwóch wartości leżących w środku, tzn. wartości zajmującej pozycję 5 i 6, czyli  $(35+36):2=35,5$  kg. Jest to więc wartość, która dzieli szereg na dwie części.

Bardziej złożone jest **wyznaczanie mediany z szeregu rozdzielczego**, np. ze zmienną skokową, a także z szeregu zawierającego przedziały klasowe. Polega ono na:

- ustaleniu pozycji, którą zajmuje mediana,
- ustaleniu klasy, w której znajduje się mediana,
- sporządzeniu szeregu skumulowanego,
- wyznaczenie mediany.

Sposób wyznaczania mediany w szeregu nieparzystym z cechą skokową przedstawiono w Tablicy 27 ( $N=79$ ).

Tablica 27

*Uczniowie kl. V według liczby zebranych książek*

| Liczba książek<br>$X$ | Liczba uczniów<br>$f$ | Szereg skumulowany<br>$f_c$ |
|-----------------------|-----------------------|-----------------------------|
| 1                     | 5                     | 5                           |
| 2                     | 20                    | 25                          |
| 3                     | 37                    | 62                          |
| 4                     | 9                     | 71                          |
| 5                     | 8                     | 79                          |
|                       | $N=79$                |                             |

Źródło: dane umowne

Mediana zajmuje w tym szeregu pozycję  $\frac{N+1}{2}$ , czyli jest to wartość zajmująca 40 pozycję. Aby ustalić klasę, w której znajduje się mediana (w tym przypadku  $i=1$ ), należy utworzyć szereg skumulowany ( $f_c$ ), z którego wynika, w której klasie znajduje się mediana – w podanym przykładzie jest to klasa trzecia, a mediana wynosi 3, czyli jest to wartość, która dzieli szereg na dwie części.

Podobnie postępuje się w przypadku wyznaczania mediany w szeregu parzystym z cechą skokową, przy czym pozycję mediany wyznacza się odpowiednio tak, jak w szeregu parzystym.

Medianę dla szeregu zawierającego przedziały klasowe można wyznaczyć z następującego wzoru:

$$M_e = x_0 + \frac{i}{f_a} \left( \frac{N}{2} - f_b \right),$$

gdzie:

- $x_0$  – dolna granica przedziału, w którym znajduje się mediana,
- $i$  – wielkość przedziału klasowego,
- $f_a$  – liczebność przedziału, w którym znajduje się mediana,
- $f_b$  – liczebność skumulowana w przedziale poprzedzającym przedział mediany.

W Tablicy 28 przedstawiono wyznaczanie mediany dla wieku,  $N = 78$  uczniów uczęszczających na zajęcia z rytmiki.

Tablica 28

*Uczestnicy zajęć z rytmiki według wieku*

| Wiek<br>$X$ | Liczebność<br>$f$ | Szereg skumulowany<br>$f_c$ |
|-------------|-------------------|-----------------------------|
| 7 – 8       | 8                 | 8                           |
| 9 – 10      | 23                | 31                          |
| 11 – 12     | 35                | 66                          |
| 13 – 14     | 12                | 78                          |
|             | $N=78$            |                             |

Źródło: dane umowne

Podobnie, jak w szeregu rozdzielczym ze zmienną skokową, zbudowano szereg skumulowany ( $f_c$ ). Następnie obliczono połowę szeregu rozdzielczego uczniów:  $\frac{N}{2}$ , ponieważ klasą mediany jest przedział, w którym znajduje się wyraz  $\frac{N}{2}$  (dla szeregów parzystych i nieparzystych). W podanym przykładzie jest to obserwacja 39. Z szeregu skumulowanego odczytano, w którym przedziale klasowym znajduje się pozycja 39 i dolną granicę tego przedziału przyjęto jako  $x_o = 11$ .

Podstawiając do wzoru uzyskano:

$$M_e = 11 + \frac{2}{35} \left( \frac{78}{2} - 31 \right) = 11 + \frac{16}{35} \approx 11,5$$

Mediana wieku wynosi 11,5 i ta wartość dzieli szereg na dwie części.

**Właściwości mediany:**

- można ją wyznaczać tylko w szeregu uporządkowanym, szczególnie przy małym  $N$ ,
- nie jest zależna od wartości skrajnych szeregu,
- może być wyznaczana z szeregów rozdzielczych otwartych, także dla liczebności względnych,
- może być wyznaczana na wykresie,
- zaleca się jej wyznaczanie w rozkładzie wykazującym dużą asymetrię.

### 2.1.4. Dominanta

**Dominanta** (modalna, moda, wartość dominująca) to wartość, która w danym rozkładzie występuje najczęściej, jest uważana za typową. Dominanta może być wyznaczana także w stosunku do zmiennych jakościowych.

W szeregu liczącym 10 wartości ( $N=10$ ): 3, 4, 5, 6, 6, 6, 6, 7, 8, 8 najczęściej występuje wartość 6 (cztery razy) i jest to dominanta tego rozkładu. Jeżeli w danym rozkładzie dwie lub więcej sąsiadujących ze sobą wartości występuje taką samą liczbę razy, to za dominantę należy uważać średnią arytmetyczną tych dwu lub więcej wartości.

W następującym szeregu: 10, 11, 13, 14, 15, 15, 15, 16, 16, 16, 17, 19, 21 ( $N=13$ ) wartości 15 i 16 występują jednakową liczbę razy (każda po trzy razy). Jako dominantę tego rozkładu należy przyjąć średnią arytmetyczną tych dwóch wartości, czyli  $(15+16):2=15,5$ .

Jeśli natomiast w jakimś rozkładzie dwie lub więcej wartości występują jednakową liczbę razy, ale nie są to wartości sąsiadujące ze sobą bezpośrednio, wówczas te wartości należy przyjąć jako dominanty, a taki rozkład określa się jako dwu- lub wielowierzchołkowy (dwu- lub wielomodalny).

W szeregu rozdzielczym zawierającym zmienną skokową jako dominantę wyznacza się wartość, która występuje największą liczbę razy, np. w Tablicy 25 przedstawiającej uczniów według liczby zebranych książek jako dominantę należy przyjąć wartość 3 (wystąpiła trzydzieści siedem razy).

Nieco inaczej ustala się **dominantę w szeregu rozdzielczym**, zawierającym przedziały klasowe. **Gdy przedziały są równe**, ma tu zastosowanie następujący wzór:

$$D = X_o + i \frac{(f_a - f_b)}{(f_a - f_b) + (f_a - f_d)},$$

gdzie:

$X_o$  – dolna granica przedziału, w którym znajduje się dominanta,

$i$  – wielkość przedziału,

$f_a$  – liczebność przedziału, w którym znajduje się dominanta,

$f_b$  – liczebność przedziału poprzedzającego przedział dominanty,

$f_d$  – liczebność przedziału następującego po przedziale dominanty.

Jeśli nie występuje przedział poprzedzający dominantę, to  $f_b = 0$  (zero). Jeśli zaś nie występuje przedział następujący po przedziale dominanty, wówczas  $f_d = 0$  (zero).

W Tablicy 29 przedstawiono obliczanie dominanty z szeregu rozdzielczego zawierającego przedziały klasowe.

Tablica 29

*Uczestnicy zajęć wyrównawczych według wieku*

| Wiek<br>$X$ | Liczebność<br>$f$ |
|-------------|-------------------|
| 7 – 8       | 4                 |
| 9 – 10      | 27                |
| 11 – 12     | 15                |
| 13 – 14     | 9                 |
|             | $N=55$            |

Źródło: dane umowne

Dominanta znajduje się w przedziale klasowym 9-10, ponieważ występuje tu największa liczebność ( $f_a$ ), a w związku z tym  $X_o=9$ . Przedział dominanty poprzedza przedział o liczebności  $f_b=4$ . Po podstawieniu do wzoru otrzymano:

$$D = 9 + 2 \frac{(27-4)}{(27-4)+(27-15)} = 9 + 2 \frac{23}{23+12} \approx 10,31$$

Dominanta wynosi 10,3 lat, co oznacza, że w zajęciach wyrównawczych dominują, tzn. najczęściej uczestniczą dzieci w wieku 10,3 lat.

**Właściwości dominanty:**

- nie musi być wyliczana, może być wyznaczana z szeregu, tablicy, a także z wykresu,
- można ją wyznaczyć z szeregu z otwartymi przedziałami, także dla liczebności względnych,
- nie jest żadną funkcją algebraiczną w różnych rozkładach wartości zmiennej,
- może przyjmować różną wielkość, w zależności od rozpiętości klas tworzonych podczas grupowania i to jest jej wadą,
- nadaje się najbardziej do charakteryzowania rozkładów wielowierchołkowych.

## 2.2. Miary rozproszenia

Przedstawione dotychczas średnie nie charakteryzują precyzyjnie uzyskanego rozkładu, a czasem nawet powodują zacieranie jego cech istotnych. W trzech szeregach szczegółowych, w trzech zbiorowościach ustalono średnią arytmetyczną i medianę:<sup>19</sup>

$$X_1: 10, 12, 13 \quad \bar{x} = 12 \quad M_e = 12$$

$$X_2: 10, 12, 14 \quad \bar{x} = 12 \quad M_e = 12$$

$$X_3: 6, 12, 18 \quad \bar{x} = 12 \quad M_e = 12$$

Średnia arytmetyczna i mediana w każdym szeregu wynosi 12, ale szeregi te nie są jednakowe. W pierwszym szeregu wartości są bardzo skoncentrowane wokół średniej arytmetycznej, w drugim rozpiętość (odległość) pomiędzy poszczególnymi wartościami i średnią arytmetyczną jest większa, a w trzecim szeregu wartości są najbardziej odległe od średniej arytmetycznej. Pomimo jednakowej średniej arytmetycznej i mediany szeregi te różnią się głównie wielkością różnic pomiędzy poszczególnymi wartościami i wartością średniej arytmetycznej, czyli różnią się rozproszeniem poszczególnych wartości od średniej arytmetycznej.

W celu mierzenia rozproszenia wykorzystywane są **miary rozproszenia**, nazywane czasem miarami dyspersji lub miarami zmienności.

Istnieją różne miary rozproszenia: obszar zmienności, odchylenie ćwiartkowe, odchylenie przeciętne, wariancja, odchylenie standardowe i współczynnik zmienności. Do częściej stosowanych miar rozproszenia należy obszar zmienności, odchylenie przeciętne, wariancja i odchylenie standardowe i te miary niżej scharakteryzowano.

**Obszar zmienności** (rozstęp) jest to różnica między największą i najmniejszą wartością zmiennej w szeregu. Najczęściej oznaczany jest literą  $R$ :

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

Obszar zmienności jest najprostszą i najbardziej prymitywną miarą rozproszenia, bardzo nieprecyzyjną, ponieważ wielkość jej zależy tylko od skrajnych wartości zmiennej, które mogą być wartościami przypadkowymi. Poprawna miara rozproszenia powinna uwzględniać wszystkie wartości w rozkładzie. W związku z tym obszar zmienności jako miara rozproszenia wykorzystywany jest tylko w celu ogólnej orientacji o właściwościach szeregu.

---

<sup>19</sup> Z. Peuker, *Statystyka ... op. cit.*, s. 110.

### 2.2.1. Odchylenie przeciętne

**Odchylenie przeciętne** to miara uwzględniająca wszystkie występujące w szeregu wartości zmiennej. Charakteryzuje przeciętny poziom odchylenia poszczególnych wartości szeregu od średniej arytmetycznej. Jest miarą mianowaną i przyjmuje zawsze wartości dodatnie. Obliczając odchylenie przeciętne należy pomijać wartości znaków algebraicznych, a działania wykonywać na liczbach bezwzględnych, ponieważ suma odchyleń (ujemnych i dodatnich) od średniej arytmetycznej wynosi zero (0) i dlatego nie mogłaby być wykorzystywana do dalszych obliczeń.

W szeregu szczegółowym odchylenie przeciętne oblicza się według następującego wzoru:

$$d = \frac{\sum(X - \bar{x})}{N},$$

gdzie:

$X$  – poszczególne wartości zmiennej,

$\bar{x}$  – średnia arytmetyczna,

$N$  – liczebność zbiorowości.

W szeregu szczegółowym, indywidualnym ( $N = 10$ ) zawierającym wiek członków koła teatralnego: 12, 14, 15, 13, 10, 18, 17, 15, 16, 10 odchylenie przeciętne oblicza się w sposób następujący:

- obliczenie średniej arytmetycznej:  $\bar{x} = \frac{140}{10} = 14$  lat,
- od poszczególnych wartości zmiennej należy odjąć wartość średniej arytmetycznej i w ten sposób otrzymuje się odchylenie, które należy zsumować (wartości znaków algebraicznych pomija się):

$$\begin{aligned}\sum(X - \bar{x}) &= (12-14) + (14-14) + (15-14) + (13-14) + (10-14) + (18-14) + \\ &+ (17-14) + (15-14) + (16-14) + (10-14) = 22\end{aligned}$$

- sumę odchyleń należy podzielić przez liczebność szeregu:

$$d = \frac{\sum(X - \bar{x})}{N} = \frac{22}{10} = 2,2$$

Przeciętne odchylenie wieku członków koła teatralnego od jego średniej arytmetycznej wynosi 2,2 lata.

Przy obliczaniu odchylenia przeciętnego **dla szeregów rozdzielczych z cechą skokową** postępuje się podobnie, przy czym po obliczeniu odchyleń należy pomnożyć je przez liczebności poszczególnych klas.

Obliczanie odchylenia przeciętnego **dla szeregów rozdzielczych zawierających przedziały klasowe** następuje według wzoru:

$$d = \frac{\sum f(x_i - \bar{x})}{N},$$

gdzie:

- $f$  – liczba obserwacji w poszczególnych przedziałach klasowych,
- $x_i$  – środek przedziału klasowego,
- $\bar{x}$  – średnia arytmetyczna.

W Tablicy 30 przedstawiono sposób obliczania odchylenia przeciętnego dla szeregu rozdzielczego zawierającego przedziały klasowe.

Tablica 30

*Uczniowie dojeżdżający do szkoły według wieku*

| Wiek<br>$X$ | Liczba uczniów<br>$f$ | Środek przedziału<br>$x_i$ | $fx_i$       | $(x_i - \bar{x})$ | $f(x_i - \bar{x})$ |
|-------------|-----------------------|----------------------------|--------------|-------------------|--------------------|
| 6 – 8       | 7                     | 7                          | 49           | 4,5               | 31,5               |
| 9 – 11      | 12                    | 10                         | 120          | 1,5               | 18,0               |
| 12 – 14     | 20                    | 13                         | 260          | 1,5               | 30,0               |
| 15 – 17     | 3                     | 16                         | 48           | 4,5               | 13,5               |
|             | $N=42$                |                            | $\Sigma=477$ |                   | $\Sigma=93,0$      |

Źródło: dane umowne

W celu obliczenia odchylenia przeciętnego dla szeregu w tab. 30 wykonano następujące operacje:

- obliczono środki przedziałów klasowych:  $x_i$ ,
- pomnożono środki przedziałów klasowych przez liczebność klas:  $fx_i$ ,
- obliczono średnią arytmetyczną:  $\bar{x} = \frac{477}{42} \approx 11,5$  lat.
- obliczono odchylenie każdego środka przedziału od średniej arytmetycznej:  $(x_i - \bar{x})$  – pomijając wartości znaków algebraicznych,

- każde odchylenie pomnożono przez liczebność odpowiedniej klasy i dokonano sumowania wszystkich odchyleń, czyli suma odchyleń wynosi:  $\sum f(x_i - \bar{x}) = 93$ ,
- podstawiając do wzoru uzyskano:

$$d = \frac{\sum f(x_i - \bar{x})}{N} = \frac{93}{42} = 2,2$$

Średni wiek dzieci dojeżdżających do szkoły wynosi 11,5 lat, a odchylenie przeciętne wieku dzieci wynosi 2,2 lata. W podanym przykładzie odchylenie przeciętne informuje, o ile wiek poszczególnych uczniów różni się przeciętnie od średniej arytmetycznej tego wieku.

Odchylenie przeciętne nie jest często stosowaną miarą rozproszenia. Ze względu na konieczność pomijania znaków algebraicznych, przy jego obliczaniu występują trudności w przekształceniach algebraicznych. Podkreślić jednak należy, że na przykładzie odchylenia przeciętnego stosunkowo łatwo można zorientować się w funkcjonowaniu miar rozproszenia obliczanych na podstawie średniej arytmetycznej.

### 2.2.2. Wariancja i odchylenie standardowe

**Wariancja** jest jedną z częściej stosowanych miar rozproszenia. Określa się ją jako sumę kwadratów odchylenia wszystkich wartości zmiennej od ich średniej arytmetycznej, podzieloną przez ich liczbę. Dla szeregu szczegółowego wariancję oblicza się według następującego wzoru:

$$s^2 = \frac{\sum (X - \bar{x})^2}{N},$$

gdzie:

- $X$  – każda wartość zmiennej,
- $\bar{x}$  – średnia arytmetyczna,
- $N$  – liczebność zbiorowości.

Wariancja jest przydatna do porównywania struktury różnych części zbiorowości logicznie ze sobą powiązanych. W przypadku braku zróżnicowania zbiorowości – gdy wszystkie wartości zmiennej równe są średniej arytmetycznej – równa jest zero (0), a im większe zróżnicowanie, tym większa wartość wariancji.

**Odchylenie standardowe** jest najczęściej stosowaną i bardzo precyzyjną miarą rozproszenia wartości zmiennej. Informuje, o ile przeciętnie, średnio odchyła się każda wartość zmiennej od **średniej arytmetycznej**. Miara ta uwzględnia wszystkie występujące w szeregu wartości zmiennej i ma duże znaczenie przy określaniu rozkładu normalnego. Jest pierwiastkiem kwadratowym z wariancji.

Dla szeregu szczegółowego odchylenie standardowe oblicza się według następującego wzoru:

$$s_x = \sqrt{\frac{\Sigma(X - \bar{x})^2}{N}},$$

gdzie:

$X$  – każda wartość zmiennej,

$\bar{x}$  – średnia arytmetyczna,

$N$  – liczebność zbiorowości.

W Tablicy 31 przedstawiono sposób obliczania odchylenia standardowego dla szeregu szczegółowego zawierającego wyniki testu z matematyki dla  $N=10$  uczniów kl. V.

Tablica 31

*Uczniowie kl. V według wyników testu z matematyki*

| Uczeń | Wyniki testu<br>$X$ | $(X - \bar{x})$ | $(X - \bar{x})^2$ |
|-------|---------------------|-----------------|-------------------|
| A     | 8                   | -3              | 9                 |
| B     | 15                  | 4               | 16                |
| C     | 11                  | 0               | 0                 |
| D     | 11                  | 0               | 0                 |
| E     | 9                   | -2              | 4                 |
| F     | 12                  | 1               | 1                 |
| G     | 14                  | 3               | 9                 |
| H     | 11                  | 0               | 0                 |
| I     | 10                  | -1              | 1                 |
| J     | 9                   | -2              | 4                 |
|       | $\Sigma=110$        |                 | $\Sigma=44$       |

Źródło: dane umowne

W celu obliczenia odchylenia standardowego dla szeregu w Tablicy 31 wykonano następujące operacje:

- obliczono średnią arytmetyczną:  $\bar{x} = \frac{110}{10} = 11$  pkt.
- obliczono odchylenie każdej wartości od średniej arytmetycznej:  $(X - \bar{x})$ ,
- każde odchylenie podniesiono do kwadratu:  $(X - \bar{x})^2$ ,
- dokonano sumowania wszystkich kwadratów odchyleń od średniej:  $\Sigma(X - \bar{x})^2 = 44$ ,
- podstawiając do wzoru otrzymano:

$$s_x = \sqrt{\frac{\Sigma(X - \bar{x})^2}{N}} = \sqrt{\frac{44}{10}} \approx 2,1$$

Odchylenie standardowe wyników testu z matematyki wynosi,  $s_x = 2,1$  pkt., przy średniej arytmetycznej  $\bar{x} = 11$  pkt. Oznacza to, że wynik testu z matematyki poszczególnych uczniów przeciętnie wahał się w granicach od: średnia arytmetyczna minus odchylenie standardowe do średnia arytmetyczna plus odchylenie standardowe, czyli od  $(11 - 2,1)$  do  $(11 + 2,1)$ , co daje wyniki testu w granicach od 8,9 pkt. do 13,1 pkt. W przedziale 8,9÷13,1 mieszczą się wyniki testu uzyskane przez 7 uczniów, co stanowi 70% badanej grupy. Na tej podstawie można stwierdzić, że rozproszenie wyników w tej grupie jest prawidłowe, tzn. brak jest wyników, które znacznie odchylają się *in plus* lub *in minus* od średniej arytmetycznej, co oznacza, że jest to rozkład zbliżony do normalnego<sup>20</sup>.

**Dla szeregu rozdzielczego** zawierającego przedziały klasowe odchylenie standardowe oblicza się według następującego wzoru:

$$s_x = \sqrt{\frac{\Sigma f(x_i - \bar{x})^2}{N}},$$

gdzie:

- $f$  – liczba obserwacji w poszczególnych przedziałach klasowych (liczebność),
- $x_i$  – środek przedziału klasowego,
- $\bar{x}$  – średnia arytmetyczna.

---

<sup>20</sup> Patrz rozdz. V, pkt. 2.

W Tablicy 32 przedstawiono sposób obliczania odchylenia standardowego dla szeregu rozdzielczego zawierającego przedziały klasowe.

Tablica 32

*Wyniki testu z ortografii uczniów kl. III*

| Wynik testu<br>$X$ | Liczba<br>uczniów<br>$f$ | $x_i$ | $fx_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $f(x_i - \bar{x})^2$ |
|--------------------|--------------------------|-------|--------|-------------------|---------------------|----------------------|
| 8 – 9              | 6                        | 8,5   | 51,0   | -6                | 36                  | 216                  |
| 10 – 11            | 10                       | 10,5  | 105,0  | -4                | 16                  | 160                  |
| 12 – 13            | 26                       | 12,5  | 32,5   | -2                | 4                   | 104                  |
| 14 – 15            | 31                       | 14,5  | 449,5  | 0                 | 0                   | 0                    |
| 16 – 17            | 22                       | 16,5  | 363,5  | 2                 | 4                   | 88                   |
| 18 – 19            | 9                        | 18,5  | 166,5  | 4                 | 16                  | 144                  |
| 20 – 21            | 8                        | 20,5  | 164,0  | 6                 | 36                  | 288                  |
|                    | 112                      |       | 1624,0 |                   |                     | $\Sigma=1000$        |

Źródło: dane umowne

W celu obliczenia odchylenia standardowego dla szeregu w tab. 32 wykonano następujące operacje:

- obliczono środki przedziałów klasowych:  $x_i$ ,
- pomnożono środki przedziałów klasowych przez liczebność klas:  $fx_i$ ,
- obliczono średnią arytmetyczną:  $\bar{x} = \frac{1624}{112} = 14,5$  pkt.,
- obliczono odchylenie każdego środka przedziału od średniej arytmetycznej:  $(x_i - \bar{x})$ ,
- każde odchylenie podniesiono do kwadratu:  $(x_i - \bar{x})^2$ ,
- kwadrat każdego odchylenia pomnożono przez liczebność każdej klasy:  $f(x_i - \bar{x})^2$ ,
- dokonano sumowania wszystkich kwadratów odchylen, pomnożonych przez liczebność klas:  $\Sigma f(x_i - \bar{x})^2 = 1000$ ,
- podstawiając do wzoru otrzymano:

$$s_x = \sqrt{\frac{\Sigma f(x_i - \bar{x})^2}{N}} = \sqrt{\frac{1000}{112}} \approx 3 \text{ pkt.}$$

Odchylenie standardowe poszczególnych wyników testu z ortografii od średniej arytmetycznej wynosi 3 pkt., co oznacza, że przeciętnie wyniki odchylają się od średniej arytmetycznej o około minus 3 pkt. i plus 3 pkt., a więc wahają się w przedziale  $11,5 \div 17,5$ . W takim przedziale znajdują się wyniki około 79 uczniów, co stanowi 70,5% badanej grupy. Można więc stwierdzić, że jest to rozkład zbliżony do normalnego, a uczniowie, którzy uzyskali wyniki w tym przedziale osiągnęli przeciętny poziom, czyli normę. Natomiast uczniowie, którzy uzyskali wyniki niższe niż 11,5 pkt., to uczniowie z rezultatami poniżej normy (średniej), a ci, którzy osiągnęli wyniki wyższe niż 17,5 pkt., stanowią grupę powyżej normy (średniej, przeciętnej).

### **Właściwości wariancji i odchylenia standardowego**

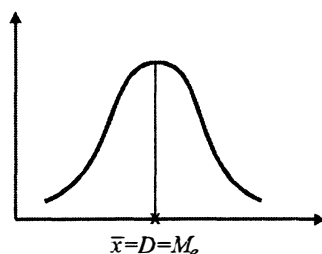
- są zależne od wszystkich występujących w rozkładzie wartości zmiennej,
- nie są uzależnione od przypadkowych, skrajnych wartości zmiennej,
- są mianowane, tzn. każda wartość tych miar wyrażona jest w tych samych jednostkach, w jakich podana jest wartość zmiennej.

## **2.3. Miary asymetrii**

Często scharakteryzowanie uzyskanego rozkładu wyników przy pomocy dotychczas przedstawionych miar jest niewystarczające. Badacz jest czasem zainteresowany tym, czy odchylenie od wartości średniej arytmetycznej jest większe w jedną lub drugą stronę, np. przy badaniach poziomu wiadomości uczniów badacz jest zainteresowany, czy liczba uczniów z wynikami wyższymi od przeciętnego jest większa czy też mniejsza od liczby uczniów z wynikami poniżej przeciętnych. Rozmiary takiego odchylenia ustala się przy pomocy miar asymetrii.

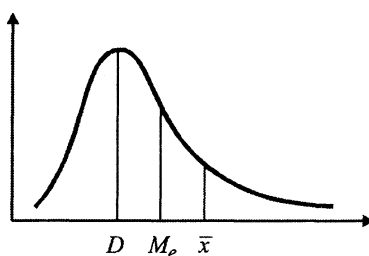
W szeregu symetrycznym liczebności są rozłożone proporcjonalnie po obu stronach średniej arytmetycznej, średnia arytmetyczna, dominanta i mediana są sobie równe, czyli  $\bar{x} = D = M_e$  (Rys. 6).

*Rys. 6. Rozkład symetryczny,  $\bar{x} = D = M_e$*



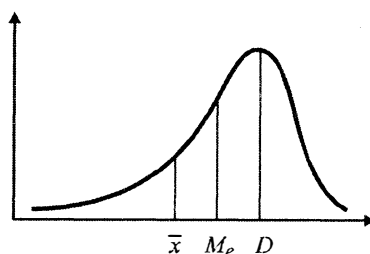
W szeregach asymetrycznych wartości zmiennej nie są rozmieszczone równomiernie i nie mają tendencji do skupiania się wokół średniej arytmetycznej, natomiast skupiają się w pobliżu dolnej lub górnej części rozkładu.

**Asymetria prawostronna** (dodatnia) występuje wówczas, gdy wartości zmiennej skupiają się w pobliżu dolnej części rozkładu i wówczas dłuższe ramię krzywej znajduje się po prawej stronie wykresu (Rys. 7).



Rys. 7. Rozkład z asymetrią prawostronną,  $\bar{x} - D > 0$

**Asymetria lewostronna** (ujemna) występuje wówczas, gdy wartości zmiennej skupiają się w pobliżu dolnej części rozkładu i wówczas dłuższe ramię krzywej znajduje się po lewej stronie wykresu (Rys. 8).



Rys. 8. Rozkład z asymetrią lewostronną,  $\bar{x} - D < 0$

Asymetrię szeregu można mierzyć poprzez ustalenie różnicy między średnią arytmetyczną i dominantą:<sup>21</sup>

$\bar{x} - D = 0$  – szereg symetryczny,

$\bar{x} - D > 0$  – szereg z asymetrią prawostronną,

$\bar{x} - D < 0$  – szereg z asymetrią lewostronną.

Różnica jest liczbą mianowaną, zależną od jednostek miary i wartości przyjmowanych przez zmienną i w związku z tym ta miara ma niewielkie zastosowanie, ponieważ trudno jest porównywać szeregi o różnych jednostkach.

Częściej stosowaną miarą asymetrii (skośności) jest **współczynnik asymetrii**:

$$A_s = \frac{\bar{x} - D}{s_x},$$

gdzie:

$\bar{x}$  – średnia arytmetyczna,

$D$  – dominanta,

$s_x$  – odchylenie standardowe.

Współczynnik asymetrii równa się zero (0) w przypadku szeregu symetrycznego, ma znak "+", przy asymetrii prawostronnej, a znak "-", przy asymetrii lewostronnej.

Dla uzyskanych wyników z ortografii uczniów kl. V obliczono współczynniki asymetrii:

- średnia arytmetyczna:  $\bar{x} = 20,5$
- dominanta:  $D = 26,8$
- odchylenie standardowe:  $s_x = 8,4$

$$A_s = \frac{\bar{x} - D}{s_x} = \frac{20,5 - 26,8}{8,4} = \frac{-6,3}{8,4} = -0,75$$

Z wartości współczynnika asymetrii wynika, że rozkład wyników testu z ortografii wykazuje wyraźną asymetrię lewostronną (ujemną), co oznacza, że w badanej zbiorowości występuje przewaga uczniów uzyskujących wyniki niższe od przeciętnych.

---

<sup>21</sup> Cz. Nowaczyk, *Podstawy ...*, op. cit., s. 87.

## 2.4. Obliczanie i interpretacja miar zmiennej

Precyzyjna analiza struktury badanej zbiorowości wymaga obliczania miar przeciętnych, rozproszenia, asymetrii wybranej cechy. Podstawą takiej analizy są wyniki pomiarów (obserwacji) przedstawione w szeregu statystycznym o określonym rozkładzie, charakteryzowanym przy pomocy parametrów opisowych.

Niżej, jako podsumowanie tego rozdziału, dla wybranych przykładów obliczono i dokonano łącznej interpretacji omówionych w skrypcie parametrów opisowych. W Tablicy 33 i 34 przedstawiono wyniki testu z ortografii uczniów kl. III ze szkół A i B. W obu badanych zbiorowościach liczebności są jednakowe:  $N_A = N_B = 58$ .

Tablica 33

Wyniki testu z ortografii uczniów kl. III ze szkoły A

| Wynik testu<br>$X$ | Liczba uczniów<br>$f$ | Szereg kumulacyjny<br>$f_c$ | Środek przedziału<br>$x_i$ | $fx_i$ | $(x_i - \bar{x})$ | $f(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $f(x_i - \bar{x})^2$ |
|--------------------|-----------------------|-----------------------------|----------------------------|--------|-------------------|--------------------|---------------------|----------------------|
| 0 - 2              | 4                     | 4                           | 1                          | 4      | -8                | -32                | 64                  | 256                  |
| 3 - 5              | 8                     | 12                          | 4                          | 32     | -5                | -40                | 25                  | 200                  |
| 6 - 8              | 18                    | 30                          | 7                          | 126    | -2                | -36                | 4                   | 72                   |
| 9 - 11             | 11                    | 41                          | 10                         | 110    | 1                 | 11                 | 1                   | 11                   |
| 12 - 14            | 9                     | 50                          | 13                         | 117    | 4                 | 36                 | 16                  | 144                  |
| 15 - 17            | 6                     | 56                          | 16                         | 96     | 7                 | 42                 | 49                  | 294                  |
| 18 - 20            | 2                     | 58                          | 19                         | 38     | 10                | 20                 | 100                 | 200                  |
|                    | $N=58$                |                             |                            | 523    |                   | 217*               | 259                 | 1177                 |

\* – przy sumowaniu pomija się znaki algebraiczne

Źródło: dane umowne

Wyniki testu z ortografii uczniów kl. III ze szkoły B

| Wynik testu<br>$X$ | Liczba uczniów<br>$f$ | Szereg kumulacyjny<br>$f_c$ | Środek przedziału<br>$x_i$ | $fx_i$ | $(x_i - \bar{x})$ | $f(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $f(x_i - \bar{x})^2$ |
|--------------------|-----------------------|-----------------------------|----------------------------|--------|-------------------|--------------------|---------------------|----------------------|
| 0 - 2              | 2                     | 2                           | 1                          | 2      | -11               | -22                | 121                 | 242                  |
| 3 - 5              | 4                     | 6                           | 4                          | 16     | -8                | -32                | 64                  | 256                  |
| 6 - 8              | 8                     | 14                          | 7                          | 56     | -5                | -40                | 25                  | 200                  |
| 9 - 11             | 10                    | 24                          | 10                         | 100    | -2                | -20                | 4                   | 40                   |
| 12 - 14            | 15                    | 39                          | 13                         | 195    | 1                 | 15                 | 1                   | 15                   |
| 15 - 17            | 11                    | 50                          | 16                         | 176    | 4                 | 44                 | 16                  | 176                  |
| 18 - 20            | 8                     | 58                          | 19                         | 152    | 7                 | 56                 | 49                  | 392                  |
|                    | $N=58$                |                             |                            | 697    |                   | 229*               | 280                 | 1321                 |

\* – przy sumowaniu pomija się znaki algebraiczne

Źródło: dane umowne

Tablica 35

Wartości parametrów opisowych dla uczniów ze szkół A i B

| Parametry opisowe      | Zastosowany wzór                                            | Wyniki uczniów ze szkoły                                        |                                                                 |
|------------------------|-------------------------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------------------------|
|                        |                                                             | A                                                               | B                                                               |
| Średnia arytmetyczna   | $\bar{x} = \frac{\sum x_i f}{N}$                            | $\bar{x} = \frac{523}{58} \approx 9$                            | $\bar{x} = \frac{697}{58} \approx 12$                           |
| Mediana                | $M_e = x_0 + i \left( \frac{N}{2} - f_b \right) / f_a$      | $M_e = 6 + \frac{3}{18} \left( \frac{58}{2} - 12 \right) = 8,8$ | $M_e = 12 + \frac{3}{15} \left( \frac{58}{2} - 24 \right) = 13$ |
| Dominanta              | $D = X_0 + i \frac{(f_a - f_b)}{(f_a - f_b) + (f_a - f_d)}$ | $D = 6 + 3 \frac{(18-8)}{(18-8) + (18-11)} = 7,8$               | $D = 12 + 3 \frac{(15-10)}{(15-10) + (15-11)} = 13,7$           |
| Odchylenie przeciętne  | $d = \frac{\sum f(x_i - \bar{x})}{N}$                       | $d = \frac{217}{58} = 3,7$                                      | $d = \frac{229}{58} = 3,9$                                      |
| Odchylenie standardowe | $s_x = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{N}}$            | $s_x = \sqrt{\frac{1177}{58}} \approx 4,5$                      | $s_x = \sqrt{\frac{1321}{58}} \approx 4,8$                      |
| Współczynnik asymetrii | $A_s = \frac{\bar{x} - D}{s_x}$                             | $A_s = \frac{9 - 7,8}{4,5} = +0,26$                             | $A_s = \frac{12 - 13,7}{4,8} = -0,35$                           |

Źródło: dane umowne

Po obliczeniu parametrów opisowych można scharakteryzować wyniki testu z ortografii uczniów kl. III ze szkół A i B dokonując między nimi pewnych porównań (Tablica 35). Z uzyskanych wartości średnich arytmetycznych wynika, że średni wynik testu w obu szkołach znacznie się różni. Przeciętnie, średnio każdy uczeń ze szkoły B uzyskał z testu o 3 punkty więcej od każdego ucznia ze szkoły A. Średniowy wynik testu jest również wyraźnie odmienny w obu grupach uczniów: 8,8 pkt. w grupie ze szkoły A i 13 pkt. w grupie ze szkoły B. Z wartości dominant wynika, że najczęściej występującym wynikiem testu u uczniów ze szkoły A było 7,8 pkt., a u uczniów ze szkoły B – 13,7 pkt.

Z przedstawionych wartości parametrów opisowych wynika, że grupa uczniów ze szkoły B reprezentuje wyższy poziom badanych wiadomości i umiejętności z ortografii niż uczniowie ze szkoły A, chociaż zróżnicowanie uzyskanych wyników, na co wskazuje wartość odchylenia przeciętnego i standardowego, jest większe w grupie uczniów ze szkoły B niż w grupie ze szkoły A. Oznacza to, że uczniowie ze szkoły B reprezentują wyższy poziom badanych wiadomości i umiejętności, ale bardziej zróżnicowany, natomiast uczniowie ze szkoły A reprezentują poziom niższy, ale bardziej wyrównany. Wartość odchylenia standardowego wskazuje, że wynik testu w grupie A odchyła się przeciętnie od średniego wyniku o 4,5 pkt., a w szkole B o 4,8 pkt.. Oba uzyskane szeregi są asymetryczne. Szereg przedstawiający wyniki uczniów ze szkoły A ma lekką asymetrię prawostronną, a szereg ze szkoły B – lekką asymetrię lewostronną.

## Pytania i zadania

1. W jaki sposób dokonuje się opisu cechy jakościowej?
2. Wskaż różnice między wskaźnikami struktury i natężenia.
3. Podaj przykład tablicy zawierającej dwie powiązane ze sobą cechy jakościowe. Wpisz umowne dane liczbowe, oblicz wskaźniki procentowe i dokonaj analizy danych zawartych w tablicy. Przedstaw odpowiedni wykres.
4. Podaj przykład tablicy zawierającej dwie powiązane ze sobą cechy ilościowe. Wpisz umowne dane liczbowe, oblicz wskaźniki procentowe i dokonaj analizy danych w tablicy. Przedstaw odpowiedni wykres.
5. Na podstawie obliczonych wskaźników procentowych dokonaj analizy danych w zamieszczonej tablicy. Przedstaw odpowiedni wykres.

*Miejsce zamieszkania a wyniki w studiach studentów studiów zaocznych*

| Miejsce zamieszkania | Wyniki w studiach |         |        |        |
|----------------------|-------------------|---------|--------|--------|
|                      | Wysokie           | Średnie | Niskie | Ogółem |
| Wieś                 | 1                 | 9       | 6      | 16     |
| Gmina                | -                 | 5       | 3      | 8      |
| Miasto do 15 tys.    | 6                 | 13      | 8      | 27     |
| Miasto ponad 15 tys. | 2                 | 30      | 22     | 54     |
| Ogółem               | 9                 | 57      | 39     | 105    |

6. Na podstawie obliczonych wskaźników procentowych dokonaj analizy danych w zamieszczonej tablicy.

*Liczba dzieci w rodzinach uczniów kl. III*

| Badani | Liczba dzieci w rodzinie |    |    |    |   |        |
|--------|--------------------------|----|----|----|---|--------|
|        | 1                        | 2  | 3  | 4  | 5 | Ogółem |
| Liczba | 47                       | 35 | 18 | 10 | 6 | 116    |
| %      |                          |    |    |    |   |        |

7. Na podstawie obliczonych wskaźników procentowych dokonaj analizy danych w tablicy. Przedstaw odpowiedni wykres.

*Sytuacja materialna a aspiracje edukacyjne uczniów kl. VIII*

| Sytuacja materialna | Poziom aspiracji edukacyjnych |           |     |    |        |
|---------------------|-------------------------------|-----------|-----|----|--------|
|                     | ZSZ                           | Technikum | L.Z | LO | Ogółem |
| Bardzo dobra        | 2                             | 5         | 2   | 4  | 13     |
| Dobra               | 8                             | 10        | 3   | 20 | 41     |
| Wystarczająca       | 3                             | 18        | 8   | 16 | 45     |
| Zła                 | 2                             | -         | -   | -  | 2      |
| Ogółem              | 15                            | 33        | 13  | 40 | 101    |

8. W jaki sposób dokonuje się opisu rozkładu cechy ilościowej?
9. Wskaż różnice między rozkładem symetrycznym i asymetrycznym?
10. Scharakteryzuj istotę i właściwości miar przeciętnych.
11. Oblicz średnią arytmetyczną, medianę i dominantę, a także przedstaw graficzny obraz dla następującego szeregu statystycznego:

a/ X: 17, 20, 34, 41, 45, 23, 31, 34, 40, 28, 18, 12, 22, 33, 46, 43, 38, 39, 16, 21, 20, 26, 25, 46, 41, 48, 26, 29, 35, 17, 29, 41, 31, 38, 15, 19, 25, 30, 42, 34.

b/ X: 20, 26, 21, 23, 21, 25, 22, 25, 22, 24, 22, 23, 22, 24, 22, 24, 24, 23, 22, 24, 23, 20, 23, 23, 23, 22, 23, 23, 23, 25.

c/ X: 31, 50, 32, 54, 53, 65, 66, 70, 71, 36, 51, 52, 60, 61, 83, 75, 76, 40, 41, 37, 38, 53, 54, 52, 56, 57, 62, 63, 76, 41, 44, 45, 55, 55, 58, 59, 46, 47, 67, 68, 72, 64, 64, 73, 74, 48, 49, 48, 49, 54.

d/ X: 1, 5, 2, 6, 6, 3, 4, 1, 5, 1, 1, 6, 3, 3, 4, 1, 2, 5, 3, 2, 3, 3, 4, 5, 1, 3, 3, 1, 2, 3, 4, 4, 5.

12. Określ zbiorowość, oblicz i dokonaj interpretacji średniej arytmetycznej, mediany i dominanty dla następujących wyników badań:

a)

| Wynik testu $X$ | Liczebność $f$ |
|-----------------|----------------|
| 18 – 22         | 7              |
| 23 – 27         | 14             |
| 28 – 32         | 29             |
| 33 – 37         | 19             |
| 38 – 42         | 11             |
|                 | $N=80$         |

b)

| Liczba dzieci $X$ | Liczebność rodzin $f$ |
|-------------------|-----------------------|
| 1                 | 11                    |
| 2                 | 29                    |
| 3                 | 18                    |
| 4                 | 10                    |
| 5 i więcej        | 4                     |
|                   | $N=72$                |

c)

| Wiek dzieci $X$ | Liczebność $f$ |
|-----------------|----------------|
| 7 – 8           | 4              |
| 9 – 10          | 11             |
| 11 – 12         | 9              |
| 13 – 14         | 3              |
|                 | $N=27$         |

d)

| Dni absencji $X$ | Liczebność nauczycieli $f$ |
|------------------|----------------------------|
| 0 – 3            | 21                         |
| 4 – 7            | 15                         |
| 8 – 11           | 10                         |
| 12 – 15          | 3                          |
|                  | $N=49$                     |

13. Oblicz średni czas potrzebny na wykonanie zadania z matematyki przez uczniów kl. III.

| Czas w min. $X$ | Liczebność uczniów $f$ |
|-----------------|------------------------|
| 15              | 1                      |
| 17              | 1                      |
| 18              | 2                      |
| 20              | 4                      |
| 21              | 1                      |
| 22              | 2                      |
|                 | $N=11$                 |

14. Scharakteryzuj istotę i właściwości miar rozproszenia.

15. Oblicz odchylenie przeciętne dla następujących wyników:

| Odległość do szkoły (w km) $X$ | Liczebność uczniów $f$ |
|--------------------------------|------------------------|
| 0 – 1                          | 14                     |
| 2 – 3                          | 27                     |
| 4 – 5                          | 11                     |
|                                | $N = 52$               |

16. Określ zbiorowość, oblicz i dokonaj interpretacji odchylenia standardowego dla następujących wyników:

| Uczeń | Liczba błędów $f$ |
|-------|-------------------|
| A     | 4                 |
| B     | 7                 |
| C     | 5                 |
| D     | 10                |
| E     | 9                 |
| F     | 12                |
|       | $N = 47$          |

| Wyniki testu – $X$ | Liczebność uczniów $f$ |
|--------------------|------------------------|
| 10 – 13            | 7                      |
| 14 – 17            | 16                     |
| 18 – 21            | 24                     |
| 22 – 25            | 32                     |
| 26 – 29            | 11                     |
| 30 – 33            | 4                      |
|                    | $N = 94$               |

17. W rezultacie analizy wyników testu z matematyki w dwóch klasach VII otrzymano następujące wyniki:

$$a/ \bar{x} = 40,5, D = 38, M_e = 39, s_x = 2,5,$$

$$b/ \bar{x} = 40,0, D = 31, M_e = 33, s_x = 8,8.$$

Dokonaj interpretacji uzyskanych miar, a także porównaj poziom osiągnięć szkolnych z matematyki w obu badanych klasach.

18. Dokonaj interpretacji wyników testu z ortografii uczniów dwóch badanych klas V:

$$a/ \bar{x} = 13,1, D = 14, M_e = 13,5, s_x = 3,2,$$

$$b/ \bar{x} = 15,6, D = 14,5, M_e = 15,0, s_x = 5,1.$$

19. Oblicz i dokonaj interpretacji współczynników asymetrii dla wyników badań z zadania 18.

## Rozdział IV

# WSPÓŁZALEŻNOŚĆ ZJAWISK W BADANIACH PEDAGOGICZNYCH

W poprzednich rozdziałach skryptu scharakteryzowano rozkłady zawierające jedną zmienną ( $X$ ), czyli rozkłady jednozmiennne. W tej części skryptu przedstawiono rozkłady dwuzmiennne, tzn. rozkłady zawierające dwie zmienne, np.  $X$  i  $Y$ , a także dokonano analizy korelacji zachodzącej między dwoma zmiennymi. Związek korelacyjny jest wyrazem zmian zachodzących jednocześnie w dwóch obserwowanych cechach (zmiennych), dotyczących tej samej jednostki<sup>22</sup>. Obserwacje dokonywane są parami. W związku z tym można sformułować pytanie, czy między zmienną  $X$  i zmienną  $Y$  wyrażającą zmiany w tej samej jednostce istnieje związek:

- związek między zmienną  $X$  – oznaczającą płeć studenta i zmienną  $Y$  oznaczającą wynik egzaminu,
- związek między wynikami testu z matematyki (zmienna  $X$ ) i fizyki (zmienna  $Y$ ),
- związek między aktywnością uczniów podczas lekcji języka polskiego w opinii nauczyciela przedstawionej w porządku rangowym (zmienna  $X$ ) i popularnością ich w klasie, przedstawioną również w kolejności rangowej (zmienna  $Y$ ).

---

<sup>22</sup> G. Clauss, H. Ebner, *Podstawy ...*, op. cit., s. 105.

W pierwszym przykładzie zmienne przedstawione są na skali nominalnej, wyrażone są określeniem słownym, nie miarą, są to cechy jakościowe: płeć ( $X$ ): kobieta, mężczyzna; wynik egzaminu ( $Y$ ): pozytywny, negatywny.

W drugim przypadku zmienne przedstawione są na skali metrycznej (można je też przedstawić na skali przedziałowej), są to cechy ilościowe: wyniki testu z matematyki ( $X$ ) i fizyki ( $Y$ ) w punktach.

W ostatnim przykładzie zmienne przedstawione są na skali rangowej (pozycyjnej) według intensywności zmiennej: opinia nauczyciela o aktywności uczniów podczas lekcji ( $X$ ) i popularność uczniów w klasie ( $Y$ ).

Sposób wykazywania istnienia związku korelacyjnego zależy od rodzaju skali, na której przedstawione są zmienne. Gdy cechy wyrażone są w dwu jakościowo różnych skalach nominalnych, wówczas związek między nimi określa się przy pomocy mierników zbieżności lub asocjacji. Jeśli obie zmienne wyrażone są w skali interwałowej, wówczas przy związku liniowym możliwe jest określenie korelacji miarowej. Jeśli natomiast zmienne uszeregowane są na skali rangowej (pozycyjnej), wówczas mamy do czynienia z korelacją rangową.

Między zmiennymi mogą zachodzić trzy rodzaje związku:

- **zgodność** – wysokim wartościom zmiennej  $X$  odpowiadają wysokie wartości zmiennej  $Y$  i odwrotnie – niskim wartościom zmiennej  $X$  odpowiadają niskie wartości zmiennej  $Y$ . W takiej sytuacji między zmiennymi istnieje **korelacja pozytywna**.
- **przeciwieństwo** – wysokim wartościom zmiennej  $X$  odpowiadają niskie wartości zmiennej  $Y$  i odwrotnie – niskim wartościom zmiennej  $X$  odpowiadają wysokie wartości zmiennej  $Y$ . Między zmiennymi istnieje **korelacja negatywna**.
- **niezależność** – wysokim wartościom zmiennej  $X$  odpowiadają raz wysokie, a innym razem niskie wartości zmiennej  $Y$  i odwrotnie. Między zmiennymi nie istnieje żaden związek, są one niezależne, nie korelują ze sobą.

Związki korelacyjne różnią się od związków funkcjonalnych.

**Związek funkcjonalny** występuje wtedy gdy każdej wartości jednej zmiennej odpowiada jedna, jednoznacznie określona wartość drugiej zmiennej, np. związek między długością boku kwadratu i jego powierzchnią, związek między ilością wykonanej pracy i zarobkiem w akordzie.

**Związek korelacyjny** występuje wtedy gdy konkretnej wartości jednej zmiennej odpowiada pewna zmieniająca się wartość drugiej zmiennej, np. związek między liczbą opuszczonych godzin w szkole i wynikami testu z jakiegoś przedmiotu. W przypadku zmiennych wykazujących związek korelacyjny na ten związek mają wpływ również inne czynniki działające na obie zmienne, np. wynik testu nie zależy wyłącznie od liczby opuszczonych godzin w szkole, ponieważ istnieją i inne czynniki wpływające na takie rezultaty, np. uzdolnienia ucznia, jego pracowitość, przyczyna nieobecności, a także takie, które czasem trudne są do przewidywania i badań.

W badaniach pedagogicznych (i nie tylko) występują związki korelacyjne między cechami mierzalnymi, ale także niemierzalnymi.

## 1. Związek cech jakościowych

W badaniach pedagogicznych często poszukuje się związków między cechami jakościowymi, wyrażonymi na skali nominalnej, np. wykształcenie rodziców i wybór typu szkoły ponadpodstawowej przez ich dzieci, rodzaj zainteresowań i forma spędzania czasu wolnego, płeć studenta i stan cywilny, itp. Związki takie wyraża się przy pomocy współczynników zbieżności cech (w niektórych opracowaniach określa się je jako współczynniki korelacji cech jakościowych).

Dwie zmienne, wyrażone na skali nominalnej, występujące w postaci dwóch wariantów, można przedstawić w postaci następującej tablicy (Tablica 36).

Tablica 36

*Płeć studenta a wynik egzaminu z przedmiotu X*

| Płeć studenta | Wynik egzaminu |           |        |
|---------------|----------------|-----------|--------|
|               | Pozytywny      | Negatywny | Ogółem |
| Kobieta       | 45             | 14        | 59     |
| Mężczyzna     | 10             | 11        | 21     |
| Ogółem        | 55             | 25        | 80     |

Źródło: dane umowne

W przypadku, gdy cechy jakościowe występują w więcej niż w dwóch wariantach, tablica może przyjąć następujący kształt (Tablica 37).

*Zainteresowania kandydatów na pedagogikę według  
typu ukończonej szkoły średniej*

| Rodzaj zainteresowań | Typ ukończonej szkoły średniej |    |           |        |
|----------------------|--------------------------------|----|-----------|--------|
|                      | LO                             | LZ | Technikum | Ogółem |
| Humanistyczne        | 41                             | 10 | 7         | 58     |
| Matematyczne         | 5                              | 4  | 3         | 12     |
| Sportowe             | 24                             | 6  | 4         | 34     |
| Muzyczne             | 25                             | 5  | 5         | 35     |
| Ogółem               | 95                             | 25 | 19        | 139    |

*Źródło: dane umowne*

Podstawą analizy związku między dwoma zmiennymi jest tablica kombinowana, przedstawiająca badaną zbiorowość z punktu widzenia tych cech jednocześnie. Niekiedy dla cech przedstawionych w tablicy, wystarczające jest obliczenie rozkładu procentowego i na podstawie tego miernika struktury, poprzez wskazanie różnicy między odpowiednimi wskaźnikami procentowymi można orientacyjnie określić związek między zmiennymi. Często jednak pożądane są bardziej precyzyjne mierniki. Wówczas zastosowanie znajduje szereg współczynników zbieżności, przy pomocy których określa się siłę związku między cechami jakościowymi. Zanim jednak badacz zdecyduje się na obliczanie siły związku między cechami jakościowymi, powinien wcześniej ustalić, czy taki związek między badanymi zmiennymi w ogóle istnieje. Tylko wówczas, jeśli występowanie takiego związku zostanie stwierdzone, można określić jego siłę, obliczając współczynniki zbieżności, np. kontyngencji C, współczynnik  $\phi$  (fi), współczynnik Q–Yule'a, współczynnik Bykowskiego i inne.

## 2. Związek cech ilościowych

Czasem w analizach badacz poszukuje związków między cechami ilościowymi, np. liczba opuszczonych godzin w szkole i wyniki testu z określonego przedmiotu, wiek i wyniki testu sprawności fizycznej, itp. W takich sytuacjach przy związku liniowym możliwe jest określenie współczynnika korelacji miarowej  $r$  Pearsona.

Jeśli natomiast badacz jest zainteresowany ustalaniem związku dla dwóch zmiennych wyrażonych na skalach rangowych (pozycyjnych), np. związek między opiniami dotyczącymi aktywności uczniów podczas lekcji, wyrażonymi przez dwóch nauczycieli na skalach rangowych, to wówczas istnieje możliwość określenia współczynnika korelacji rangowej  $R_s$  Spearmana.

## 2.1. Współczynnik korelacji Pearsona

Występowanie związku między zmiennymi ilościowymi lub też jego brak najprościej można ustalić poprzez przedstawienie obu zmiennych w tablicy uwzględniającej ich współwystępowanie lub też przedstawienie ich na wykresie w postaci graficznej.

Występowanie dwóch zmiennych ilościowych można przedstawić w następującej tablicy:

Tablica 38

*Związek między wynikami testu X i Y*

| Osoba | Wynik testu<br>X | Wynik testu<br>Y |
|-------|------------------|------------------|
| A     | 15               | 6                |
| B     | 18               | 7                |
| C     | 11               | 10               |
| D     | 9                | 6                |
| E     | 11               | 10               |

Źródło: dane umowne

Gdy obserwacji jest wiele, wówczas zaleca się stosowanie tablicy wielopolowej, np. Tablica 39.

Tablica 39

*Związek między wynikami testu X i Y*

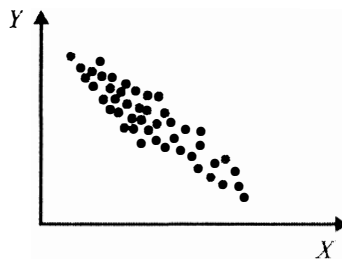
| Wyniki testu Y | Wyniki testu X |         |         |         |        | Ogółem |
|----------------|----------------|---------|---------|---------|--------|--------|
|                | 21 - 20        | 19 - 17 | 16 - 14 | 13 - 11 | 10 - 8 |        |
| 5 - 7          | -              | -       | -       | 1       | 3      | 4      |
| 8 - 10         | -              | -       | 1       | 4       | 2      | 7      |
| 11 - 13        | -              | 4       | 9       | 7       | -      | 20     |
| 14 - 16        | 4              | 5       | 3       | -       | -      | 12     |
| 17 - 19        | 2              | 1       | -       | -       | -      | 3      |
| Ogółem         | 6              | 10      | 13      | 12      | 5      | 46     |

Źródło: dane umowne

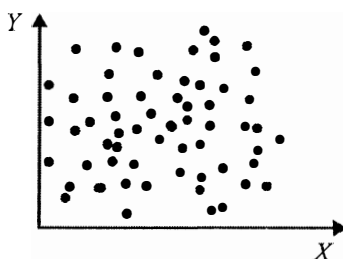
Istnienie związku korelacyjnego można stwierdzić także poprzez nanie-  
sienie wartości zmiennej  $X$  i  $Y$  na wykres.



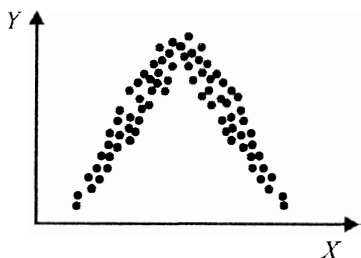
a) Wysoka korelacja dodatnia



b) Wysoka korelacja ujemna



c) Korelacja prawie równa zeru



d) Korelacja nieliniowa



e) Korelacja nieliniowa

Rys. 9. Wykresy korelacji

Podobnie jak z układu danych w tabelcy, tak i z układu punktów na wykresie wynika, jaki występuje związek korelacyjny: dodatni, ujemny czy też jego brak.

**Korelacja dodatnia** występuje wówczas gdy wraz ze wzrostem jednej zmiennej ( $X \uparrow$ ) następuje wzrost wartości drugiej zmiennej ( $Y \uparrow$ ) i odwrotnie – wraz ze spadkiem jednej zmiennej ( $X \downarrow$ ) następuje również spadek drugiej zmiennej ( $Y \downarrow$ ) (Rys. 9a).

**Korelacja ujemna** występuje wówczas gdy wzrostowi wartości jednej zmiennej ( $X \uparrow$ ) odpowiada spadek wartości drugiej zmiennej ( $Y \downarrow$ ) i odwrotnie (Rys. 9b).

Korelacja może przybierać postać związku liniowego, gdy wartości zmiennej  $X$  i  $Y$  naniesione na wykres układają się wzdłuż linii prostej (przekątnej) – im bardziej są zbliżone do tej linii, tym związek jest silniejszy, a im bardziej oddalają się, tym związek jest słabszy (Rys. 9a i b). Korelacja prawie równa zeru (0) wystąpi wówczas gdy np. wartościom zmiennej  $X$  odpowiadają raz wysokie raz średnie lub niskie wartości zmiennej  $Y$  (Rys. 9c).

**Korelacja nieliniowa** występuje wówczas, gdy wartości zmiennej  $X$  i  $Y$  naniesione na wykres rozkładają się tak, jak na Rys. 9d i 9e.

Jeśli występuje potrzeba precyzyjnego określenia związku między zmiennymi ilościowymi, można stopień tego związku określić przy pomocy współczynnika korelacji  $r$  Pearsona. Współczynnik korelacji  $r$  Pearsona można stosować, gdy są spełnione następujące warunki:

1. obie zmienne są zmiennymi ilościowymi,
2. związek między zmiennymi jest liniowy,
3. obie zmienne powinny mieć rozkład normalny. W praktyce występują rozkłady jedynie zbliżone do rozkładu normalnego, czyli takie, w których wartości zmiennej  $X$  i  $Y$  tworzą rozkłady jednowierzchołkowe i możliwie symetryczne. Jeśli nie ma potrzeby dokładnego określenia „normalności” rozkładu, to wystarczające może być przedstawienie rozkładów zmiennej  $X$  i  $Y$  na wykresach. Potwierdzeniem, że rozkłady są zbliżone w swym kształcie do normalnego, będą wykresy w kształcie pagórka o łagodnych zboczach.

Do obliczenia współczynnika korelacji stosuje się różne wzory, równoznaczne pod względem matematycznym. Można stosować je zamiennie. O wyborze odpowiedniego wzoru przesądzają głównie względy praktyczne, cele, którymi kieruje się badacz podczas dokonywania analizy statystycznej uzyskanych wyników, a także liczba pomiarów ( $N$ ).

W badaniach pedagogicznych stosunkowo często wykorzystywany jest wzór na obliczanie współczynnika korelacji z danych pierwotnych, czyli z takich, które uzyskano w badaniu, bez ich przetworzenia. Jest to sposób nie

wymagający obliczania średnich arytmetycznych i odchyłeń standardowych, bardzo wygodny szczególnie wówczas gdy  $N$  i wartości pomiarowe nie są zbyt duże.

Wzór na obliczanie współczynnika  $r$  Pearsona z danych nie pogrupowanych ma następującą postać :

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

gdzie:

$X$  – wartości zmiennej  $X$ ,

$Y$  – wartości zmiennej  $Y$ ,

$N$  – liczba par wyników.

Współczynnik korelacji Pearsona ( $r$ ) określa kierunek i siłę związku między badanymi zmiennymi o charakterze ilościowym. Może przybierać wartości w przedziale liczbowym od  $-1$  do  $+1$ . Jeżeli wartość współczynnika korelacji zbliża się do  $0$ , oznacza to, że związek pomiędzy badanymi zmiennymi prawie nie istnieje, jeżeli zbliża się do  $+1$  lub  $-1$  – występuje związek bardzo silny, przyjmujący kierunek dodatni lub ujemny.

Interpretacja współczynnika korelacji zależy w dużym stopniu od tego, do czego ma on służyć lub od powodów, dla których go obliczono. Interpretacja jest tu kwestią względną. Analizując jednak korelację szeroko, nie wchodząc w szczegóły jej zastosowania, można za J.P. Guilfordem określić dla różnych wielkości  $r$  różne stopnie związku:<sup>23</sup>

|              |                            |
|--------------|----------------------------|
| poniżej 0,20 | – związek słaby,           |
| 0,20 - 0,40  | – korelacja niska,         |
| 0,40 - 0,70  | – korelacja umiarkowana,   |
| 0,70 - 0,90  | – korelacja wysoka,        |
| 0,90 - 1,00  | – korelacja bardzo wysoka. |

Taką samą interpretację można zastosować do  $r$  dodatniego i ujemnego. Interpretacja  $r$ , jak już wskazano, zależy w dużym stopniu od rozpatrywanego problemu. Jeśli np. współczynnik  $r = 0,40$  i oznacza związek między stanem wiadomości uczniów i kwalifikacjami pedagogicznymi ich nauczycieli, to taką korelację można by uznać za stosunkowo silną. Gdyby natomiast współczynnik  $r = 0,40$  i oznaczał związek między wynikami w rozwiązywaniu zadania tego samego rodzaju w pierwszej

<sup>23</sup> J.P. Guilford, *Podstawowe ...*, op. cit., s. 157.

i drugiej serii badań, wówczas trzeba by uznać taką korelację za dziwnie „niską”. Aby ocenić wielkość danego współczynnika, należy go porównać z przeciętnymi wielkościami otrzymywanymi w odniesieniu do tego samego problemu w innych badaniach. Ponadto należy zawsze uwzględnić liczbę pomiarów ( $N$ ) stanowiących podstawę do obliczeń. W związku z tym, bez uwzględnienia tych warunków nie należy formułować jednoznacznych wniosków odnośnie tego, że obliczony w konkretnym przypadku współczynnik korelacji  $r$  oznacza związek „słaby”, „niski” lub „wysoki”<sup>24</sup>.

Współczynniki korelacji nie mogą być uważane za wielkości wyrażone w procentach, a więc  $r = 0,80$  nie oznacza, że wartości obu zmiennych w 80 przypadkach na 100 możliwych będą się pokrywać.

Niżej przedstawiono sposób obliczania współczynnika korelacji Pearsona z danych pierwotnych. W przeprowadzonych badaniach uwarunkowań osiągnięć szkolnych z języka polskiego w kl. III w miejscowości X, podjęto próbę ustalenia występowania związku korelacyjnego między liczbą przeczytanych książek w ciągu semestru (zmienna  $X$ ) i ocenami z języka polskiego uzyskanymi na koniec semestru (zmienna  $Y$ ). W badaniach uczestniczyło 20 losowo wybranych osób (Tablica 40).

Tablica 40

Obliczanie współczynnika korelacji  $r$  z danych pierwotnych

| Lp. | Liczba książek<br>$X$ | Ocena z języka<br>polskiego<br>$Y$ | $X^2$               | $Y^2$              | $XY$               |
|-----|-----------------------|------------------------------------|---------------------|--------------------|--------------------|
| 1   | 7                     | 2                                  | 49                  | 4                  | 14                 |
| 2   | 20                    | 5                                  | 400                 | 25                 | 100                |
| 3   | 11                    | 4                                  | 121                 | 16                 | 44                 |
| 4   | 20                    | 6                                  | 400                 | 36                 | 120                |
| 5   | 4                     | 1                                  | 16                  | 1                  | 4                  |
| 6   | 17                    | 3                                  | 289                 | 9                  | 36                 |
| 7   | 12                    | 3                                  | 144                 | 9                  | 51                 |
| 8   | 19                    | 4                                  | 361                 | 16                 | 76                 |
| 9   | 8                     | 5                                  | 64                  | 25                 | 40                 |
| 10  | 20                    | 5                                  | 400                 | 25                 | 100                |
| 11  | 16                    | 4                                  | 196                 | 16                 | 64                 |
| 12  | 6                     | 2                                  | 36                  | 4                  | 12                 |
| 13  | 13                    | 4                                  | 169                 | 16                 | 52                 |
| 14  | 20                    | 6                                  | 400                 | 36                 | 120                |
| 15  | 15                    | 4                                  | 225                 | 16                 | 60                 |
| 16  | 13                    | 3                                  | 169                 | 9                  | 39                 |
| 17  | 11                    | 4                                  | 121                 | 16                 | 44                 |
| 18  | 15                    | 5                                  | 225                 | 25                 | 75                 |
| 19  | 10                    | 3                                  | 100                 | 9                  | 30                 |
| 20  | 8                     | 3                                  | 64                  | 9                  | 24                 |
|     | $\Sigma X = 265$      | $\Sigma Y = 76$                    | $\Sigma X^2 = 4009$ | $\Sigma Y^2 = 322$ | $\Sigma XY = 1105$ |

<sup>24</sup> G. Clauss, H. Ebner, *Podstawy ...*, op. cit., s. 128.

Podstawiając do wzoru otrzymano:

$$r = \frac{20 \cdot 1105 - 265 \cdot 76}{\sqrt{[20 \cdot 4009 - (265)^2][20 \cdot 322 - (76)^2]}} = + 0,760$$

Obliczony współczynnik korelacji jest dodatni i wynosi 0,760. Związek między liczbą przeczytanych książek przez badanych uczniów w ciągu semestru i oceną z języka polskiego jest znaczny, a ścisłość tych powiązań bardzo wysoka. Oznacza to, że uczniowie, którzy więcej czytali, uzyskiwali wyższe oceny z języka polskiego i jednocześnie ci, którzy czytali mniej, uzyskiwali oceny niższe<sup>25</sup>. Pamiętać jednak należy, że uwarunkowania osiągnięć szkolnych są bardzo złożone i należałoby badać również i inne zmienne: tempo czytania, rozumienie treści, motywację i warunki do nauki, itp. oraz poszukiwać między nimi związków, np. wykorzystując korelację cząstkową i wieloraką.

Współczynnik korelacji  $r$  Pearsona może być również obliczony z pogrupowanych wartości zmiennej  $X$  i  $Y$ , przedstawionych w szeregach rozdzielczych.

## 2.2. Współczynnik korelacji rangowej Spearmana

Występowanie związku lub jego brak między zmiennymi wyrażonymi na skali rangowej można przedstawić w postaci tabelarycznej. W Tablicy 41 przedstawiono związek między pozycją zajmowaną przez ucznia w klasie w zakresie jego popularności i pozycją wyznaczoną przez nauczyciela z tytułu jego aktywności podczas lekcji.

*Tablica 41*

*Związek między popularnością ucznia w klasie  
a opinią nauczyciela o jego aktywności*

| Uczeń | Popularność w klasie<br>$X$ | Opinia nauczyciela<br>$Y$ |
|-------|-----------------------------|---------------------------|
| A     | 1                           | 4                         |
| B     | 2                           | 5                         |
| C     | 3                           | 2                         |
| D     | 4                           | 1                         |
| :     | :                           | :                         |

*Źródło: dane umowne*

---

<sup>25</sup> Sprawdzenie istotności współczynnika korelacji – *patrz* rozdz. VI, pkt. 1.2.

Związek między zmiennymi wyrażonymi na skali rangowej może być również przedstawiony na wykresie. Jeśli istnieje potrzeba dokładnego określenia związku między zmiennymi na skali rangowej, wówczas ma zastosowanie współczynnik korelacji rangowej (pozycyjnej) Spearmana. Ma zastosowanie również wtedy gdy obie zmienne lub któraś z nich są wartościami, które można porangować i jednocześnie liczba obserwacji nie jest duża ( $N < 30$ ).

W badaniach pedagogicznych współczynnik korelacji rangowej ma praktyczne zastosowanie do poszukiwania związku między np. pozycjami, które zajmują uczniowie z punktu widzenia ich wyników w nauce a pozycjami, które zajmują w rankingu popularności w klasie; opinią o nauczycielach wyrażoną na skali rangowej przez dyrektora szkoły i wizytatora; oceną z jakiegoś przedmiotu a opinią nauczyciela tego przedmiotu o aktywności ucznia podczas lekcji itp.

Współczynnik korelacji rangowej interpretuje się analogicznie jak współczynnik korelacji Pearsona, zawiera się on również w zakresie od +1 do -1, ale w tym wypadku zmienne wyrażane są na skali rangowej (pozycyjnej).

Wzór na obliczanie współczynnika korelacji rangowej ma następującą postać:

$$R_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)},$$

gdzie:

$d_i$  – różnica między rangami ( $d_i = R_1 - R_2$ ),

$N$  – liczebność zbiorowości.

W Tablicy 42 przedstawiono przykład przedstawiania związku między wynikami w nauce uczniów kl. III (średnia ocen) oraz ich popularnością w klasie.

Obliczanie współczynnika korelacji rangowej ( $R_s$ )

| Lp. | Średnia ocen | Popularność<br>$R_1$ | $R_2$ | $d_i = R_1 - R_2$ | $d_i^2$              |
|-----|--------------|----------------------|-------|-------------------|----------------------|
| 1   | 4,0          | 15                   | 9     | 6                 | 36                   |
| 2   | 3,5          | 10                   | 12,5  | -2,5              | 6,25                 |
| 3   | 4,0          | 8                    | 9     | -1                | 1                    |
| 4   | 5,7          | 2                    | 3     | -1                | 1                    |
| 5   | 3,5          | 9                    | 12,5  | -3,5              | 12,25                |
| 6   | 3,2          | 11                   | 14,5  | -3,5              | 12,25                |
| 7   | 3,2          | 12                   | 14,5  | -2,5              | 6,25                 |
| 8   | 6,0          | 1                    | 1     | 0                 | 0                    |
| 9   | 5,8          | 3                    | 2     | 1                 | 0                    |
| 10  | 5,2          | 7                    | 4     | 3                 | 9                    |
| 11  | 4,0          | 4                    | 9     | -5                | 25                   |
| 12  | 5,1          | 6                    | 5     | 1                 | 1                    |
| 13  | 4,5          | 13                   | 6     | 7                 | 49                   |
| 14  | 3,7          | 5                    | 11    | -6                | 36                   |
| 15  | 4,2          | 14                   | 7     | 7                 | 49                   |
| 16  | 3,0          | 16                   | 16    | 0                 | 0                    |
|     |              |                      |       |                   | $\Sigma d_i^2 = 244$ |

Obliczeń dokonuje się w następujący sposób :

- Wyniki popularności uczniów w klasie przedstawione są na skali rangowej ( $R_1$ ).
- Średnie ocen uczniów należy porangować ( $R_2$ ), przypisując rangę 1 wartości najwyższej. Jeśli występują np. dwa takie same wyniki, którym należałoby przyporządkować kolejne rangi, np. 12 i 13, wówczas obliczamy średnią obu rang i obu wynikom przypisujemy rangę 12,5. W tej sytuacji kolejny wynik otrzyma rangę 14. Podobnie postępuje się w przypadku trzech i więcej jednakowych wyników.
- Dla każdej pary rang ustalamy ich różnicę:  $d_i = R_1 - R_2$ ,
- Podnosimy do kwadratu różnicę rang:  $d_i^2$ ,
- Sumujemy kwadraty różnic, aby otrzymać  $\Sigma d_i^2$  podstawiamy do wzoru:

$$R_s = 1 - \frac{6 \Sigma d_i^2}{N(N^2 - 1)} = 1 - \frac{6 \cdot 244}{16(16^2 - 1)} = 0,64$$

Z wartości współczynnika korelacji rangowej,  $R_s = 0,64$  wynika, że zachodzi umiarkowany związek między wynikami w nauce ucznia i jego popularnością w klasie. Dodatnia wartość współczynnika korelacji rangowej oznacza, że często wysokiej pozycji pod względem wyników w nauce odpowiada wysoka pozycja w zakresie popularności w klasie i odwrotnie. Podkreślić jednak należy, że na pozycję ucznia w klasie III w zakresie jego popularności mają wpływ również i inne czynniki, a wyniki w nauce są tylko jednym spośród nich.

Nie należy przeceniać znaczenia analizy związków korelacyjnych. Nic nie dzieje się bez przyczyny, a poszczególne zjawiska, zdarzenia, stany rzeczy są powiązane w różnorodny sposób, warunkują się nawzajem. Wśród czynników determinujących poszczególne zjawiska i zdarzenia występują bardziej i mniej istotne, a także o jednostronnej i dwustronnej zależności. Badacz powinien określać istotne uwarunkowania badanych zjawisk oraz występujący typ zależności i wówczas mierzyć ścisłość związku między zmiennymi.

## Pytania i zadania

1. Omów rodzaje związków między zmiennymi.
2. Związek funkcjonalny i korelacyjny – wskaż różnice.
3. Przedstaw sposób prezentacji związku między cechami jakościowymi.
4. Jak można określić związek między cechami jakościowymi?
5. W kwestionariuszu skierowanym do uczniów kl. VII, jedno z pytań sformułowano następująco: czy uczestniczysz w zajęciach jakiegokolwiek koła zainteresowań? Uzyskano następujące odpowiedzi: spośród 46 dziewcząt, 20 udzieliło odpowiedzi pozytywnej (tak), a spośród 32 chłopców, 17 udzieliło odpowiedzi negatywnej (nie). Ustal, czy płeć uczniów ma związek z uczestnictwem w kołach zainteresowań.
6. Przedstaw sposoby prezentacji związku między cechami ilościowymi.
7. Co jest graficznym obrazem związku korelacyjnego?
8. Jakie warunki powinny być spełnione, aby można było obliczać współczynnik korelacji  $r$  Pearsona?
9. Jaka jest interpretacja współczynnika korelacji  $r$  Pearsona?
10. W rezultacie badań frekwencji w szkole (liczba opuszczonych godzin w szkole –  $X$ ) i jej związku z wynikami semestralnymi z matematyki uczniów kl. V ( $Y$ ) uzyskano:

| Uczeń | 1  | 2 | 3  | 4  | 5  | 6 | 7 | 8 | 9  | 10 | 11 | 12 | 13 | 14 | 15 |
|-------|----|---|----|----|----|---|---|---|----|----|----|----|----|----|----|
| $X$   | 12 | 9 | 27 | 38 | 10 | 4 | 8 | 4 | 17 | 40 | 5  | 29 | 50 | 28 | 10 |
| $Y$   | 4  | 4 | 3  | 2  | 5  | 6 | 6 | 4 | 2  | 1  | 4  | 3  | 1  | 3  | 5  |

Ustal kształt rozkładu wyników dla zmiennej  $X$  i  $Y$ , podejmij decyzję o obliczeniu lub też nie współczynnika korelacji  $r$  Pearsona. W przypadku obliczania  $r$  dokonaj interpretacji.

11. Wyniki testu z fizyki uczniów kl. VIII ( $X$ ) zestawiono z ocenami uzyskanymi przez tych uczniów za semestr I ( $Y$ ) i uzyskano następujące wyniki:

| Uczeń | 1 | 2  | 3  | 4 | 5  | 6  | 7  | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-------|---|----|----|---|----|----|----|---|---|----|----|----|----|----|----|----|
| $X$   | 5 | 14 | 25 | 9 | 18 | 13 | 10 | 8 | 9 | 14 | 19 | 6  | 10 | 4  | 7  | 15 |
| $Y$   | 2 | 4  | 6  | 3 | 5  | 4  | 3  | 4 | 3 | 4  | 5  | 2  | 3  | 1  | 2  | 4  |

Ustal kształt rozkładu wyników dla zmiennej  $X$  i  $Y$ , podejmij decyzję o obliczeniu lub też nie współczynnika  $r$  Pearsona. W przypadku obliczania  $r$  dokonaj interpretacji.

12. Przedstaw sposoby prezentacji związków między zmiennymi wyrażonymi na skali rangowej.
13. Kiedy można obliczać współczynnik korelacji rangowej  $R_s$ ?
14. Jaka jest interpretacja współczynnika korelacji rangowej  $R_s$ ?
15. Określ związek między wynikami testu z historii ( $X$ ) a oceną zainteresowania tym przedmiotem, wyrażoną przez nauczyciela ( $Y$ )?

a)

| Lp. | $X$ | $Y$ |
|-----|-----|-----|
| 1   | 10  | 4   |
| 2   | 20  | 3   |
| 3   | 13  | 8   |
| 4   | 16  | 5   |
| 5   | 13  | 10  |
| 6   | 12  | 7   |
| 7   | 18  | 2   |
| 8   | 23  | 1   |
| 9   | 7   | 9   |
| 10  | 13  | 6   |

b)

| Lp. | $X$ | $Y$ |
|-----|-----|-----|
| 1   | 18  | 1   |
| 2   | 11  | 10  |
| 3   | 8   | 2   |
| 4   | 11  | 8   |
| 5   | 8   | 9   |
| 6   | 13  | 4   |
| 7   | 15  | 3   |
| 8   | 11  | 7   |
| 9   | 13  | 6   |
| 10  | 14  | 5   |

16. Dla wyników badań zamieszczonych w zadaniu 10 oblicz związek między pozycją uzyskaną przez poszczególnych badanych uczniów w zakresie liczby opuszczonych godzin w szkole ( $X$ ) a pozycją, którą uzyskali w klasie w zakresie wyników semestralnych z matematyki ( $Y$ ).

## Rozdział V

# PODSTAWY WNIOSKOWANIA STATYSTYCZNEGO

W części pierwszej skryptu przedstawiono parametry opisowe zbiorowości statystycznej, czyli metody dokonywania charakterystyki badanej zbiorowości. Jak już wcześniej wskazano, najczęściej badaniu podlega nie cała populacja, lecz jej część pobrana w specjalny sposób, a nazywana próbą. Często potrzebą badacza jest, aby na podstawie informacji uzyskanych z badanej próby formułować sądy o całej populacji. Miary charakteryzujące badaną próbę, np. średnia arytmetyczna, odchylenie standardowe to **statystyki**, a te same miary odnoszące się do całej populacji generalnej to **parametry**. Proces formułowania sądów o całej populacji generalnej na podstawie miar charakteryzujących badaną próbę to **wnioskowanie statystyczne**. Wyróżnia się dwa sposoby wnioskowania:

- estymacja parametrów (szacowanie),
- weryfikacja hipotez.

Niżej przedstawiono tylko podstawy estymacji, nieco obszerniej omówiono zagadnienie weryfikacji hipotez.

# 1. Zmienna losowa i jej rodzaje

Podstawą wnioskowania jest rachunek prawdopodobieństwa. Rachunek prawdopodobieństwa jako podstawa statystyki matematycznej zajmuje się analizą prawidłowości zachodzenia tzw. zdarzeń losowych, czyli takich, które zależą od przypadku. Zdarzeniem losowym będzie np. wylosowanie podczas egzaminu spośród dwóch pytań – jednego pytania. Przed losowaniem nie można określić, które pytanie zostanie wylosowane, pierwsze czy drugie. Można tylko określić zbiór tzw. **zdarzeń elementarnych**, czyli takich, z których jedno może zaistnieć wykluczając zaistnienie drugiego.

Wynikiem jakiegoś losowania może być zdarzenie mieszczące się w zbiorze zdarzeń elementarnych. Każdemu takiemu zdarzeniu można przyporządkować pewną liczbę rzeczywistą w taki sposób, że każdemu zdarzeniu odpowiada tylko jedna liczba, ale zdarzeń elementarnych może być wiele. Stąd każde zdarzenie elementarne jest reprezentowane przez liczbę przybierającą różne wartości, nazywaną **zmienną losową**<sup>26</sup>.

Jeśli spośród 120 uczniów wylosowano 20, to wylosowanie konkretnego ucznia jest zdarzeniem losowym. Zapytano ich o wiek, podawany w liczbach (lata, miesiące, dni). Właśnie te liczby są zmiennymi losowymi, ponieważ przypadek zdecydował, jakie one konkretnie były oraz że wylosowano uczniów właśnie w takim, a nie innym wieku.

Zmienne losowe można podzielić na: zmienne losowe skokowe (dyskretne) i zmienne losowe ciągłe, podobnie jak w przypadku podziału cech.

**Zmienna losowa skokowa** to taka zmienna, która przyjmuje skończoną lub przeliczalną liczbę wartości w postaci liczb całkowitych lub naturalnych, a wystąpienie każdej z tych wartości jest zdarzeniem o określonym prawdopodobieństwie. Najbardziej znanym rozkładem zmiennej losowej skokowej jest rozkład dwumianowy.

**Zmienna losowa ciągła** to taka zmienna, która przyjmuje wartości mieszczące się w pewnym przedziale liczbowym (skończonym lub nieskończonym), ale nieprzeliczalnym. Jej właściwość stanowi to, że poszczególnym zdarzeniom elementarnym nie można przypisać prawdopodobieństwa ich występowania, ponieważ suma tych prawdopodobieństw zawsze jest równa nieskończoności<sup>27</sup>.

Najbardziej znanym i stosowanym rozkładem zmiennej losowej ciągłej jest rozkład normalny.

---

<sup>26</sup> L.A. Gruszczyński. *Elementy statystyki dla socjologów*. Katowice 1986. s. 80-81.

<sup>27</sup> *Tamże*, s. 81 i 88.

## 2. Rozkład normalny

Duże znaczenie rozkładu normalnego wynika z faktu, że wiele zjawisk społecznych i przyrodniczych kształtuje się według tego rozkładu.

Graficznym obrazem zmiennej o rozkładzie normalnym jest krzywa normalna, o kształcie dzwonowatym, symetrycznym, jednomodalnym, dla której średnia arytmetyczna, mediana i dominanta są równe. Należy jednak wskazać, że żaden zbiór danych nigdy nie odpowiada idealnie rozkładowi normalnemu. Odchylenie krzywej otrzymane z rozkładu empirycznego od krzywej normalnej może dotyczyć asymetrii rozkładu (Rys. 2a, b, s. 64) lub stopnia jego spłaszczenia (Rys. 5a, b, s. 65).

Równanie krzywej normalnej ma postać:

$$y = \frac{N}{\delta\sqrt{2\Pi}} e^{\frac{-x^2}{2\delta^2}},$$

gdzie:

$y$  – liczebność odpowiadająca poszczególnym wartościom zmiennej  $X$ ,

$N$  – ogólna liczebność zbiorowości,

$\delta$  – odchylenie standardowe zmiennej  $X$ ,

$\Pi$  – stała wartość ( $\approx 3,1416$ ),

$e$  – podstawa logarytmów naturalnych ( $\approx 2,718$ ).

$x$  – odchylenie pomiaru od średniej arytmetycznej ( $X - \bar{x}$ ).

Dla wyznaczenia krzywej normalnej potrzebne są wartości  $N$  oraz  $\delta$ . Posiadając takie dane dla szeregu rozdzielczego można określić całą krzywą normalną, najlepiej dobraną do danych empirycznych, którymi dysponuje badacz. Krzywa normalna, będąca graficznym obrazem rozkładu normalnego, nazywana także krzywą Gaussa, ma kilka ważnych własności<sup>28</sup>:

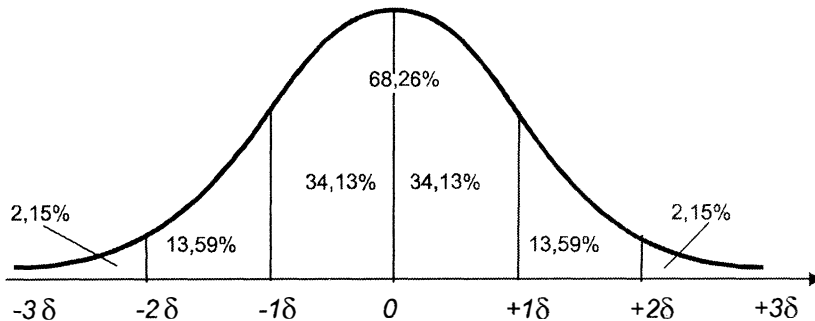
- rozkład normalny opisany jest w przedziale nieokreślonym ( $-\infty < X < +\infty$ ); w rozkładzie empirycznym zmienność poszczególnych obserwacji występuje w obszarze zmienności;
- rozkład normalny jest rozkładem symetrycznym, o ściśle oznaczonym kształcie; największą częstotliwość występowania mają zdarzenia o średniej wartości badanej zmiennej; częstotliwość występowania zdarzeń

---

<sup>28</sup> Cz. Nowaczyk, *Podstawy ...., op. cit.*, s. 125-126.

maleje wraz ze wzrostem odchylenia zmiennej losowej od jej średniej arytmetycznej;

- wewnątrz przedziału od  $\bar{x} - \delta$  do  $\bar{x} + \delta$  krzywa normalna jest wypukła, a na zewnątrz jest wklęsła;
- jeżeli całe pole pod krzywą rozkładu normalnego przyjąć jako 100%, to wartość odchylenia standardowego dzieli je w ściśle określony sposób (Rys. 10).



Rys. 10. Procent powierzchni pod krzywą normalną przy różnych odchyleniach standardowych

Całkowita powierzchnia pod krzywą normalną wynosi  $N$ , czyli jest równa liczbie wszystkich obserwacji. Jeśli przyjmie się powierzchnię pod krzywą rozkładu normalnego jako 100% lub 1, to z analizy proporcji powierzchni między średnią arytmetyczną i odchyleniem standardowym wynika, że przedział rozciągający się od średniej arytmetycznej minus jedno odchylenie standardowe do średniej arytmetycznej plus jedno odchylenie standardowe obejmuje około 2/3 globalnej liczby obserwacji z populacji, czyli między średnią minus jedno odchylenie standardowe ( $-1\delta$ ) mieści się 1/3 wszystkich obserwacji, a także między średnią i plus jedno odchylenie standardowe ( $+1\delta$ ) mieści się 1/3 wszystkich obserwacji. Można to sprawdzić, korzystając z Tablicy I (zamieszczonej w aneksie) przez wyszukanie proporcji powierzchni między średnią arytmetyczną i  $\delta$  (tj.  $z$ , czyli wynik standaryzowany  $z = \frac{x}{\delta} = 1$ ).

Powierzchnia podana w Tablicy I wynosi 0,3413, co w procentach daje 34,13% obserwacji, czyli jest to około 1/3. Z Tablicy I wynika również, że między średnią i punktem odległym o  $2\delta$  ( $z = 2$ ), czyli w obszarze średnia minus dwa odchylenia ( $-2\delta$ ) oraz średnia plus dwa odchylenia ( $+2\delta$ ) można się spodziewać 0,4772 całej powierzchni albo 47,72% obserwacji. Od średniej do

$3\delta$  ( $z=3$ ) znajduje się 49,87% obserwacji, czyli w zakresie od  $-3\delta$  do  $+3\delta$  można się spodziewać 99,74% obserwacji. Oznacza to, że w zbiorowości liczącej  $N=10000$  poza rozpiętością od  $-3\delta$  do  $+3\delta$  można się spodziewać tylko 26 obserwacji ( $10000 - 9974 = 26$ )<sup>29</sup>.

Jak już wskazano, rozkład normalny jest rozkładem teoretycznym. Występujące w badaniach pedagogicznych i psychologicznych rozkłady empiryczne są zazwyczaj rozkładami zbliżonymi do rozkładu normalnego lub też odbiegają od niego. Różnice między rozkładem wartości empirycznych (rozkład badanej próby) i rozkładem normalnym mogą być rezultatem:

- nie przeprowadzonego poprawnie grupowania (jeśli np. większość jednostek koncentruje się w jednej klasie lub są klasy puste);
- błędów popełnionych przy pobieraniu próby, np. tendencyjność, powodująca zniekształcenie rozkładu, jego asymetrię, będąca skutkiem niejednakowego prawdopodobieństwa dostania się do próby poszczególnych jednostek;
- czasem niedostatecznie licznej próby.

### 3. Estymacja parametrów

**Estymacja** statystyczna jest rodzajem wnioskowania polegającym na szacowaniu parametrów populacji generalnej na podstawie statystyk obliczonych dla wylosowanej z tej populacji próby. Wyróżnia się dwa rodzaje estymacji:

- estymację punktową,
- estymację przedziałową.

**Estymacja punktowa** polega na znalezieniu konkretnej liczby dla każdego szacowanego parametru. Jednak przy szacowaniu nieznanego parametru można popełnić pewien błąd wynikający z faktu, że np. wartość średniej arytmetycznej z próby różni się od średniej z populacji. Błąd ten jest uwzględniany przy szacowaniu i nosi nazwę **błędu standardowego** szacunku.

---

<sup>29</sup> J.P. Guilford, *Podstawowe .... op. cit.*, s. 149-150.

**Estymacja przedziałowa** polega na wyznaczaniu pewnego przedziału liczbowego, wewnątrz którego znajduje się – z pewnym prawdopodobieństwem – estymowany parametr. Przedział ten nosi nazwę **przedziału ufności**, a prawdopodobieństwo tego, że mieści się w nim wartość szacowanego parametru, to tzw. **współczynnik ufności**. Przedział ufności można zwiększyć lub zmniejszyć, a tym samym zmniejszyć lub zwiększyć współczynnik ufności, dlatego mówi się, że określany jest przedział ufności, który z pewnym prawdopodobieństwem będzie zawierał szacowany parametr. Im dłuższy przedział ufności, tym mniejsza precyzja szacowania parametru, ale za to większa jego pewność.

Sposoby szacowania średniej arytmetycznej, odchylenia standardowego i innych parametrów precyzyjnie przedstawiono w cytowanych już opracowaniach<sup>30</sup>.

## 4. Etapy wnioskowania statystycznego

W badaniach wyczerpujących postępowanie badawcze ogranicza się do opisu statystycznego, rezultatem którego są **stwierdzenia kategoryczne**. Natomiast w przypadku badań reprezentacyjnych, po dokonaniu opisu statystycznego wyników próby przeprowadza się wnioskowanie statystyczne umożliwiające uogólnienie uzyskanych wyników na całą populację generalną. Rezultatem takiego uogólnienia są **stwierdzenia hipotetyczne**, przyjmowane z określonym prawdopodobieństwem niepewności, dotyczące całej populacji.

Wnioskowanie statystyczne obejmuje kilka etapów:<sup>31</sup>

- I. **Opis populacji generalnej.** Zawiera określenie zakresu populacji, wyszczególnienie liczby i rodzaju zmiennych (zależnych, niezależnych i innych), charakteryzujących populację, określenie typu skal pomiarowych, na których wyrażone zostaną wyniki obserwacji próby, a także przyjęcie odpowiednich założeń odnośnie rozkładu wyróżnionych zmiennych.

---

<sup>30</sup> G. Claus, H. Ebner, *Podstawy ...*, *op. cit.*; Cz. Nowaczyk, *Podstawy ...*, *op. cit.*

<sup>31</sup> A. Góralski, *Metody ...*, *op. cit.*, s. 143-158.

- II. **Opis próby.** Zawiera sposób pobrania próby i udokumentowanie jej reprezentatywności, a także opis właściwości próby – jej liczebność, wyliczenie zmiennych i właściwych im skal pomiarowych, charakterystykę wyników, określenie wartości miar zmiennych, typów ich rozkładów.
- III. **Sformułowanie hipotezy statystycznej.** To etap formułowania pewnych przypuszczeń (w postaci hipotez zerowych i alternatywnych), które mogą dotyczyć istotności różnic uzyskanych wyników, niezależności badanych cech, a także zgodności rozkładu empirycznego z rozkładem normalnym lub rozkładem tej samej zmiennej w innych populacjach.
- IV. **Weryfikacja hipotez statystycznych.** Jest to etap polegający na zastosowaniu specjalnych testów statystycznych, przy pomocy których dokonuje się oceny stopnia wiarygodności hipotez statystycznych, w rezultacie czego następuje ich odrzucenie lub przyjęcie.
- V. **Sformułowanie wniosku statystycznego.** Jest to etap kończący wnioskowanie statystyczne, w którym następuje porównanie wartości funkcji testu z wartością krytyczną (z odpowiedniej tablicy) i sformułowanie wniosku, mogącego przyjąć następującą postać:
- odrzucam  $H_0$  (i przyjmuję  $H_1$ ) lub stwierdzam brak podstaw do odrzucenia  $H_0$ , albo
  - odrzucam  $H_0$  (i przyjmuję  $H_1$ ) lub przyjmuję  $H_0$  (i odrzucam  $H_1$ ).

## 5. Hipotezy statystyczne

**Hipoteza statystyczna** to każde takie przypuszczenie, które dotyczy postaci rozkładu lub wartości parametrów rozkładu pewnej zmiennej losowej, które może być weryfikowane statystycznie, tzn. w oparciu o wyniki zaobserwowane w próbie losowej.<sup>32</sup>

Z takiej definicji hipotezy statystycznej wynikają dwa rodzaje hipotez:

---

<sup>32</sup> Z. Pawłowski, *Wstęp do statystyki matematycznej*, Warszawa 1966, s. 324.

- hipotezy parametryczne, dotyczące parametrów zbiorowości,
- hipotezy nieparametryczne, dotyczące postaci rozkładu zmiennej losowej.

Formułowane w badaniach hipotezy robocze są zaprzeczeniem hipotez statystycznych, nazywanych inaczej hipotezami zerowymi. **Hipoteza zerowa** ( $H_0$ ) jest przypuszczeniem, które należy zweryfikować i odnosi się zawsze do populacji, jest zawsze tak sformułowana, aby jej odrzucenie oznaczało dla badacza przyjęcie **hipotezy alternatywnej** – roboczej ( $H_1$ ). Weryfikacji hipotezy alternatywnej ( $H_1$ ) dokonuje się pośrednio, ponieważ bezpośredniemu sprawdzeniu poddaje się hipotezę zerową ( $H_0$ ). Dopiero przyjęcie lub odrzucenie hipotezy zerowej ( $H_0$ ) powoduje konsekwencje dla hipotezy alternatywnej ( $H_1$ ) nakazujące jej przyjęcie lub odrzucenie.

Weryfikacja hipotezy zerowej ( $H_0$ ) przy pomocy odpowiedniego testu statystycznego prowadzi do jej przyjęcia lub odrzucenia. Zarówno w przypadku jej przyjęcia, jak i odrzucenia istnieje niebezpieczeństwo, że jest to decyzja błędna. Badacz naraża się tutaj na popełnienie błędów:

- **błędu pierwszego rodzaju**, polegającego na odrzuceniu  $H_0$ , pomimo że jest ona prawdziwa; prawdopodobieństwo popełnienia tego błędu określane jest jako  $\alpha$  (alfa),
- **błędu drugiego rodzaju**, polegającego na przyjęciu  $H_0$ , pomimo że jest ona fałszywa; prawdopodobieństwo popełnienia tego błędu określa się jako  $\beta$  (beta).

Z punktu widzenia interesów badacza szczególnie ważny jest błąd pierwszego rodzaju. W związku z tym większość testów statystycznych wymaga jedynie określenia błędów pierwszego rodzaju, ponieważ błąd drugiego rodzaju jest uwzględniony przez dobór odpowiednich wartości krytycznych testu.

**Test statystyczny** to każda jednoznacznie zdefiniowana reguła postępowania określająca warunki, przy których należy sprawdzaną hipotezę przyjąć lub odrzucić<sup>33</sup>. Należy zwrócić uwagę, że w rezultacie takiej weryfikacji uzyskuje się ocenę stopnia wiarygodności hipotezy, która pozwala tę hipotezę przyjąć lub odrzucić, nie oznacza to jednak, że hipoteza jest udowodniona lub nie jest udowodniona. Odpowiednio do rodzaju hipotez wyróż-

---

<sup>33</sup> Z. Pawłowski, *Wstęp ...*, op. cit., s. 325.

nia się dwa rodzaje testów statystycznych, przy stosowaniu których następuje weryfikacja hipotez statystycznych:

- **testy parametryczne** (dotyczące hipotez o parametrach populacji i zawierające założenie o postaci rozkładu),
- **testy nieparametryczne** (dotyczące hipotez o rozkładzie populacji i nie zawierające założeń odnośnie rozkładu – są niezależne od rozkładu).

Przy decyzji o przyjęciu lub odrzuceniu  $H_0$ , bez względu na rodzaj stosowanego testu istnieje możliwość popełnienia błędu  $\alpha$  lub  $\beta$ . W praktyce najczęściej przed przystąpieniem do badań ustala się prawdopodobieństwo popełnienia błędu ( $\alpha$ ), czyli tzw. poziom istotności ( $\alpha$ ).

**Poziom istotności**, nazywany w skrócie poziomem  $\alpha$  jest prawdopodobieństwem uzyskania z testu statystycznego takiej wartości, która nakazuje badaczowi odrzucenie hipotezy zerowej ( $H_0$ ), pomimo że w rzeczywistości hipoteza zerowa jest prawdziwa<sup>34</sup>.

W pedagogice i w psychologii zazwyczaj przyjmuje się  $\alpha = 0,05$ , czyli 5%, co oznacza, że na 100 przypadków może być 5 błędów polegających na tym, że badacz odrzucił słuszną hipotezę zerową. Czasem stawiane są wymagania wyższe i przyjmuje się poziom istotności  $\alpha = 0,01$  (1%) lub nawet  $\alpha = 0,001$ .

Wnioski, których podstawą jest  $\alpha = 0,05$  i  $\alpha = 0,01$ , określa się jako **istotne statystycznie**, a takie, przy których  $\alpha = 0,001$ , nazywa się bardzo istotnymi statystycznie.

Wybór odpowiedniego poziomu istotności często uzależniony jest od rodzaju badanego problemu, a także od konsekwencji wynikających z tytułu podjęcia określonej decyzji wobec hipotezy zerowej ( $H_0$ ).

Ze względu na typ problemu badawczego wyróżnia się trzy rodzaje hipotez statystycznych: istotności różnic, niezależności i zgodności.

**Hipotezy istotności różnic** formułuje się w celu ustalenia, czy uzyskane statystyki charakteryzujące próby losowe różnią się między sobą w sposób istotny, czy też różnice te są przypadkowe. O różnicy istotnej można mówić wtedy gdy statystyki z dwu lub więcej prób losowych odbiegają od siebie tak dalece, że prowadzą do szacunkowego określenia różnych parametrów czyli prawdopodobnie nie pochodzą z tej samej populacji. Różnica przypadkowa

---

<sup>34</sup> J. Brzeziński, *Elementy metodologii badań psychologicznych*, Warszawa 1984, s. 151.

występuje wówczas gdy różnice między statystykami z prób losowych pozwalają na przypuszczenie, że próby te pochodzą z tej samej populacji<sup>35</sup>.

Hipoteza zerowa w hipotezie istotności różnic wyraża przypuszczenie, że dwie próby losowe różnią się między sobą nie w sposób istotny, lecz przypadkowy, a badane próby pochodzą z tej samej zbiorowości generalnej i w związku z tym występujące różnice między statystykami są przypadkowe:

$$H_0: \lambda = \lambda_0$$

$$H_1: \lambda \neq \lambda_0 \text{ ale może też być } H_1: \lambda > \lambda_0 \text{ lub } H_1: \lambda < \lambda_0,$$

gdzie:  $\lambda$  – wartość pewnej miary charakteryzującej model populacji generalnej,  
 $\lambda_0$  – postulowana przez  $H_0$  liczbową wartość tej miary.

Przykład:

$H_0$ : Różnica między średnimi arytmetycznymi wyników testu X uczniów kl. A<sub>1</sub> i A<sub>2</sub> jest nieistotna (wynosi zero) i powstała na skutek przypadku, czyli  $H_0: \bar{x}_1 - \bar{x}_2 = 0$ .  $\bar{x}_1 = \bar{x}_2$ .

$H_1$ : Różnica między średnimi arytmetycznymi wyników testu X uczniów kl. A<sub>1</sub> i A<sub>2</sub> jest istotna i wskazuje, że średnie te nie pochodzą z tej samej populacji, czyli  $H_1: x_1 \neq x_2$ .

Z tak sformułowanej hipotezy alternatywnej ( $H_1$ ) nie wynika, której badanej klasy średnia arytmetyczna jest wyższa. W takim przypadku mamy do czynienia z tzw. dwustronnym stawianiem problemu, lub odnosząc to do sposobu weryfikacji hipotezy, z tzw. dwustronnym testem sprawdzającym.

Jeśli natomiast hipotezie zerowej,  $H_0: \bar{x}_1 - \bar{x}_2 = 0$  przeciwstawimy hipotezę alternatywną  $H_1: x_1 < x_2$  lub  $H_1: x_1 > x_2$ , to mamy wówczas do czynienia z tzw. testem jednostronnym.

---

<sup>35</sup> Cz. Nowaczyk, *Podstawy ...*, op. cit., s. 143.

Inny przykład:

**test dwustronny:**

$H_0$ : średnie wynagrodzenie nauczycieli szkół podstawowych na wsi jest równe  
średniemu wynagrodzeniu nauczycieli szkół podstawowych w mieście.

$H_1$ : średnie wynagrodzenie nauczycieli szkół podstawowych na wsi nie  
jest równe średniemu wynagrodzeniu nauczycieli szkół podstawowych  
w mieście.

**test jednostronny:**

$H_0$ : taka sama, jak przy teście dwustronnym.

$H_1$ : średnie wynagrodzenie nauczycieli szkół podstawowych na wsi jest  
wyższe lub niższe od średniego wynagrodzenia nauczycieli szkół podsta-  
wowych w mieście.

Sposób sformułowania hipotezy alternatywnej uzależniony jest od wiedzy badacza na temat problemu badawczego z literatury problemu, a nie może opierać się na wynikach opisu badanej próby.

**Hipotezy niezależności** formułowane są wówczas, gdy bada się współzależność (współzmiennność) co najmniej dwóch cech charakteryzujących populację generalną i należy rozstrzygnąć, czy współzmiennność ta istnieje. Hipoteza zerowa ( $H_0$ ) w hipotezie niezależności najczęściej jest formułowana w następujący sposób i zakłada, że wartości liczbowe miary współzmienności<sup>36</sup> rozpatrywanego zespołu cech określone dla populacji generalnej są równe zero.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

gdzie:

$\rho$  – oznacza właściwą dla danego problemu miarę  
współzmienności.

Miary współzmienności dwóch cech przybierają zazwyczaj wartości z przedziału od -1 do +1, co powoduje, że mogą tu mieć zastosowanie testy dwustronne i jednostronne, czyli

$$-1 \leq \rho < 0,$$

$$+1 \geq \rho > 0$$

---

<sup>36</sup> Np. współczynnik korelacji.

Niżej podano dwa przykłady formułowania hipotezy zerowej i alternatywnej dla hipotez niezależności:

1.  $H_0$ : Pomiędzy poziomem wiadomości z matematyki mierzonym testem X a poziomem wiadomości z fizyki mierzonym testem Y, w populacji uczniów kl. VIII brak jest związku, czyli

$$\rho = 0.$$

$H_1$ : Pomiędzy poziomem wiadomości z matematyki i fizyki uczniów klasy VIII występuje związek, czyli

$$\rho \neq 0 \text{ lub } H_1: -1 \leq \rho < 0 \text{ lub } H_1: +1 \geq \rho > 0.$$

2.  $H_0$ : Nie istnieje związek między płcią i typem wybranej szkoły ponadpodstawowej w populacji absolwentów kl. VIII, czyli

$$\rho = 0.$$

$H_1$ : Istnieje związek między płcią i typem wybranej szkoły ponadpodstawowej wśród absolwentów kl. VIII, czyli

$$\rho \neq 0 \text{ lub } H_1: \rho > 0.$$

W każdym przykładzie sposób formułowania hipotezy alternatywnej uzależniony jest od wiedzy, którą badacz posiada przed pobraniem próby. pochodzącej ze znajomości literatury problemu.

**Hipotezy zgodności** formułowane są w dwóch postaciach. Pierwsza wystąpi, gdy badacz zainteresowany jest ustaleniem typu rozkładu zmiennej losowej charakteryzującej populację generalną. Hipoteza zerowa zakłada wówczas, że otrzymany rozkład zmiennej jest zgodny z rozkładem normalnym:

$H_0$ : Rozkład empiryczny jest zgodny z rozkładem normalnym.

$H_1$ : Rozkład empiryczny nie jest zgodny z rozkładem normalnym.

Przykład:

$H_0$ : Rozkład wyników testu z ortografii w populacji uczniów kl. III jest zgodny z rozkładem normalnym.

$H_1$ : Rozkład wyników testu z ortografii wśród uczniów kl. III nie jest zgodny z rozkładem normalnym.

Druga postać hipotezy zgodności występuje, gdy badacz przeprowadza porównanie rozkładu tej samej zmiennej w dwu różnych populacjach, np. rozkład wyników testu dojrzałości szkolnej w populacji dzieci zamieszkałych na wsi i w mieście lub dwu różnych zmiennych w tej samej populacji, np. rozkład wyników testu nauczycielskiego z fizyki i wyników testu standaryzowanego z fizyki w populacji uczniów kl. VII.

Również w przypadku hipotez zgodności formułowanie hipotez alternatywnych uzależnione jest od wiedzy badacza na temat badanego problemu przed pobraniem próby.

## Rozdział VI

# WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Weryfikacja hipotez statystycznych następuje w rezultacie stosowania testów statystycznych. Jak już wskazano, w wyniku tego rodzaju weryfikacji uzyskuje się ocenę stopnia wiarygodności hipotezy, zezwalającą na jej przyjęcie lub odrzucenie, ale nie dającą podstaw, aby ją uważać za udowodnioną lub nie.

Testy statystyczne można podzielić na dwie kategorie:

- testy parametryczne, które stosowane są przy weryfikacji hipotez dotyczących parametrów rozkładu cechy w populacji i wymagają założenia o normalności rozkładu,
- testy nieparametryczne, które nie wymagają żadnych założeń odnośnie rozkładu zmiennej i są niezależne od rozkładu wartości cechy.

## 1. Parametryczne testy istotności

Parametryczne testy istotności stosowane są m. in. w celu sprawdzenia hipotez dotyczących równości wartości średnich arytmetycznych w dwóch populacjach o rozkładzie normalnym. Stosuje się je w takich sytuacjach, jak np. porównywanie średnich arytmetycznych uzyskanych w nowej i starej metodzie nauczania czy porównywanie średnich arytmetycznych przed i po działaniu

jakiegoś czynnika, a także sprawdzania hipotez dotyczących braku różnicy pomiędzy średnią z populacji a średnią z próby.

Podczas weryfikacji hipotez parametrycznych często stosowane są następujące testy:

- dla różnych prób, ale najczęściej dla małych – test  $T$ -Studenta,
- dla dużej próby – statystyka  $Z$  (zmienna standaryzowana),
- dla różnych prób – test  $F$ .

Niżej przedstawiono zastosowanie testu  $T$ -Studenta, a także omówiono sprawdzanie istotności współczynnika korelacji miarowej i rangowej. Charakterystykę i stosowanie pozostałych wymienionych testów, a także innych, można znaleźć w cytowanej już literaturze<sup>37</sup>.

### 1.1. Test $T$ -Studenta

Jest testem istotności różnic między dwoma średnimi wyrażonymi na skali przedziałowej. Stosowany jest dla różnych prób, najczęściej dla małych, gdy  $N < 30$ , przy czym im mniejsza jest próba, im mniej stopni swobody, tym bardziej rozkład  $T$ -Studenta staje się spiętrzony w środku. Gdy liczba stopni swobody zwiększa się, rozkład  $T$ -Studenta zbliża się do normalnego.

**Mianem stopni swobody** ( $df$ ) określa się liczbę niezależnych obserwacji (pomiarów) w próbie, tzn. takich, które mogą mieć dowolną wartość. Jeżeli jednak znana jest średnia arytmetyczna tych obserwacji, to nie wszystkie obserwacje mogą mieć dowolne wartości, ale tylko  $N-1$  obserwacji. Np. mając trzy obserwacje o wartości: 3, 4, 5, jeśli znamy ich średnią arytmetyczną i dwie pierwsze obserwacje, to trzecią możemy wyliczyć na tej podstawie, a więc nie jest ona już niezależna i należy ją odjąć przy obliczaniu stopni swobody.

Test  $T$ -Studenta stosowany jest wtedy gdy dokonuje się weryfikacji hipotez dotyczących całej populacji, a warunkiem jego stosowania jest rozkład normalny lub zbliżony do normalnego.

Test  $T$ -Studenta ma najczęstsze zastosowanie do ustalania istotności różnic między dwoma średnimi obliczonymi dla dwóch rozkładów zmiennej ilościowej. Rozkłady zmiennych mogą być próbami niezależnymi (nieskorelowanymi), tzn. mogą pochodzić z dwóch różnych populacji, np. z populacji

---

<sup>37</sup> Por. Cz. Nowaczyk, *Podstawy ..., op. cit.*; G. Clauss, H. Ebner, *Podstawy ..., op. cit.*

dziewcząt i chłopców, dzieci ze szkół miejskich i wiejskich, a także mogą to być próby zależne (skorelowane), tzn. oba rozkłady pochodzą od tych samych badanych.

Hipoteza zerowa ( $H_0$ ) hipotezy istotności różnic zakłada:

$H_0$ : Dwie próby różnią się między sobą nie w sposób istotny, tylko przypadkowy. Badane próby pochodzą z tej samej populacji i w związku z tym obserwowane różnice między statystykami (miarami charakteryzującymi te próby, np.  $\bar{x}$ ,  $s_x$ ) należy tłumaczyć oddziaływaniem czynników przypadkowych.

$H_1$ : Dwie próby różnią się między sobą w sposób istotny, a nie przypadkowy. Badane próby pochodzą z różnych populacji i w związku z tym występują różnice pomiędzy statystykami (miarami charakteryzującymi te próby, np.  $\bar{x}$ ,  $s_x$ ).

W celu przyjęcia lub odrzucenia hipotezy zerowej ( $H_0$ ) przy zastosowaniu testu *T*-Studenta należy obliczyć wartość *t*-empiryczną ( $t_{emp.}$ ) dla wartości badanych zmiennych i porównać ją z wartością *t*-teoretyczną, krytyczną ( $t_{teoret.}$ ), zawartą w odpowiedniej tablicy (Tablica II, aneks) i zastosować następujące kryterium:

$$\begin{aligned} |t_{emp.}| < t_{teoret.; \alpha; df} & - H_0 \text{ należy przyjąć} \\ |t_{emp.}| \geq t_{teoret.; \alpha; df} & - H_0 \text{ należy odrzucić} \end{aligned}$$

W rezultacie porównania obu wartości *t* (teoretycznej i empirycznej) następuje podjęcie decyzji o przyjęciu  $H_0$  lub jej odrzuceniu, a przyjęciu hipotezy alternatywnej ( $H_1$ ).

### 1.1.1. Istotność różnic w próbach niezależnych

Jak już wskazano, próby niezależne to takie, które pochodzą z dwóch różnych populacji. W badaniach pedagogicznych mogą to być dwie próby pobrane np. z populacji dzieci uczęszczających do klasy zerowej w szkole i z populacji dzieci uczęszczających do klasy zerowej w przedszkolu, z populacji dzieci zamieszkających na wsi i w mieście, z populacji studentów studiów dziennych i zaocznych itp.

Gdy dysponujemy średnimi dla takich dwóch prób, to są one średnimi nieskorelowanymi, ponieważ pochodzą z prób niezależnych.

Przy sprawdzaniu istotności różnic między średnimi nieskorelowanymi, pochodzącymi z prób niezależnych ma zastosowanie następujący wzór:

$$t_{emp.} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{N_1 + N_2 - 2}\right) \left(\frac{N_1 + N_2}{N_1 N_2}\right)}},$$

gdzie:

$\bar{x}_1, \bar{x}_2$  – średnie arytmetyczne próby pierwszej i drugiej,

$N_1, N_2$  – liczebności próby pierwszej i drugiej,

$\sum x_1^2, \sum x_2^2$  – suma kwadratów odchyleń każdego wyniku od średniej arytmetycznej w próbie pierwszej i drugiej.

Przykład:

W rezultacie pomiaru testem dojrzałości szkolnej intelektualnej 12 dzieci zamieszkałych na wsi ( $N_1$ ) i 14 zamieszkałych w mieście ( $N_2$ ), uzyskano następujące wyniki dla dzieci zamieszkałych na wsi:  $X_1$ : 9, 11, 6, 6, 7, 9, 10, 7, 8, 10, 6, 7. Dla dzieci z miasta:  $X_2$ : 13, 6, 8, 10, 10, 11, 9, 8, 8, 12, 6, 8, 7, 10. Ustal, czy występujące różnice między średnimi wynikami uzyskanymi przez dzieci zamieszkałe na wsi i w mieście są istotne czy przypadkowe.

$H_0$ : Obserwowane różnice między średnimi wynikami uzyskanymi z testu dojrzałości szkolnej intelektualnej przez dzieci zamieszkałe na wsi i w mieście nie są istotne lecz przypadkowe, a badane próby pochodzą z tej samej populacji, czyli

$$H_0: \mu_1 = \mu_2, \quad \mu - \text{oznacza średnią z populacji.}$$

$H_1$ : Obserwowane różnice są istotne i powstały w wyniku tego, że badane próby pochodzą z różnych populacji, czyli

$$H_1: \mu_1 \neq \mu_2.$$

W tab. 43 przedstawiono sposób obliczania wartości  $t_{emp.}$ , a niżej dokonano zweryfikowania sformułowanej hipotezy zerowej.

Obliczanie  $t_{emp.}$  dla średnich nieskorelowanych

| $X_1$           | $x_1 = X_1 - \bar{x}_1$ | $x_1^2$           | $X_2$            | $x_2 = X_2 - \bar{x}_2$ | $x_2^2$           |
|-----------------|-------------------------|-------------------|------------------|-------------------------|-------------------|
| 9               | 1                       | 1                 | 13               | 4                       | 16                |
| 11              | 3                       | 9                 | 6                | -3                      | 9                 |
| 6               | -2                      | 4                 | 8                | -1                      | 1                 |
| 6               | -2                      | 4                 | 10               | 1                       | 1                 |
| 7               | -1                      | 1                 | 10               | 1                       | 1                 |
| 9               | 1                       | 1                 | 11               | 2                       | 4                 |
| 10              | 2                       | 4                 | 9                | 0                       | 0                 |
| 7               | -1                      | 1                 | 8                | -1                      | 1                 |
| 8               | 0                       | 0                 | 8                | -1                      | 1                 |
| 10              | 2                       | 4                 | 12               | 3                       | 9                 |
| 6               | -2                      | 4                 | 6                | -3                      | 9                 |
| 7               | -1                      | 1                 | 8                | -1                      | 1                 |
|                 |                         |                   | 7                | -2                      | 4                 |
|                 |                         |                   | 10               | 1                       | 1                 |
| $\Sigma X_1=96$ |                         | $\Sigma x_1^2=34$ | $\Sigma X_2=126$ |                         | $\Sigma x_2^2=58$ |

$$\bar{x}_1 = \frac{\Sigma X_1}{N_1} = \frac{96}{12} = 8$$

$$\bar{x}_2 = \frac{\Sigma X_2}{N_2} = \frac{126}{14} = 9$$

Kolejność obliczeń:

- – średnia arytmetyczna dla  $X_1$  ( $\bar{x}_1 = 8$ ),
- $x_1$  – różnica między każdą wartością  $X_1$  i średnią arytmetyczną:  

$$x_1 = X_1 - \bar{x}_1,$$
- $x_1^2$  – kwadrat każdego  $x_1$ ,
- – średnia arytmetyczna dla  $X_2$  ( $\bar{x}_2 = 9$ ),
- $x_2$  – różnica między każdą wartością  $X_2$  i średnią arytmetyczną:  

$$x_2 = X_2 - \bar{x}_2,$$
- $x_2^2$  – kwadrat każdego  $x_2$ ,
- – podstawiając do wzoru na  $t_{emp.}$  otrzymano:

$$t_{emp.} = \frac{|8-9|}{\sqrt{\left(\frac{34+58}{12+14-2}\right)\left(\frac{12+14}{12 \cdot 14}\right)}} = \frac{1}{\sqrt{0,599}} = 1,3$$

W Tablicy II (aneks) odczytujemy wartość  $t_{teoret.}$  na poziomie istotności  $\alpha = 0,05$  i  $\alpha = 0,01$ . Liczba stopni swobody dla  $t$  w próbach niezależnych:

$$df = N_1 + N_2 - 2,$$

w związku z tym:

$$t_{teoret.; 0,05; df=24} = 2,06 \qquad t_{teoret.; 0,01; df=24} = 2,80,$$

W prezentowanym przykładzie  $t_{emp.} = 1,3$ , a więc:

$$t_{emp.} = 1,3 < t_{teoret.; 0,05; df=24} = 2,06 \qquad t_{emp.} = 1,3 < t_{teoret.; 0,01; df=24} = 2,80$$

Oznacza to, że  $t_{emp.} < t_{teoret.}$  i sformułowaną hipotezę zerową ( $H_0$ ) należy przyjąć, czyli obserwowane różnice między wynikami testu dojrzałości szkolnej intelektualnej dzieci zamieszkałych na wsi i w mieście nie są istotne, tylko przypadkowe, a badane próby pochodzą z tej samej populacji.

### 1.1.2. Istotność różnic w próbach zależnych

Próby zależne (skorelowane) to takie, w których dwie próby uzyskane są z tych samych jednostek, pochodzą od tych samych badanych, lecz są rezultatem badań przeprowadzonych w różnych warunkach, przy wykorzystaniu różnych metod. W związku z tym uzyskane średnie są zależne, ponieważ pochodzą z prób dobranych parami, np. wyniki testu osiągnięć szkolnych uzyskane przez uczniów przed eksperymentem sprawdzającym skuteczność nowej metody nauczania i wyniki tych samych uczniów uzyskane z testu po zakończeniu eksperymentu; wyniki badań umiejętności cichego czytania ze zrozumieniem w wybranej grupie uczniów kl. III przed rozpoczęciem zajęć reedukacyjnych i wyniki tych samych uczniów po zakończeniu tych zajęć; wyniki poziomu uspołecznienia uczniów kl. III (w punktach) przed poddaniem ich specjalnym oddziaływaniom wychowawczym oraz wyniki uzyskane przez tych samych uczniów po półrocznym okresie tych oddziaływań itp.

Przy sprawdzaniu istotności różnic między średnimi w próbach zależnych (skorelowanych) ma zastosowanie następujący wzór:

$$t_{emp.} = \frac{|\bar{d}|}{\sqrt{\frac{\sum x_d^2}{N(N-1)}}},$$

gdzie:

- $\bar{d}$  – średnia arytmetyczna różnic  $N$  par wyników,
- $\sum x_d^2$  – suma kwadratów odchyłeń między parami wyników,
- $N$  – liczba par wyników.

**Przykład:**

W rezultacie badań 10 uczniów kl. III ( $N$ ) testem osiągnięć szkolnych uzyskano następujące wyniki:  $X_1$ : 13, 11, 9, 15, 15, 7, 9, 11, 11, 9. Po dwumiesięcznym okresie wprowadzania nowych metod objaśniania i kontroli wykonania pracy domowej, ta sama grupa uczniów uzyskała następujące wyniki testu osiągnięć szkolnych:  $X_2$ : 14, 21, 17, 13, 28, 13, 21, 19, 21, 13. Określ, czy obserwowane różnice między uzyskanymi średnimi są istotne czy przypadkowe.

$H_0$ : Obserwowane różnice między średnimi wynikami nie są istotne. Wyniki pomiarów  $X_1$  i  $X_2$  uzyskane u tych samych jednostek, w różnym czasie, pochodzą z populacji, których średnie są równe,

$$\text{czyli } H_0: \mu_1 - \mu_2 = 0$$

$H_1$ : Średnie nie są równe,  $\bar{x}_1 \neq \bar{x}_2$ ,

$$\text{czyli } H_1: \mu_1 - \mu_2 \neq 0$$

W tab. 44 przedstawiono sposób obliczania  $t_{emp.}$  dla średnich skorelowanych, a niżej dokonano weryfikacji sformułowanej hipotezy zerowej.

Kolejność obliczeń:

- $\bar{d}$  – średnia arytmetyczna dla  $X_1$  i  $X_2$ ,
- $d$  – różnica między poszczególnymi wartościami  $X_1$  i  $X_2$ ,
- $\bar{d}$  – średnia arytmetyczna różnic między  $X_1$  i  $X_2$ ,
- $x_d$  – odchylenie każdej różnicy  $X_1 - X_2$  od średniej różnic ( $\bar{d}$ ),
- $\sum x_d^2$  – suma kwadratów różnic odchyłeń między parami wyników.

Podstawiając do wzoru otrzymano:

$$t_{emp.} = \frac{|-7|}{\sqrt{\frac{208}{10(10-1)}}} = 4,60$$

Z tablicy II (aneks) odczytano wartość  $t_{teoret.}$  na poziomie  $\alpha = 0,05$  i  $\alpha = 0,01$ . Liczba stopni swobody dla  $t$  w próbach zależnych (skorelowanych):  $df = N-1$  i w związku z tym:

$$t_{teoret.; 0,05; df=9} = 2,26 \quad t_{teoret.; 0,01; df=9} = 3,25$$

W analizowanym przykładzie wartość  $t_{emp.} = 4,60$ , czyli

$$t_{emp.} = 4,60 > t_{teoret.; 0,01; df=9} = 3,25$$

Wynika stąd, że  $t_{emp.} > t_{teoret.}$ , a więc sformułowaną hipotezę zerową należy odrzucić, a przyjąć hipotezę alternatywną ( $H_1$ ) z prawdopodobieństwem popełnienia błędu 0,01. Oznacza to, że wyniki testu osiągnięć szkolnych uzyskane przez badaną grupę osób po wprowadzeniu nowych sposobów objaśniania i kontroli wykonania pracy domowej różnią się w stopniu istotnym od wyników uzyskanych wcześniej.

Tablica 44

Obliczanie  $t_{emp.}$  dla średnich skorelowanych

| Badany | $X_1$            | $X_2$            | $d = X_1 - X_2$  | $x_d = d - \bar{d}$ | $x_d^2$              |
|--------|------------------|------------------|------------------|---------------------|----------------------|
| 1      | 13               | 14               | -1               | 6                   | 36                   |
| 2      | 11               | 21               | -10              | -3                  | 9                    |
| 3      | 9                | 17               | -8               | -1                  | 1                    |
| 4      | 15               | 13               | 2                | 9                   | 81                   |
| 5      | 15               | 28               | -13              | -6                  | 36                   |
| 6      | 7                | 13               | -6               | 1                   | 1                    |
| 7      | 9                | 21               | -12              | -5                  | 25                   |
| 8      | 11               | 19               | -8               | -1                  | 1                    |
| 9      | 11               | 21               | -10              | -3                  | 9                    |
| 10     | 9                | 13               | -4               | 3                   | 9                    |
| $N=10$ | $\Sigma X_1=110$ | $\Sigma X_2=180$ | $\Sigma d = -70$ | $\Sigma x_d = 0,0$  | $\Sigma x_d^2 = 208$ |
|        | $\bar{x}_1 = 11$ | $\bar{x}_2 = 18$ | $\bar{d} = -7$   |                     |                      |

Testy parametryczne, których wybranym przykładem jest test  $T$ -Studenta, mają zastosowanie do wnioskowania w zakresie parametrów populacji generalnej. Wymagają one jednak – jak już wskazano – założenia o normalnym lub zbliżonym do takiego rozkładzie populacji. Zastosowanie tych testów do populacji o rozkładach innych może prowadzić do błędów we wnioskowaniu. W związku z tym w stosunku do populacji, których rozkład nie jest znany lub bardzo odbiega od normalnego mają zastosowanie testy nieparametryczne.

## 1.2. Sprawdzanie istotności współczynników korelacji

Jak już wskazano, współczynniki korelacji mogą przyjmować wartości w przedziale liczbowym od  $-1$  do  $+1$ . Czasem jednak może się zdarzyć, że w jakiejś próbie losowej badacz uzyska bardzo niską wartość współczynnika korelacji, np.  $r = 0,15$  i można przypuszczać, że obliczony współczynnik powstał przypadkowo, nie jest istotny. W takiej sytuacji istnieje potrzeba zweryfikowania (sprawdzenia) istotności współczynnika korelacji.

W literaturze przedmiotu przedstawione są różne sposoby ustalania istotności współczynnika korelacji. Istotność tę, głównie w przypadku małej próby, można sprawdzić stosując test  $T$ -Studenta według następującego wzoru:

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{N-2},$$

gdzie:

$r$  – współczynnik korelacji,

$N$  – liczba par obserwacji, liczba stopni swobody dla korelacji:

$df = N-2$ .

$H_0$ : dla współczynników korelacji zakłada, że w populacji, z której pochodzą próby, brak jest korelacji, czyli  $H_0: \sigma = 0$

$\sigma$  – oznacza współczynnik korelacji liniowej w populacji

$H_1: \sigma \neq 0$

W rezultacie badań stwierdzono, że związek między wynikami testu i oceną z matematyki  $N=20$  uczniów kl. V wynosi  $r=0,68$ . Ustal, czy współczynnik ten jest istotny statystycznie.

$H_0$ : Związek cech nie istnieje, czyli  $\sigma = 0$

$H_1$ :  $\sigma \neq 0$

Po podstawieniu do wzoru otrzymano:

$$t = \frac{0,68}{\sqrt{1 - (0,68)^2}} \sqrt{20 - 2} = 3,94$$

Z tablicy II (aneks) odczytano wartość  $t_{teoret.}$  dla poziomu istotności 0,01, przy  $df = 20 - 2 = 18$ :

$$t_{teoret.; 0,01; df=18} = 2,88, \text{ a więc } t_{emp.} = 3,94 > t_{teoret.} = 2,88$$

W związku z tym hipotezę zerową należy odrzucić, a przyjąć hipotezę alternatywną, z prawdopodobieństwem popełnienia błędu 0,01. Oznacza to, że gdyby przebadano 100 takich przypadków, to tylko w jednym nie byłoby takiego związku.

W praktyce stosowane są również znacznie prostsze sposoby sprawdzania istotności współczynnika korelacji, czyli sprawdzania czy otrzymany współczynnik istotnie różni się od zera. Umożliwia to tablica (Tablica III, aneks) zawierająca współczynniki korelacji istotne na poziomie 5% i 1% (0,05 i 0,01).

Sprawdzenie polega na ustaleniu liczby stopni swobody dla obliczonej wartości liczbowej współczynnika korelacji (dla każdego problemu korelacji liczba stopni swobody  $df = N - 2$ ) i odczytaniu z tej tablicy wartości najniższego istotnego współczynnika korelacji na poziomie 0,05 (5%). Obliczony współczynnik jest wówczas istotny statystycznie, jeśli jego liczbową wartość jest równa albo wyższa od wartości współczynnika w Tablicy III, na poziomie 0,05.

Przykłady:

W wyniku badań próby o liczebności  $N = 20$  uczniów stwierdzono, że związek między liczbą przeczytanych książek i oceną z języka polskiego wynosi  $r = 0,76$ . Sprawdź, czy uzyskany współczynnik korelacji jest istotny statystycznie.

$H_0$ :  $\sigma = 0$ , nie istnieje związek między liczbą przeczytanych książek i oceną z języka polskiego.

$H_1$ :  $\sigma \neq 0$

Weryfikacja hipotezy zerowej:

- $r_{emp.} = 0,76$ ,  $N = 20$ ,  $df = N - 2 = 20 - 2 = 18$ ,
- $r_{teoret.}$  odczytane z Tablicy III (aneks):

$$r_{teoret.; 0,05; df=18} = 0,444 \quad r_{teoret.; 0,01; df=18} = 0,561$$

- porównanie  $r_{emp.}$  i  $r_{teoret.}$ :  $r_{emp.} = 0,76 > r_{teoret.; 0,01; df=18} = 0,561$

W związku z tym hipotezę zerową należy odrzucić, a przyjąć alternatywną z prawdopodobieństwem popełnienia błędu 0,01, czyli obliczony współczynnik korelacji  $r = 0,76$  jest istotny statystycznie.

W rezultacie badania próby o liczebności  $N = 82$  uczniów stwierdzono, że związek między wynikami testu dojrzałości szkolnej intelektualnej i ocenami z matematyki w kl. II wynosi:  $r = 0,27$ . Sprawdź, czy uzyskany współczynnik korelacji jest istotny statystycznie.

$H_0$ : Nie istnieje związek między wynikami testu dojrzałości szkolnej intelektualnej i ocenami z matematyki w kl. II, czyli  $H_0: \sigma = 0$

$H_1$ : Związek istnieje, czyli  $H_1: \sigma \neq 0$ .

Weryfikacja hipotezy zerowej:

- $r_{emp.} = 0,27$ ,  $N = 82$ ,  $df = N - 2 = 82 - 2 = 80$ ,
- $r_{teoret.}$  odczytano z tablicy III (aneks):

$$r_{teoret.; 0,05; df=80} = 0,217 \quad r_{teoret.; 0,01; df=80} = 0,283$$

- porównanie  $r_{emp.}$  i  $r_{teoret.}$ :

$$r_{emp.} = 0,27 > r_{teoret.; 0,05; df=80} = 0,217 \text{ i jednocześnie}$$

$$r_{emp.} = 0,27 < r_{teoret.; 0,01; df=80} = 0,283$$

W związku z tym hipotezę zerową zakładającą brak związku należy odrzucić na poziomie 0,05 i przyjąć hipotezę alternatywną zakładającą, że związek istnieje. Obliczony współczynnik korelacji  $r = 0,27$  jest istotny statystycznie na poziomie 0,05 (5%).

W podobny sposób, korzystając z odpowiedniej tablicy (Tablica IV, aneks), można sprawdzić istotność współczynnika korelacji rangowej Spearmana ( $R_s$ ). Sprawdzenie następuje w rezultacie odczytania z tablicy liczebności próby ( $N$ ), dla której obliczono współczynnik korelacji rangowej i odpowiadających jej wartości krytycznych (teoretycznych) dla współczynnika na poziomie 0,05 lub 0,01. Obliczony współczynnik korelacji rangowej ( $R_s$ ) jest wówczas istotny statystycznie, jeśli jego liczbowa wartość jest równa albo wyższa od wartości współczynnika w tablicy IV na poziomie 0,05.

#### Przykłady:

W próbie o liczebności  $N=16$  uczniów stwierdzono, że związek między pozycjami zajmowanymi przez poszczególnych uczniów z punktu widzenia ich wyników w nauce i pozycjami, które zajmują w rankingu popularności w klasie wynosi:  $R_s=0,63$ . Sprawdź, czy uzyskany współczynnik korelacji rangowej Spearmana  $R_s$  jest istotny statystycznie.

$H_0$ : Nie istnieje związek między pozycjami zajmowanymi przez uczniów z punktu widzenia ich wyników w nauce i pozycjami, które zajmują w rankingu popularności w klasie, czyli  $H_0: \sigma_s = 0$ .

$\sigma_s$  - oznacza współczynnik korelacji rangowej w populacji.

$H_1$ : Związek istnieje, czyli  $H_1: \sigma_s \neq 0$ .

#### Weryfikacja hipotezy zerowej:

- $R_{s_{emp.}} = 0,63$ ,  $N = 16$ ,  $df = N - 1 = 15$ ,
- $R_{s_{teoret.}}$  odczytano z tablicy IV (aneks):

$$R_{s_{teoret.}; 0,05; df=15} = 0,425 \quad R_{s_{teoret.}; 0,01; df=15} = 0,601$$

- porównanie  $R_{s_{emp.}}$  i  $R_{s_{teoret.}}$ :

$$R_{s_{emp.}} = 0,63 > R_{s_{teoret.}; 0,05; df=15} = 0,425,$$

a także

$$R_{s_{emp.}} = 0,63 > R_{s_{teoret.}; 0,01; df=15} = 0,601$$

W wyniku porównania obu wartości  $R_s$  hipotezę zerową, zakładającą brak związku należy odrzucić, a przyjąć hipotezę alternatywną, wskazującą że związek istnieje. Obliczony współczynnik korelacji rangowej  $R_s = 0,63$  jest istotny statystycznie na poziomie 0,01.

W próbie o liczebności  $N = 8$  uczniów stwierdzono, że związek między pozycją zajmowaną w osiągnięciach sportowych a pozycją zajmowaną z punktu widzenia wyników w nauce wynosi  $R_s = 0,63$ . Sprawdź, czy uzyskany współczynnik korelacji rangowej jest istotny statystycznie.

$H_0$ : Nie istnieje związek między pozycją zajmowaną przez ucznia w osiągnięciach sportowych i pozycją zajmowaną z punktu widzenia wyników w nauce, czyli  $H_0: \sigma_s = 0$ .

$H_1$ : Związek istnieje, czyli  $H_1: \sigma_s \neq 0$ .

Weryfikacja hipotezy zerowej:

- $R_{s_{emp.}} = 0,63$ ,  $N = 8$ ,  $df = N - 1 = 7$ ,
- $R_{s_{teoret.}}$  odczytano z Tablicy IV (aneks):

$$R_{s_{teoret.}; 0,05; df=7} = 0,643 \quad R_{s_{teoret.}; 0,01; df=7} = 0,833$$

- porównanie  $R_{s_{emp.}}$  i  $R_{s_{teoret.}}$ :

$$R_{s_{emp.}} = 0,63 < R_{s_{teoret.}; 0,05; df=7} = 0,643,$$

a także

$$R_{s_{emp.}} = 0,63 < R_{s_{teoret.}; 0,01; df=7} = 0,833$$

W wyniku porównania obu wartości  $R_s$  hipotezę zerową, zakładającą brak związku, należy przyjąć, a odrzucić hipotezę alternatywną, zakładającą, że związek istnieje. Obliczony współczynnik korelacji rangowej  $R_s = 0,63$  nie jest istotny statystycznie. Wynika stąd, że nawet jakaś względnie wysoka wartość współczynnika korelacji rangowej (miarowej także) przy bardzo małych rozmiarach próby nie upoważnia do stwierdzenia o istnieniu istotnego statystycznie związku między rangami.

## 2. Testy nieparametryczne

W badaniach pedagogicznych często występują sytuacje, w których badacz nie zna rozkładu zmiennej w populacji generalnej, z której pobrano próbę, i nie może stosować testów parametrycznych (np. *T*-Studenta lub innych). Wówczas wykorzystywane są testy nieparametryczne. Mają one częste zastosowanie w badaniach pedagogicznych, nie wymagają żadnych założeń co do rozkładu, można je stosować wobec zmiennych wyrażonych na wszystkich skalach pomiarowych.

Należy jednak wskazać, że testy nieparametryczne wykazują większą ogólność i ich moc jest mniejsza od mocy testów parametrycznych. Oznacza to, że w przypadku posługiwania się testem nieparametrycznym w celu odrzucenia pewnej hipotezy zerowej należy poddać badaniom próbę o większej liczebności niż ta, która wystarcza, by odrzucić hipotezę zerową przy pomocy testu parametrycznego<sup>38</sup>.

**Testy nieparametryczne najczęściej stosowane są do:**

- badania zgodności rozkładu empirycznego z rozkładem teoretycznym (normalnym),
- sprawdzania niezależności dwóch cech,
- porównywania niezależnych od siebie prób losowych – test U-Manna-Whitneya i Kołmogorowa-Smirnowa (wymagają założenia, że badana zmienna jest ciągła).

### 2.1. Test $\chi^2$

Jest jednym z najczęściej stosowanych testów w badaniach pedagogicznych, ma liczne i różnorodne zastosowania. Nie wymaga założenia o normalności rozkładu, ale liczebność badanej próby powinna być stosunkowo duża. Stosowany jest do badania zgodności rozkładu empirycznego z teoretycznym i sprawdzania niezależności dwóch cech (zmiennych).

Przed stosowaniem tego testu (a także innych) należy sformułować hipotezę zerową i hipotezę alternatywną. W celu przyjęcia lub odrzucenia hipotezy zerowej przy zastosowaniu testu  $\chi^2$  należy obliczyć wartość empiryczną dla rozkładu badanej zmiennej lub zmiennych i porównać z wartością teoretyczną (krytyczną), odczytaną z tablicy (Tablica V, aneks) przy uwzględnieniu od-

---

<sup>38</sup> A. Góralski, *Metody ...*, op. cit., s. 157.

powiedniego poziomu istotności i liczby stopni swobody. Wówczas można zastosować następujące kryterium:

$$\chi^2_{emp.} < \chi^2_{teoret.:\alpha.;df} - H_0 \text{ należy przyjąć}$$

$$\chi^2_{emp.} \geq \chi^2_{teoret.:\alpha.;df} - H_0 \text{ należy odrzucić}$$

Oznacza to, że jeśli  $\chi^2_{emp.}$  jest mniejsze od  $\chi^2_{teoret.}$ , to sformułowaną hipotezę zerową należy przyjąć i odrzucić hipotezę alternatywną. Natomiast, jeśli  $\chi^2_{emp.}$  jest większe lub równe wartości  $\chi^2_{teoret.}$ , hipotezę zerową należy odrzucić i przyjąć hipotezę alternatywną.

## 2.2. Test zgodności $\chi^2$

Test zgodności  $\chi^2$  stosowany jest do sprawdzania zgodności rozkładu empirycznego (uzyskanego w badaniach) z rozkładem normalnym.

Próba, której rozkład zmiennej ma być oceniany, powinna mieć dużą liczebność, ponieważ jej rozkład przedstawia się w postaci szeregu rozdzielczego, zawierającego przedziały klasowe, każdy o określonej liczebności. Test  $\chi^2$  stosuje się, gdy każda liczebność przedziału klasowego jest większa lub równa 5. Jeśli tak nie jest, to dokonuje się ich łączenia.

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_o},$$

gdzie:

$f_e$  – liczebności empiryczne kolejnych przedziałów klasowych,

$f_o$  – liczebności oczekiwane (teoretyczne) kolejnych przedziałów klasowych,

$df$  – liczba stopni swobody, w tym przypadku:  $df = k - 3$  ( $k$  – liczba klas), ponieważ liczba parametrów określających rozkład normalny to dwa (średnia arytmetyczna i odchylenie standardowe), a trzeci związany jest z liczebnością próby.

W przypadku testu zgodności  $\chi^2$  hipoteza zerowa i alternatywna są następujące:

$H_0$ : Rozkład empiryczny jest zgodny z rozkładem normalnym.

$H_1$ : Rozkład empiryczny nie jest zgodny z rozkładem normalnym.

Liczebności oczekiwane zgodnie z hipotezą to liczebności, jakich można by się spodziewać, gdyby rozkład empiryczny był zgodny z rozkładem normalnym.

Zastosowanie testu zgodności  $\chi^2$  do sprawdzania normalności rozkładu przedstawiono na przykładzie danych w Tablicy 45.

Tablica 45

*Zastosowanie testu  $\chi^2$  do sprawdzania normalności rozkładu  
wyników testu z ortografii w kl. V*

| Wyniki<br>testu<br>$X$ | $x_i$ | $f_e$  | $x = x_i - \bar{x}$ | $z = \frac{x}{s_x}$ | $y$ (z Tab.) | $f_o = y \frac{N_i}{s_x}$ | $f_e \cdot f_o$ | $(f_e \cdot f_o)^2$ | $\frac{(f_e - f_o)^2}{f_o}$ |
|------------------------|-------|--------|---------------------|---------------------|--------------|---------------------------|-----------------|---------------------|-----------------------------|
| 3 - 5                  | 4     | 8      | -5,25               | -1,41               | 0,1476       | 6,2                       | 1,8             | 3,24                | 0,52                        |
| 6 - 8                  | 7     | 18     | -2,25               | -0,60               | 0,3332       | 14,0                      | 4,0             | 16,00               | 1,14                        |
| 9 - 11                 | 10    | 11     | 0,75                | 0,20                | 0,3910       | 16,5                      | -5,5            | 30,25               | 1,83                        |
| 12 - 14                | 13    | 9      | 3,75                | 1,00                | 0,2420       | 10,2                      | -1,2            | 1,44                | 0,14                        |
| 15 - 17                | 16    | 6      | 6,75                | 1,81                | 0,0776       | 3,3                       | 2,7             | 7,29                | 2,20                        |
|                        |       | $N=52$ |                     |                     |              | $\Sigma=52$               |                 |                     | $\Sigma=5,83$               |

$$\bar{x} = 9,25$$

$$s_x = 3,71$$

Aby dokonać obliczeń w Tablicy 45, należy:

- przedstawić dane empiryczne w postaci przedziałów klasowych, obliczyć środki przedziałów ( $x_i$ ) i ustalić liczebności klas ( $f_e$ ),
- obliczyć średnią arytmetyczną ( $\bar{x}$ ) i odchylenie standardowe ( $s_x$ ),
- znaleźć odchylenie każdego środka przedziału klasowego od średniej arytmetycznej: ( $x = x_i - \bar{x}$ ),
- obliczyć wyniki standaryzowane  $z$ , gdzie  $z = \frac{x}{s_x}$ . Otrzymane dla po-

szczególnych przedziałów  $x = x_i - \bar{x}$  podzielić przez wartość odchylenia standardowego ( $s_x$ ),

- dla każdego wyniku standaryzowanego  $z$  odczytać z Tablicy I (aneks) wartości rzędnej  $y$ . Gdy w tablicy nie ma podanej wartości  $y$ , obliczyć dokonując interpolacji,
- dokonać obliczenia wartości oczekiwanych ( $f_o$ ) poprzez zastosowanie następującego wzoru :

$$f_o = \left( \frac{Ni}{s_x} \right) y ,$$

gdzie:

- $N$  – liczebność próby,
- $i$  – wielkość przedziału klasowego,
- $s_x$  – odchylenie standardowe,
- $y$  – wartość z Tablicy I (aneks).

Najpierw oblicza się  $\frac{Ni}{s_x}$ , ponieważ jest to stała dla wszystkich przedziałów, następnie mnoży się ją przez każdą wartość  $y$  i otrzymuje się  $f_o$  dla poszczególnych przedziałów.

W podanym przykładzie:

$$f_o = \left( \frac{Ni}{s_x} \right) y = \frac{52 \cdot 3}{3,71} y = 42,05 \cdot 0,1476 = 6,2$$

- obliczyć różnicę między liczebnościami empirycznymi ( $f_e$ ) i oczekiwanymi ( $f_o$ ):  $f_e - f_o$ ,
- podnieść do kwadratu różnicę ( $f_e - f_o$ ), czyli:  $(f_e - f_o)^2$ ,
- kwadrat różnicy podzielić przez liczebności oczekiwane:

$\frac{(f_e - f_o)^2}{f_o}$ . Suma  $\sum \frac{(f_e - f_o)^2}{f_o}$  dla wszystkich przedziałów jest wartością  $\chi^2_{emp.} = 5,83$ .

Weryfikacja sformułowanej hipotezy zerowej przebiega następująco:

- ustalenie liczby stopni swobody dla  $\chi^2$ :  $df = k - 3$ , w podanym przykładzie  $k = 5$ , a więc  $df = 5 - 3 = 2$ ,
- z tablicy V (aneks) odczytano wartość  $\chi^2_{teoret.}$  dla poziomu istotności 0,05:

$$\chi^2_{teoret.:0.05;df=2} = 5,991$$

- porównanie  $\chi^2_{emp.}$  i  $\chi^2_{teoret.}$ :

$$\chi^2_{emp.} = 5,83 < \chi^2_{teoret.; 0,05; df=2} = 5,991$$

W związku z tym, że  $\chi^2_{emp.}$  jest mniejsze od  $\chi^2_{teoret.}$ , sformułowaną hipotezę zerową należy przyjąć, a odrzucić alternatywną, czyli uzyskany rozkład wyników testu z ortografii w kl.V jest zgodny z rozkładem normalnym i można obliczać interesujące badacza parametry.

Należy podkreślić, że hipoteza zerowa występuje tu w nieco innym aspekcie niż w przypadku istotności różnic. W przypadku istotności różnic badacz jest zainteresowany tym, aby hipotezę zerową odrzucić, natomiast w przypadku porównywania rozkładu empirycznego z normalnym badacz jest zainteresowany przyjęciem hipotezy zerowej.

## 2.3. Test niezależności $\chi^2$

**Test niezależności  $\chi^2$**  ma zastosowanie do sprawdzania niezależności cech dla dwóch zmiennych dwudzielnych, dychotomicznych (tablice czteropolowe), a także dla dwóch zmiennych wielodzielnych (tablice wielopolowe), przy czym zmienne te są najczęściej zmiennymi jakościowymi (niemierzalnymi). Stosowanie tego testu wymaga próby o dużej liczebności.

### 2.3.1. Test niezależności $\chi^2$ (tablice czteropolowe)

Test niezależności  $\chi^2$  ma zastosowanie do sprawdzania niezależności cech dla dwóch zmiennych dychotomicznych, czyli takich, które mają po dwa warianty występowania, np. podczas badań uczniów kl. VIII (dziewcząt i chłopców) uzyskano następujące rezultaty: spośród 40 dziewcząt 13 wyraziło zainteresowanie sportem, a w grupie 48 chłopców – 30. Dysponując takimi dwoma zmiennymi, jak płeć i zainteresowanie sportem można ustalić, czy istnieje między nimi związek (Tablica 46).

Tablica 46

Zainteresowanie sportem uczniów kl. VIII według płci

| Płeć       | Zainteresowanie sportem |     |        |
|------------|-------------------------|-----|--------|
|            | Tak                     | Nie | Ogółem |
| Chłopcy    | 30                      | 18  | 48     |
| Dziewczęta | 13                      | 27  | 40     |
| Ogółem     | 43                      | 45  | 88     |

Źródło: dane umowne

Hipotezy zerowa i alternatywna są następujące:

$H_0$ : Nie istnieje związek między zmiennymi (obie zmienne są od siebie niezależne), nie istnieje związek między płcią i zainteresowaniem sportem.

$H_1$ : Związek między zmiennymi istnieje.

Weryfikacja hipotezy zerowej następuje w rezultacie stosowania testu  $\chi^2$ :

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_o},$$

gdzie:

$f_e$  – liczebności empiryczne,

$f_o$  – liczebności oczekiwane,

$$df = (w-1)(k-1),$$

$w$  – liczba wierszy w tablicy,

$k$  – liczba kolumn w tablicy.

Aby zweryfikować hipotezę zerową, należy najpierw obliczyć liczebności oczekiwane ( $f_o$ ), które należałoby wstawić do poszczególnych pól tablicy. jeżeli hipoteza zerowa jest prawdziwa:

$$f_o = \frac{\sum w \cdot \sum k}{N},$$

gdzie:

$\sum w$  – suma wiersza, w którym znajduje się liczebność,

$\sum k$  – suma kolumny, w której znajduje się liczebność,

$N$  – liczebność próby.

Dla podanego przykładu obliczenia są następujące:

- chłopcy zainteresowani sportem:  $f_o = \frac{48 \cdot 43}{88} = 23,45$
- chłopcy nie interesujący się sportem:  $f_o = \frac{48 \cdot 45}{88} = 24,54$
- dziewczęta zainteresowane sportem:  $f_o = \frac{40 \cdot 43}{88} = 19,54$
- dziewczęta nie interesujące się sportem:  $f_o = \frac{40 \cdot 45}{88} = 20,45$

Kolejność dalszych obliczeń przedstawiono w tablicy pomocniczej:

*Tablica pomocnicza*

| $f_e$  | $f_o$               | $f_e - f_o$ | $(f_e - f_o)^2$ | $\frac{(f_e - f_o)^2}{f_o}$ |
|--------|---------------------|-------------|-----------------|-----------------------------|
| 30     | 23,4                | 6,6         | 43,56           | 1,86                        |
| 18     | 24,5                | -6,5        | 42,25           | 1,72                        |
| 13     | 19,5                | -6,5        | 42,25           | 2,16                        |
| 27     | 20,4                | 6,6         | 42,56           | 2,13                        |
| $N=88$ | $\Sigma \approx 88$ |             |                 | $\Sigma = 7,87$             |

Obliczona wartość  $\chi^2_{emp.} = 7,87$ .

Kolejność działań prowadząca do weryfikacji sformułowanej hipotezy zerowej jest następująca:

- ustalenie liczby stopni swobody:  $df = (w-1)(k-1) = (2-1)(2-1) = 1$ ,
- $\chi^2_{teoret.}$  odczytano z Tablicy V (aneks), na poziomie istotności 0,05 i 0,01, dla  $df = 1$ :

$$\chi^2_{teoret.; 0,05; df=1} = 3,84 \quad \chi^2_{teoret.; 0,01; df=1} = 6,63$$

- porównanie  $\chi^2_{emp.}$  i  $\chi^2_{teoret.}$ :

$$\chi^2_{emp.} = 7,87 > \chi^2_{teoret.; 0,05; df=1} = 3,84,$$

a także

$$\chi^2_{emp.} = 7,87 > \chi^2_{teoret.; 0,01; df=1} = 6,63$$

Wniosek:

$\chi^2_{emp.}$  jest większe od  $\chi^2_{teoret.}$  (nawet przy wyższym poziomie istotności). Oznacza to, że sformułowaną hipotezę zerową należy odrzucić, a przyjąć hipotezę alternatywną, z której wynika, że związek między badanymi zmiennymi: płcią i zainteresowaniem sportem występuje.

W przypadku, gdy obie zmienne, między którymi poszukuje się związku, są zmiennymi dwudzielnymi (dychotomicznymi) i występują w tablicy czteropolowej, można zastosować ogólniejszą formułę do ustalania wartości  $\chi^2$ .

Zakłada się w tym celu, że zmienna  $X$  występuje w kategoriach  $x_1$  i  $x_2$ , a druga zmienna  $Y$  w kategoriach  $y_1$  i  $y_2$ . Poszczególne liczebności wpisuje się do pól tablicy odpowiednio oznaczając:  $a, b, c, d$  (Tablica 47)<sup>39</sup>.

Tablica 47

Schemat tablicy czteropolowej

| Zmienna $Y$ | Zmienna $X$ |       | Suma wierszy |
|-------------|-------------|-------|--------------|
|             | $x_1$       | $x_2$ |              |
| $y_1$       | $a$         | $b$   | $(a+b)$      |
| $y_2$       | $c$         | $d$   | $(c+d)$      |
| Suma kolumn | $a+c$       | $b+d$ | $a+b+c+d=N$  |

Źródło: G. Clauss, H. Ebner, *Podstawy ..., op. cit., s. 273.*

W celu ustalenia wartości  $\chi^2$  ma tu zastosowanie następujący wzór:

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)},$$

gdzie:

$a, b, c, d$  – liczebności empiryczne uzyskane w badaniu,

$N$  – liczebność próby.

<sup>39</sup> G. Clauss, H. Ebner, *Podstawy ..., op. cit., s. 273.*

Stosując tę formułę dla podanego wcześniej przykładu uzyskano:

$$\chi^2 = \frac{88(30 \cdot 27 - 18 \cdot 13)^2}{(48 \cdot 40 \cdot 43 \cdot 45)} = \frac{88 \cdot 576^2}{3715200} = 7,86$$

Występująca różnica wynika z dokonanych zaokrągleń do jednego miejsca po przecinku przy obliczaniu liczebności oczekiwanych.

Test ten dla tablic czteropolowych (dwa wiersze i dwie kolumny – zmienne dwudzielne) można stosować wówczas, gdy liczebność empiryczna nie jest w żadnym polu mniejsza niż 5, a próba losowa powinna zawierać co najmniej 40 pomiarów<sup>40</sup>.

Jeśli, w którymś polu liczebność jest niższa niż 5, wówczas należy zastosować poprawkę Yatesa.<sup>41</sup> W tym celu zmniejsza się o 0,5 każdą liczebność empiryczną ( $f_e$ ) większą od odpowiadającej jej liczebności oczekiwanej ( $f_o$ ) i jednocześnie zwiększa się o 0,5 każdą liczebność empiryczną ( $f_e$ ) mniejszą od liczebności oczekiwanej ( $f_o$ ):

*Tablica bez poprawki Yatesa*

| Zmienna<br>Y | Zmienna X |           | Ogółem    |
|--------------|-----------|-----------|-----------|
|              | $x_1$     | $x_2$     |           |
| $y_2$        | 10 (6,5)  | 4 (7,5)   | 14 (14,0) |
| $y_1$        | 3 (6,5)   | 11 (7,5)  | 14 (14,0) |
| Ogółem       | 13 (13,0) | 15 (15,0) | 28 (28,0) |

*Tablica z poprawką Yatesa*

| Zmienna<br>Y | Zmienna X |            | Ogółem    |
|--------------|-----------|------------|-----------|
|              | $x_1$     | $x_2$      |           |
| $y_2$        | 9,5 (6,5) | 4,5 (7,5)  | 14 (14,0) |
| $y_1$        | 3,5 (6,5) | 10,5 (7,5) | 14 (14,0) |
| Ogółem       | 13 (13,0) | 15 (15,0)  | 28 (28,0) |

Źródło: G. Clauss, H. Ebner, *Podstawy ....*, op. cit., s. 273-274.

Poprawkę Yatesa można zastosować również bezpośrednio na uzyskanych liczebnościach empirycznych, a obliczając  $\chi^2$  stosuje się następujący wzór:

$$\chi^2 = \frac{\left( |ad - bc| - \frac{N}{2} \right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

<sup>40</sup> G. Clauss, H. Ebner, *Podstawy ....*, op. cit., s. 273.

<sup>41</sup> *Tamże*, s. 273.

W rezultacie zastosowania poprawki wartość  $\chi^2$  zmniejsza się, co ma istotne znaczenie przy podejmowaniu decyzji o przyjęciu lub odrzuceniu hipotezy zerowej.

Stwierdzenie występowania związku między dwiema zmiennymi dychotomicznymi przez odrzucenie hipotezy zerowej daje podstawy do ustalania siły tego związku.

Miernikiem siły związku między zmiennymi dychotomicznymi, przedstawionymi w tablicy czteropolowej (Tablica 47) jest współczynnik  $\varphi$ :

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}},$$

gdzie:

$a, b, c, d$  – liczebności empiryczne uzyskane w badaniu.

Współczynnik  $\varphi$  może teoretycznie przyjmować wartości leżące w przedziale liczbowym od  $-1$  do  $+1$ .<sup>42</sup> Znak „ $-$ ” lub „ $+$ ” przed analizowanym współczynnikiem, nie wskazuje bezpośrednio na kierunek związku obu zmiennych. Znak ten zależy od tego, jak porządkuje się w tablicy czteropolowej poszczególne kategorie cech. W tym zakresie dla danych jakościowych nie istnieją żadne wiążące zasady postępowania. Przy wartościach wpisanych w pola tablicy można osiągnąć te krańcowe liczby ( $-1$  lub  $+1$ ) tylko w przypadkach, gdy  $b = c = 0$  lub  $a = d = 0$ , czyli gdy dwa przeciwległe pola są puste. Informację o ścisłości związku między zmiennymi opiera się na wartości absolutnej (bezwzględnej) współczynnika.

W związku z tym, że między  $\varphi$  i  $\chi^2$  istnieje bezpośredni związek, siłę związku między zmiennymi dwudzielnymi można przedstawić również według wzoru:

$$\varphi = \sqrt{\frac{\chi^2}{N}},$$

gdzie:

$\chi^2$  – wartość empiryczna.

$N$  – liczebność próby.

<sup>42</sup> G. Clauss, H. Ebner, *Podstawy ...*, op. cit., s. 292-293.

W podanym wcześniej przykładzie (s. 141-143), w którym stwierdzono występowanie związku między płcią uczniów kl. VIII i zainteresowaniem sportem ( $\chi^2_{emp.} = 7,87, N = 88$ ), siła tego związku wynosi:

$$\varphi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{7,87}{88}} = 0,29$$

lub

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{30 \cdot 27 - 18 \cdot 13}{\sqrt{(48 \cdot 40 \cdot 43 \cdot 45)}} = \frac{576}{1,927} = 0,29$$

Obliczona wartość  $\varphi = 0,29$  wskazuje na występowanie niewielkiego związku między płcią uczniów kl. VIII i przejawianiem zainteresowania sportem. Jako wyizolowana z kontekstu wartość współczynnika oznacza niewiele. Większe znaczenie ma, gdy porówna się go z innymi wartościami tego współczynnika, np. występującymi w innych środowiskach (wiejskim, małomiasteczkowym), klasach itp.

Współczynnik  $\varphi$  może być obliczany nie tylko wówczas gdy dokonuje się analizy danych z kwestionariusza, np. pytań, które mają tylko dwie możliwości odpowiedzi (tak, nie lub innych), ale także wówczas, gdy uzyskuje się wypowiedzi dotyczące pozytywnego lub negatywnego stanowiska w jakiejś kwestii czy rezultatu czegoś.

W kwestionariuszu do badania sposobów przygotowania się studentów do egzaminu z przedmiotu X zapytano studentów: Jaki sposób przygotowania do egzaminu z przedmiotu X preferujesz? Odpowiedź na to pytanie przedstawiono uwzględniając inną zmienną – rezultat egzaminu: pozytywny, negatywny (Tablica 48).

Tablica 48

Wyniki egzaminu z przedmiotu X  
według sposobu przygotowania

| Sposób<br>przygotowania | Rezultat egzaminu |           |        |
|-------------------------|-------------------|-----------|--------|
|                         | Pozytywny         | Negatywny | Ogółem |
| Grupowy                 | 23                | 7         | 30     |
| Indywidualny            | 13                | 17        | 30     |
| Ogółem                  | 36                | 24        | 60     |

Źródło: dane umowne

Hipoteza zerowa i alternatywna są następujące:

$H_0$ : Nie istnieje związek między sposobem przygotowania się studentów do egzaminu i jego rezultatem, czyli obie zmienne są od siebie niezależne.

$H_1$ : Istnieje związek między sposobem przygotowania się studentów do egzaminu i jego rezultatem, obie zmienne są od siebie zależne.

Weryfikacja sformułowanej hipotezy zerowej jest następująca:

- obliczenie wartości  $\chi^2$  według następującego wzoru dla tablicy czteropolowej:

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{60(23 \cdot 17 - 7 \cdot 13)^2}{30 \cdot 30 \cdot 36 \cdot 24} = 6,94$$

- ustalenie liczby stopni swobody:  $df = (w-1)(k-1) = (2-1)(2-1) = 1$
- odczytano wartość  $\chi^2_{teoret.}$  na poziomie istotności 0,05 i 0,01 z Tablicy V (aneks):

$$\chi^2_{teoret.; 0,05; df=1} = 3,84 \quad \chi^2_{teoret.; 0,01; df=1} = 6,63$$

- porównanie wartości  $\chi^2_{emp.}$  i  $\chi^2_{teoret.}$ :

$$\chi^2_{emp.} = 6,94 > \chi^2_{teoret.; 0,05; df=1} = 3,84,$$

także

$$\chi^2_{emp.} = 6,94 > \chi^2_{teoret.; 0,01; df=1} = 6,63$$

Wniosek:

$\chi^2_{emp.}$  jest większe od  $\chi^2_{teoret.}$ . Oznacza to, że sformułowaną hipotezę zerową należy odrzucić, a przyjąć hipotezę alternatywną, czyli występuje związek między sposobem przygotowania studenta do egzaminu z przedmiotu X i rezultatem tego egzaminu.

Stwierdzenie występowania związku daje podstawę do obliczenia jego siły:

$$\varphi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{6,94}{60}} = 0,34$$

Obliczona wartość współczynnika  $\varphi = 0,34$  wskazuje na występowanie niewielkiego związku między sposobem przygotowania studentów do egzaminu z przedmiotu X i rezultatem tego egzaminu. Jak już wskazywano, wartość tego współczynnika sama dostarcza niepełnej informacji, dopiero w rezultacie porównania jej z wartościami uzyskanymi w innych badaniach, np. innych przedmiotów, innych grup studentów miałyby ona dużo większe znaczenie.

### 2.3.2. Test niezależności $\chi^2$ (tablice wielopolowe)

**Test niezależności  $\chi^2$**  ma stosunkowo częste zastosowanie do sprawdzania niezależności cech dla dwóch zmiennych wielodzielnych, tzn. takich, w których każda ze zmiennych ma więcej niż dwa warianty. Występują one wówczas w tablicy wielopolowej, np. jeśli badacz jest zainteresowany ustaleniem związku między pochodzeniem społecznym i poziomem osiągnięć szkolnych uczniów kl. III (Tablica 49).

Tablica 49

*Poziom osiągnięć szkolnych uczniów kl. III  
według pochodzenia społecznego*

| Pochodzenie społeczne | Poziom osiągnięć szkolnych |        |       |        |
|-----------------------|----------------------------|--------|-------|--------|
|                       | Wysoki                     | Średni | Niski | Ogółem |
| Robotnicze            | 15                         | 17     | 6     | 38     |
| Chłopskie             | 4                          | 8      | 18    | 30     |
| Inteligentkie         | 9                          | 15     | 9     | 33     |
| Ogółem                | 28                         | 40     | 33    | 101    |

Źródło: dane umowne

Hipoteza zerowa i alternatywna są następujące:

$H_0$ : Nie istnieje związek między zmiennymi, czyli obie zmienne są od siebie niezależne, nie istnieje zatem związek między pochodzeniem społecznym i poziomem osiągnięć szkolnych uczniów kl. III.

$H_1$ : Związek między zmiennymi istnieje.

Weryfikacja hipotezy zerowej następuje tu w rezultacie zastosowania testu  $\chi^2$ :<sup>43</sup>

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_o},$$

gdzie:

$f_e$  – liczebności empiryczne,

$f_o$  – liczebności oczekiwane,

$$df = (w-1)(k-1),$$

$w$  – liczba wierszy w tablicy,

$k$  – liczba kolumn w tablicy.

Aby zweryfikować sformułowaną hipotezę zerową, należy najpierw obliczyć liczebności oczekiwane ( $f_o$ ), które należałoby wstawić do poszczególnych pól tablicy, jeżeli hipoteza zerowa jest prawdziwa:

$$f_o = \frac{\sum w \cdot \sum k}{N},$$

gdzie:

$\sum w$  – suma wiersza, w którym znajduje się liczebność,

$\sum k$  – suma kolumny, w której znajduje się liczebność,

$N$  – liczebność próby.

Kolejność obliczeń dla podanego przykładu przedstawiono w tablicy pomocniczej:

---

<sup>43</sup> Weryfikacji hipotezy zerowej dokonuje się tu w podobny sposób, jak w przypadku tablicy czteropolowej (patrz s. 141 i następne), ale w celu przejrzystości wywodu i dla wygody Czytelnika przytoczono raz jeszcze potrzebne wzory i ich oznaczenia.

Tablica pomocnicza

| $f_e$   | $f_o$                            | $f_e - f_o$ | $(f_e - f_o)^2$ | $\frac{(f_e - f_o)^2}{f_o}$ |
|---------|----------------------------------|-------------|-----------------|-----------------------------|
| 15      | $\frac{38 \cdot 28}{101} = 10,5$ | 4,5         | 20,25           | 1,93                        |
| 17      | $\frac{38 \cdot 40}{101} = 15,1$ | 1,9         | 3,61            | 0,23                        |
| 6       | $\frac{38 \cdot 33}{101} = 12,4$ | -6,4        | 40,96           | 3,30                        |
| 4       | $\frac{30 \cdot 28}{101} = 8,3$  | -4,3        | 18,49           | 2,22                        |
| 8       | $\frac{30 \cdot 40}{101} = 11,9$ | -3,9        | 15,21           | 1,27                        |
| 18      | $\frac{30 \cdot 33}{101} = 9,8$  | 8,2         | 67,24           | 6,86                        |
| 9       | $\frac{33 \cdot 28}{101} = 9,1$  | -0,1        | 0,01            | 0,00                        |
| 15      | $\frac{33 \cdot 40}{101} = 13,1$ | 1,9         | 3,61            | 0,27                        |
| 9       | $\frac{33 \cdot 33}{101} = 10,8$ | -1,8        | 3,24            | 0,30                        |
| $N=101$ | $\Sigma \approx 101,0$           |             |                 | $\Sigma = 16,38$            |

Obliczona wartość  $\chi^2_{emp.} = 16,38$ .

Kolejność działań prowadząca do weryfikacji sformułowanej hipotezy zerowej jest następująca:

- ustalenie liczby stopni swobody:  $df = (w-1)(k-1) = (3-1)(3-1) = 4$ ,
- z Tablicy V (aneks) odczytano wartość  $\chi^2_{teoret.}$  na poziomie istotności 0,05 i 0,01:

$$\chi^2_{teoret.; 0,05; df=4} = 9,488 \quad \chi^2_{teoret.; 0,01; df=4} = 13,277$$

- porównanie wartości  $\chi^2_{emp.}$  i  $\chi^2_{teoret.}$ :

$$\chi^2_{emp.} = 16,38 > \chi^2_{teoret.; 0,05; df=4} = 9,488$$

i

$$\chi^2_{emp.} = 16,38 > \chi^2_{teoret.; 0,01; df=4} = 13,277$$

Wniosek:

$\chi^2_{emp.}$  jest większe od  $\chi^2_{teoret.}$ . Oznacza to, że sformułowaną hipotezę zerową należy odrzucić, a przyjmując alternatywną, z której wynika, że występuje związek między pochodzeniem społecznym uczniów kl. III a poziomem ich osiągnięć szkolnych.

Stwierdzenie występowania związku między zmiennymi upoważnia badacza do obliczania jego siły. Siła związku między zmiennymi przedstawionymi w tablicach wielopolowych wyrażana jest poprzez obliczanie **współczynnika kontyngencji** (zbieżności)  $C$  według następującego wzoru:<sup>44</sup>

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

gdzie:

$\chi^2$  – wartość empiryczna,  
 $N$  – liczebność próby.

Współczynnik kontyngencji  $C$  może przyjmować tylko wartości dodatnie, przedstawia siłę związku między zmiennymi, nie wskazuje jego kierunku. Może być stosowany przy obliczaniu siły związku między zmiennymi występującymi w tablicach wielopolowych dowolnej wielkości (najmniejsza liczba pól powinna wynosić więcej niż 4) i dowolnego kształtu (kwadratowych lub prostokątnych).

Maksymalna wartość omawianego współczynnika zależy od liczby pól w tablicy będącej podstawą obliczeń. Im większa ich liczba, tym bardziej współczynnik  $C_{max}$  zbliża się do +1 i jednocześnie, im mniejsza liczba pól, tym niższa jest maksymalnie osiągnięta wartość  $C$ .

Jeżeli  $k = w$ , tzn. liczba kolumn i wierszy w tablicy są równe, to  $C_{max}$  można wyznaczyć za pomocą wzoru:

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

gdzie:

$k$  – oznacza liczbę wierszy lub kolumn w tablicy.

---

<sup>44</sup> G. Clauss, H. Ebner, *Podstawy ..., op. cit.*, s. 301.

Zamieszczona niżej Tablica 50 przedstawia maksymalne wartości współczynnika  $C$  w różnych tablicach wielopolowych.

*Tablica 50*

*Maksymalne wartości „C” w zależności od liczby wierszy i kolumn w tablicy*

| $k$        | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| $C_{max.}$ | 0,707 | 0,816 | 0,866 | 0,894 | 0,913 | 0,926 | 0,935 | 0,943 |

Źródło: G. Clauss, H. Ebner, *Podstawy ...*, op. cit., s. 303.

Dla tablic prostokątnych (liczba wierszy nierówna liczbie kolumn) współczynnik  $C$  można określić tylko szacunkowo, poprzez obliczenie go z wielkości podanych w tablicy, np. w tablicy 2 x 3 (dwa wiersze i trzy kolumny lub odwrotnie) maksymalna wartość współczynnika  $C$  wynosi :

$$C_{max.} = \frac{0,707 + 0,816}{2} = 0,762$$

Dysponując wartością współczynnika zbieżności  $C$  obliczoną dla konkretnego przykładu oraz  $C_{max.}$ , można obliczyć wartość skorygowaną współczynnika kontyngencji  $C_{kor.}$ :

$$C_{kor.} = \frac{C}{C_{max.}}$$

W każdym przypadku  $C_{kor.}$  prowadzi do podwyższenia wartości współczynnika  $C$ .

Dla podanego wcześniej przykładu (s. 148-151), w którym poszukiwano związku między pochodzeniem społecznym uczniów kl. III a poziomem ich osiągnięć szkolnych, dokonano weryfikacji hipotezy zerowej i ustalono, stosując test  $\chi^2$ , że związek między zmiennymi występuje.

Stwierdzenie występowania związku między zmiennymi upoważnia badacza do ustalenia jego siły. W podanym przykładzie  $\chi^2_{emp.} = 16,38$ ,  $N = 101$ , a więc siła związku między zmiennymi wyrażona przy pomocy współczynnika zbieżności  $C$  wynosi:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{16,38}{16,38 + 101}} = 0,373$$

W analizowanym przykładzie dane z badań przedstawione są w tablicy dziewięciopolowej (Tablica 49 – trzy wiersze i trzy kolumny), tzn.  $k=w=3$ , wobec tego maksymalna wartość współczynnika  $C$  przy najsilniejszym związku może wynosić  $C_{max.} = 0,816$  (Tablica 50). W związku z tym :

$$C_{kor.} = \frac{C}{C_{max.}} = \frac{0,373}{0,816} = 0,457$$

Siła związku między pochodzeniem społecznym i poziomem osiągnięć szkolnych w kl. III wyrażona jest współczynnikiem  $C = 0,457$ , co oznacza, że można określić ją jako umiarkowaną. Poziom osiągnięć szkolnych jest w umiarkowanym stopniu związany z pochodzeniem społecznym uczniów kl. III.

Interpretując ten wynik podkreślić należy (jak już wcześniej wskazywano), że istnieje jeszcze wiele innych czynników wpływających na poziom osiągnięć szkolnych uczniów, a współczynniki rozpatrywane w oderwaniu od całości zagadnienia znaczą bardzo niewiele. Dopiero porównanie kilku wartości współczynnika z innych prób może dostarczyć interesujących spostrzeżeń.

## Pytania i zadania

1. Wskaż różnice między testami parametrycznymi i nieparametrycznymi.
2. Kiedy ma zastosowania test  $T$ -Studenta?
3. Wskaż różnice między próbami niezależnymi i zależnymi. Podaj przykłady.
4. Jak jest sformułowana hipoteza zerowa i alternatywna w hipotezie istotności różnic?
5. Podaj kryterium stosowane przy weryfikacji hipotezy zerowej w teście  $T$ -Studenta.
6. W rezultacie testu z języka angielskiego  $N = 18$  uczniów kl. VII uzyskało następujące wyniki w punktach: 15, 13, 8, 14, 14, 14, 13, 18, 9, 15, 14, 16, 17, 16, 12, 11, 9, 14. W tym samym teście  $N = 15$  uczniów innej kl. VII (z rozszerzonym programem języka angielskiego) uzyskało następujące wyniki: 13, 14, 11, 14, 14, 11, 12, 19, 18, 10, 10, 19, 18, 17, 19. Ustal, czy występujące różnice między wynikami uzyskanymi przez uczniów obu badanych klas są istotne, czy przypadkowe.

7. W rezultacie sprawdzianu ortograficznego  $N = 13$  uczniów kl. III popełniło następujące liczby błędów: 8, 7, 9, 12, 5, 7, 9, 12, 9, 4, 7, 7, 7. Po zastosowaniu szeregu ćwiczeń ortograficznych przeprowadzono ten sam sprawdzian i uczniowie ci popełnili następującą liczbę błędów: 6, 5, 4, 8, 7, 8, 10, 9, 8, 4, 6, 5, 4. Ustal, czy występujące różnice między średnią liczbą błędów popełnionych przez uczniów w obu sprawdzianach są istotne, czy przypadkowe. Dokonaj interpretacji uzyskanych wyników.
8. Przed przeprowadzeniem eksperymentu w klasie eksperymentalnej w rezultacie testu z matematyki dla  $N = 18$  uczniów uzyskano następujące wyniki: 9, 12, 25, 14, 17, 21, 20, 16, 18, 13, 17, 28, 19, 17, 12, 20, 17, 16. Po przeprowadzeniu eksperymentu, w którym jako czynnik eksperymentalny wprowadzono pracę z nowym podręcznikiem, uzyskano dla tych samych uczniów następujące wyniki: 14, 16, 21, 18, 15, 28, 22, 18, 19, 11, 19, 30, 20, 18, 18, 19, 16, 18. Ustal, czy występujące różnice między średnimi wynikami testu z matematyki przed eksperymentem i po jego zakończeniu są istotne, czy przypadkowe. Dokonaj interpretacji uzyskanych wyników.
9. W jakim celu sprawdza się istotność współczynnika korelacji  $r$  ?
10. Na wybranych przez siebie przykładach sprawdź istotność obliczonego współczynnika korelacji  $r$ :
- a/  $r = + 0,57$  ,  $N = 20$ .                      b/  $r = - 0,57$  ,  $N = 20$ .  
c/  $r = + 0,27$  ,  $N = 30$ .                      d/  $r = + 0,23$  ,  $N = 92$ .
11. Na wybranych przez siebie przykładach sprawdź istotność współczynnika korelacji rangowej  $R_s$ :
- a/  $R_s = - 0,61$  ,  $N = 22$ .                      b/  $R_s = + 0,37$  ,  $N = 26$ .
12. Kiedy mają zastosowanie testy nieparametryczne?
13. Jak jest sformułowana hipoteza zerowa w teście zgodności  $\chi^2$ ?
14. Dla zamieszczonych danych dokonaj weryfikacji hipotezy zgodności:

| Wynik testu $X$ | $f_e$    |
|-----------------|----------|
| 20 - 24         | 9        |
| 25 - 29         | 20       |
| 30 - 34         | 31       |
| 35 - 39         | 17       |
| 40 - 44         | 8        |
| 45 - 49         | 4        |
|                 | $N = 89$ |

15. Kiedy ma zastosowanie test niezależności  $\chi^2$ ?

16. Jak jest sformułowana hipoteza zerowa w teście niezależności  $\chi^2$ ?

17. Dla zamieszczonych danych dokonaj weryfikacji hipotezy niezależności.

W przypadku stwierdzenia związku, ustal jego siłę i dokonaj interpretacji:

| Płeć      | Stosunek do zawodu nauczycielskiego |           | Ogółem |
|-----------|-------------------------------------|-----------|--------|
|           | Pozytywny                           | Negatywny |        |
| Kobieta   | 35                                  | 15        | 50     |
| Mężczyzna | 9                                   | 29        | 38     |
| Ogółem    | 44                                  | 44        | 88     |

| Płeć        | Udział w zajęciach wyrównawczych |     | Ogółem |
|-------------|----------------------------------|-----|--------|
|             | Tak                              | Nie |        |
| Dziewczynka | 3                                | 21  | 24     |
| Chłopiec    | 14                               | 4   | 18     |
| Ogółem      | 17                               | 25  | 42     |

18. Dla niżej zamieszczonych danych zweryfikuj hipotezę niezależności. Ustal, czy istnieje związek między wykształceniem matek i wynikami w nauce ich dzieci – uczniów kl. II.

| Wykształcenie matek | Rodzaj wyników w nauce |         |        |
|---------------------|------------------------|---------|--------|
|                     | Wysokie                | Średnie | Niskie |
| podstawowe          | 4                      | 13      | 5      |
| zawodowe            | 10                     | 16      | 8      |
| średnie             | 18                     | 13      | 5      |
| wyższe              | 10                     | 7       | –      |

Weryfikacji dokonaj na poziomie istotności 0,01. W przypadku stwierdzenia związku między zmiennymi oblicz jego siłę i dokonaj interpretacji.

19. W kształtowaniu i rozwijaniu zainteresowań dzieci dużą rolę pełnią rodzice. Dla niżej zamieszczonych wyników badań zweryfikuj hipotezę niezależności. Ustal, czy istnieje związek między rodzajem zainteresowań rodziców i zainteresowaniami ich dzieci.

| Zainteresowania rodziców | Zainteresowania uczniów |    |    |    |
|--------------------------|-------------------------|----|----|----|
|                          | A                       | B  | C  | D  |
| A                        | 19                      | 8  | 5  | 6  |
| B                        | 14                      | 40 | 20 | 10 |
| C                        | 17                      | 25 | 42 | 18 |
| D                        | 12                      | 10 | 7  | 31 |

Weryfikacji dokonaj na poziomie istotności 0,01. W przypadku stwierdzenia związku między zmiennymi oblicz jego siłę i dokonaj jego interpretacji.

20. Na podstawie niżej zamieszczonych danych zweryfikuj hipotezę o niezależności formy spędzania czasu wolnego przez dzieci w młodszym wieku szkolnym od ich płci.

| Płeć        | Forma spędzania czasu wolnego |              |              |              |      |
|-------------|-------------------------------|--------------|--------------|--------------|------|
|             | Oglądanie TV                  | Gry i zabawy | Czytelnictwo | Pomoc w domu | Inne |
| Dziewczynka | 15                            | 11           | 23           | 19           | 6    |
| Chłopiec    | 15                            | 20           | 14           | 17           | 7    |

Weryfikacji dokonaj na poziomie istotności 0,05. W przypadku stwierdzenia związku między zmiennymi oblicz jego siłę i dokonaj interpretacji.

21. Na wybranych przez siebie przykładach, dla podanych wyników badań, dokonaj weryfikacji hipotezy niezależności oraz ustal poziom istotności. W przypadku stwierdzenia związku między zmiennymi oblicz jego siłę i dokonaj interpretacji:

$$a/ \chi^2_{emp.} = 10,28, \quad df = (w-1)(k-1) = 2, \quad N = 72.$$

$$b/ \chi^2_{emp.} = 1,12, \quad df = (w-1)(k-1) = 1, \quad N = 92.$$

$$c/ \chi^2_{emp.} = 13,45, \quad df = (w-1)(k-1) = 6, \quad N = 105.$$

# PODSUMOWANIE – WYKORZYSTANIE TECHNIK KOMPUTEROWYCH W PRZETWARZANIU DANYCH

W skrypcie przedstawiono jedynie elementarne i wybrane zagadnienia ze statystyki dla pedagogów, dotyczące grupowania i prezentacji materiału empirycznego, opisu zbiorowości statystycznej, współzależności zjawisk, elementów wnioskowania statystycznego, a także weryfikacji hipotez statystycznych. Zaprezentowano niektóre schematy postępowania w dokonywaniu opisu i analizy statystycznej zgromadzonego materiału empirycznego.

W zakończeniu należy stwierdzić, że znajomość przedstawionych treści może być przydatna studentowi w przygotowywaniu pracy magisterskiej.

Jak już wskazano w skrypcie, typ operacji na zbiorze danych empirycznych uzależniony jest od postawionego wcześniej problemu – pytania, które wyznacza kierunek poszukiwań badawczych oraz właściwości skali pomiarowej, w której zarejestrowano dane empiryczne. Czynności mające na celu udzielenie odpowiedzi na postawiony problem na podstawie zestawu operacji na danych empirycznych, wyrażonych w określonej skali pomiarowej nazywa się **przetwarzaniem wyników badań**.

W przypadku operacji bardzo prostych znajdują zastosowanie konwencjonalne procedury przetwarzania danych (ich niektóre przykłady zawiera przedstawiony tekst).

Jeśli natomiast liczba danych empirycznych jest duża, a typ wykonywanych operacji skomplikowany, można skorzystać z gotowych zestawów programów komputerowych (statystycznych, graficznych, edytorskich itp.) instalowanych

w mikrokomputerach klasy IBM PC, które w istotnym stopniu ułatwiają i przyspieszają realizację potrzeb badacza.

W procedurach przetwarzania wyników badań w naukach społecznych wykorzystuje się następujące typy oprogramowania:

1. programy numeryczne w środowisku komputerów klasy IBM PC (służą do przetwarzania danych ilościowych i jakościowych uzyskanych w badaniach empirycznych);
2. programy edytorskie (służą do przetwarzania danych tekstowych, mogą być wykorzystywane w poszczególnych etapach tworzenia tekstu, jak i w końcowym opracowywaniu wyników);
3. programy graficzne (służą do przetwarzania danych na obrazy płaskie i grafikę przestrzenną);
4. programy wspomagające i usługowe (służą do łączenia różnych programów, np. numerycznych z graficznymi lub edytorskimi).

W celu przetwarzania wyników badań w pedagogice, psychologii i innych naukach społecznych opracowano specjalne zestawy programów. Najbardziej znanym, należącym do programów numerycznych jest SPSS (ang. *Statistical Package for the Social Sciences*), czyli pakiet programów statystycznych dla nauk społecznych oraz programy STATGRAF i STATISTICA.

W skład standardowego pakietu programów statystycznych SPSS/PC wchodzi m. in. następujące zestawy procedur statystycznych: analiza wariancji, analiza korelacji, badanie tablic kontyngencji, wyliczanie statystyk z próby, analiza dyskryminacji, analiza czynnikowa, tworzenie histogramów, testy nieparametryczne, modele hierarchiczne, analiza regresji.<sup>45</sup>

Dane dla systemu SPSS/PC przygotowuje się najczęściej przy użyciu dowolnego edytora lub programu dającego zbiór tekstowy (ASCII), w szczególności nadają się tu programy dBase, Framework, Symphony, Lotus 123 lub NE, PE, a nawet dowolny edytor wbudowany w język oprogramowania<sup>46</sup>.

A oto zalecany sposób pracy z systemem SPSS/PC w pedagogice<sup>47</sup>:

1. Zaplanowanie i przeprowadzenie badań empirycznych oraz przygotowanie danych pod kątem analizy numerycznej SPSS/PC. Wcho-

---

<sup>45</sup> A. Niemiec, *Oprogramowanie mikrokomputera IBM PC -- System statystyczny SPSS/PC*, Wrocław 1992, s. 5.

<sup>46</sup> J. Gnitecki, *Pomiar i przetwarzanie wyników badań w pedagogice empirycznej*, Poznań 1992, s. 167.

<sup>47</sup> A. Niemiec, *Oprogramowania ...*, *op. cit.*, s. 6-10.

- dają tu w zakres takie standardowe czynności jak: sformułowanie problemów badawczych, dobór modelu badawczego, ustalenie listy zmiennych, zdefiniowanie zmiennych, opracowanie wskaźników, skonstruowanie i zweryfikowanie narzędzi badawczych, przeprowadzenie badań i uporządkowanie zbioru danych.
2. Przygotowanie danych i przygotowanie zbioru tekstowego w formacie dBase, Framework, Lotus 123.
  3. Pierwsza sesja z SPSS:
    - przygotowanie nagłówka,
    - nadanie nazw zmiennym, obliczenie dla danych wartości zmiennych pomocniczych,
    - wstępne przetwarzanie danych,
    - wykrywanie błędów, analiza danych odstających, brakujących.
  4. Utworzenie zbioru roboczego:
    - określenie środowiska w postaci wewnętrznej systemu SPSS.
    - przygotowanie zadania dla systemu (NE, PE, SK).
  5. Analizy statystyczne:
    - pobranie podzbioru danych (zlecenie GET...DROP...),
    - wykonanie analizy.
  6. Edycja zbioru wyników.
  7. Zakończenie pracy nad problemem badawczym:
    - skasowanie zbiorów roboczych,
    - archiwowanie wyników.

Przykłady procedur przetwarzania wyników badań w pedagogice można znaleźć np. w cytowanym już opracowaniu J. Gniteckiego.

Aktualnie istnieją duże możliwości wykorzystania komputerów w przetwarzaniu wyników badań w pedagogice i w psychologii. Zastosowanie znajdują szczególnie, chociaż nie tylko, przy prowadzeniu badań wieloczynnikowych uwzględniających interakcję między nimi. Elektroniczna technika obliczeniowa umożliwia sprawną i precyzyjną analizę rezultatów takich badań.

## Bibliografia

- Blalock H., *Statystyka dla socjologów*, Warszawa 1977.
- Brzeziński J., *Metodologia badań psychologicznych*, Warszawa 1997.
- Brzeziński J., *Elementy metodologii badań psychologicznych*, Warszawa 1984.
- Clauss G., Ebner H., *Podstawy statystyki dla psychologów, pedagogów i socjologów*, PZWS, Warszawa 1972.
- Cleeg F., *Po prostu statystyka*, WSiP, Warszawa 1994.
- Ferguson G.A., Takane Y., *Analiza statystyczna w psychologii i pedagogice*, Warszawa 1997.
- Gnitecki J., *Pomiar i przetwarzanie wyników badań w pedagogice empirycznej*, UAM, Poznań 1992.
- Góralski A., *Metody opisu i wnioskowania statystycznego w psychologii i pedagogice*, Warszawa 1987.
- Góralski A. (red.), *Metody badań pedagogicznych w zarysie*, WSPS, Warszawa 1994.
- Gruszczyński L.A., *Elementy statystyki dla socjologów*, Katowice 1986.
- Guilford J.P., *Podstawowe metody statystyczne w psychologii i pedagogice*, Warszawa 1960.
- Jaworska A., Michalicka M., *Zastosowanie niektórych testów nieparametrycznych w badaniach psychologicznych i pedagogicznych*, [w:] Wołoszynowa L. (red.), *Metody badań psychologicznych*, s. III, t. 3, Warszawa 1978.
- Lewicki Cz., *Zbiór zadań ze statystyki dla pedagogów*, Wydawnictwo FOSZE, Rzeszów 1993.

- Łachowski J., *Metody statystyczne w naukach pedagogicznych*, WSP, Bydgoszcz 1978.
- Łaniec J. D., *Elementy statystyki dla pedagogów*, WSP, Olsztyn 1994.
- Mayntz R., Holm K., Hubner P., *Wprowadzenie do metod socjologii empirycznej*, Warszawa 1985.
- Niemiec A., *Oprogramowanie mikrokomputera IBM PC – system statystyczny SPSS/PC*, Wrocław 1992.
- Niemierko B., *Testy osiągnięć szkolnych. Podstawowe pojęcia i technika obliczeniowa*, Warszawa 1975.
- Nowaczyk Cz., *Podstawy metod statystycznych dla pedagogów*, Warszawa 1985.
- Podoski K., *Metody statystyczne. Podstawowe pojęcia i zastosowanie w naukach społecznych*, WSP, Gdańsk 1967.
- Peuker Z., *Statystyka*, WSiP, Warszawa 1987.
- Puchalski T., *Statystyka. Wykład podstawowych zagadnień*, Warszawa 1978.
- Rogoziński Z., *Statystyka opisowa i indukcyjna*, Warszawa 1971.
- Rogoziński Z., *Wstęp do statystyki społecznej*, Warszawa 1969.
- Skorny A., *Prace magisterskie z psychologii i pedagogiki*, WSiP, Warszawa 1984.
- Sztumski J., *Wstęp do metod i technik badań społecznych*, Katowice 1995.



## **ANEKS**

# **TABLICE STATYSTYCZNE**



# **Tablica I. Rozkład normalny**

## POWIERZCHNIA POD KRZYWĄ NORMALNĄ

| $z$  | Powierzchnia<br>od średniej do $z$ | Powierzchnia<br>większej części | Powierzchnia<br>mniejszej części | $y$    |
|------|------------------------------------|---------------------------------|----------------------------------|--------|
| 0,00 | 0,0000                             | 0,5000                          | 0,5000                           | 0,3989 |
| 0,05 | 0,0199                             | 0,5199                          | 0,4801                           | 0,3984 |
| 0,10 | 0,0398                             | 0,5398                          | 0,4602                           | 0,3970 |
| 0,15 | 0,0596                             | 0,5596                          | 0,4404                           | 0,3945 |
| 0,20 | 0,0793                             | 0,5793                          | 0,4207                           | 0,3910 |
| 0,25 | 0,0987                             | 0,5937                          | 0,4013                           | 0,3867 |
| 0,30 | 0,1179                             | 0,6179                          | 0,3821                           | 0,3814 |
| 0,35 | 0,1368                             | 0,6368                          | 0,3632                           | 0,3752 |
| 0,40 | 0,1554                             | 0,6554                          | 0,3446                           | 0,3683 |
| 0,45 | 0,1736                             | 0,6736                          | 0,3264                           | 0,3605 |
| 0,50 | 0,1915                             | 0,6915                          | 0,3085                           | 0,3521 |
| 0,55 | 0,2088                             | 0,7088                          | 0,2912                           | 0,3429 |
| 0,60 | 0,2257                             | 0,7257                          | 0,2743                           | 0,3332 |
| 0,65 | 0,2422                             | 0,7422                          | 0,2578                           | 0,3230 |
| 0,70 | 0,2580                             | 0,7580                          | 0,2420                           | 0,3123 |
| 0,75 | 0,2734                             | 0,7734                          | 0,2266                           | 0,3011 |
| 0,80 | 0,2881                             | 0,7881                          | 0,2119                           | 0,2897 |
| 0,85 | 0,3023                             | 0,8023                          | 0,1977                           | 0,2780 |
| 0,90 | 0,3159                             | 0,8159                          | 0,1841                           | 0,2661 |
| 0,95 | 0,3289                             | 0,8289                          | 0,1711                           | 0,2541 |
| 1,00 | 0,3413                             | 0,8413                          | 0,1587                           | 0,2420 |
| 1,05 | 0,3531                             | 0,8531                          | 0,1469                           | 0,2299 |
| 1,10 | 0,3643                             | 0,8643                          | 0,1357                           | 0,2179 |
| 1,15 | 0,3749                             | 0,8749                          | 0,1251                           | 0,2059 |
| 1,20 | 0,3849                             | 0,8849                          | 0,1151                           | 0,1942 |
| 1,25 | 0,3944                             | 0,8944                          | 0,1056                           | 0,1826 |
| 1,30 | 0,4032                             | 0,9032                          | 0,0968                           | 0,1714 |
| 1,35 | 0,4115                             | 0,9115                          | 0,0885                           | 0,1604 |
| 1,40 | 0,4192                             | 0,9192                          | 0,0808                           | 0,1497 |
| 1,45 | 0,4265                             | 0,9265                          | 0,0735                           | 0,1394 |
| 1,50 | 0,4332                             | 0,9332                          | 0,0668                           | 0,1295 |
| 1,55 | 0,4394                             | 0,9394                          | 0,0606                           | 0,1200 |
| 1,60 | 0,4452                             | 0,9452                          | 0,0548                           | 0,1109 |
| 1,65 | 0,4505                             | 0,9505                          | 0,0495                           | 0,1123 |
| 1,70 | 0,4554                             | 0,9554                          | 0,0446                           | 0,0940 |

**Rozkład normalny – powierzchnia pod krzywą normalną**

| $z$  | Powierzchnia<br>od średniej do $z$ | Powierzchnia<br>większej części | Powierzchnia<br>mniejszej części | $y$         |
|------|------------------------------------|---------------------------------|----------------------------------|-------------|
| 1,75 | 0,4599                             | 0,9599                          | 0,0401                           | 0,0863      |
| 1,80 | 0,4641                             | 0,9641                          | 0,0359                           | 0,0790      |
| 1,85 | 0,4678                             | 0,9678                          | 0,0322                           | 0,0721      |
| 1,90 | 0,4713                             | 0,9713                          | 0,0287                           | 0,0656      |
| 1,95 | 0,4744                             | 0,9744                          | 0,0256                           | 0,0596      |
| 2,00 | 0,4772                             | 0,9772                          | 0,0228                           | 0,0540      |
| 2,05 | 0,4798                             | 0,9798                          | 0,0202                           | 0,0488      |
| 2,10 | 0,4821                             | 0,9821                          | 0,0179                           | 0,0440      |
| 2,15 | 0,4842                             | 0,9842                          | 0,0158                           | 0,0396      |
| 2,20 | 0,4861                             | 0,9861                          | 0,0139                           | 0,0355      |
| 2,25 | 0,4878                             | 0,9878                          | 0,0122                           | 0,0317      |
| 2,30 | 0,4893                             | 0,9893                          | 0,0107                           | 0,0283      |
| 2,35 | 0,4906                             | 0,9906                          | 0,0094                           | 0,0252      |
| 2,40 | 0,4918                             | 0,9918                          | 0,0082                           | 0,0224      |
| 2,45 | 0,4929                             | 0,9929                          | 0,0071                           | 0,0198      |
| 2,50 | 0,4938                             | 0,9938                          | 0,0062                           | 0,0175      |
| 2,60 | 0,4953                             | 0,9953                          | 0,0047                           | 0,0136      |
| 2,70 | 0,4965                             | 0,9965                          | 0,0035                           | 0,0104      |
| 2,80 | 0,4974                             | 0,9974                          | 0,0026                           | 0,0079      |
| 2,90 | 0,4981                             | 0,9981                          | 0,0019                           | 0,0060      |
| 3,00 | 0,49865                            | 0,99865                         | 0,00135                          | 0,0044      |
| 3,10 | 0,49903                            | 0,99903                         | 0,00097                          | 0,0033      |
| 3,20 | 0,49931                            | 0,99931                         | 0,00069                          | 0,0024      |
| 3,40 | 0,49966                            | 0,99966                         | 0,00034                          | 0,0012      |
| 3,60 | 0,49984                            | 0,99984                         | 0,00016                          | 0,00061     |
| 3,80 | 0,499928                           | 0,999928                        | 0,000072                         | 0,00029     |
| 4,00 | 0,4999683                          | 0,9999683                       | 0,0000317                        | 0,00013     |
| 4,50 | 0,4999966                          | 0,9999966                       | 0,0000034                        | 0,000015    |
| 5,00 | 0,49999971                         | 0,99999971                      | 0,00000029                       | 0,0000015   |
| 6,00 | 0,49999999                         | 0,99999999                      | 0,000000001                      | 0,000000006 |

 $z$  – zmienna standaryzowana $y$  – rzędna punktu, wartość względna funkcji gęstości  $x/\delta$ 

Źródło: J.P. Guilford, *Podstawowe metody statystyczne w psychologii i pedagogice*,  
Warszawa 1964.

Tablica II. Rozkład T-Studenta

PRAWDOPODOBIENSTWO BŁĘDU DLA TESTU DWUSTRONNEGO ( $\alpha$  w procentach)

| df       | $\alpha$ |      |      |      |       |      |       |       |
|----------|----------|------|------|------|-------|------|-------|-------|
|          | 50       | 25   | 10   | 5    | 2     | 1    | 0,2   | 0,1   |
| 1        | 1,00     | 2,41 | 6,31 | 12,7 | 31,82 | 63,7 | 318,3 | 637,0 |
| 2        | 0,816    | 1,60 | 2,92 | 4,30 | 6,97  | 9,92 | 22,33 | 31,6  |
| 3        | 0,765    | 1,42 | 2,35 | 3,18 | 4,54  | 5,84 | 10,22 | 12,9  |
| 4        | 0,741    | 1,34 | 2,13 | 2,78 | 3,75  | 4,60 | 7,17  | 8,61  |
| 5        | 0,727    | 1,30 | 2,01 | 2,57 | 3,37  | 4,03 | 5,89  | 6,86  |
| 6        | 0,718    | 1,27 | 1,94 | 2,45 | 3,14  | 3,71 | 5,21  | 5,96  |
| 7        | 0,711    | 1,25 | 1,89 | 2,36 | 3,00  | 3,50 | 4,79  | 5,40  |
| 8        | 0,706    | 1,24 | 1,86 | 2,31 | 2,90  | 3,36 | 4,50  | 5,04  |
| 9        | 0,703    | 1,23 | 1,83 | 2,26 | 2,82  | 3,25 | 4,30  | 4,78  |
| 10       | 0,700    | 1,22 | 1,81 | 2,23 | 2,76  | 3,17 | 4,14  | 4,59  |
| 11       | 0,697    | 1,21 | 1,80 | 2,20 | 2,72  | 3,11 | 4,03  | 4,44  |
| 12       | 0,695    | 1,21 | 1,78 | 2,18 | 2,68  | 3,05 | 3,93  | 4,32  |
| 13       | 0,694    | 1,20 | 1,77 | 2,16 | 2,65  | 3,01 | 3,85  | 4,22  |
| 14       | 0,692    | 1,20 | 1,76 | 2,14 | 2,62  | 2,98 | 3,79  | 4,14  |
| 15       | 0,691    | 1,20 | 1,75 | 2,13 | 2,60  | 2,95 | 3,73  | 4,07  |
| 16       | 0,690    | 1,19 | 1,75 | 2,12 | 2,58  | 2,92 | 3,69  | 4,01  |
| 17       | 0,689    | 1,19 | 1,74 | 2,11 | 2,75  | 2,90 | 3,65  | 3,96  |
| 18       | 0,688    | 1,19 | 1,73 | 2,10 | 2,55  | 2,88 | 3,61  | 3,92  |
| 19       | 0,688    | 1,19 | 1,73 | 2,09 | 2,54  | 2,86 | 3,58  | 3,88  |
| 20       | 0,687    | 1,18 | 1,73 | 2,09 | 2,53  | 2,85 | 3,55  | 3,85  |
| 21       | 0,686    | 1,18 | 1,72 | 2,08 | 2,52  | 2,83 | 3,53  | 3,82  |
| 22       | 0,686    | 1,18 | 1,72 | 2,07 | 2,51  | 2,82 | 3,51  | 3,79  |
| 23       | 0,685    | 1,18 | 1,71 | 2,07 | 2,50  | 2,81 | 3,49  | 3,77  |
| 24       | 0,685    | 1,18 | 1,71 | 2,06 | 2,49  | 2,80 | 3,47  | 3,74  |
| 25       | 0,684    | 1,18 | 1,71 | 2,06 | 2,49  | 2,80 | 3,45  | 3,72  |
| 26       | 0,684    | 1,18 | 1,71 | 2,06 | 2,48  | 2,78 | 3,44  | 3,71  |
| 27       | 0,684    | 1,18 | 1,71 | 2,05 | 2,47  | 2,77 | 3,42  | 3,69  |
| 28       | 0,683    | 1,17 | 1,70 | 2,05 | 2,47  | 2,76 | 3,41  | 3,67  |
| 29       | 0,683    | 1,17 | 1,70 | 2,05 | 2,46  | 2,76 | 3,40  | 3,66  |
| 30       | 0,683    | 1,17 | 1,70 | 2,04 | 2,46  | 2,76 | 3,39  | 3,65  |
| 40       | 0,681    | 1,17 | 1,68 | 2,02 | 2,42  | 2,70 | 3,31  | 3,55  |
| 60       | 0,679    | 1,16 | 1,67 | 2,00 | 2,39  | 2,66 | 3,23  | 3,46  |
| 120      | 0,677    | 1,16 | 1,66 | 1,98 | 2,36  | 2,62 | 3,17  | 3,37  |
| $\infty$ | 0,674    | 1,15 | 1,64 | 1,96 | 2,33  | 2,58 | 3,09  | 3,29  |
| df       | 25       | 12,5 | 5,0  | 2,5  | 1,0   | 0,5  | 0,1   | 0,05  |
|          | $\alpha$ |      |      |      |       |      |       |       |

( $\alpha$  w procentach)

PRAWDOPODOBIENSTWO BŁĘDU DLA TESTU JEDNOSTRONNEGO

Źródło: G. Clauss, H. Ebner, *Podstawy statystyki dla psychologów, pedagogów i socjologów*, PZWS Warszawa 1972, s. 388.

### Tablica III. Współczynniki korelacji

ISTOTNE NA POZIOMIE 5% i 1% DLA RÓŻNYCH LICZB STOPNI SWOBODY

| <i>df</i> | prawdopodobieństwo |       |
|-----------|--------------------|-------|
|           | 5%                 | 1%    |
| 1         | 0,997              | 1,000 |
| 2         | 0,950              | 0,990 |
| 3         | 0,878              | 0,959 |
| 4         | 0,811              | 0,917 |
| 5         | 0,754              | 0,874 |
| 6         | 0,707              | 0,834 |
| 7         | 0,666              | 0,798 |
| 8         | 0,632              | 0,765 |
| 9         | 0,602              | 0,735 |
| 10        | 0,576              | 0,708 |
| 11        | 0,553              | 0,684 |
| 12        | 0,532              | 0,661 |
| 13        | 0,514              | 0,641 |
| 14        | 0,497              | 0,623 |
| 15        | 0,482              | 0,606 |
| 16        | 0,468              | 0,590 |
| 17        | 0,456              | 0,575 |
| 18        | 0,444              | 0,561 |
| 19        | 0,433              | 0,549 |
| 20        | 0,423              | 0,537 |
| 21        | 0,413              | 0,526 |
| 22        | 0,404              | 0,515 |
| 23        | 0,396              | 0,505 |

| <i>df</i> | prawdopodobieństwo |       |
|-----------|--------------------|-------|
|           | 5%                 | 1%    |
| 24        | 0,388              | 0,496 |
| 25        | 0,381              | 0,487 |
| 26        | 0,374              | 0,478 |
| 27        | 0,367              | 0,470 |
| 28        | 0,361              | 0,463 |
| 29        | 0,355              | 0,456 |
| 30        | 0,349              | 0,449 |
| 35        | 0,325              | 0,418 |
| 40        | 0,304              | 0,393 |
| 45        | 0,288              | 0,372 |
| 50        | 0,273              | 0,354 |
| 60        | 0,250              | 0,325 |
| 70        | 0,233              | 0,302 |
| 80        | 0,217              | 0,283 |
| 90        | 0,205              | 0,267 |
| 100       | 0,195              | 0,254 |
| 125       | 0,174              | 0,228 |
| 150       | 0,159              | 0,208 |
| 200       | 0,138              | 0,181 |
| 300       | 0,133              | 0,148 |
| 400       | 0,098              | 0,128 |
| 500       | 0,088              | 0,115 |
| 1000      | 0,062              | 0,081 |

Źródło: J.P. Guilford, *Podstawowe metody statystyczne w psychologii i pedagogice*.  
Warszawa 1964, s. 543.

## Tablica IV. Wartość współczynnika korelacji korejności

ISTOTNE NA POZIOMIE 0,05 i 0,01 (TEST JEDNOSTRONNY)

| <i>N</i> | 0,05  | 0,01  |
|----------|-------|-------|
| 5        | 0,900 | 1,00  |
| 6        | 0,829 | 0,943 |
| 7        | 0,714 | 0,893 |
| 8        | 0,643 | 0,833 |
| 9        | 0,600 | 0,783 |
| 10       | 0,564 | 0,764 |
| 12       | 0,506 | 0,712 |
| 14       | 0,456 | 0,645 |
| 16       | 0,425 | 0,601 |
| 18       | 0,399 | 0,543 |
| 20       | 0,377 | 0,534 |
| 22       | 0,359 | 0,508 |
| 24       | 0,343 | 0,485 |
| 26       | 0,330 | 0,465 |
| 28       | 0,317 | 0,448 |
| 30       | 0,306 | 0,432 |

Źródło: G. Clauss, H. Ebner, *Podstawy statystyki dla psychologów, pedagogów i socjologów*, PZWS Warszawa 1972, s. 287.

**Tablica V. Wartość  $\chi^2$**

| df | $\alpha$ |         |         |        |        |        |        |        |        |        |        |        |        |        |
|----|----------|---------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|    | 0,99     | 0,98    | 0,95    | 0,90   | 0,80   | 0,70   | 0,50   | 0,30   | 0,20   | 0,10   | 0,05   | 0,02   | 0,01   | 0,001  |
| 1  | 0,03157  | 0,03628 | 0,00393 | 0,0158 | 0,0642 | 0,148  | 0,455  | 1,074  | 1,642  | 2,706  | 3,841  | 5,412  | 6,635  | 10,827 |
| 2  | 0,0201   | 0,0404  | 0,103   | 0,211  | 0,446  | 0,713  | 1,386  | 2,408  | 3,219  | 4,605  | 5,991  | 7,824  | 9,210  | 13,815 |
| 3  | 0,115    | 0,185   | 0,325   | 0,584  | 1,005  | 1,424  | 2,366  | 3,665  | 4,642  | 6,251  | 7,815  | 9,837  | 11,348 | 16,268 |
| 4  | 0,297    | 0,429   | 0,711   | 1,064  | 1,649  | 2,195  | 3,357  | 4,878  | 5,989  | 7,779  | 9,488  | 11,668 | 13,277 | 18,465 |
| 5  | 0,554    | 0,752   | 1,145   | 1,610  | 2,343  | 3,000  | 4,351  | 6,064  | 7,289  | 9,236  | 11,070 | 13,388 | 15,086 | 20,517 |
| 6  | 0,872    | 1,134   | 1,635   | 2,204  | 3,070  | 3,828  | 5,348  | 7,231  | 8,558  | 10,645 | 12,592 | 15,033 | 16,812 | 22,457 |
| 7  | 1,239    | 1,564   | 2,167   | 2,833  | 3,822  | 4,671  | 6,346  | 8,383  | 9,803  | 12,017 | 14,067 | 16,622 | 18,475 | 24,322 |
| 8  | 1,646    | 2,032   | 2,733   | 3,490  | 4,594  | 5,527  | 7,344  | 9,524  | 11,030 | 13,362 | 15,507 | 18,168 | 20,090 | 26,125 |
| 9  | 2,088    | 2,532   | 3,325   | 4,168  | 5,380  | 6,393  | 8,343  | 10,656 | 12,242 | 14,684 | 16,919 | 19,679 | 21,666 | 27,877 |
| 10 | 2,558    | 3,059   | 3,940   | 4,856  | 6,179  | 7,267  | 9,342  | 11,781 | 13,442 | 15,987 | 18,307 | 21,161 | 23,209 | 29,588 |
| 11 | 3,053    | 3,609   | 4,575   | 5,578  | 6,989  | 8,148  | 10,341 | 12,899 | 14,631 | 17,275 | 19,675 | 22,618 | 24,725 | 31,264 |
| 12 | 3,571    | 4,178   | 5,226   | 6,304  | 7,807  | 9,034  | 11,340 | 14,011 | 15,812 | 18,549 | 21,026 | 24,054 | 26,217 | 32,909 |
| 13 | 4,107    | 4,765   | 5,892   | 7,042  | 8,634  | 9,926  | 12,340 | 15,119 | 16,985 | 19,812 | 22,362 | 25,472 | 27,688 | 34,528 |
| 14 | 4,660    | 5,368   | 6,571   | 7,790  | 9,467  | 10,821 | 13,339 | 16,222 | 18,151 | 21,064 | 23,685 | 26,873 | 29,141 | 36,123 |
| 15 | 5,229    | 5,985   | 7,261   | 8,547  | 10,307 | 11,721 | 14,339 | 17,322 | 19,311 | 22,307 | 24,996 | 28,259 | 30,578 | 37,697 |
| 16 | 5,812    | 6,614   | 7,962   | 9,312  | 11,152 | 12,624 | 15,338 | 18,418 | 20,465 | 23,542 | 26,296 | 29,633 | 32,000 | 39,252 |
| 17 | 6,408    | 7,255   | 8,672   | 10,058 | 12,002 | 13,531 | 16,338 | 19,511 | 21,615 | 24,769 | 27,587 | 30,995 | 33,409 | 40,790 |
| 18 | 7,015    | 7,906   | 9,390   | 10,865 | 12,857 | 14,440 | 17,338 | 20,601 | 22,760 | 25,989 | 28,869 | 32,346 | 34,805 | 42,312 |
| 19 | 7,633    | 8,567   | 10,117  | 11,651 | 13,716 | 15,352 | 18,338 | 21,689 | 23,900 | 27,204 | 30,144 | 33,687 | 36,191 | 43,820 |
| 20 | 8,260    | 9,237   | 10,851  | 12,443 | 14,578 | 16,266 | 19,337 | 22,775 | 25,038 | 28,412 | 31,410 | 35,020 | 37,566 | 45,315 |

Źródło: A. Góralski, *Metody opisu i wnioskowania statystycznego w psychologii*, Warszawa 1987, s. 327.