

4.2. Statystyczne opracowanie zebranego materiału

Zebrany i pogrupowany materiał badawczy należy poddać analizie statystycznej w celu dokonania pełnej i szczegółowej charakterystyki interesujących badacza zjawisk.

Zdaniem K. Zająca (1974, s. 14), „dokonując za pomocą metod statystycznych pewnych uogólnień w opisie statystycznym, wprowadzamy jednocześnie porządek w pozornym chaosie przypadkowych zmian, a więc wykrywamy prawidłowości występujące w zjawiskach masowych”.

W zależności od badanych zjawisk w analizie wyników stosuje się odpowiednią metodę statystyczną. Zastosowanie wybranej metody powinno zawsze łączyć się z właściwą interpretacją wyników (potwierdza to jej zrozumienie). Metody statystyczne służą do porządkowania materiału, do jego opisu traktowanego jako przygotowanie do analizy jakościowej i wnioskowania. Wykaz możliwych metod i technik opracowania zebranego materiału zamieszczono w tabeli 5.

Tabela 5. Metody i techniki opracowywania materiału empirycznego

Metoda	Technika
Metody statystyczne/ /kwantytatywne	Rozkład liczbowy – przedziały klasowe Miary tendencji centralnej: średnia arytmetyczna, modalna, mediana Miary zmienności: rozpiętość, odchylenia standardowe, proporcje, procenty, korelacje Miary rzetelności (błędy wnioskowania statystycznego) Testy istotności statystycznej

Źródło: opracowanie własne.

Pierwszą czynnością w opracowaniu statystycznym jest ujęcie w liczby podstawowej charakterystyki badań, odnoszącej się do wielkości próby pod kątem cech demograficznych i społecznych.

Podstawowe znaczenie dla późniejszego wyboru metod ma precyzyjne zdefiniowanie cech statystycznych oraz określenie **skal pomiaru**. Szeroko akceptowane jest przyjęcie czterech skal pomiaru: **nominalnej, porządkowej, interwałowej i ilorazowej**. Skala pomiaru determinuje wybór metody przy analizie współzależności zjawisk.

Według T. Pilcha (1995, s. 98) „skala jest szeregiem zdań ułożonych według określonego porządku wyczerpujących możliwe określenia badanego zjawiska, cechy lub układu”.

Skala zakłada istnienie pewnego ciągu logicznego stopni oznaczających natężenie pewnego zjawiska. Skala powinna być:

- *wyczerpująca* – uwzględniać wszystkie możliwe określenia danej sytuacji,
- *rzetelna* – dająca możliwość uzyskania takich samych wyników przy kolejnych badaniach tej samej próby,
- *trafna* – poszczególne stopnie skali powinny być logicznie dopasowane do rzeczywistości.

Na skali nominalnej możemy obliczyć zmienną najczęściej występującą, rozkład częstości dla danej zmiennej, jaki procent z całości stanowi dany pomiar. Przykładem zmiennej obliczanej na skali nominalnej jest np. płeć (kobieta, mężczyzna), pora roku (wiosna, lato, jesień, zima), kolor oczu (piwne, niebieskie, itp.). Skala nominalna może być zastosowana do wstępnej klasyfikacji zbioru badanego i jest raczej narzędziem badania niż jego efektem.

Zmienne na **skali porządkowej** mogą dotyczyć np. wykształcenia (podstawowe, zawodowe, średnie, wyższe), wielkości miejscowości, w których badani mieszkają (wieś, miasto), stanowiska pracy (pracownik biurowy, dyrektor, pracownik fizyczny). Służy do ułożenia w porządku elementów jakiegoś zbioru ze względu na wybraną cechę. Dla zmiennej na skali porządkowej możemy obliczyć zmienną najczęściej występującą, uszeregować zmienne od najniższej do najwyższej, natomiast nie możemy obliczyć odchylenia standardowego ani wariancji.

Skala interwałowa zawiera informacje z dwóch poprzednio wymienionych skal, czyli wyróżnia elementy zbioru, układa elementy zbioru w porządku od największego do najmniejszego. Oprócz tego określa odległość (interwał) między elementami zbioru na umownej skali liniowej.

Zmienne na **skali ilorazowej** mogą dotyczyć wieku, wzrostu, czasu trwania leczenia, czasu trwania nauki. Na skali ilorazowej możemy wyznaczyć dominantę, medianę, odchylenie standardowe, wariancję i wiele innych statystyk.

Po ustaleniu skal pomiaru możliwe jest dokonanie dalszych obliczeń statystycznych za pomocą miar średnich (średnia arytmetyczna, modalna, kwantyle) i miar zróżnicowania (zmienności, rozproszenia, dyspersji).

Miary położenia wskazują miejsce wartości najlepiej reprezentującej wszystkie wielkości badanej cechy. Są to średnie (arytmetyczne, geometryczne, harmoniczne) o symbolu \bar{x} , modalna o symbol M_o i kwantyle. Najczęściej stosowana jest **średnia arytmetyczna**, która stanowi sumę poszczególnych wartości badanej cechy wszystkich jednostek badanej próby podzieloną przez liczebność tej próby. Na wartość średniej arytmetycznej duży wpływ wywierają najmniejsze i największe wartości badanej cechy, czyli wartości skrajne.

Modalna jest to najczęściej występująca wartość badanej cechy, dlatego nazywana jest dominantą (oznaczamy ją symbolem D_o). Stosujemy ją wtedy, gdy chcemy za pomocą jednej liczby wyrazić wariant lub

wartość cechy najbardziej typowe, najczęściej spotykane. Sposób wyznaczania dominanty zależy od typu szeregu statystycznego, z którym mamy do czynienia. Jeżeli dane przedstawione są w postaci szeregu szczegółowego lub rozdzielczego, dominanta jest wartością cechy, której odpowiada największa liczebność, wobec czego wystarczy ją wskazać. Jeżeli dane są wprowadzone do arkusza kalkulacyjnego Excel, dominantę wyznaczamy, wybierając funkcję „wys. najczęściej”.

Mediana (oznaczamy ją symbolem Me) jest to wartość znajdująca się w środku, gdy zbiorowość jest uporządkowana pod względem wielkości jej elementów, tzn. od wartości najmniejszej do wartości największej. Mediana dzieli zbiorowość na dwie równe części w ten sposób, że połowa jednostek statystycznych ma wartość cechy niższą lub równą medianie, a połowa ma wartości równe lub wyższe od Me. Aby wyznaczyć medianę w arkuszu kalkulacyjnym Excel, wybieramy funkcję „mediana”.

Jeżeli konieczna jest bardziej szczegółowa analiza właściwości strukturalnych, oprócz mediany, która jest kwartylem drugim, znajdują zastosowanie kwartyle pierwszy i trzeci.

Kwartyle należą do miar statystycznych ogólnie zwanych **kwantylami**, które dzielą zbiorowość statystyczną w określonej proporcji.

Kwartyl pierwszy (symbol Q1) jest wartością cechy, która dzieli próbę tak, że 25% jednostek ma wartości od niej nie większe, a 75% – nie mniejsze.

Kwartyl trzeci (symbol Q3) jest wartością cechy jednostki, która dzieli próbę tak, że 75% jednostek ma od niej wartości nie większe, a 25% nie mniejsze.

Kwartyle wyznacza się w sposób analogiczny jak w przypadku mediany, z tym że należy uwzględnić numer kwantyla.

Miary zróżnicowania (zmienności, rozproszenia, dyspersji) informują, jak duże są różnice (odchylenia) między wartościami cech poszczególnych a średnią, najczęściej arytmetyczną. Do tej grupy miar należą rozstęp, wariancja i odchylenie standardowe.

Rozstęp wyraża różnicę między największą a najmniejszą wartością cechy.

Wariancja jest średnią arytmetyczną kwadratów odchyłeń poszczególnych wartości cechy od średniej arytmetycznej całej próby.

Odchylenie standardowe (symbol SD) jest najczęściej stosowaną miarą obok średniej. SD jest pierwiastkiem kwadratowym z wariancji. Wystarczy wiedzieć, że odchylenie standardowe jest tym większe, im cechy jednostek są bardziej zróżnicowane, i odwrotnie – małe odchylenie standardowe świadczy o niskim zróżnicowaniu poszczególnych wartości badanej cechy. Wyliczenie średniej wartości cechy i rozproszenia wartości nie pozwala ocenić asymetrii rozkładu wartości.

Dla uchwycenia pełnego obrazu różnic między szeregami danej cechy stosuje się **współczynnik asymetrii**, symbol As. Współczynnik asymetrii równy zeru wskazuje, że rozkład wartości jest symetryczny, współczyn-

nik dodatni oznacza asymetrię prawostronną, a ujemny – asymetrię lewostronną. Jeśli próbę studentów podzielimy na trzy podgrupy: studenci I, II, III roku studiów, to może okazać się, że rozkłady wartości badanej cechy (oceny egzaminów z pielęgniarstwa) się różnią.

W celu stwierdzenia występowania współzależności między dwiema cechami wykorzystujemy analizę korelacyjną, oceniając siłę i kierunek zależności. Wykrycie zależności jest możliwe, jeżeli materiał empiryczny jest przedstawiony w postaci szeregu korelacyjnego, inaczej prostego, lub tablicy korelacyjnej. Szereg korelacyjny zawiera uporządkowane parami warianty dwóch cech jednocześnie, stanowi zestawienie dwóch szeregów szczegółowych (tabela 6).

Tabela 6. Schemat szeregu korelacyjnego

Wariant zmiennej zależnej (X)	Wariant zmiennej niezależnej (Y)
X ₁	Y ₁
X ₂	Y ₂
..	..
..	..
X ₀	Y ₀

Źródło: Starzyńska W.: *Analiza współzależności dwóch cech*. W: *Podstawy statystyki* (red. W. Starzyńska). Edukacja Difin, Warszawa 2006.

Wybór miernika współzależności zależy od tego, jak dane są uporządkowane: czy w szeregu korelacyjnym (patrz tabela), czy w tablicy korelacyjnej. Istnieje wiele miar współzależności dla szeregu korelacyjnego. W tym rozdziale zostaną omówione dwa: współczynnik korelacji liniowej Pearsons'a i współczynnik korelacji rang Spearmana oraz empiryczna funkcja regresji zmiennej zależnej względem zmiennej niezależnej. Wymienione współczynniki pozwalają badać korelacje między dwoma cechami mierzalnymi, natomiast funkcja regresji pozwala na ustalenie związków przyczynowo-skutkowych.

Współczynnik korelacji liniowej Pearsona (r_{yx}) umożliwia zbadanie zależności między dwiema cechami mierzalnymi, między którymi ma miejsce zależność liniowa, tzn. jednostkowym przyrostem jednej cechy towarzyszy, średnio biorąc, stały przyrost lub spadek drugiej cechy. Współczynnik oblicza się według wzoru lub za pomocą arkusza Excel i funkcji „wsp. korelacji”. Po uaktywnieniu tego polecenia zaznaczamy jako „tablice” obszary (kolumny) ze zmiennymi (w dowolnej kolejności, np. tablicy 1 odpowiadają wartości zmiennej Y, a tablicy 2 – wartości zmiennej X). Przed automatycznym obliczeniem współczynnika należy sprawdzić, czy spełnione są założenia jego wyznaczania.

Współczynnik liniowy Pearsona przyjmuje zawsze wartości z przedziału od -1 do $+1$. Dodatni znak współczynnika korelacji wskazuje na dodatnią korelację między zmiennymi, co oznacza, że wzrostowi jednej cechy towarzyszy na ogół wzrost wartości drugiej cechy. Ujemny znak współczynnika korelacji wskazuje na ujemną korelację między zmiennymi, co oznacza, że wzrostowi jednej cechy towarzyszy na ogół spadek wartości drugiej cechy. Współczynnik korelacji równy 0 świadczyć może o zupełnym braku związku korelacyjnego między badanymi zmiennymi.

Orientacyjnie przyjmuje się, że siła korelacji między dwoma cechami jest:

- niewyraźna, gdy $|r_{yx}| < 0,2$,
- wyraźna, ale niska, gdy $0,2 < |r_{yx}| < 0,4$,
- umiarkowana, gdy $0,4 < |r_{yx}| < 0,7$,
- znacząca, gdy $0,7 < |r_{yx}| < 0,9$,
- bardzo silna, gdy $|r_{yx}| > 0,9$.

Współczynnik korelacji rang Spearmana służy do badania zależności dwóch cech przedstawionych w szeregu korelacyjnym. Cechy mogą być mierzalne i niemierzalne, ale dające się uporządkować przez nadanie im kolejnego numeru. To uporządkowanie określa się mianem rangi, a procedurę rangowaniem. Współczynnik rang Spearmana oblicza się ze wzoru. Przyjmuje on wartość z przedziału od -1 do 1 . Im wartość jest bliższa -1 , tym silniejsza ujemna korelacja między badanymi cechami, a z kolei im bliższa $+1$, tym silniejsza jest korelacja dodatnia. Wartość bliska 0 wskazuje na słabą zależność.

W celu ustalenia wpływu zmiennych uznanych za niezależne na zmienną uznaną za zależną wykorzystuje się analizę regresji i oblicza współczynnik regresji. Współczynnik regresji dodatni oznacza, że wraz ze wzrostem zmiennej niezależnej o jednostkę zmienna zależna wzrośnie o b jednostek. Funkcja regresji jest rosnąca. Współczynnik regresji ujemny wskazuje na to, że pod wpływem wzrostu zmiennej niezależnej o jednostkę zmienna zależna zmaleje średnio o b jednostek. Funkcja regresji jest malejąca. Współczynnik regresji równy 0 świadczy o tym, że zmienna niezależna nie wywiera żadnego wpływu na zmienną zależną. Funkcja regresji jest stała (Zajac 1974; Dudkiewicz 2000; Starzyńska 2006).

Te podstawowe informacje na temat opisu statystycznego mają jedynie ułatwić zrozumienie artykułów w piśmiennictwie naukowym, do czego znajomość wzorów i umiejętność obliczeń nie jest konieczna. Natomiast osoby zainteresowane poszerzeniem wiedzy z zakresu statystyki odsyłamy do piśmiennictwa (patrz: Podgórski 2001; Wierziński 2000; Józwiak, Podgórski 1997; Starzyńska 2000; Apanowicz, Kubielski 1999; Ferguson, Takane 1999).