Główny Urząd Statystyczny

Podstawowe miary statystyczne:

Miary położenia

Materiały dla nauczyciela

MIARY POŁOŻENIA

MIARA 1

ŚREDNIA ARYTMETYCZNA

Miara klasyczna

MIARA 2

DOMINANTA

Miara pozycyjna

MIARA 3

MEDIANA

Miara pozycyjna

MIARA 4

KWARTYLE

Miara pozycyjna

MIARA 5

DECYLE

Miara pozycyjna

1. Wprowadzenie do statystyki opisowej

Termin statystyka współcześnie ma kilka znaczeń:

- zbiór danych liczbowych, przedstawiających kształtowanie się określonych zjawisk i procesów,
- wszelkie prace związane z gromadzeniem i opracowywaniem danych liczbowych,
- charakterystyki opisowe obliczane ze zbiorowości próbnych np. średnia arytmetyczna,
- dyscyplina naukowa mająca własne metody badawcze – nauka o ilościowych metodach badania prawidłowości występujących w zjawiskach masowych scharakteryzowanych za pomocą liczb.

1.1. Podstawowe pojęcia

Zbiorowość statystyczna (populacja generalna) jest to zbiór dowolnych elementów podobnych pod względem określonych właściwości i poddanych badaniu statystycznemu. Przykład: ludność Polski

Próba statystyczna – podzbiór populacji generalnej. Przykład: ludność województwa mazowieckiego.

Jednostka statystyczna – to element (obiekt) zbiorowości statystycznej. Przykład: jedna osoba.

Cecha statystyczna jest to właściwość obiektu tworzącego zbiorowość statystyczną.

Cechy statystyczne dzielą się na:

- **mierzalne** (ilościowe) – podawane liczbowo, przykład: wiek,
- **niemierzalne** (jakościowe) – podawane opisowo, przykład: płeć.

Cechy **ilościowe** można podzielić na:

- **skokowe** (dyskretne) - przyjmują wartości ze skończonych i przeliczalnych przedziałów liczbowych, ale bez wartości pośrednich, przykład: liczba gospodarstw domowych.
- **ciągłe** – przyjmują każdą wartość z określonego przedziału, przykład: wzrost.

1.2. Opracowanie i prezentacja materiału statystycznego

Opracowanie materiału statystycznego polega w uproszczeniu na przekształceniu danych statystycznych na informacje. Można tego dokonać poprzez grupowanie danych – polega na podziale zbiorowości na możliwie jednorodne grupy zgodnie z przyjętymi kryteriami. Jeżeli dokonuje się grupowania ze względu na jedną cechę to jest to grupowania proste, a gdy stosujemy więcej cech to jest to grupowanie złożone. W wyniku grupowania statystycznego tworzone są szeregi statystyczne (w przypadku jednej cechy) oraz tablice statystyczne (w przypadku co najmniej dwóch cech).

Szereg statystyczny jest to ciąg wariantów cechy uporządkowanych rosnąco lub malejąco, pogrupowany według określonych kryteriów. Szeregi dzielimy na:

- **wyliczające (szczegółowe)** – uporządkowane rosnąco lub malejąco wartości cechy. Szeregi te są niewygodne w wykrywaniu prawidłowości statystycznych z uwagi na to, że na ogół są długie,
- **rozdzielcze (strukturalne)** – zawierają ciągi wartości badanej cechy wraz z przypisanymi im liczebnościami.

2. Miary położenia

Miary położenia wskazują wokół jakich wartości skupia się rozkład analizowanych zmiennych. Dzieli się je na **klasyczne** i **pozycyjne**. Miary klasyczne to średnie: arytmetyczna, harmoniczna i geometryczna. Do miar pozycyjnych należy dominanta (modalna, wartość najczęstsza) oraz kwantyle. Wśród kwantyli najczęściej stosowane są: kwartyle (dzielące zbiorowość na cztery części pod względem liczebności), kwintyle (dzielące zbiorowość na pięć części), decyle (dzielące zbiorowość na dziesięć części) oraz percentyle (dzielące zbiorowość na sto części).

Obydwie grupy średnich (klasyczne i pozycyjne) nie wykluczają się, ale nawzajem uzupełniają. Każda z nich opisuje bowiem poziom wartości cechy z innego punktu widzenia.

2.1. Miary położenia klasyczne

2.1.1. Średnia arytmetyczna

Jest ilorazem sumy wartości zmiennej i liczebności badanej zbiorowości. Z szeregów wyliczających średnią arytmetyczną obliczamy następująco:

$$\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{N} = \frac{\sum_{j=1}^{N} x_j}{N},$$

gdzie:

$\bar{x}$ – symbol średniej arytmetycznej,

x_j – warianty cechy mierzalnej,

N – liczebność badanej zbiorowości.

Średnia ta nazywana jest **średnią arytmetyczną nieważoną (prostą, zwykłą)**, w przeciwieństwie do średniej arytmetycznej ważonej obliczanej z szeregów rozdzielczych punktowych i przedziałowych.

Jeżeli warianty zmiennej w badanej zbiorowości występują z różną częstotliwością, wówczas oblicza się średnią arytmetyczną ważoną. Wagami są liczebności (częstości) odpowiadające poszczególnym wariantom zmiennej.

Wzór na obliczenie średniej arytmetycznej z szeregów rozdzielczych punktowych przyjmuje następującą postać:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_k n_k}{N} = \frac{\sum_{i=1}^{k} x_i n_i}{N},$$

gdzie: n_i $(i = 1, 2, \ldots, k)$ - liczebność jednostek odpowiadająca poszczególnym wariantom zmiennej, N – ogólna liczebność badanej zbiorowości.

W szeregach rozdzielczych przedziałowych wartości zmiennej w każdej klasie nie są jednoznacznie określone, ale zawarte są w przedziale od – do. Dolną granicę przedziału klasowego zazwyczaj oznacza się symbolem x_{oi}, górną zaś - x_{1i}. W celu obliczenia średniej arytmetycznej z szeregu rozdzielczego przedziałowego należy uprzednio wyznaczyć środki przedziałów. Środki przedziałów klasowych – oznaczone symbolem $\dot{x}_i$ - obliczamy następująco:

$$\dot{x}_i = \frac{x_{0i} + x_{1i}}{2}; (i = 1, 2, \ldots, k).$$

Wzór na średnią arytmetyczną z szeregu rozdzielczego przedziału jest następujący:

$$\bar{x} = \frac{\dot{x}_1 n_1 + \dot{x}_2 n_2 + \cdots + \dot{x}_k n_k}{N} = \frac{\sum_{i=1}^{k} \dot{x}_i n_i}{N}.$$

Jeżeli zamiast liczebności absolutnych w obliczeniach zostaną wykorzystane wskaźniki struktury, to wzór na średnią arytmetyczną przyjmuje postać:

$$\bar{x} = \sum_{i=1}^{k} \dot{x}_i w_i,$$

gdzie: $w_i = \frac{n_i}{N}$, oraz $w_i \in (0; 1)$ oraz $\sum_{i=1}^{k} w_i = 1$.

Jeżeli zamiast liczebności absolutnych w obliczeniach zostaną wykorzystane procentowe wskaźniki częstości, to wzór na średnią arytmetyczną przyjmuje postać:

$$\bar{x} = \frac{\sum_{i=1}^{k} \dot{x}_i d_i}{100}.$$

gdzie: $d_i = \frac{n_i}{N} \times 100 = w_i \times 100$, oraz $d_i \in (0; 100)$ oraz $\sum_{i=1}^{k} d_i = 100$.

Przy obliczaniu średniej arytmetycznej z szeregów rozdzielczych przedziałowych arbitralnie wybiera się wartości reprezentujące poszczególne klasy. Oznacza to przyjęcie założenia o równomiernym rozkładzie jednostek w poszczególnych przedziałach klasowych. Założenie to nie zawsze jest jednak spełnione. Obliczona średnia arytmetyczna będzie wówczas wielkością przybliżoną. Ogólnie można stwierdzić, że dokładność średniej arytmetycznej obliczonej z szeregów rozdzielczych przedziałowych będzie tym większa, im przedziały klasowe będą węższe.

Średniej arytmetycznej nie można obliczyć z szeregu rozdzielczego o otwartych przedziałach klasowych. Jeśli otwarte przedziały klasowe mają niewielkie liczebności, to przed obliczeniem średniej można je umownie domknąć. Przyjmuje się, że otwarty przedział można domknąć, gdy liczba jednostek w tym przedziale nie przekracza 5% liczebności. Ze względu na to, że główną przyczyną niedomknięcia przedziałów jest wysokie rozproszenie skrajnych wartości cechy, trudno jest trafnie wybrać wartości reprezentujące otwarte klasy.

Średnia arytmetyczna posiada wiele własności. M.in:

- jako miara klasyczna jest wypadkową działania wszystkich wartości badanej cechy i spełnia nierówność: $x_{min} < \bar{x} < x_{max}$,
- suma odchyleń poszczególnych wartości zmiennej od średniej arytmetycznej wynosi zero, czyli:
-
  - $\sum_{j=1}^{N}(x_j - \bar{x}) = 0$ w przypadku szeregu wyliczającego,
  - $\sum_{i=1}^{k}(x_i - \bar{x})\, n_i = 0$ w przypadku szeregu rozdzielczego punktowego,
  - $\sum_{i=1}^{k}(\dot{x}_i - \bar{x})\, n_i = 0$ w przypadku szeregu rozdzielczego przedziałowego,
- jeśli pomnożymy średnią arytmetyczną przez ogólną liczebność badanej zbiorowości, to otrzymamy sumę wartości wszystkich jednostek, tzn. $N\bar{x} = \sum_{i=1}^{N} x_i$,
- średnia arytmetyczna sumy (różnicy) zmiennych równa się sumie (różnicy) ich średnich arytmetycznych,
- jeśli wszystkie wartości zmiennej powiększymy (pomniejszymy, podzielimy lub pomnożymy) o pewną stałą c, to średnia arytmetyczna będzie równa sumie (różnicy, ilorazowi lub iloczynowi) średniej arytmetycznej i stałej c:
  - $\frac{1}{N}\sum_{i=1}^{N}(x_i + c) = \bar{x} + c$,
- na poziom średniej arytmetycznej silny wpływ wywierają wartości ekstremalne (skrajne), przy czym wpływ ten jest silniejszy w przypadku wysokich wartości zmiennej;
- średnia arytmetyczna – jako wypadkowa wszystkich zaobserwowanych wartości cechy – jest wielkością abstrakcyjną. Oznacza to, że w niektórych przypadkach może ona przyjmować wartości w ogóle nie występujące w zbiorowości,
- średnia arytmetyczna jest miarą prawidłową tylko w odniesieniu do zbiorowości jednorodnych, o niewielkim zróżnicowaniu wartości zmiennej u poszczególnych jednostek. W miarę wzrostu asymetrii i dyspersji rozkładu, a także w rozkładach bi- i wielomodalnych należy do opisu wykorzystywać przeciętne pozycyjne.

2.1.2. Średnia harmoniczna

Jest odwrotnością średniej arytmetycznej z odwrotności wartości zmiennych. Z szeregów wyliczających średnią harmoniczną obliczamy ze wzoru:

$$H = \frac{N}{\sum_{j=1}^{N}\frac{1}{x_j}},$$

gdzie: H jest symbolem średniej harmonicznej.

Przy obliczaniu średniej harmonicznej z szeregów rozdzielczych (punktowych i przedziałowych) zachodzi konieczność stosowania wag (czyli uwzględniania liczebności). Z szeregów rozdzielczych punktowych średnią harmoniczną obliczamy następująco:

$$H = \frac{N}{\sum_{i=1}^{k} \frac{1}{x_i} n_i}.$$

Z szeregów rozdzielczych przedziałowych średnią harmoniczną obliczamy według tego samego wzoru, z tym że konkretne warianty cechy (x_i) zastępujemy środkami przedziałów klasowych ($\dot{x}_i$).

Średnią harmoniczną stosuje się wówczas, gdy wartości zmiennej podane są w jednostkach względnych, wagi zaś – w jednostkach względnych liczników. Przykładowo można tu wymienić takie zmienne, jak:

- prędkość pojazdu (zmienna – km/h, waga – liczba kilometrów),
- gęstość zaludnienia (zmienna – mieszkańcy/km^2, waga – liczba mieszkańców).

2.1.3. Średnia geometryczna

Jest pierwiastkiem k-tego stopnia z iloczynu k zmiennych, czyli dla szeregów wyliczających:

$$G = \sqrt[k]{x_1 x_2 \dots x_k} = \sqrt[k]{\prod_{i=1}^{k} x_i},$$

gdzie: π jest znakiem iloczynu określonej liczby wyrazów.

Średnią geometryczną dla szeregu rozdzielczego punktowego oblicza się ze wzoru:

$$G = \sqrt[N]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}} = \sqrt[N]{\prod_{i=1}^{k} x_i^{n_i}},$$

gdzie:

$N = \sum_{i=1}^{k} n_i.$

Średnią geometryczną dla szeregów rozdzielczych przedziałowych oblicza się według tego samego wzoru z tym, że konkretne wartości zmiennej $(x_1, x_2, \dots, x_k)$ zastępuje się środkami przedziałów klasowych $\dot{x}_i$ $(i = 1, 2, \dots, k)$, gdzie k oznacza liczbę przedziałów klasowych w danym szeregu rozdzielczym.

Średnia geometryczna znajduje zastosowanie przy badaniu średniego tempa zmian zjawisk, których rozwój przedstawiony jest w postaci szeregów dynamicznych.

2.2. Miary położenia pozycyjne

2.2.1. Dominanta

Dominanta (modalna, wartość najczęstsza) jest to najczęściej powtarzająca się wartość zmiennej w szeregu statystycznym. Określa ona najbardziej typową wartość zmiennej w badanej zbiorowości. Charakterystyczną cechą dominanty jest możliwość jej wyznaczenia zarówno z szeregów dotyczących cechy mierzalnej, jak i niemierzalnej. Wartość dominanty można ustalić jedynie dla rozkładów jednomodalnych. W szeregach wyliczających i rozdzielczych punktowych dominanta jest tą wartością cechy, której odpowiada największa liczebność.

W szeregach rozdzielczych przedziałowych bezpośrednio można określić tylko przedział, w którym znajduje się dominanta. Jest to przedział o największej liczebności. Konkretną wartość liczbową, należącą do tego przedziału i będącą dominantą, wyznacza się za pomocą następującego wzoru interpolacyjnego:

$$D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} i_D,$$

gdzie:

D – symbol dominanty,

x_D- dolna granica klasy, w której znajduje się dominanta,

n_D- liczebność przedziału dominanty,

n_{D-1}- liczebność przedziału poprzedzającego przedział dominanty,

n_{D+1}- liczebność przedziału następującego po przedziale dominanty,

i_D- interwał (rozpiętość) przedziału dominanty.

Przy wyznaczaniu dominanty za pomocą powyższego wzoru wymagane jest, aby rozpiętości przedziału dominanty i obu przedziałów bezpośrednio z nim sąsiadujących były jednakowe.

Należy pamiętać, że w przypadku gdy najliczniejszy jest przedział pierwszy bądź ostatni to dominanty nie wyznaczamy.

Jeżeli liczebność przedziałów przed i za przedziałem dominanty są jednakowe, to dominanta jest równa środkowi klasy dominującej. W przypadku, gdy liczebność klasy poprzedzającej jest większa od liczebności klasy następującej, to dominanta będzie wartością bliższą dolnej granicy przedziału, w którym jest zawarta i odwrotnie.

Z szeregów rozdzielczych przedziałowych dominantę można również wyznaczyć graficznie. Graficzne wyznaczenie modalnej sprowadza się do wykreślenia histogramu liczebności i połączenia dwoma odcinkami wierzchołków najwyższego prostokąta, po przekątnej z najbliższymi wierzchołkami sąsiednich prostokątów. Rzut punktu przecięcia tych odcinków na oś odciętych wskazuje wartość dominanty.

2.2.2. Kwantyle

Kwantyle są to takie wartości cechy, które dzielą badaną zbiorowość na określone, części pod względem liczebności. Spośród kwantyli najczęściej używanymi miarami są **kwartyle**, wśród których wyróżniamy kwartyl pierwszy (dolny), kwartyl drugi (mediana) oraz kwartyl trzeci (górny). Każdy z kwartyli dzieli zbiorowość uporządkowaną niemalejąco pod względem liczebności na dwie części, przy czym:

- **kwartyl pierwszy** dzieli zbiorowość na dwie części w ten sposób, że 25% jednostek zbiorowości ma wartości zmiennej mniejsze lub równe kwartylowi pierwszemu, a 75% - równe lub większe od tego kwartyla,
- **mediana** dzieli zbiorowość na dwie części w ten sposób, że 50% jednostek ma wartości mniejsze lub równe medianie oraz 50% - równe lub większe od mediany,
- **kwartyl trzeci** dzieli zbiorowość na dwie części w ten sposób, że 75% jednostek ma wartości zmiennej mniejsze lub równe kwartylowi trzeciemu, a 25% - równe lub większe od kwartyla trzeciego.

Z szeregów wyliczających najczęściej wyznacza się **medianę**. W przypadku, gdy liczba obserwacji jest nieparzysta, medianą jest wartość środkowa. Jeśli natomiast liczba jednostek zbiorowości jest parzysta – mediana jest średnią arytmetyczną dwóch środkowych wartości zmiennej. Sposób wyznaczania mediany z szeregów wyliczających można zapisać następującymi wzorami:

$$Me = x_{\frac{N+1}{2}}, gdy\ N\ jest\ nieparzyste,$$

$$Me = \frac{1}{2}\left(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}\right), gdy\ N\ jest\ parzyste.$$

Wyznaczanie mediany z szeregu rozdzielczego punktowego sprowadza się do wskazania jednostki środkowej i odczytania wariantu cechy odpowiadającej tej jednostce. Określenie środkowej jednostki ułatwia kumulacja liczebności, polegająca na kolejnym, narastającym sumowaniu liczebności dotyczących poszczególnych wariantów badanej zmiennej. Z szeregów rozdzielczych przedziałowych kwartyle wyznaczamy wykorzystując następujące wzory interpolacyjne:

$$Q_1 = x_{Q_1} + \frac{\frac{N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_1}} i_{Q_1},$$

$$Q_2 = Me = x_{Me} + \frac{\frac{N}{2} - \sum_{i=1}^{k-1} n_i}{n_{Me}} i_{Me},$$

$$Q_3 = x_3 + \frac{\frac{3N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_3}} i_{Q_3},$$

gdzie:

$Q_1, Q_2(Me), Q_3$ – odpowiednio kwartyl pierwszy, kwartyl drugi (mediana) oraz kwartyl trzeci,

x_{Q_1}, x_{Me}, x_{Q_3} – dolne granice przedziałów, w których znajdują się odpowiednio kwartyl pierwszy, kwartyl drugi (mediana) oraz kwartyl trzeci,

$\sum_{i=1}^{k-1} n_i$ – suma liczebności od klasy pierwszej poprzedzającej tę, w której znajdują się odpowiednio kwartyl pierwszy, mediana i kwartyl trzeci,

n_{Q_1}, n_{Me}, n_{Q_3}- liczebności przedziałów, w których znajdują się odpowiednio kwartyl pierwszy, mediana i kwartyl trzeci,

i_{Q_1}, i_{Me}, i_{Q_3}- interwały przedziałów, w których znajdują się odpowiednio kwartyl pierwszy, mediana i kwartyl trzeci.

Pierwszą czynnością związaną z wyznaczaniem kwartyli jest **kumulacja liczebności** (absolutnych bądź odsetków). Następnie wyznaczamy pozycję poszczególnych kwartyli w szeregu, tzn. $\frac{N}{4}, \frac{N}{2}, \frac{3N}{4}$.

Z reguły poszczególne kwartyle znajdują się w różnych przedziałach.

Kwartyle są dogodnymi parametrami w analizie struktury. Mogą być bowiem wykorzystywane w tych przypadkach, w których niemożliwe jest obliczenie z danego szeregu średniej arytmetycznej (otwarte przedziały klasowe, ekstremalne wartości), a także dominanty (nierówne rozpiętości przedziałów, silna asymetria rozkładów).

Decyle i centyle (percentyle) wyznacza się podobnie jak kwartyle. **Decyle** dzielą zbiorowość uporządkowaną na 10 części pod względem liczebności. Decyl trzeci np. to taka wartość cechy, że 30% wszystkich jednostek zbiorowości ma wartości równe od niej lub niższe, a 70% równe lub wyższe. Decyli jest 9, a piąty decyl jest medianą.

Centyle (percentyle) dzielą zbiorowość uporządkowaną na 100 części pod względem liczebności. Centyli jest 99, a pięćdziesiąty centyl jest równy medianie. Na przykład trzydziesty dziewiąty centyl jest taką wartością, że 39% wszystkich jednostek badanej zbiorowości ma wartości od niej niższe lub równe, a 61% jednostek – wartości równe lub wyższe.

2.2.3. Zależności miar tendencji centralnej

Średnia arytmetyczna, dominanta i mediana, jako miary tendencji centralnej, są powiązane ze sobą odpowiednimi zależnościami, które można wyrazić równościami lub nierównościami (decyduje tu typ rozkładu empirycznego). Na przykład, w przypadku rozkładu umiarkowanie asymetrycznego zachodzi między nimi następujący związek:

$$\bar{x} - D = 3(\bar{x} - Me).$$

Wzór ten nosi nazwę **wzoru Pearsona**. Ze wzoru – przy odpowiednich przekształceniach – można np. wyznaczyć nieznaną średnią, gdy znana jest mediana i dominanta.

Zadanie 2.1.

Wartości temperatur (w stopniach C) zaobserwowanych w dniu 7 listopada 2016 r. o godzinie 12.00 w miastach wojewódzkich były następujące: $8, 6, 7, 6, 6, 8, 7, 7, 7, 6, 6, 7, 6, 5, 6, 7$ (źródło: *www.twojapogoda.pl*)

Określić zbiorowość, jednostkę i cechę statystyczną.

Obliczyć średnią arytmetyczną oraz podać miary pozycyjne (dominantę, medianę).

Rozwiązanie 2.1:

- Zbiorowość statystyczna: miasta wojewódzkie
- Jednostka statystyczna: miasto wojewódzkie
- Cecha statystyczna (zmienna): temperatura (w stopniach C) zaobserwowana w dniu 7 listopada 2016 r. o godzinie 12.00
- Szereg szczegółowy (uporządkowany): 5, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8
- Średnia arytmetyczna:

$$\bar{x} = \frac{\sum_{j=1}^{N} x_j}{N} = \frac{105}{16} = 6{,}5625 \approx 6{,}6$$

Średnia temperatura w dniu 7 listopada 2016 r. o godzinie 12.00 w miastach wojewódzkich wyniosła 6,6° C.

- Mediana:

Ogólny wzór na medianę dla szeregu szczegółowego:

$$Me = x_{\frac{N+1}{2}}, gdy\ N\ jest\ nieparzyste,$$

$$Me = \frac{1}{2}\left(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}\right), gdy\ N\ jest\ parzyste$$

W zadaniu $n = 16$ zatem

$Me = \frac{1}{2}(x_8 + x_9) = \frac{1}{2}(6 + 7) = 6{,}5°\text{C}$

W połowie miast wojewódzkich temperatura w dniu 7 listopada 2016 r. o godzinie 12.00 była nie wyższa niż 6,5°C, a w połowie nie niższa niż 6,5°C.

- Dominanta:

Jest to wartość najczęściej występująca w szeregu czasowym – w tym przypadku jest to 6°C.

Wśród miast wojewódzkich dominowały te w których temperatura w dniu 7 listopada 2016 r. o godzinie 12.00 wynosiła 6°C.

Zadanie 2.2.

Przeprowadzono test zadania egzaminacyjnego, aby określić przeciętny czas potrzebny na jego rozwiązanie. Czas rozwiązywania tego zadania (w minutach) w grupie testowej został przedstawiony w tablicy 6.

Tablica 1. Dane do zad. 2.2.

Czas rozwiązywania zadania	*4*	*5*	*6*	*7*	*8*	*9*
Liczba osób	*6*	*4*	*3*	*2*	*2*	*1*

Źródło: dane umowne

Określić zbiorowość, jednostkę i cechę statystyczną.

Obliczyć średnią arytmetyczną oraz podać miary pozycyjne (dominantę, medianę).

Rozwiązanie 2.2:

- Zbiorowość statystyczna: osoby z grupy testującej czas rozwiązywania zadania egzaminacyjnego
- Jednostka statystyczna: osoba testująca zadanie egzaminacyjne
- Cecha statystyczna (zmienna): czas rozwiązywania zadania egzaminacyjnego (w minutach)

Tablica 2. Obliczenia pomocnicze do zad. 2.2.

x_i	n_i	$x_i n_i$	n_{isk}
4	*6*	*24*	*6*
5	*4*	*20*	*10*
6	*3*	*18*	*13*
7	*2*	*14*	*15*
8	*2*	*16*	*17*
9	*1*	*9*	*18*
Σ	*18*	*101*	-

- Średnia arytmetyczna: $\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{N} = \frac{101}{18} \approx 5{,}61$

Średni czas rozwiązywania zadań w grupie testowej wyniósł 5,61 minuty.

- Mediana:

W zadaniu $n = 18$ zatem $Me = \frac{1}{2}(x_9 + x_{10}) = \frac{1}{2}(5 + 5) = 5$

Połowa osób testujących zadanie rozwiązała go w czasie nie dłuższym niż 5 minut, a połowa w czasie nie krótszym niż 5 minut.

- Dominanta:

Jest to wartość najczęściej występująca w szeregu czasowym – w tym przypadku jest to 4.

Wśród osób testujących zadanie dominowały osoby, które rozwiązały je w czasie czterech minut.

Zadanie 2.3.

Przebadano rodziny pewnej wspólnoty mieszkaniowej badając średni dochód na osobę w tych rodzinach w tysiącach złotych. Wyniki zestawiono poniżej:

Tablica 3. Dane do zadania 2.3.

Dochód na osobę w rodzinie w tys. zł	*Liczba rodzin*
do 0,4	*25*
0,4-0,8	*50*
0,8-1,2	*40*
1,2-1,6	*35*
1,6-2	*30*

Źródło: dane umowne

Określić zbiorowość, jednostkę i cechę statystyczną. Obliczyć średnią arytmetyczną oraz podać miary pozycyjne (dominantę, medianę, kwartyle i decyle).

Rozwiązanie 2.3:

- Zbiorowość statystyczna: rodziny pewnej wspólnoty mieszkaniowej
- Jednostka statystyczna: rodzina pewnej wspólnoty mieszkaniowej
- Cecha statystyczna (zmienna): dochód na osobę w tyś. zł w rodzinie badanej wspólnoty mieszkaniowej

Tablica 4. Obliczenia pomocnicze do zad. 2.3.

i	x_{i0}	x_{i1}	$\dot{x}_i = \frac{x_{0i}+x_{1i}}{2}$	n_i	$\dot{x}_i n_i$	n_{isk}
1	*0,0*	*0,4*	*0,2*	*25*	*5*	*25*
2	*0,4*	*0,8*	*0,6*	*50*	*30*	*75*
3	*0,8*	*1,2*	*1,0*	*40*	*40*	*115*
4	*1,2*	*1,6*	*1,4*	*35*	*49*	*150*
5	*1,6*	*2,0*	*1,8*	*30*	*54*	*180*
Σ	-	-	-	*180*	*178*	-

- Średnia arytmetyczna:

$$\bar{x} = \frac{\sum_{i=1}^{k} \dot{x}_i n_i}{N} = \frac{178}{180} \approx 0{,}99\ tys.zł.$$

Średni dochód na osobę w tyś. zł w rodzinach pewnej wspólnoty mieszkaniowej wyniósł 0,99 tys. zł (990 zł).

- Mediana:

$$Q_2 = Me = x_{Me} + \frac{\frac{N}{2} - \sum_{i=1}^{k-1} n_i}{n_{Me}} i_{Me},$$

W pierwszej kolejności wyznaczamy przedział w którym znajduje się mediana.

W kolumnie z liczebnościami skumulowanymi (n_{isk}) szukamy pierwszej wartości większej bądź równej $\frac{N}{2} = \frac{180}{2} = 90$ - jest to przedział trzeci.

$$Me = 0{,}8 + \frac{90 - 75}{40} 0{,}4 = 0{,}95 \; tys. zł,$$

Połowa rodzin pewnej wspólnoty mieszkaniowej ma dochód na osobę nie większy niż 950 zł, a połowa nie mniejszy niż 950 zł.

- Kwartyl I:

$$Q_1 = x_{Q_1} + \frac{\frac{N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_1}} i_{Q_1}$$

W pierwszej kolejności wyznaczamy przedział w którym znajduje się kwartyl I.

W kolumnie z liczebnościami skumulowanymi (n_{isk}) szukamy pierwszej wartości większej bądź równej $\frac{N}{4} = \frac{180}{4} = 45$ - jest to przedział drugi.

$$Q_1 = 0{,}4 + \frac{45 - 25}{50} 0{,}4 = 0{,}56 \; tys. zł,$$

25 % rodzin pewnej wspólnoty mieszkaniowej ma dochód na osobę nie większy niż 560 zł, a 75 % rodzin ma dochód na osobę nie mniejszy niż 560 zł.

- Kwartyl III:

$$Q_3 = x_3 + \frac{\frac{3N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_3}} i_{Q_3},$$

W pierwszej kolejności wyznaczamy przedział w którym znajduje się kwartyl III.

W kolumnie z liczebnościami skumulowanymi (n_{isk}) szukamy pierwszej wartości większej bądź równej $\frac{3N}{4} = \frac{3*180}{4} = 135$ - jest to przedział czwarty.

$$Q_3 = 1{,}2 + \frac{135 - 115}{35} 0{,}4 \approx 1{,}43 \; tys. zł,$$

75 % rodzin pewnej wspólnoty mieszkaniowej ma dochód na osobę nie większy niż 1430 zł, a 25 % rodzin ma dochód na osobę nie mniejszy niż 1430 zł.

- Dominanta:

Jest to wartość najczęściej występująca w szeregu czasowym, a ponieważ w zadaniu dane mamy zebrane w postaci szeregu z przedziałami klasowymi to możemy wskazać przedział w którym dominanta się znajduje. Natomiast jej wartość musimy interpolować.

$$D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} i_D,$$

Dominanta występuje w najliczniejszym przedziale tj. w przedziale drugim.

$$D = 0{,}4 + \frac{50 - 25}{(50 - 25) + (50 - 40)} 0{,}4 \approx 0{,}69,$$

Wśród rodzin pewnej wspólnoty mieszkaniowej dominują rodziny z dochodem na osobę wynoszącym 690 zł.

- Przykład obliczania decyla rzędu drugiego:

$$Dr = x_r + \frac{\frac{rN}{10} - \sum_{i=1}^{k-1} n_i}{n_{D_r}} i_{D_r},$$

W pierwszej kolejności wyznaczamy przedział w którym znajduje się D_2.

W kolumnie z liczebnościami skumulowanymi (n_{isk}) szukamy pierwszej wartości większej bądź równej $\frac{rN}{10} = \frac{2*180}{10} = 36$ - jest to przedział drugi.

$$D_2 = 0{,}4 + \frac{36 - 25}{50} 0{,}4 \approx 0{,}49\ tys.zł,$$

20% rodzin pewnej wspólnoty mieszkaniowej ma dochód na osobę nie większy niż 490 zł, a 80% rodzin ma dochód na osobę nie mniejszy niż 490 zł.

Tablica 5. Wartości decyli w zad. 2.3.

Decyle	*Wartości decyli w zł*
D1	*0,29*
D2	*0,49*
D3	*0,63*
D4	*0,78*
D5	*0,95*
D6	*1,13*
D7	*1,33*
D8	*1,53*
D9	*1,76*

Główny
Urząd Statystyczny