

Marzanna Gawryluk



**Wykorzystanie statystyki opisowej
na lekcji biologii w liceum**



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej

Marzanna Gawryluk



Białystok 2021



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej

Marzanna Gawryluk

Redakcja merytoryczna

Lider projektu: Powszechne Towarzystwo Oświatowe „Edukacja Narodowa” im. Ks. Grzegorza Piramowicza w Białymstoku

Partner projektu: Narodowe Forum Doradztwa Kariery

Partnerzy merytoryczni projektu:
Uniwersytet w Białymstoku,
Podlaski Kurator Oświaty,
Miejski Ośrodki Doskonalenia Nauczycieli w Białymstoku,
Poradnia Psychologiczna nr 2 w Białymstoku, II Społeczne Liceum Ogólnokształcące im. Jana Pawła II w Białymstoku

Publikacja jest rozpowszechniana na zasadach wolnej licencji Creative Commons – Użycie niekomercyjne 4.0 Polska (CC-BY-NC)



Fundusze Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita Polska

Unia Europejska
Europejski Fundusz Społeczny



Publikacja przygotowana w ramach projektu:



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Spis treści

Wprowadzenie	5
1. Podstawowe pojęcia statystyczne	7
1.1 Zbiorowość statystyczna	8
1.2 Jednostka statystyczna	8
1.3 Cechy statystyczne i ich warianty.....	8
2. Metody i organizacja badań statystycznych	10
2.1 Metody statystyczne	11
2.2 Ocena statystyczna.....	11
2.3 Organizacja badania statystycznego	12
2.4 Grupowanie materiału	14
3. Metody prezentacji danych statystycznych.....	15
3.1. Tabelaryczna prezentacja danych statystycznych	16
3.1.1. Szeregi statystyczne	16
3.1.2. Szereg rozdzielczy	16
3.1.3. Szereg rozdzielczy z przedziałami klasowymi	19
3.2. Graficzna interpretacja danych statystycznych	23
4. Podstawowe wiadomości z analizy statystycznej	25
4.1. Cel analizy statystycznej.....	26
4.2. Analiza natężenia	26
4.3. Analiza struktury	26
4.4. Analiza tendencji centralnej.....	28
4.4.1. Średnie klasyczne i ich własności	29
4.4.2. Średnie pozycyjne.....	33
4.4.3. Miary zmienności	38
4.4.4. Współczynniki zmienności.....	42
4.4.5. Miary zmienności	43
Bibliografia	45
Nasze publikacje	47

Wprowadzenie



Wprowadzenie



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Wprowadzenie

Jednym z podstawowych celów nauki jest przewidywanie i wyjaśnianie wyników obserwacji zdarzeń oraz relacji przyczynowych zachodzących między nimi. Pomocna w tych zagadnieniach jest statystyka. Znajduje ona zastosowanie w wielu gałęziach nauki i życia związanych nie tylko z matematyką. Zasady rządzące statystyką odkryto już 3 – 4 tysiące lat p.n.e. kiedy to Sumerowie na glinianych tabliczkach zapisywali wyniki z przeprowadzonych spisów ludności. W Egipcie przeprowadzano spisy co najmniej 3 tysiące lat p.n.e. wraz ze spisem majątków. Służyły one nie tylko do poznania majątku Egipcjan ale również pomagały w przewidywaniu nastrojów panujących w społeczeństwie w zakresie funkcjonowania władzy. Uzyskane w ten sposób dane pozwalały na lepsze zarządzanie państwem oraz podejmowanie lepszych decyzji. Kilkakrotnie o spisach ludności wspomina Biblia: W księdze liczb (Numerii) mówi się o kilku spisach ludności. Pierwszy z nich przeprowadził Mojżesz z Aaronem na rozkaz Pana, spisując „wszystkich mężczyzn od dwudziestu lat zdolnych do walki”. Również w Chinach - w latach 1368 – 1644 władcy z dynastii Ming wprowadzili system spisów ludności co 10 lat. Na przełomie XV i XVI wieku w ustawodawstwie synodalnym kilku krajów europejskich zalecano prowadzenie przez proboszczów ksiąg metrykalnych chrztów, ślubów i zmarłych a w 1563 r. Sobór trydencki powyższe zalecenia zmienił na obowiązek. Dotyczyło to chrztów i ślubów. Potem w 1614 r. tak zwany „Rytuał rzymski” dołączył do tego obowiązek rejestracji pochówków (księga zgonów). Dało to początek nowoczesnej rejestracji ruchu naturalnego ludności dla celów analizy demograficznej. W Polsce pierwszy powszechny spis ludności zarządził Sejm Czteroletni w 1790r. Do tej pory statystyka ma podobne znaczenie i używa się jej do prowadzenia rozmaitych spisów np. spisu ludności, czy badania nastrojów społecznych w związku z określonym zjawiskiem. Statystyka zajmuje szerokie zastosowanie w demografii, czyli nauce o społeczności ludzkiej, jej powstawaniu, rozwoju i strukturze.

Statystyką posługują nie tylko demografowie czy matematycy. Znalazła ona szerokie zastosowanie w ekonomii, a szczególnie w jej dziale – ekonometrii, która bada związki zachodzące między zjawiskami a zmiennymi ekonomicznymi. Celem ekonometrii jest analiza teorii ekonomicznych, przewidywanie zachodzących procesów ekonomicznych.

Statystyką zainteresowali się również biolodzy, którzy wykorzystują statystykę do badania zmienności organizmów i próby zrozumienia zjawisk w otaczającym nas świecie. Wyniki badań są opracowywane przy użyciu metod statystycznych, a następnie wykorzystuje się je w innych naukach, np. fizjologii, medycynie, czy genetyce.

Celem artykułu jest pokazanie zastosowania statystyki opisowej w biologii, jednocześnie przy wykorzystaniu programu Exel

Rozdział pierwszy



Podstawowe pojęcia statystyczne



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Rozdział pierwszy: Podstawowe pojęcia statystyczne

1.1 Zbiorowość statystyczna

Aby móc wykonać **badanie statystyczne** pierwszym krokiem, jaki musimy uczynić, jest zdefiniowanie populacji, na temat której chcemy formułować sądy. **Populacją generalną (zbiorowością statystyczną lub masą statystyczną)** nazywamy zbiór dowolnych elementów (osób, przedmiotów, faktów), które są do siebie podobne, ale nie identyczne pod względem określonych cech poddanych badaniom statystycznym. Musimy jednak pamiętać o tym aby populacja generalna była zawsze określona w sposób jednoznaczny, nie budzący wątpliwości. Ponieważ populacje, które podlegają badaniom często są bardzo liczne wybieramy tylko jej część i nazywamy ją **próbą**. Wówczas na podstawie znajomości cech odpowiednio wybranej próby pewnej zbiorowości będziemy wysnuwali wnioski dotyczące rozważanych cech dla pozostałych, niezbadanych elementów tej populacji generalnej.

1.2 Jednostka statystyczna

W skład zbiorowości statystycznej wchodzi jednostki zwane **jednostkami statystycznymi (jednostkami badania lub jednostkami obserwacji)** – są to najmniejsze obiekty dostarczające informacji statystycznej. Elementy te nie są identyczne, jednak mają wiele cech wspólnych, które pozwalają zakwalifikować je do jednej zbiorowości statystycznej. O wyborze jednostki statystycznej w danej zbiorowości decyduje cel badania oraz konieczność zapewnienia dużej dokładności badania statystycznego.

1.3 Cechy statystyczne i ich warianty

Elementy badanej populacji generalnej mogą mieć różne właściwości, które podlegają obserwacji. Właściwości te nazywamy **cechami statystycznymi**. Do konkretnego badania wybieramy tylko te najważniejsze dla analizowanego problemu. Po wytypowaniu cech, które nas interesują musimy podjąć decyzję jak będziemy mierzyć wartości tych cech w trakcie obserwacji.

Stevens w 1951 roku zaproponował następujące cztery skale pomiarowe¹:

- a) **skalę nominalną** - która pozwala nam tylko na pogrupowanie badanych obiektów pod względem cech, których nie można zmierzyć i wyrazić za pomocą liczb (czyli mają charakter jakościowy) np. płeć: samica, samiec; zamieszkanie: miasto, mała miejscowość lub wieś, czy kolor oczu
- b) **skalę porządkową** - pozwala ona uporządkować obiekty wg wartości badanej cechy np. skala natężenia choroby. Podobnie jak dla skali nominalnej, nie są możliwe działania arytmetyczne na danych opisanych na skali porządkowej

¹ Brzeziński, Jerzy, (2004) Metodologia badań psychologicznych. Wydawnictwo Naukowe PWN, Warszawa, str. 188



- c) **skalę interwałową** - pozwalającą na stwierdzenie o ile jednostek natężenie (wielkość) badanej cechy dla obiektu A jest większe (mniejsze) od natężenia (wielkości) tejże cechy dla obiektu B np. temperatura ciała A równa 40°C jest większa o 10°C od temperatury ciała B równej 30°C . Działania arytmetyczne – dodawanie i odejmowanie są możliwe.
- d) **skalę ilorazową** - pozwalającą na stwierdzenie krotności obserwowanej różnicy natężenia (wielkości) badanej cechy, np. wiek, waga oraz parametry biochemiczne itd.

Zestawienie własności skal pokazuje poniższa tabela 1

Tabela 1. Charakterystyka kompetencji kluczowych i odpowiadających im najważniejszych umiejętności oraz zawierających się w nich wymaganiach szczegółowych z biologii

	Czy obiekt X jest różny od obiektu Y	Czy obiekt X jest lepszy od obiektu Y	O ile obiekt X jest lepszy od obiektu Y	Ile razy obiekt X jest lepszy od obiektu Y
Skala nominalna	TAK	NIE	NIE	NIE
Skala porządkowa	TAK	TAK	NIE	NIE
Skala przedziałowa	TAK	TAK	TAK	NIE

W konkretnym badaniu badacz musi określić, jakiego rodzaju skali będzie używał dla badanej cechy. Konsekwencją wyboru określonej skali jest dostosowanie odpowiednich narzędzi statystycznych używanych do analizy zgromadzonych danych.

W statystyce opisowej mamy również podział cech na 2 grupy: cechy mierzalne i

niemierzalne. Mogą występować u poszczególnych jednostek statystycznych w różnej formie (cechy niemierzalne), lub mogą przyjmować różne wartości (cechy mierzalne), które nazywamy wariantami cechy statystycznej. Cechy mierzalne dzielimy na:

- a) **zmienne dyskretne (skokowe)**, których zbiór wartości jest zbiorem skończonym, albo nieskończonym przeliczalnym, którego elementy można ustawić w ciągu np.: liczba osób w gospodarstwie domowym;
- b) **zmienne ciągłe**, które z definicji przyjmują wartości będące przedziałem liczbowym.

Rozdział drugi



Metody i organizacja badań statystycznych



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Rozdział drugi: Metody i organizacja badań statystycznych

2.1 Metody statystyczne

Wybierając metodę badania należy uwzględnić wiele czynników. Dzielimy je na trzy grupy. Do pierwszej z nich należą **czynniki natury statystycznej**, czyli rodzaj zbiorowości statystycznej, temat badania, cel badania i stopień szczegółowości badania. Do **czynników pozastatystycznych** zliczamy przede wszystkim środki finansowe jakimi dysponujemy, nasze możliwości techniczne i materiałowe oraz ograniczona liczba osób, które mogą być zaangażowane do przeprowadzenia badania. Ważny jest też **czynnik czasowy** czyli długość okresu jakim dysponujemy na przeprowadzenie badania.

Innym kryterium podziału badań statystycznych jest **podział ze względu na liczbę jednostek zbiorowości statystycznej**. Wyróżniamy tutaj **badanie pełne**, które obejmuje wszystkie elementy zbiorowości, a wyniki obserwacji stanowią charakterystykę zbiorowości. Drugim rodzajem jest **badanie częściowe** – obejmuje tylko pewną część elementów zbiorowości. Obserwacja częściowa powinna być realizowana w taki sposób, aby na podstawie jej wyników można było wywnioskować o całej populacji.

W wielu sytuacjach przeprowadzenie badania pełnego jest zbyt drogie, a przebadanie wszystkich jednostek zbiorowości – niemożliwe. W takich przypadkach wykonujemy badanie częściowe – polegające na tym, że w trakcie badania poddajemy obserwacji tylko pewną część zbiorowości statystycznej. Musimy wówczas określić **próbę** - czyli podzbiór elementów zbiorowości generalnej, która podlega badaniu ze względu na ustaloną cechę. **Liczebnością zbiorowości** nazywamy liczbę jednostek w zbiorowości. W biologii ze względu na charakterystykę jednostek statystycznych wykorzystujemy najczęściej badania częściowe. Do nich zaliczamy:

- a) **metodę monograficzną** – polegającą na wszechstronnym opisie i szczegółowej analizie pojedynczej jednostki lub niewielkiej liczby charakterystycznych jednostek badanej zbiorowości statystycznej. Na podstawie otrzymanych wyników uogólniamy wiadomości na całą populację.
- b) **metodę reprezentacyjną** – która polega na tym, że w celu zbadania własności całej zbiorowości statystycznej wybiera się do badania tylko pewną liczbę jednostek statystycznych reprezentujących badaną zbiorowość zwaną próbą

2.2 Ocena statystyczna

W trakcie badania niektórych zjawisk może się okazać, że materiał jest niekompletny i nie można zbadać interesującej nas zbiorowości statystycznej stosując metodę badania pełnego, czy badania częściowego. W praktyce oznacza to, że materiał, jakim dysponujemy, dotyczący badanej zbiorowości, nie wystarcza do pełnej realizacji założonego celu badania. W takich



sytuacjach ustalamy interesujące nas dane liczbowe, dotyczące rozważanej zbiorowości, korzystając z postępowania zwanego szacunkiem statystycznym lub inaczej: oceną statystyczną. **Szacunkiem statystycznym (oceną statystyczną)** nazywamy takie postępowanie, w którym:

- a) na podstawie realizacji znanych nam cech zbiorowości ustalamy wartości innych, nieznanymi cech tej samej zbiorowości;
- b) na podstawie cech znanej zbiorowości ustalamy te same lub pokrewne cechy rości nieznannej.

Przeprowadzenie szacunku jest możliwe wówczas, gdy między znanymi a nieznanymi cechami występują określone relacje ilościowe opisane wzorem matematycznym oraz gdy dobrze znamy pod względem merytorycznym problematykę, której szacunek dotyczy, tzn. prowadząc szacunek bazujemy na poprawnych założeniach dotyczących rozważanej problematyki. Szacunki statystyczne szczególnie często stosowane są w statystyce historycznej, a zwłaszcza w demografii historycznej. W tej dziedzinie nie ma możliwości przeprowadzenia żadnego badania bezpośredniego. Znane są szacunki liczby ludności Polski za czasów Kazimierza Wielkiego na podstawie wykazów świętopietrza.

Prezentowanie danych, to przedstawienie licznych zbiorów danych w postaci ułatwiającej ich ocenę i analizę. Można tu zastosować prezentację w formie tabeli, graficznie lub opisu słownego. Wykorzystywane do tego celu środki określa się mianem **metod opisu statystycznego**. W wielu przypadkach zebranie wszystkich danych nie jest możliwe i należy wypowiadać się o badanym zjawisku na podstawie zebranych danych częściowych.

2.3 Organizacja badania statystycznego

O powodzeniu całego badania w dużej mierze decyduje odpowiednie przygotowanie. Na tym etapie formułujemy koncepcje badania statystycznego. Podstawową jest dokładne **określenie ogólnych oraz szczegółowych celów badania i hipotez roboczych**. Ustalenie celu badania umożliwia sprecyzowanie wymagań dotyczących niezbędnych informacji (materiału statystycznego), co pozwala uniknąć chaosu informacyjnego spowodowanego pojawieniem się zbędnych lub przypadkowych informacji. Wyniki badania statystycznego uzależnione są również od właściwego **określenia zbiorowości statystycznej pod względem czasowym**. Wyróżniamy tu dwie możliwości: badanie stanu zbiorowości w określonym momencie lub badanie stanu zbiorowości w pewnym okresie. Po ustaleniu celu badania, określeniu zbiorowości statystycznej, jednostki i wybraniu metody badania przystępujemy do gromadzenia niezbędnych danych statystycznych, czyli informacji indywidualnych o każdej jednostce zbiorowości. Może to być jedna z metod badania pełnego albo badania częściowego. W przypadku **badania pełnego** są to dane o każdej jednostce zbiorowości. W przypadku **badania częściowego** - o każdej jednostce próby. Zbiór tych informacji nazywamy **materiałem statystycznym**. *Materiał statystyczny dzielimy na pierwotny i wtórny. Pierwotny*, gdy informacje statystyczne wykorzystywane w badaniu



uzyskujemy w trakcie specjalnie przeprowadzonego odrębnego badania, a źródła, z których one pochodzą nazywamy **źródłami pierwotnymi**. **Materiał wtórny** – to materiał zaczerpnięty z poza statystycznych źródeł, a następnie wykorzystany w badaniach statystycznych, zaś źródło, z którego on pochodzi – **źródłem wtórnym**. Korzystanie z materiału statystycznego wtórnego ma swoje zalety i wady.

Do zalet należy przede wszystkim obniżenie kosztów związanych ze zbieraniem materiału statystycznego do badania. Wady to wątpliwa wartość poznawcza i wiarygodność materiału wtórnego, który może być „zanieczyszczony” informacjami zbędnymi bądź zawierać informacje mniej szczegółowe.

Kolejnym krokiem jest ustalenie formularza i instrukcji służących do gromadzenia materiału statystycznego. **Formularzem statystycznym** – nazywamy zbiór spisanych w odpowiedniej formie na papierze lub nośniku magnetyczny pytań dotyczących badanych cech rozważanej zbiorowości statystycznej. Wypełnienie formularza statystycznego wymaga podania wyjaśnień dotyczących sposobu udzielania poprawnych odpowiedzi, które są niezbędne, aby otrzymać jednolite odpowiedzi i uniknąć błędnej interpretacji pytań przez osoby wypełniające formularz. W tym celu do formularza dołącza się **instrukcję statystyczną**, która powinna być zwięzła i czytelna. W zebranych materiale statystycznym, zwanym materiałem surowym, często występują różnego rodzaju braki i błędy. Dlatego przed wykorzystaniem tego materiału w badaniu statystycznym poddajemy go kontroli. Błędy w materiale statystycznym mogą być **przypadkowe** – do których zaliczamy przede wszystkim omyłki liczbowe, błędy zwane czeskimi wynikające głównie z nieuwagi, braku umiejętności podawania prawidłowych odpowiedzi, oraz **systematyczne** – popełniane rozmyślnie przez wypełniającego formularz, bądź popełniane nieumyślnie – polegają na podawaniu faktów celowo zniekształconych. Jedną z podstawowych spraw związanych z realizacją badania statystycznego jest dokładność ostatecznych wyników badania.

Całe badanie prowadzimy w ten sposób, aby w pełni zrealizować cel badania jednoczesnym zagwarantowaniem najwyższej dokładności wyników, jaką można uzyskać w danych warunkach. Uzyskanie absolutnej dokładności wyników jest wyjątkowo trudne, a w wielu wypadkach niemożliwe.

Każde dobrze zaprojektowane badanie statystyczne składa się z etapów:

- I etap - zdefiniowanie zbiorowości, jednostki i cech statystycznych, które będą podlegały badaniu;
- II etap - określenie odpowiedniej metody badania statystycznego, która decyduje o jakości badania;
- III etap - podanie źródła pozyskania danych;
- IV etap - opracowanie formularza statystycznego i makiety tablic wynikowych;
- V etap - zapewnienie odpowiedniej kontroli materiału statystycznego
- VI etap – prezentacja danych statystycznych



2.4 Grupowanie materiału

W badaniach statystycznych nie ograniczamy się do konkretnych jednostek, lecz na podstawie danych dotyczących badanych jednostek formułujemy uogólnienia dotyczące całej zbiorowości. Samo zebranie materiału statystycznego z różnych źródeł nie wystarcza do przeprowadzenia badania. Należy jeszcze ten materiał opracować.

Pierwszym krokiem w opracowywaniu materiału statystycznego jest grupowanie statystyczne. Przechodzimy w nim od materiałów informujących o poszczególnych jednostkach statystycznych do materiałów dających obraz całej zbiorowości oraz jej poszczególnych części. Grupowanie statystyczne polega na podziale badanej zbiorowości statystycznej na jednorodne lub względnie jednorodne części (grupy, klasy) według cech, które charakteryzują najbardziej istotne jej strony z punktu widzenia celu badania. Dzięki prawidłowo i wszechstronnie przeprowadzonemu grupowaniu statystycznemu osiągamy liczne korzyści, wśród których najważniejsze to: możliwość prawidłowego obliczenia uogólnionych charakterystyk zbiorowości statystycznej; odzwierciedlenie struktury badanej zbiorowości i możliwość ustalenia związków między zjawiskami.

Rozdział trzeci



Metody prezentacji danych statystycznych



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Rozdział trzeci: Metody prezentacji danych statystycznych

Zgromadzony materiał statystyczny, zawierający informacje dotyczące poszczególnych jednostek badanej zbiorowości służy do opisu i poznania całej zbiorowości, tzn. identyfikacji i ewentualnej kwantyfikacji jej ogólnych właściwości. Dlatego, zanim właściwości te ujawnimy i poznamy, zgromadzony materiał statystyczny musimy opracować, a następnie zaprezentować w formie, która zapewni pełną realizację celów badania statystycznego. Opracowanie materiału statystycznego sprowadza się do uporządkowania i pogrupowania materiału tak, aby możliwe było zaprezentowanie wyników badania w odpowiedniej formie, czyli **prezentacji danych statystycznych**. Podstawowe metody prezentacji danych statystycznych i odpowiadające im formy, to:

- przedstawienie danych w tablicach lub tabelach statystycznych (forma tabelaryczna),
- przedstawienie graficzne danych w postaci wykresów statystycznych (forma graficzna),
- włączenie danych do tekstu (forma opisowa).

Generalną zasadą jest taki wybór metody prezentacji, aby zapewnić osiągnięcie wszystkich celów badania oraz przejrzystość, czytelność, a dzięki temu możliwość popularyzacji i analizy prezentowanego materiału statystycznego, czyli danych wynikowych badania statystycznego.

3.1. Tabelaryczna prezentacja danych statystycznych

3.1.1. Szeregi statystyczne

Szeregiem statystycznym nazywamy ciąg wielkości statystycznych uporządkowanych wg określonych kryteriów. Przedstawienie szeregu statystycznego zależy od rodzaju badania, rodzaju cechy statystycznej, czy sposobu pomiaru, oraz liczby obserwacji.

Oznaczmy przez x_1, x_2, \dots, x_n będzie n – elementową próbką, co oznaczamy x_i

dla $i = 1, 2, \dots, n$, gdzie i – oznacza numer obserwacji.

Szeregiem szczegółowym – nazywamy uporządkowany ciąg wartości badanej cechy np.: gdy badamy niewielką liczbę jednostek. Jednostki możemy uporządkować następująco:

- rosnąco** $x_1 \leq x_2 \leq \dots \leq x_n$;
- malejąco** $x_1 \geq x_2 \geq \dots \geq x_n$.

Najczęściej porządkujemy szeregi szczegółowe rosnąco.

3.1.2. Szereg rozdzielczy

Dla wygody obliczeń i przejrzystości danych tworzymy szereg rozdzielczy. **Szereg rozdzielczy**, to zbiorowość statystyczna podzielona na części zwane **klasami** wg określonej cechy jakościowej lub ilościowej z przyporządkowaniem liczebności każdej z klas.



Zestawienie wyników w postaci szeregu rozdzielczego z cechą mierzalną nazywamy **rozkładem empirycznym**. Wyróżniamy dwa rodzaje szeregów rozdzielczych:

- a) Szereg rozdzielczy punktowy
- b) Szereg rozdzielczy z przedziałami klasowymi.

Najprostszym z nich jest **szereg rozdzielczy punktowy**, który buduje się dla cechy skokowej w taki sposób, że wartości cechy (która jest uporządkowana rosnąco) przypisujemy liczebność tej cechy, co przedstawia tabela, w której przez k – określona została liczba klas, przez x_i – wartość cechy, zaś przez n_i – liczba elementów mających wartość x_i :

Tabela 2. Konstrukcja szeregu rozdzielczego punktowego

Numer klasy	Wartość cechy	Liczba elementów
	x_i	n_i
1	x_1	n_1
\vdots	\vdots	\vdots
k	x_k	n_k
Suma		N

Przyjrzyjmy się zatem poniższemu przykładowi:

Przykład 1²: Zbadano masę ciała poczwerek mącznika (*tenebrio molitor*) w miligramach, z hodowli prowadzonej na razowej mące pszennej w temperaturze 26°C. Otrzymano następujące wyniki: 148, 148, 136, 152, 142, 130, 176, 152, 150, 140, 123, 113, 133, 117, 126, 129, 219, 156, 160, 123.

Zbudujmy szereg rozdzielczy punktowy. W tym celu porządkujemy rosnąco naszą masę ciała poczwerek mącznika i zliczamy ilość powtórzeń. Nasze obliczenia dla ułatwienia będziemy wykonywali w programie Excel.

² Łomnicki, Jerzy „Wprowadzenie do statystyki dla przyrodników” PWN Warszawa 2003 s.11



Tabela 3. Konstrukcja szeregu rozdzielczego punktowego do przykładu 1

Numer klasy	Wartość cechy	Liczba elementów
1	113	1
2	117	1
3	123	2
4	126	1
5	129	1
6	130	1
7	133	1
8	136	1
9	140	1
10	142	1
11	148	2
12	150	1
13	152	2
14	156	1
15	160	1
16	176	1
17	219	1
Suma		20

Należy zawsze pamiętać aby zsumować liczebności naszych klas. W tym celu klikamy w ostatni prostokąt w wykorzystujemy AUTOSUMOWANIE wybierając znak sumy

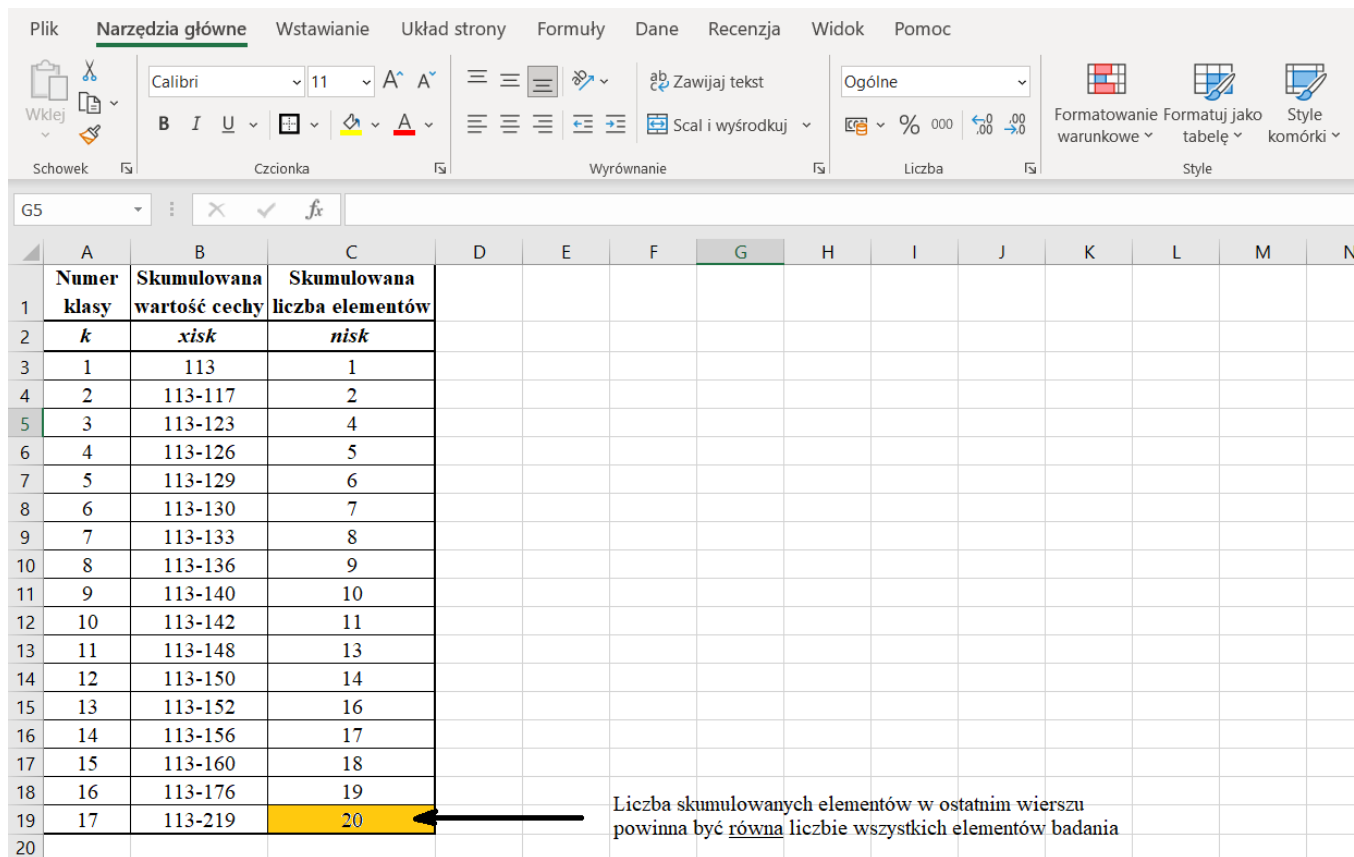
Szereg rozdzielczy skumulowany uzyskuje się poprzez przyporządkowanie kolejnym wartościom cechy odpowiadające im liczebności skumulowane. Szeregi skumulowane informują dla ilu jednostek badanej zbiorowości cecha przyjmuje wartości nie większe od górnej granicy poszczególnego przedziału klasowego.

Tabela 4. Konstrukcja szeregu rozdzielczego punktowego

Wartość cechy	Skumulowana liczba elementów
X_1	n_1
$X_1 - X_2$	$n_1 + n_2$
\vdots	\vdots
$X_1 - X_k$	$n_1 + n_2 + \dots + n_k = N$

Zbudujmy zatem tabelkę szeregu rozdzielczego punktowego z przykładu 1 i zinterpretujmy otrzymane wyniki:

Tabela 5. Skumulowany szereg rozdzielczy punktowy z przykładu 1



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Numer klasy	Skumulowana wartość cechy	Skumulowana liczba elementów											
2	<i>k</i>	<i>xisk</i>	<i>nisk</i>											
3	1	113	1											
4	2	113-117	2											
5	3	113-123	4											
6	4	113-126	5											
7	5	113-129	6											
8	6	113-130	7											
9	7	113-133	8											
10	8	113-136	9											
11	9	113-140	10											
12	10	113-142	11											
13	11	113-148	13											
14	12	113-150	14											
15	13	113-152	16											
16	14	113-156	17											
17	15	113-160	18											
18	16	113-176	19											
19	17	113-219	20											
20														

W statystyce oprócz działań ważna jest interpretacja otrzymanych wyników. Można powiedzieć, że masę od 113 do 117 miligramów osiągają 2 poczwarki, zaś wagę od 113 do 140 miligramów miało 10 poczwarek mącznika.

3.1.3. Szereg rozdzielczy z przedziałami klasowymi

Jeżeli nasza liczba wyników otrzymanych w badaniu jest większa niż 30 elementów do interpretacji otrzymanych wyników budujemy szereg rozdzielczy z przedziałami klasowymi. W tym celu dane szczegółowe również porządkujemy od wartości najmniejszej do największej. Następnie krok po kroku postępujemy zgodnie z modelem, w którym krok po kroku wyznaczamy:

- Rozstęp,
- Klasę
- Długość przedziału
- Rozpiętość przedziału klasowego
- Budujemy tabelkę.



Rozstępem badanej cechy X w próbie nazywamy różnicę między największą a najmniejszą liczbą ciągu. Rozstęp jest więc długością najkrótszego przedziału, w którym mieszczą się wszystkie wartości próbki:

$$R = x_{max} - x_{min},$$

gdzie x_{max} – największa wartość cechy;

x_{min} – najmniejsza wartość cechy.

Klasa, to przedziały o jednakowej długości. Mamy kilka sposobów wyznaczania jej liczby:

- a) $k = \sqrt{N}$, gdzie $N = \sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k$
- b) $k \leq 5 \ln N$
- c) $k = 1 + 3,322 \cdot \ln N$, gdzie k – liczba przedziałów, N – liczebność zbiorowości.

Długość klasy – jest to różnica pomiędzy górną i dolną różnicą klasy:

$$h \approx \frac{R}{k}, \quad \text{gdzie } R - \text{rozstęp; } k - \text{liczba przedziałów.}$$

Uwaga: Wyznaczając z powyższego wzoru długość klasy zawsze zaokrąglamy do całości w górę tak aby spełniony był warunek: $kh \geq R$.

Rozpiętością przedziału klasowego (interwał przedziałowy) nazywamy różnicę między górną i dolną granicą i -tego przedziału klasowego:

$h_i = x_{1i} - x_{0i}$, gdzie x_{1i} – koniec i -tego przedziału; x_{0i} – początek i -tego przedziału.

Tabela 5. Konstrukcja przedziału rozdzielczego

Numer klasy	Długość przedziału	Liczba elementów
	$x_{0i} - x_{1i}$	n_i
1	$x_{01} - x_{11}$	n_1
\vdots	\vdots	\vdots
k	$x_{0k} - x_{1k}$	n_k
Suma		N



Tworząc szereg rozdzielczy z przedziałami klasowymi należy pamiętać o tym, że początek przedziału jest tym samym co dolna granica przedziału, zaś koniec przedziału – górną granicą; przedziały mogą się stykać, ale nie zachodzić na siebie; przedziały są zawsze jednostronnie domknięte. W celu zbudowania i opracowania szeregu rozdzielczego z przedziałami klasowymi przyjrzyjmy się poniższemu przykładowi:

Przykład 2: Zbadano masę 32 zwierząt (w kg) leczonych w pewnej klinice. Otrzymano następujące wyniki: 49, 65, 80, 48, 56, 74, 56, 87, 40, 85, 60, 44, 50, 52, 66, 80, 79, 51, 44, 49, 58, 66, 70, 40, 80, 81, 55, 62, 40, 50, 64, 87.

Należy wybrać z naszego szeregu szczegółowego kolejno:

- a) Rozstęp, w tym celu znajdujemy wśród danych wagę najmniejszą i największą:

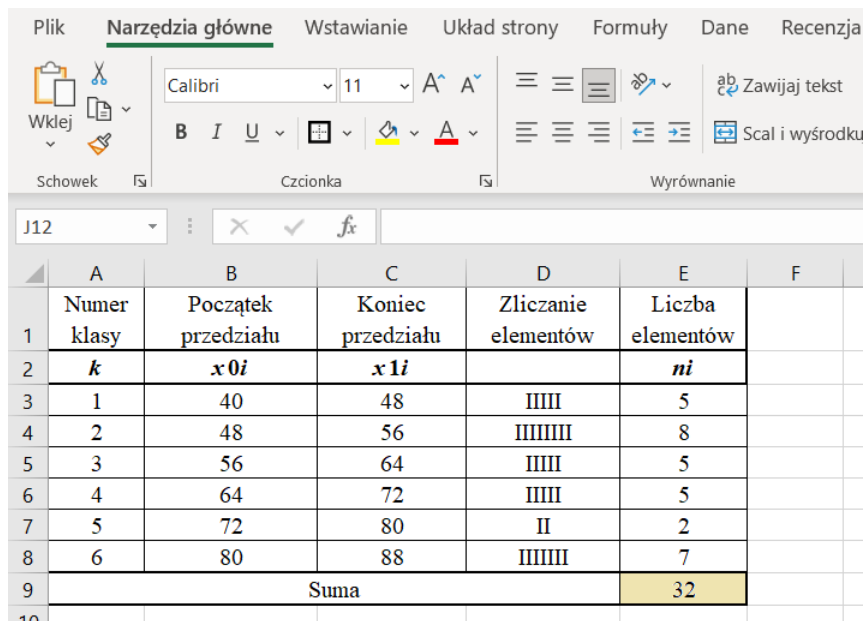
$x_{min}=40$ kg i $x_{max} = 87$ kg, zatem

$$R = x_{max} - x_{min} = 87 - 40 = 47$$

- b) Liczbę klas najczęściej wyznaczamy ze wzoru $k = \sqrt{N} = \sqrt{32} = 5,66 \approx 6$ należy pamiętać, że liczb klas jest liczbą naturalną i zaokrąglamy do góry zawsze do całości
- c) Długość klasy: $h \approx \frac{R}{k} = \frac{47}{6} \approx 7,8(3) \approx 8$ do całości przybliżono dla wygody obliczeń
- d) Rozpiętość przedziału klasowego wyznaczamy ze wzoru $h_i = x_{1i} - x_{0i}$. Początkiem pierwszego przedziału klasowego jest zawsze wartość najmniejsza z naszego szeregu szczegółowego, czyli $x_{0i} = x_{min} = 40$, zaś koniec naszego przedziału ze wzoru: $x_{11} = x_{01} + h = 40 + 8 = 48$. Kolejno początkiem długości drugiego przedziału jest koniec przedziału pierwszego, a koniec obliczamy ze wzoru $x_{12} = 48 + 8 = 56$. Czynność oczywiście powtarzamy dla wszystkich 8 klas. Możemy tak zrobić, ponieważ nasz przedział klasowy jest zawsze lewostronnie domknięty, a prawostronnie otwarty.

Pozostało nam już tylko zbudowanie tabelki:

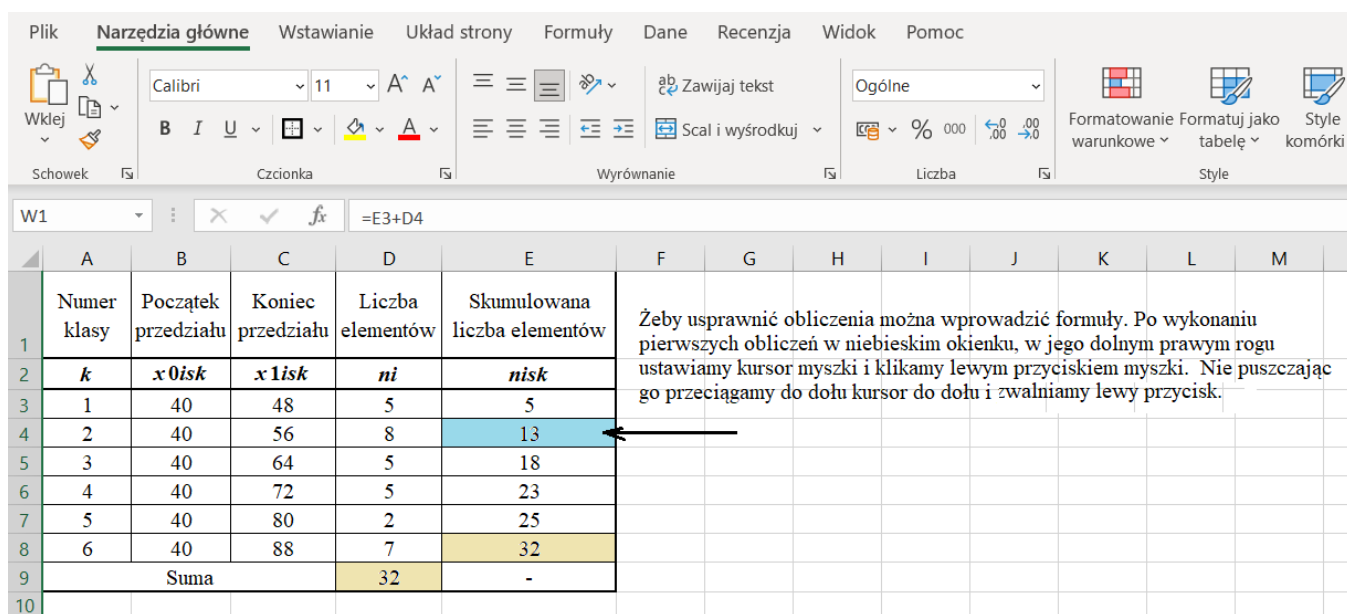
Tabela 6. Konstrukcja szeregu rozdzielczego z przedziałami klasowymi dla przykładu 2



	A	B	C	D	E	F
1	Numer klasy	Początek przedziału	Koniec przedziału	Zliczanie elementów	Liczba elementów	
2	k	x_{0i}	x_{1i}		n_i	
3	1	40	48	IIII	5	
4	2	48	56	IIIIIIII	8	
5	3	56	64	IIII	5	
6	4	64	72	IIII	5	
7	5	72	80	II	2	
8	6	80	88	IIIIIIII	7	
9	Suma				32	

Możemy również zbudować szereg rozdzielczy z przedziałami klasowymi skumulowany. Przedstawia go poniższa tabela:

Tabela 7. Skumulowany szereg rozdzielczy z przedziałami klasowymi



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Numer klasy	Początek przedziału	Koniec przedziału	Liczba elementów	Skumulowana liczba elementów	<p>Żeby usprawnić obliczenia można wprowadzić formuły. Po wykonaniu pierwszych obliczeń w niebieskim okienku, w jego dolnym prawym rogu ustawiamy kursor myszki i klikamy lewym przyciskiem myszki. Nie puszczając go przeciągamy do dołu kursor do dołu i zwalniamy lewy przycisk.</p>							
2	k	x_{0isk}	x_{1isk}	n_i	n_{isk}								
3	1	40	48	5	5								
4	2	40	56	8	13								
5	3	40	64	5	18								
6	4	40	72	5	23								
7	5	40	80	2	25								
8	6	40	88	7	32								
9	Suma				32	-							



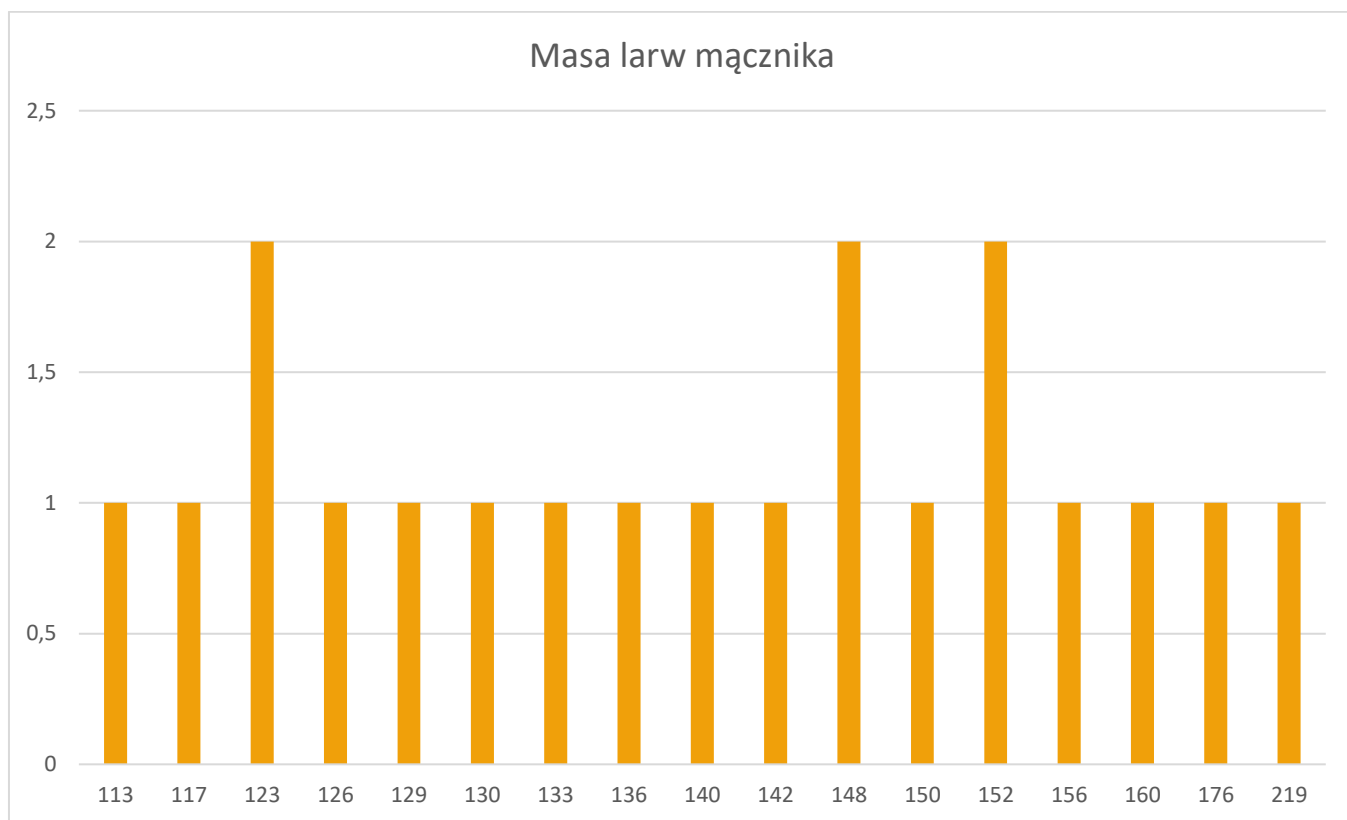
3.2. Graficzna interpretacja danych statystycznych

Graficzna metoda prezentacji danych statystycznych (materiałów liczbowych) polega na sporządzaniu różnego rodzaju wykresów. Wykresy zyskały szerokie zastosowanie, zarówno przy prezentacji surowego, jak i opracowanego materiału statystycznego, przede wszystkim dzięki temu, że: prezentują one badane zjawisko bardziej bezpośrednio, przejrzysto i obrazowo niż liczby, szybciej przemawiają do wyobraźni odbiorcy i pozwalają zorientować się w całości rozpatrywanych zagadnień, są doskonałym narzędziem popularyzowania i propagowania treści w nich zawartych.

W badaniach statystycznych najczęściej stosowanymi metodami graficznymi są: metoda liniowa, metoda powierzchniowa, metoda obrazkowa, metoda ilościowa i ilościowo-symbolowa, kartogramy oraz wykresy w układzie współrzędnych. Najprostszym sposobem graficznej prezentacji wielkości statystycznych jest **metoda liniowa**, w której prezentowane wielkości przedstawiane są w postaci pionowych lub poziomych odcinków. Poszczególne odcinki są obrazem, w odpowiedniej skali, rozważanych wielkości statystycznych. Zwykle przy liniach (z prawej strony linii lub nad liniami) podaje się liczby określające dokładny poziom ilustrowanej wielkości statystycznej. Podawanie tych liczb nie jest konieczne, ale zaleca się taką formę opisu wtedy, wykres ma być jedynym źródłem informacji dla odbiorcy. Rysunek musi być uzupełniony skalą, czyli linią, za pomocą której można określić wielkości zilustrowane poszczególnymi liniami wykresu. **Metoda powierzchniowa** polega na stosowaniu wykresów powierzchniowych, charakteryzujących zbiorowość lub zjawiska za pomocą powierzchni figur płaskich. Stosowana jest w prezentacji struktury zbiorowości zjawiska, do obrazowania szeregów rozdzielczych oraz prezentacji wielkości (liczebności) niektórych zjawisk. Generalnie obowiązującą zasadą w stosowaniu wykresów powierzchniowych jest zachowanie proporcjonalności powierzchni figur do prezentowanych wielkości. Do najczęściej wykorzystywanych wykresów przedstawiających szeregi rozdzielcze należą wykresy słupkowe tzw. histogramy i wykresy liniowe tzw. diagramy, oraz krzywe liczebności lub częstości.

Histogram – to zbiór prostokątów, których podstawy zaznaczone na osi odciętych stanowią rozpiętości poszczególnych przedziałów klasowych, natomiast wysokości są określone na osi rzędnych przez liczebności (częstości) odpowiadające poszczególnym przedziałom klasowym lub przez gęstości, liczebności (częstości).

Możemy graficznie przedstawić masę 20 poczwerek mącznika w postaci histogramu liczebności i częstości. Zaczniemy od histogramu liczebności.



Wykres 1. Histogram liczebności wagi larw mącznika

Jak można odczytać z histogramu dla wagi 123, 148 i 152 miligramów mamy po dwie larwy, a pozostałe wartości występują tylko jednorazowo.

Diagram – to łamana powstała przez połączenie punktów, których współrzędnymi są środki przedziałów klasowych i odpowiadające im liczebności (częstości lub gęstości). Diagramy dzielimy na diagram szeregów czasowych umożliwiające przedstawienie obserwacji w czasie, diagram częstości i diagram liczebności.

Łamana częstości – na osi poziomej odkładamy przedziały klasowe i środki klasowe, a na osi pionowej liczebności (częstości). Gdy łączymy punkty $(x_1^* - h, 0), \dots, (x_k^* + h, 0)$, to mamy łamaną częstości.

Dystrybuanta empiryczna – to przyporządkowanie kolejnym wartościom cechy statystycznej (zmiennej) odpowiadające im częstości skumulowane (względnie liczebności skumulowane). Dystrybuanta to łamana powstała przez połączenie punktów, których współrzędne to górne granice przedziałów klasowych i odpowiadające im liczebności (częstości) skumulowane.

Rozdział czwarty



Podstawowe wiadomości z analizy statystycznej



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Rozdział czwarty: Podstawowe wiadomości z analizy statystycznej

4.1. Cel analizy statystycznej

Opracowany i odpowiednio przedstawiony w postaci tablic, wykresów, opisów czy form mieszanych materiał statystyczny jest jedynie punktem wyjściowym w realizacji głównego i ostatecznego celu badania statystycznego – analizy danych statystycznych i wyciągnięcia na podstawie tej analizy wniosków i uogólnień dotyczących badanego zjawiska.

4.2. Analiza natężenia

W wyniku obserwacji statystycznej otrzymamy obraz badanych zjawisk wyrażony w **liczbach bezwzględnych**, które otrzymujemy w wyniku zliczania poszczególnych jednostek lub sumowania wartości cech mierzalnych. Są to **liczby mianowane**, czyli wyrażane w określonych jednostkach miary (waga w kg, liczba w sztukach, powierzchnia w m²). Wartość poznawcza liczb bezwzględnych jest znacząca, bowiem określają one wielkości badanych zjawisk, charakteryzują zasoby produkcyjne, wyniki działalności gospodarczej i społecznej, są wielkościami wykorzystywanymi w procesie wykonywania podstawowych założeń realizowanych programów. W analizie statystycznej wykorzystujemy również **liczby względne (stosunkowe)**, które są ilorazami liczb bezwzględnych opisujących zjawiska związane ze sobą.

Liczby względne w analizie statystycznej mają szczególne znaczenie: charakteryzują natężenie (intensywność) zjawisk oraz charakteryzują strukturę zjawiska.

4.3. Analiza struktury

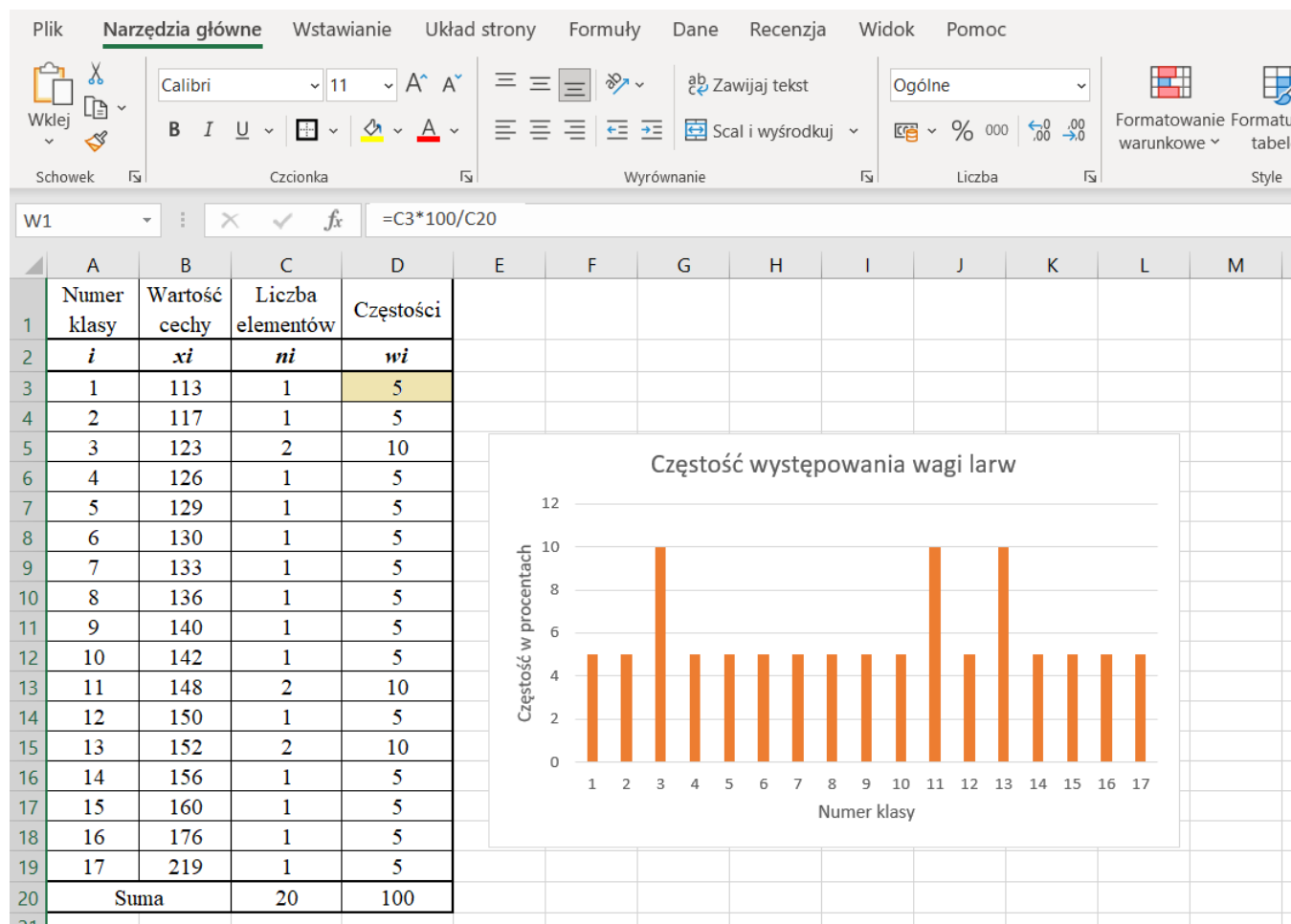
Struktura zbiorowości określona jest przez podział badanej zbiorowości statystycznej na grupy jednostek różniących się od siebie wartościami rozważanej cechy mierzalnej lub wariantami cechy niemierzalnej. Liczbową formą opisu struktury są liczby względne, powszechnie stosowane w analizie statystycznej wskaźniki struktury.

Wskaźnikiem struktury nazywamy częstość występowania danego wariantu cechy. Jest, to stosunek liczby jednostek o danej wartości cechy do liczebności całej zbiorowości co zapisujemy:

$$w_i = \frac{n_i}{N} \cdot 100\%, \quad \text{gdzie } i = 1, 2, \dots, k.$$

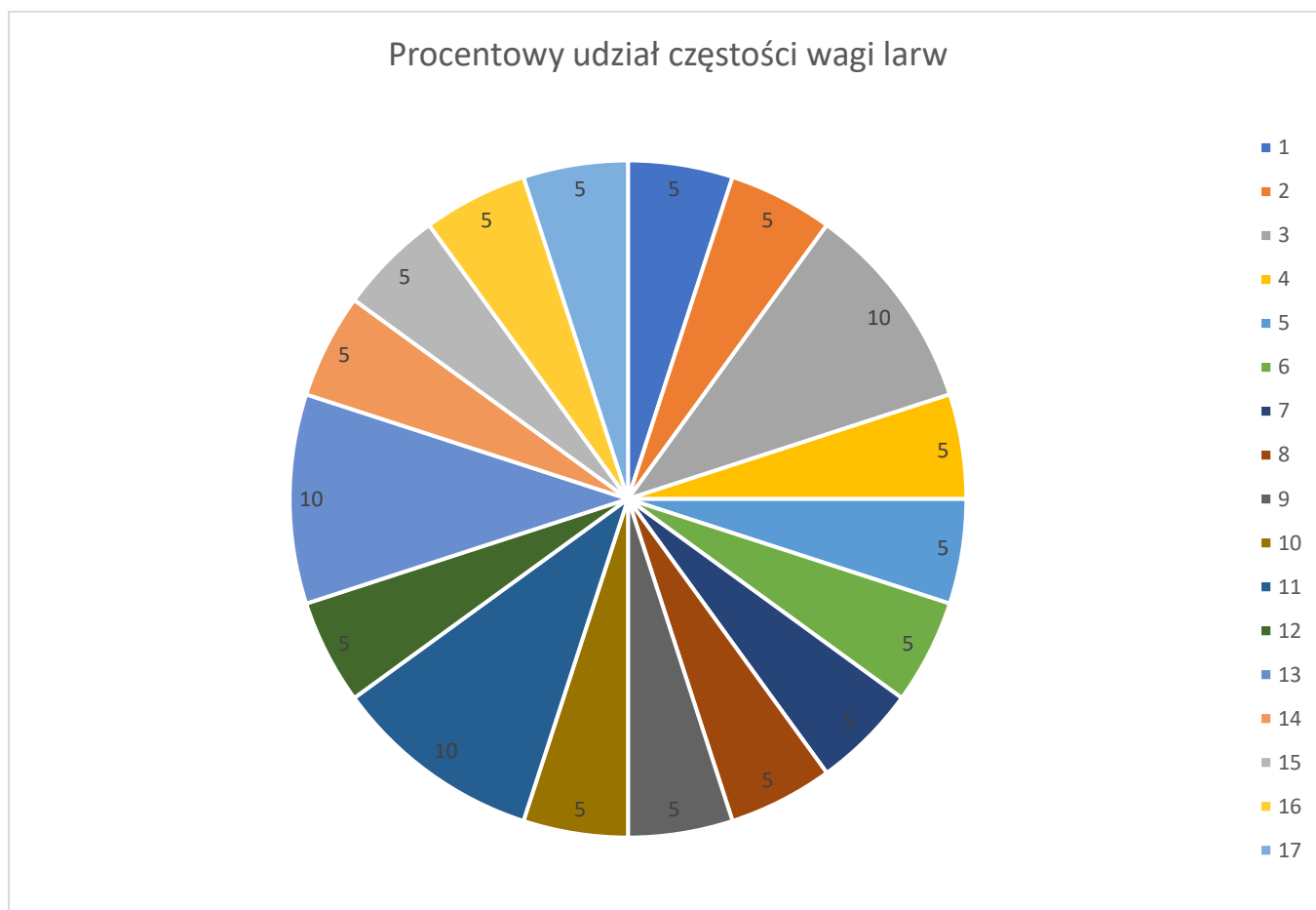
Należy pamiętać, że suma częstości musi być równa jedynce: $\sum_{i=1}^k w_i = 1$ 00.

Wyznamy dla naszego przykładu 1 częstość występowania zjawiska a następnie narysujemy histogramu częstości.



Rysunek 1. Tabela i histogram częstości wagi larw

W powyższej tabeli przedstawiona została częstość w procentach wagi larw mącznika. Po 10% (najwięcej) było larw o wadze 123, 148 i 152 miligramów. Taką tabelę można też przedstawić za pomocą grafu kołowego:



Wykres 2. Procentowy udział wagi larw mącznika

Wskaźnik struktury skumulowany wyraża się wzorem:

$$w_{i_{sk}} = \frac{n_{i_{sk}}}{N}, \quad \text{gdzie } i = 1, 2, \dots, k.$$

Uwaga: Należy zawsze pamiętać o tym, że w szeregu rozdzielczym zwykłym liczebności maleją i rosną natomiast w szeregu skumulowanym liczebności zawsze rosną.

4.4. Analiza tendencji centralnej

Analiza badanych zbiorowości statystycznych z określonego rzeczowego punktu widzenia na podstawie cech mierzalnych wymaga ustalenia dla tych zbiorowości średniego poziomu wartości tych cech.

Średnia (miary położenia) – to miara odzwierciedlająca przeciętny poziom cechy mierzalnej jednostek zbiorowości statystycznej, charakteryzuje centralnie położoną wartość, dookoła której skupiają się jednostki zbiorowości. Jest, to zatem wielkość abstrakcyjna charakteryzująca poziom badanego zjawiska.



Miary tendencji centralnej dzielimy na:

- a) **średnie klasyczne** – otrzymywane drogą rachunkową z uwzględnieniem wszystkich wartości szeregu statystycznego. Do średnich klasycznych zaliczamy **średnią arytmetyczną, średnią harmoniczną i średnią arytmetyczną**. Skupimy się w artykule tylko na średniej arytmetycznej
- b) **średnie pozycyjne** – to rzeczywiste wartości cechy występujące w uporządkowanym rosnąco szeregu statystycznym wybrane ze względu na zajmowaną pozycję w tym szeregu. Do średnich pozycyjnych zaliczamy:
 - o **dominantę (modę)**;
 - o **kwartył pierwszy (kwartył dolny)**;
 - o **kwartył drugi (medianę)**;
 - o **kwartył trzeci (kwartył górny)**.

4.4.1. Średnie klasyczne i ich własności

Średnia arytmetyczna – jest sumą wartości cechy poszczególnych jednostek populacji, podzieloną przez ich ilość. Jeżeli populacja ma n elementów o wartościach cechy x_1, x_2, \dots, x_n ,

to zgodnie z określeniem średniej arytmetycznej, którą oznaczmy przez \bar{x} (lub symbolem m). Średnia arytmetyczna jest określona wzorami:

- a) **Średnia arytmetyczna nieważona (prosta, zwykła)** w szeregu szczegółowym:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{\sum_{i=1}^N x_i}{N},$$

gdzie x_i – wariant cechy mierzalnej; N – liczebność badanej zbiorowości.

Uwaga: Jeżeli warianty zmiennej występują z różną częstotliwością, to obliczamy średnią arytmetyczną ważoną. Wagami są wówczas liczebności odpowiadające poszczególnym wariantom. Z taką sytuacją mamy do czynienia w szeregach rozdzielczych punktowych i w szeregach rozdzielczych z przedziałami klasowymi.

- b) **Średnia arytmetyczna ważona** – wyznacza się w wg wzoru:

- o dla szeregu rozdzielczego punktowego:

$$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{N}; \quad \bar{x} = \sum_{i=1}^k x_i \cdot w_i;$$

- o dla szeregu rozdzielczego z przedziałami klasowymi:



$$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{N}; \quad \bar{x} = \sum_{i=1}^k x_i \cdot w_i$$

Tabela 8. Wyznaczanie średniej wagi larw mącznika

	A	B	C	D	E	F	G	H	I	J
1	Numer klasy	Wartość cechy	Liczba elementów	Obliczenia pomocnicze						
2	<i>i</i>	<i>x_i</i>	<i>n_i</i>	<i>x_i·n_i</i>						
3	1	113	1	113	Wprowadzamy funkcję, przeciągamy ją na wszystkie klasy i na końcu sumujemy					
4	2	117	1	117						
5	3	123	2	246						
6	4	126	1	126						
7	5	129	1	129						
8	6	130	1	130						
9	7	133	1	133						
10	8	136	1	136						
11	9	140	1	140						
12	10	142	1	142						
13	11	148	2	296						
14	12	150	1	150						
15	13	152	2	304						
16	14	156	1	156						
17	15	160	1	160						
18	16	176	1	176						
19	17	219	1	219						
20	Suma		20	2873						
21										

Korzystając ze wzoru na średnią arytmetyczną otrzymamy $\bar{x} = \frac{\sum_{i=1}^k x_i \cdot n_i}{N} = \frac{2873}{20} = 143,65$

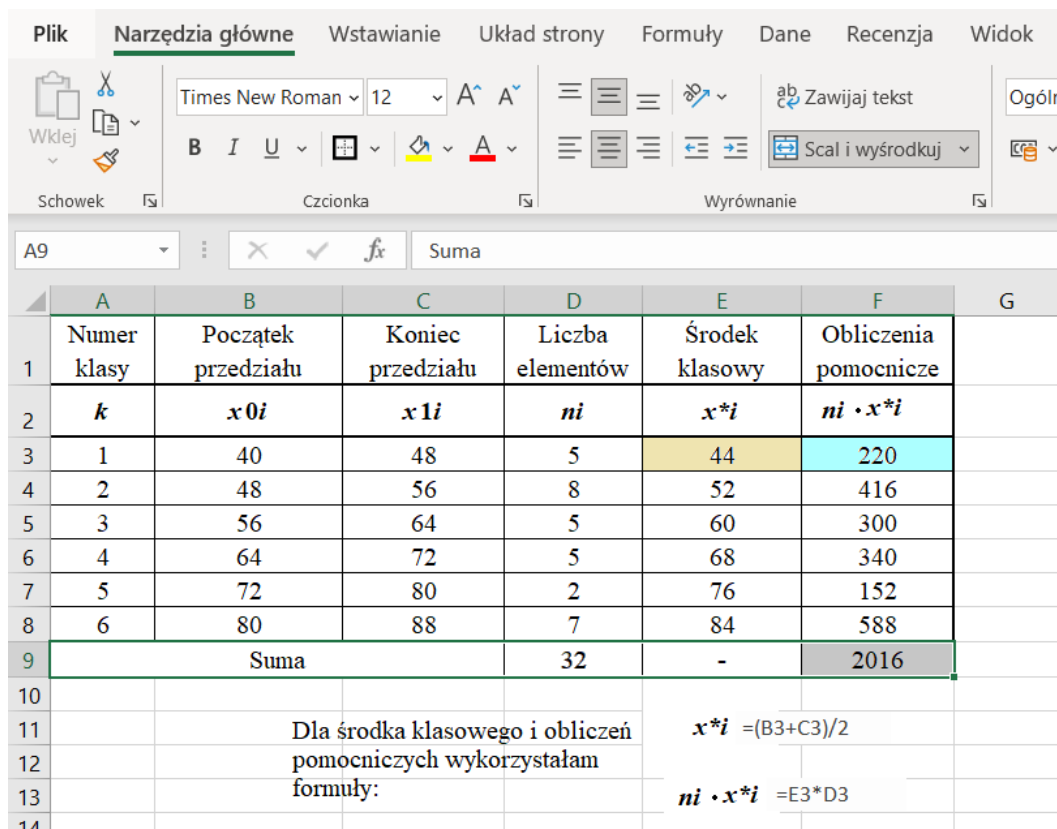
Możemy zatem powiedzieć, że średnia waga larwy mącznika wynosi 143,65 miligrama.

Dla szeregu rozdzielczego z przedziałami klasowymi przed obliczeniem średniej arytmetycznej musimy wyznaczyć środek klasowy.

Środek klasowy (środek przedziału) – to średnia arytmetyczna dolnej i górnej granicy przedziału, który można obliczyć za pomocą wzoru:

$$x_i^* = \frac{x_{0i} + x_{1i}}{2}$$

Tabela 9. Wyznaczenie średniej wagi dla szeregu rozdzielczego z przedziałami klasowymi



	A	B	C	D	E	F	G
1	Numer klasy	Początek przedziału	Koniec przedziału	Liczba elementów	Środek klasowy	Obliczenia pomocnicze	
2	k	x_{0i}	x_{1i}	n_i	x^*i	$n_i \cdot x^*i$	
3	1	40	48	5	44	220	
4	2	48	56	8	52	416	
5	3	56	64	5	60	300	
6	4	64	72	5	68	340	
7	5	72	80	2	76	152	
8	6	80	88	7	84	588	
9	Suma			32	-	2016	
10							
11		Dla środka klasowego i obliczeń pomocniczych wykorzystałam formuły:			$x^*i = (B3+C3)/2$		
12					$n_i \cdot x^*i = E3*D3$		
13							
14							

Średnia arytmetyczna jest zatem równa $x = \frac{\sum_{i=1}^k x_i^* \cdot n_i}{N} = \frac{2016}{32} = 63$. Zatem średnia waga pacjentów u weterynarza wynosiła 63 kg.

Dla każdej średniej arytmetycznej zachodzi kilka własności, ale najważniejszą z nich jest to ta, że średnia jest między wartością najmniejszą i największą szeregu statystycznego. Jest wielkością abstrakcyjną jaką miałyby wszystkie jednostki zbiorowości przy ustalonej sumie cechy gdyby nie występowała zmienność. Dla szeregów rozdzielczych obliczamy średnią arytmetyczną tylko wtedy, gdy mają jednostronnie domknięte przedziały klasowe oraz gdy mają taką samą długość. Średnia harmoniczna stosowana jest do szeregów rozdzielczych charakteryzują się tym, że pierwsza kolumna – to warianty interesującej nas cechy, a w drugiej kolumnie nie liczebności (jak to było przy wyznaczaniu średniej arytmetycznej) lecz inne wartości logiczne związane z wariantami cechy. Są to tzw. szeregów quasi – rozdzielcze.

Średnią harmoniczną o wartościach cechy x_1, x_2, \dots, x_n różnych od zera ($x_i \neq 0$) nazywamy odwrotnością średniej arytmetycznej z odwrotności wartości zmiennych. Średnią harmoniczną stosujemy gdy wartości cechy podane są w przeliczeniu na stałą jednostkę innej zmiennej



(jednostkach względnych) w postaci wskaźnika natężenia = wskaźnik struktury skumulowanej natomiast wagami są jednostki liczników tych cech:

- | | |
|-------------------------------------------------|--------------|
| a) prędkość [km/h] | waga = km |
| b) prędkość [min/szt.] | waga = min |
| c) gęstość zaludnienia [osoba/km ²] | waga = osoba |
| d) spożycie artykuły X na 1 osobę [kg/osoba] | waga = kg |
| e) cena jednostkowa [zł/szt.] | waga = zł |

Średnia harmoniczna jest określona wzorami:

- a) **Średnia harmoniczna nieważona** w szeregu prostym:

$$H = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}};$$

- b) **Średnia harmoniczna ważona**:

- o dla szeregu rozdzielczego punktowego:

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

- o dla szeregu rozdzielczego z przedziałami klasowymi:

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

Średnia geometryczna jest pierwiastkiem n -tego stopnia z iloczynu n wartości danej zmiennej. Wykorzystujemy, gdy badamy średnie tempo zmian zjawiska gdy ujęcie zjawisk jest dynamiczne. Wykorzystujemy następujący wzór:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_k} = \sqrt[n]{\prod_{i=1}^k (x_i)},$$

gdzie x_i – oznacza wielkość (dynamikę) zjawiska; \prod – znak iloczynu.

Jeżeli wartości zmienne występują z różną częstotliwością, średnią geometryczną można wyznaczyć za pomocą wzoru:



$$\bar{x}_g = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_k^{n_k}} = \sqrt[n]{\prod_{i=1}^k (x_i)^{n_i}}$$

Średnia harmoniczna i średnia geometryczna te są rzadziej wykorzystywane w problemach statystycznych w biologii i medycynie.

4.4.2. Średnie pozycyjne

Jedyną poważniejszą wadą średniej arytmetycznej jest to, że duży wpływ na nią wywierają skrajne wartości cechy (niejednokrotnie przypadkowo włączone do próby). Przykładowo w pewnej firmie usługowej zarobki ośmiu zatrudnionych pracowników osiągnęły 800 zł miesięcznie. Księgowa i kierownik zespołu otrzymali po 3 000 zł, a właściciel wypłacił sobie 10 000 zł. Średnia zarobków w tej firmie wygląda zachęcająco (około 2036 zł). Jakże myląca i niepełna jest ta informacja. Sytuację tę lepiej opisują inne (poniżej opisane) miary, zwane miarami pozycyjnymi.

Dominanta zmiennej skokowej, to wartość cechy, która w danym rozkładzie empirycznym najczęściej się powtarza, czyli jest to wartość o największej liczebności, lub częstości.

Powracając do naszego przykładu zapisanego w postaci szeregu rozdzielczego punktowego nie jestem w stanie znaleźć wagi występującej najczęściej, ponieważ dla wagi 123, 148 i 152 miligramów występują po dwie larwy. Stąd mamy wniosek, że wartość dominująca może pojawić się nam w badaniu, ale wcale nie musi. Przyjrzyjmy się teraz dominancie dla szeregu rozdzielczego z przedziałami klasowymi.

Dominanta zmiennej ciągłej, to liczba, wokół której skupiona jest (lokalnie) największa liczba obserwacji tej zmiennej. Dominanta wyraża się wzorami:

$$D = x_{0D} + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D ;$$

gdzie **D** – symbol dominanty; **x_{0D}** – początek przedziału w którym znajduje się dominanta; **n_D** – liczebność przedziału w którym znajduje się dominanta; **n_{D+1}** – liczebność przedziału następującego po przedziale z dominantą; **n_{D-1}** – liczebność przedziału poprzedzającego przedział z dominantą; **h_D** – długość przedziału w którym znajduje się dominanta.

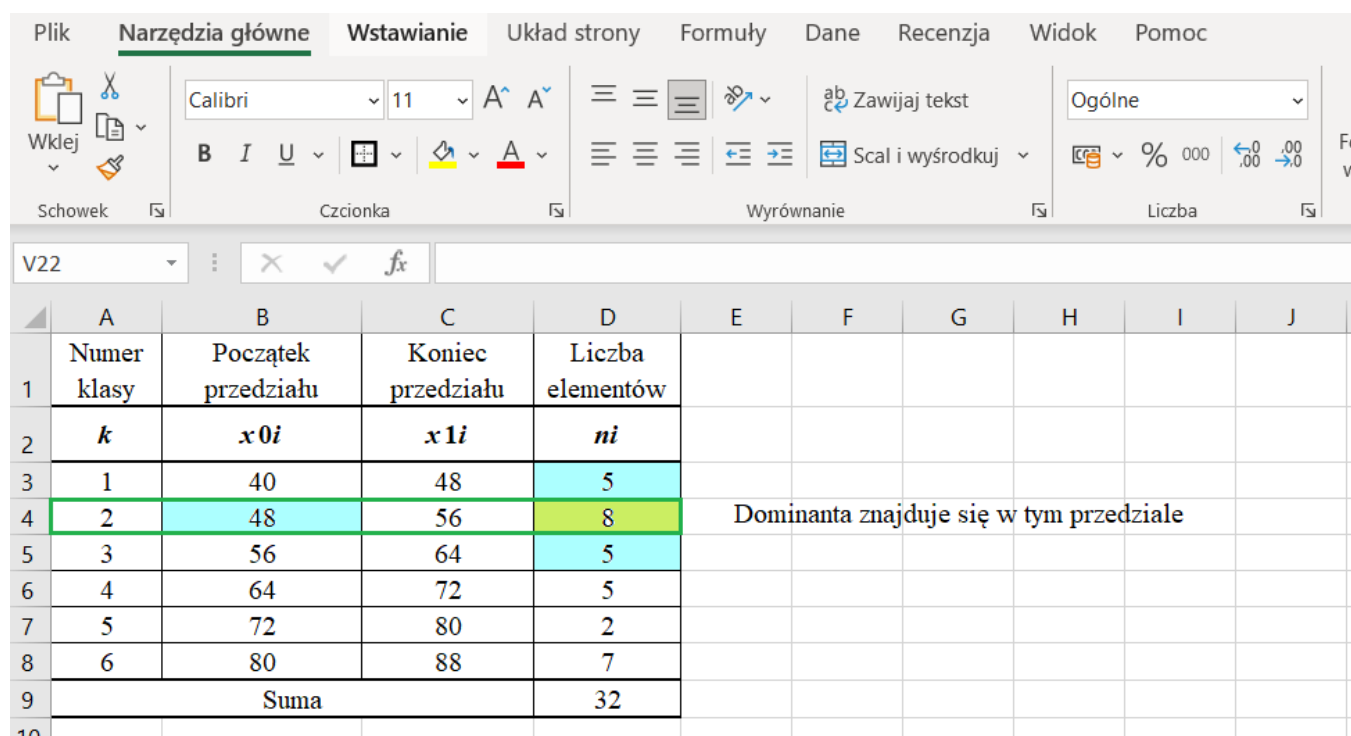
$$D = x_{0D} + \frac{w_D - w_{D-1}}{(w_D - w_{D-1}) + (w_D - w_{D+1})} \cdot h_D ;$$

gdzie **D** – symbol dominanty; **x_{0D}** – początek przedziału w którym znajduje się dominanta; **w_D** – częstość przedziału w którym znajduje się dominanta; **w_{D+1}** – częstość przedziału następującego po przedziale z dominantą; **w_{D-1}** – częstość przedziału poprzedzającego przedział z dominantą; **h_D** – długość przedziału w którym znajduje się dominanta.

Wyznaczanie dominanty w szeregach rozdzielczych ma sens tylko wtedy, gdy występuje tylko **jedno** wyraźnie zaznaczone maksimum. Przedział, w którym występuje dominanta, oraz dwa sąsiadujące z nią przedziały muszą mieć taką samą rozpiętość. Wartość dominanty zależy od liczebności przedziału i zawsze znajduje się w wybranym przedziale. W odróżnieniu od średniej arytmetycznej dominanta nie jest czuła na krańcowe wartości przedziałów. Przy obliczaniu dominanty liczebności możemy zastąpić częstościami. Gdy zbiory są mało liczne nie powinniśmy wyznaczać dominanty.

Wyznamy zatem dominantę dla szeregu rozdzielczego z przedziałami klasowymi dla przykładu 2. Przyjrzymy się zatem tabelce:

Tabela 10. Wyznaczanie dominanty dla przykładu 2



	A	B	C	D	E	F	G	H	I	J
	Numer klasy	Początek przedziału	Koniec przedziału	Liczba elementów						
1	k	x_{0i}	x_{1i}	n_i						
2	1	40	48	5						
3	2	48	56	8	Dominanta znajduje się w tym przedziale					
4	3	56	64	5						
5	4	64	72	5						
6	5	72	80	2						
7	6	80	88	7						
8	Suma			32						

Wyznamy potrzebne wartości do podstawienia we wzorze: $x_{0D}=48$, $n_D=8$, $n_{D+1}=5$, $n_{D-1}=5$, $h_D=8$ i podstawmy do wzoru

$$D = x_{0D} + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} \cdot h_D = 48 + \frac{8-5}{(8-5)+(8-5)} \cdot 8 = 52.$$

Najważniejsza jest interpretacja otrzymanego wyniku. Najwięcej przebadano zwierząt o wadze 52 kg.

Następną miarą pozycyjną są kwartyle. Mamy trzy najczęściej wykorzystywane.

Kwartył pierwszy – to wartość cechy, która dzieli zbiorowość na dwie części w taki sposób, że 25% elementów ma wartości nie większe od kwartyła pierwszego i druga część stanowiąca 75% ma wartości nie mniejsze od kwartyła pierwszego. Wyznamy ze wzorów:



$$Q_1 = x_{0Q_1} + \frac{N_{Q_1} - \sum_{i=1}^{k-1} n_i}{n_{Q_1}} \cdot h_{Q_1}, \text{ gdzie } N_{Q_1} = \frac{1}{4} N$$

Q_1 – symbol pierwszego kwartyla; x_{0Q_1} – początek przedziału w którym znajduje się kwartył pierwszy; N_{Q_1} – pozycja kwartyla pierwszego; n_{Q_1} – liczebność przedziału z kwartyłem

pierwszym; $\sum_{i=1}^{k-1} n_i$ – suma liczebności od klasy pierwszej do tej, w której znajduje się kwartył pierwszy; h_{Q_1} – długość przedziału w którym znajduje się kwartył pierwszy.

$$Q_1 = x_{0Q_1} + \frac{\frac{1}{4} - \sum_{i=1}^{k-1} w_i}{w_{Q_1}} \cdot h_{Q_1},$$

Gdzie Q_1 – symbol kwartyla pierwszego; x_{0Q_1} – początek przedziału w którym znajduje się kwartył pierwszy; $\frac{1}{4}$ – pozycja kwartyla pierwszego; w_{Q_1} – częstość przedziału z kwartyłem

pierwszym; $\sum_{i=1}^{k-1} w_i$ – suma częstości od klasy pierwszej do tej, w której znajduje się kwartył pierwszy; h_{Q_1} – długość przedziału w którym znajduje się kwartył pierwszy.

Kwartył drugi (Mediana) – to wartość cechy, która dzieli zbiorowość na dwie części w taki sposób, że połowa ma wartości nie większe od mediany i druga połowa ma wartości nie mniejsze od mediany:

$$Me = x_{0Me} + \frac{N_{Me} - \sum_{i=1}^{k-1} n_i}{n_{Me}} \cdot h_{Me}, \text{ gdzie } N_{Me} = \frac{1}{2} N$$

Me – symbol mediany; x_{0Me} – początek przedziału z medianą; N_{Me} – pozycja mediany; n_{Me} –

liczebność przedziału z medianą; $\sum_{i=1}^{k-1} n_i$ – suma liczebności od klasy pierwszej do tej, w której znajduje się mediana; h_{Me} – długość przedziału z medianą.

$$Me = x_{0Me} + \frac{\frac{1}{2} - \sum_{i=1}^{k-1} w_i}{w_{Me}} \cdot h_{Me};$$



Me – symbol mediany; x_{0Me} – początek przedziału z medianą; $\frac{1}{2}$ – pozycja mediany; W_{Me} –

częstość przedziału z medianą; $\sum_{i=1}^{k-1} w_i$ – suma częstości od klasy pierwszej do tej, w której znajduje się mediana; h_{Me} – długość przedziału z medianą.

Uwaga: W przypadku szeregów wyliczających, składających się z niewielkiej liczby jednostek, medianę obliczamy za pomocą wzory:

$$Me = \begin{cases} x_{\frac{N+1}{2}}, & N - \text{nieparzyste} \\ \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}), & N - \text{parzyste} \end{cases}$$

Do wyznaczenia mediany wykorzystamy szereg rozdzielczy punktowy skumulowany. Skorzystamy ze wzoru znajdującego się wyżej. Nasza ilość otrzymanych wyników jest równa 20 i jest to liczba parzysta, zatem mediana jest równa:

$$Me = \frac{1}{2} \left(x_{\frac{20}{2}} + x_{\frac{20}{2}+1} \right) = \frac{1}{2} (x_{10} + x_{11}) = \frac{1}{2} (140 + 142) = 141$$

Oznacza to dla nas, że połowa larw mącznika osiągała wagę co najmniej równą 141 miligramów, a druga połowa co najwyżej 141 miligramów.

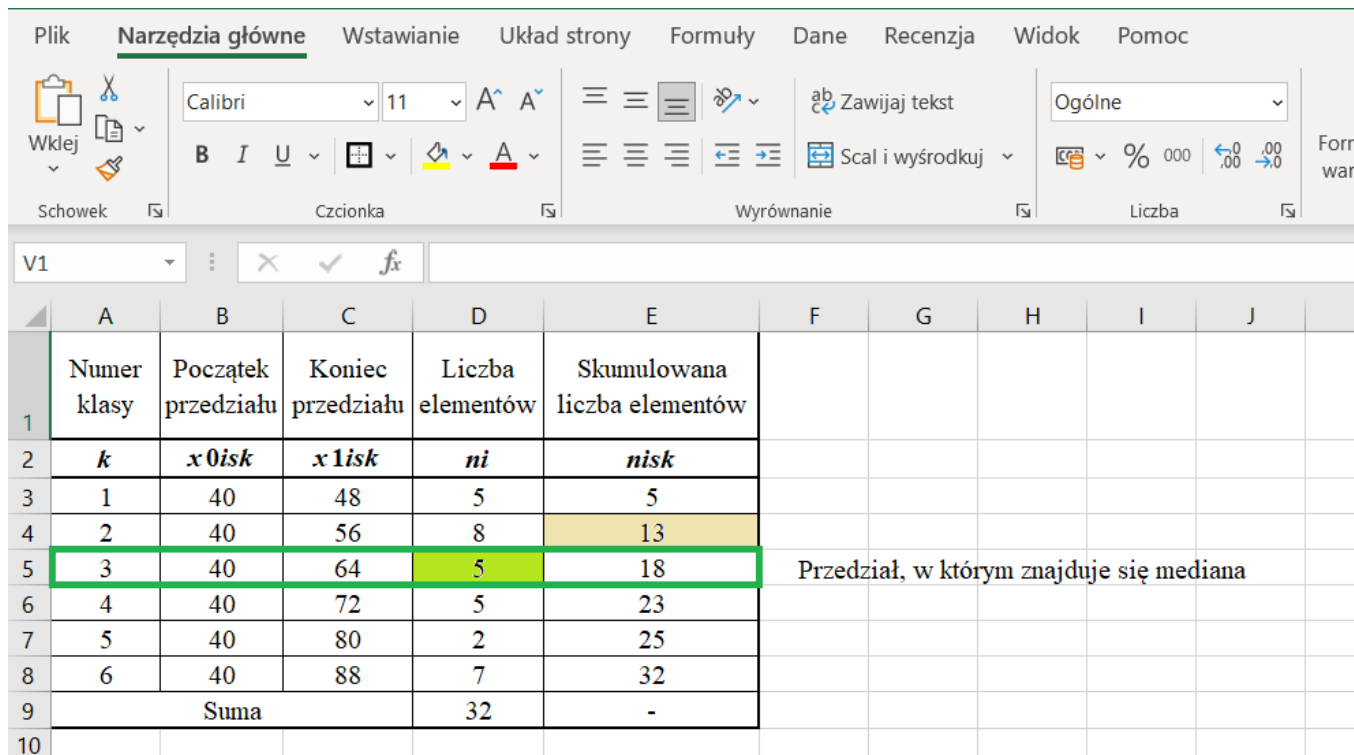
Jeżeli rozwiązywalibyśmy nasze zadanie wykorzystując szereg rozdzielczy z przedziałami klasowymi, to musielibyśmy wyznaczyć z tabelki 11 potrzebne wartości:

Na początek wyznaczmy przedział, w którym znajduje się nasza mediana ze wzoru $N_{Me}=0,5 \cdot 32=16$. Mediana znajduje się zatem w trzecim przedziale i teraz odpowiednio

wybieramy wartości $n_{Me}=5$, $\sum_{i=1}^{k-1} n_i =13$, $h_{Me}=8$.

$$Me = x_{0Me} + \frac{N_{Me} - \sum_{i=1}^{k-1} n_i}{n_{Me}} \cdot h_{Me} = 56 + \frac{16-13}{5} \cdot 8 = 60,8$$

Tabela 11. Mediana dla szeregu rozdzielczego z przedziałami klasowymi



	A	B	C	D	E	F	G	H	I	J
1	Numer klasy	Początek przedziału	Koniec przedziału	Liczba elementów	Skumulowana liczba elementów					
2	k	x_{0isk}	x_{1isk}	n_i	n_{isk}					
3	1	40	48	5	5					
4	2	40	56	8	13					
5	3	40	64	5	18	Przedział, w którym znajduje się mediana				
6	4	40	72	5	23					
7	5	40	80	2	25					
8	6	40	88	7	32					
9	Suma			32	-					
10										

Oznacza to dla nas, że połowa badanych zwierząt osiągała wagę co najmniej równą 60,8 kg miligramów, a druga połowa co najwyżej 60,8 kg..

Kwartył trzeci – to wartość cechy, która dzieli zbiorowość na dwie części w taki sposób, że 75% elementów ma wartości nie większe od kwartyła trzeciego i druga część stanowiąca 25% ma wartości nie mniejsze od kwartyła trzeciego:

$$Q_3 = x_{0Q_3} + \frac{N_{Q_3} + \sum_{i=1}^{k-1} n_i}{n_{Q_3}} \cdot h_{Q_3}, \text{ gdzie } N_{Q_3} = \frac{3}{4}N$$

Q_3 – symbol kwartylna trzeciego; x_{0Q_3} – początek przedziału w którym znajduje się kwartył

trzeci; N_{Q_3} – pozycja kwartylna trzeciego; n_{Q_3} – liczebność przedziału z kwartyłem trzecim; $\sum_{i=1}^{k-1} n_i$

– suma liczebności od klasy pierwszej do tej, w której znajduje się kwartył pierwszy; h_{Q_3} – długość przedziału z kwartyłem trzecim.

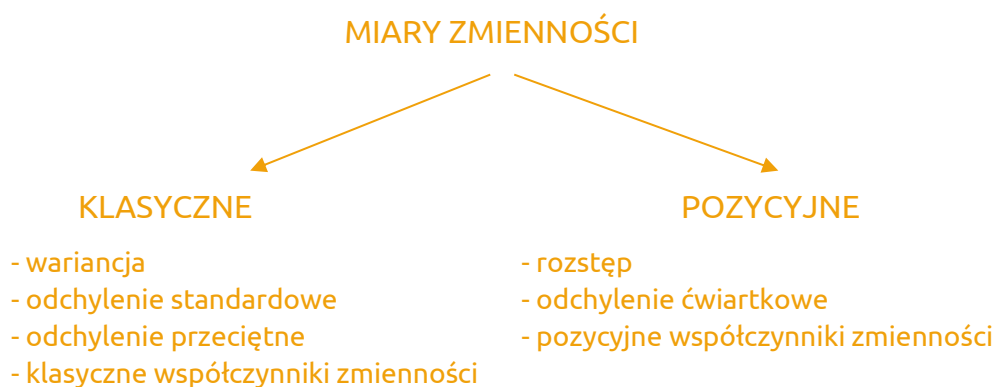
$$Q_3 = x_{0Q_3} + \frac{\frac{3}{4} - \sum_{i=1}^{k-1} w_i}{w_{Q_3}} \cdot h_{Q_3},$$



gdzie Q_3 – symbol kwartyła trzeciego; x_{0Q_3} – początek przedziału w którym znajduje się kwartył trzeci; $\frac{3}{4}$ – pozycja kwartyła trzeciego; w_{Q_3} – częstość przedziału z kwartyłem trzecim; $\sum_{i=1}^{k-1} w_i$ – suma częstości od klasy pierwszej do tej, w której znajduje się kwartył trzeci; h_{Q_3} – długość przedziału w którym znajduje się kwartył trzeci.

4.4.3. Miary zmienności

Dyspersja (rozproszenie) to zróżnicowanie jednostek zbiorowości statystycznej ze względu na wartość badanej cechy. Siłę dyspersji oceniamy za pomocą pozycyjnych i klasycznych miar zmienności.



Najprostszą miarą zmienności, którą już poznaliśmy jest rozstęp. **Rozstęp** – pokazuje nam różnicę między największą i najmniejszą wartością w badanej zbiorowości. Rozstęp wyraża się wzorem:

$$R = x_{max} - x_{min}.$$

Rozstęp międzykwartyłowy – jest różnicą między kwartyłem trzecim, a kwartyłem pierwszym. Określa on długość obszaru zmienności cechy, w którym znajduje się 50% wartości cechy. W ten sposób odrzucamy wartości skrajne:

$$I = Q_3 - Q_1.$$

Odchylenie przeciętne średnia arytmetyczna bezwzględnych wartości (modułów) cechy od jej średniej arytmetycznej. Określa ono o ile jednostki danej zbiorowości różnią się średnio ze względu na wartości cechy od średniej arytmetycznej tej zmiennej. Do obliczania odchylenia przeciętnego wykorzystujemy następujące wzory:

a) **Odchylenie przeciętne** w szeregu prostym:

$$d = \frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}|$$



b) **Odchylenie przeciętne** – wyznacza się w szeregach rozdzielczych punktowych i w szeregach rozdzielczych z przedziałami klasowymi wg wzoru:

- o dla szeregu rozdzielczego punktowego:

$$d = \frac{\sum_{i=1}^k |x_i - \bar{x}| \cdot n_i}{N}; \quad d = \sum_{i=1}^k |x_i - \bar{x}| \cdot w_i$$

- o dla szeregu rozdzielczego z przedziałami klasowymi:

$$d = \frac{\sum_{i=1}^k |x_i^* - \bar{x}| \cdot n_i}{N}; \quad d = \sum_{i=1}^k |x_i^* - \bar{x}| \cdot w_i$$

Ponieważ odchylenie przeciętne jest miarą statystyczną rzadziej wykorzystywaną niż odchylenie standardowe wyznaczmy tą miarę jedynie dla szeregu rozdzielczego punktowego. Przyjrzyjmy się tabelce z przykładu 1:

Tabela 12. Odchylenie przeciętne dla przykładu 1

Plik Narzędzia główne Wstawianie Układ strony Formuły Dane Recenzja W							
Times New Roman 12 A ⁺ A ⁻ Zawijaj tekst							
B I U W Czcionka Wyrównanie							
17 =MODUŁ.LICZBY(E4)							
	A	B	C	D	E	F	G
1	Numer klasy	Wartość cechy	Liczba elementów	Obliczenia pomocnicze			
2	<i>i</i>	<i>x_i</i>	<i>n_i</i>	<i>x_i*n_i</i>	<i>x_i-x̄</i>	<i> x_i-x̄ </i>	<i> x_i-x̄ n_i</i>
3	1	113	1	113	-30,65	30,65	30,65
4	2	117	1	117	-26,65	26,65	26,65
5	3	123	2	246	-20,65	20,65	41,3
6	4	126	1	126	-17,65	17,65	17,65
7	5	129	1	129	-14,65	14,65	14,65
8	6	130	1	130	-13,65	13,65	13,65
9	7	133	1	133	-10,65	10,65	10,65
10	8	136	1	136	-7,65	7,65	7,65
11	9	140	1	140	-3,65	3,65	3,65
12	10	142	1	142	-1,65	1,65	1,65
13	11	148	2	296	4,35	4,35	8,7
14	12	150	1	150	6,35	6,35	6,35
15	13	152	2	304	8,35	8,35	16,7
16	14	156	1	156	12,35	12,35	12,35
17	15	160	1	160	16,35	16,35	16,35
18	16	176	1	176	32,35	32,35	32,35
19	17	219	1	219	75,35	75,35	75,35
20	Suma		20	2873	-	-	336,3

Odchylenie przeciętne jest zatem równe $d = \frac{336,3}{20} = 16,82$ zatem waga przeciętnej larwy waha się od 143,16 miligramów o 16,82, czyli wynosi od 126,34 miligramów do 159,98 miligramów. Odchylenie przeciętne jest jednak mniej miarodajną wartością niż odchylenie standardowe używane najczęściej w statystyce. Aby wyznaczyć odchylenie standardowe musimy policzyć wariancję.

Wariancja – to średnia arytmetyczna kwadratów odchyłeń poszczególnych wartości cechy od średniej arytmetycznej zbiorowości. Wykorzystujemy do obliczania wariancji następujące wzory:

a) **Wariancja** w szeregu prostym:

$$s^2 = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2$$

b) **Wariancja** – wyznacza się w szeregach rozdzielczych punktowych i w szeregach rozdzielczych z przedziałami klasowymi wg wzoru:

o dla szeregu rozdzielczego punktowego:

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{N}; \quad s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot w_i}{1}$$

o dla szeregu rozdzielczego z przedziałami klasowymi:

$$s^2 = \frac{\sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot n_i}{N}; \quad s^2 = \frac{\sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot w_i}{1}$$

Wariancję traktujemy jako miarę umożliwiającą obliczenie innej bardziej ważnej miary statystycznej, czyli odchylenia standardowego

Odchylenie standardowe – to pierwiastek kwadratowy z sumy kwadratów odchyłeń poszczególnych wartości cechy x_i od średniej arytmetycznej zbiorowości. Jest to pierwiastek z wariancji. Odchylenie standardowe wyznaczany ze wzorów:

a) **Odchylenie standardowe** w szeregu prostym:

$$s = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2}$$

b) **Odchylenie standardowe** – wyznacza się w szeregach rozdzielczych punktowych i w szeregach rozdzielczych z przedziałami klasowymi wg wzoru:

o dla szeregu rozdzielczego punktowego:



$$s = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i}{N}}; \quad s = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot w_i}{1}}$$

- o dla szeregu rozdzielczego z przedziałami klasowymi:

$$s = \sqrt{\frac{\sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot n_i}{N}}; \quad s = \sqrt{\frac{\sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot w_i}{1}}$$

Odchylenie standardowe określa, o ile wszystkie jednostki danej zbiorowości różnią się przeciętnie od średniej arytmetycznej badanej cechy. Obok średniej arytmetycznej najczęściej stosowanym parametrem statystycznym. Jest to wielkość obliczana na podstawie wszystkich obserwacji danego szeregu. Im zbiorowość jest bardziej zróżnicowana, tym większe jest odchylenie standardowe. Dzięki odchyleniu standardowemu możemy wyznaczyć typowy obszarze zmienności mieści się około 67% wszystkich jednostek badanej zbiorowości.

Typowy obszar zmienności – możemy wyznaczyć go za pomocą odchylenia standardowego według wzoru:

$$\bar{x} - s < x_{typ} < \bar{x} + s$$

Reguła trzech sigm:

Z odchyleniem standardowym wiąże się reguła trzech sigm, która mówi o tym, że:

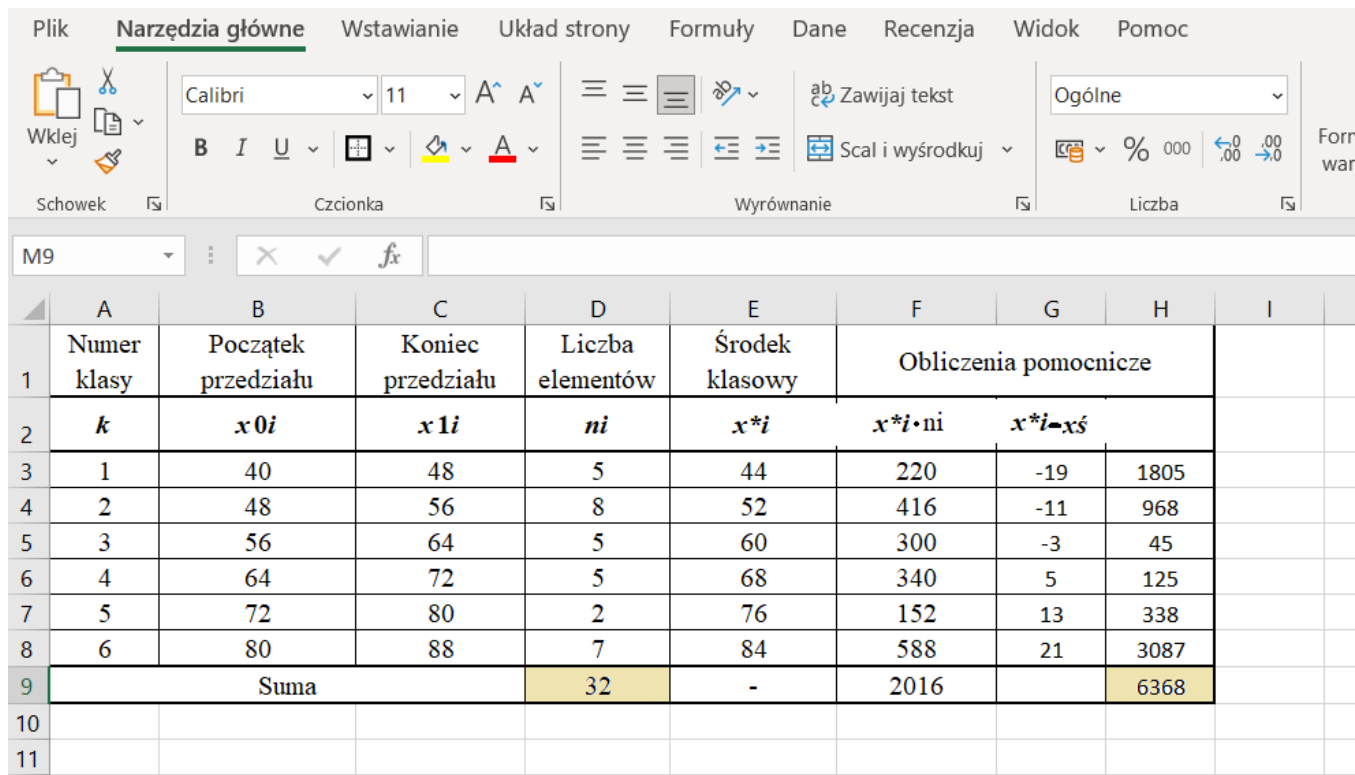
$(\bar{x} - s < x_{typ} < \bar{x} + s)$ mieści się około 67 % wszystkich jednostek badanej zbiorowości;

$(\bar{x} - 2s < x_{typ} < \bar{x} + 2s)$ mieści się około 95 % wszystkich jednostek badanej zbiorowości;

$(\bar{x} - 3s < x_{typ} < \bar{x} + 3s)$ mieści się około 99,7 % wszystkich jednostek badanej zbiorowości.

Wyznamy odchylenie standardowe dla przykładu drugiego:

Tabela 13. Wariancja w przykładzie 2



	A	B	C	D	E	F			G	H	I
1	Numer klasy	Początek przedziału	Koniec przedziału	Liczba elementów	Środek klasowy	Obliczenia pomocnicze					
2	k	x_{0i}	x_{1i}	n_i	x^*_{i}	$x^*_{i} \cdot n_i$	$x^*_{i} - x^s$				
3	1	40	48	5	44	220	-19		1805		
4	2	48	56	8	52	416	-11		968		
5	3	56	64	5	60	300	-3		45		
6	4	64	72	5	68	340	5		125		
7	5	72	80	2	76	152	13		338		
8	6	80	88	7	84	588	21		3087		
9	Suma			32	-	2016			6368		
10											
11											

$$S = \sqrt{\frac{6368}{32}} = 14$$

Wykorzystując regułę trzech sigm możemy powiedzieć, że 67% zwierząt miała wagę od 49 do 77 kilogramów.

Odchylenie ćwiartkowe – nazywamy inaczej odchyleniem między **kwartylowym**. Jest to połowa różnicy między kwartylem pierwszym i trzecim:

$$Q = \frac{Q_3 - Q_1}{2}$$

Uwaga: Pomiędzy odchyleniami ćwiartkowym, standardowym i przeciętnym obliczonymi dla tego samego szeregu zachodzi relacja:

$$Q < s < d.$$

4.4.4. Współczynniki zmienności

Współczynnik zmienności – to iloraz bezwzględnej miary zmienności cechy i średniej wartości cechy. Współczynniki zmienności wyznaczamy wzorów:

$$V_s = \frac{s}{x} \cdot 100\%$$



$$V_d = \frac{d}{x} \cdot 100\%$$

$$V_Q = \frac{Q}{Me} \cdot 100\% ;$$

$$V_{Q_1, Q_2} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \cdot 100\%$$

Celem wyznaczenia współczynnika jest stwierdzenie ile procent elementów ma wartości typowe. Jeżeli spojrzymy na przed momentem policzone dane do przykładu 2 to

$$V_s = \frac{s}{\bar{x}} \cdot 100\% = 22,2\%$$

Czyli 22,2% zwierząt ma wagę przeciętną.

4.4.5. Miary zmienności

Z punktu widzenia potrzeb analizy statystycznej ważny jest nie tylko przeciętny poziom i wewnętrzne zróżnicowanie, ale również to, czy przeważająca liczba jednostek znajduje się powyżej, czy poniżej przeciętnego poziomu badanej cechy. Rozkłady jednej zmiennej różnią się między sobą kierunkiem i siłą asymetrii. Poziom rozkładu najłatwiej jest określić poprzez porównanie dominanty, mediany i średniej arytmetycznej.

Wskaźnik skośności (asymetrii) – to bezwzględna miara asymetrii posiadająca miano badanej cechy. Wskazuje on kierunek asymetrii. Wyznacza się za pomocą następujących wzorów:

$$W_{sk} = \bar{x} - D$$

$$W_{sk} = (Q_3 - Me) - (Me - Q_1)$$

Uwaga:

- Wskaźnik skośności używany jest do porównywania asymetrii w zbiorowościach, w których wartość zmiennej jest wyrażona w różnych jednostkach miary;
- Wskaźnik skośności określa jedynie kierunek asymetrii, nie wykazuje jej siły, gdyż jest wielkością nienormowaną.

Współczynnik asymetrii (skośności) służy do określania siły i kierunku asymetrii. Współczynnik skośności obliczamy za pomocą wzorów:

mierzalne: $A_s = \frac{\bar{x} - D}{s} ; \quad A_d = \frac{\bar{x} - D}{d}$

klasyczny: $A^k = \frac{\mu_3}{s^3}$



$$A_Q = \frac{Q_3 + Q_1 - 2Me}{2Q}$$

pozycyjny:

Współczynniki skośności zawsze są z przedziału $\langle -1, 1 \rangle$, a przy skrajnej asymetrii może przekroczyć -1 , lub 1 . Współczynnik asymetrii to miara niemianowana, ale unormowana, co umożliwia na porównywanie asymetrii różnych rozkładów; Współczynniki mierzalne traktujemy jako wykluczające się, gdyż odpowiadają na podobne pytania; Współczynnik pozycyjny należy traktować jako miarę uzupełniającą, ponieważ określa kierunek i siłę asymetrii jednostek znajdujących się w drugiej i trzeciej ćwiartce obszaru zmienności; Współczynnik pozycyjny wykorzystujemy zwykle, gdy rozkład empiryczny nie spełnia warunków niezbędnych do obliczenia dominanty i średniej arytmetycznej.

Wykorzystując otrzymane dane w naszych zadaniach mogę wyznaczyć współczynnik dla przykładu 2

$$A_s = \frac{\bar{x} - D}{s} = \frac{63 - 52}{14} = 0,79.$$

Mamy odczynienia z asymetrią dużą prawostronna, co oznacza, że częściej występuje waga zwierząt poniżej 63 kg, niż powyżej.

Bibliografia



Bibliografia



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej



Bibliografia

1. J. Józwiak, J. Podgórski: Statystyka od podstaw, PWE, Warszawa 1992
2. H. Kassyk-Rokicka: Statystyka nie jest trudna, Część 1: Mierniki statystyczne, PWE, Warszawa, 1994
3. H. Kassyk-Rokicka: Statystyka. Zbiór zadań
4. S. Stasiewicz, Z. Rusnak, U. Siedlecka, Statystyka, Elementy teorii i zadania, Wydawnictwo AE, Wrocław 1995
5. M. Sobczyk: Statystyka, PWN, Warszawa 1995
6. J. Łomnicki „Wprowadzenie do statystyki dla przyrodników” PWN Warszawa 2003

**Narodowe Forum
Doradztwa Kariery**



Nasze publikacje
pozostałe publikacje serii
przygotowane w partnerstwie projektu



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej




Zapraszamy do lektury naszych kolejnych publikacji z serii: **Monika Zińczuk**: Rozwój Kompetencji Kluczowych (osobistych, społecznych i w zakresie umiejętności uczenia się) w trosce o minimalizację niepowodzeń szkolnych uczniów; **Alina Stankiewicz**: Kompetencje kluczowe w nauczaniu - uczeniu się biologii; **Karolina Nikołajuk**: Innowacyjne metody wspierające lekcje chemii w szkole ponadpodstawowej; **Barbara Dudel, Izabela Sietejko**: Metoda WebQuest w procesie rozwijania kompetencji kluczowych uczniów klasy trzeciej szkoły podstawowej; **Brygida Kaliszewicz**: Zrozumieć i pomóc - wspieranie uczniów z zaburzeniami przetwarzania sensorycznego w nabywaniu kompetencji kluczowych; **Marta Małaszkiwicz**: Mnemotechniki i metody aktywizujące wspomagające zapamiętywanie słownictwa poprzez kojarzenie i wizualizowanie z włączeniem emocji, zmysłów i ruchu oraz twórczego działania; **Dorota Popławska**: Grafika komputerowa i animacja w szkole ponadpodstawowej






Barbara Dudel
Izabela Sietejko



Metoda WebQuest
w procesie rozwijania kompetencji kluczowych
uczniów klasy trzeciej szkoły podstawowej



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej

Brygida Kaliszewicz



Zrozumieć i pomóc
- wspieranie uczniów z zaburzeniami przetwarzania sensorycznego
w nabywaniu kompetencji kluczowych



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej

Marta Małaszkiwicz



Mnemotechniki i metody aktywizujące
wspomagające zapamiętywanie słownictwa
poprzez kojarzenie i wizualizowanie z włączeniem emocji,
zmysłów i ruchu oraz twórczego działania



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej

Dorota Poptawska



Grafika komputerowa i animacja
w szkole ponadpodstawowej



białostocka szkoła ćwiczeń
kompetencje kluczowe w praktyce szkolnej