

RAPORT Z LABORATORIUM NR.2

Porządkowanie danych w szereg rozdzielczy

Łukasz Bajkowski
150048

Zadanie 1

Celem zadania 1 było wczytanie z pliku dane.xlsx danych dotyczące wysokości i masy dwunastolatków (wiek 12.00-12.99)

Zadanie 2

Celem zadania 2 było wyznaczenie współczynnika Rohrera z danych uzyskanych w zadaniu 1. Wzór na uzyskanie tych danych to:

$$\text{Rohrer} = \frac{\text{masa} \cdot 100000}{\text{wys}^3}$$

Zadanie 3

W przypadku współczynnika Rohrera, który jest wynikiem dzielenia masy przez trzecią potęgę wzrostu, można uznać go za zmienną ciągłą, ponieważ teoretycznie może przyjąć nieskończenie wiele różnych wartości.

Co do skali pomiarowej, w przypadku współczynnika Rohrera, jest to zmienna ilorazowa, ponieważ ma sensowny punkt zerowy (gdy masy lub wzrostu są równe zero). Skala ilorazowa umożliwia mierzenie proporcji między wartościami, co jest zgodne z definicją współczynnika Rohrera.

Zadanie 4

W pierwszej kolejności, żeby móc wyliczyć x_{01} oraz h trzeba było obliczyć podstawowe wartości, które umożliwiły dalszą pracę, a mianowicie:

- Minimum = 0.9092030935624626
- Maximum = 1.9099545905019033
- Liczbę danych = 132
- Rozstęp = 1.0007514969394407

Następnie, w trakcie dalszych obliczeń, ustalono dwie istotne wartości robocze: x_{01} i h . Te parametry odegrały kluczową rolę w procesie podejmowania decyzji dotyczących ostatecznych wartości, które będą miały istotne znaczenie w kolejnych etapach pracy. Końcowe wyniki obliczeń dla tych wartości prezentowały się w sposób następujący:

- Wyliczone $x_{01} = 0.8656509748671193$
- Wyliczone $h = 0.08710423739068678$

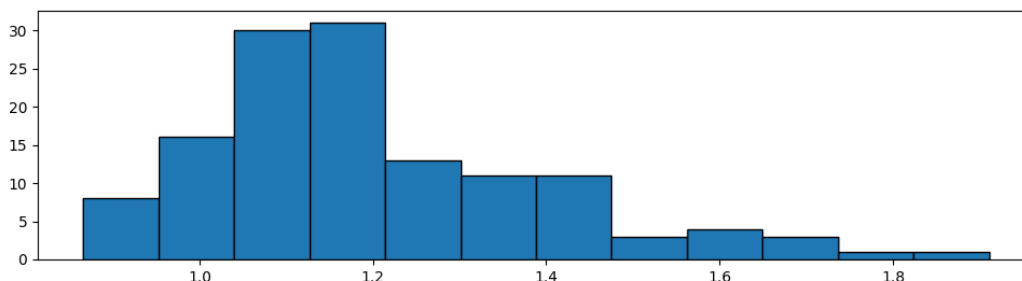
Dzięki wartościom roboczym udało się podjąć ostateczne decyzje dotyczące parametrów x_{01} oraz h . W ramach badania, te wartości zostały ustawione na:

- $x_{01} = 0.87$
- $h = 0.09$

Mając już ustalone wartości x_{01} oraz h , możliwe było przejście do realizacji kolejnych podpunktów, które były zaplanowane w ramach tego zadania.

A. Dla wybranych parametrów wyznacz szereg skumulowany, prawdopodobieństwo empiryczne i dystrybuantę empiryczną.

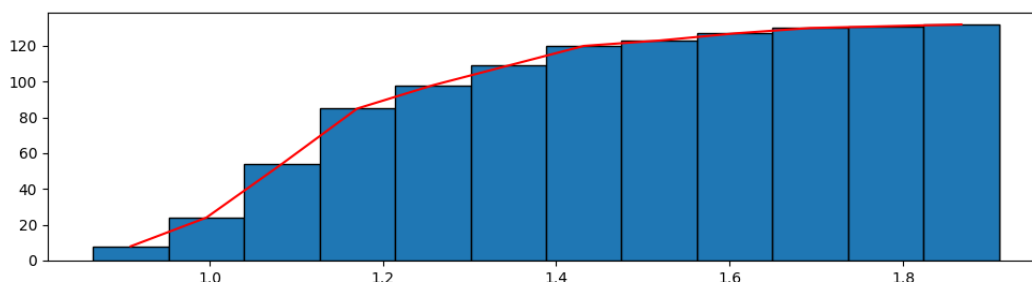
Dla wcześniej ustalonych parametrów x_{01} oraz h , otrzymano następujące wartości dla szeregu skumulowanego, prawdopodobieństwa empirycznego oraz dystrybuanty empirycznej.



RYSUNEK 1 HISTOGRAM SZEREGU

Z powyższego histogramu możemy odczytać, że najmniej dwunastolatków ma wartość Rohrera wynoszącą około 1.8 - 1.9 a najwięcej dwunastolatków ma wartość Rohrera w granicach 1 - 1.2.

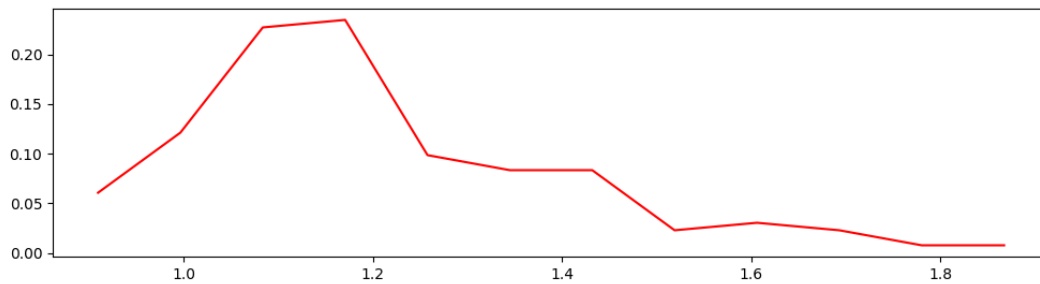
Szereg skumulowany: [8, 24, 54, 85, 98, 109, 120, 123, 127, 130, 131, 132]



RYSUNEK 2 HISTOGRAM DLA SZEREGU SKUMULOWANEGO

Histogram dla szeregu skumulowanego obrazuje liczbę obserwacji w poszczególnych klasach, włączając w to kategorie poprzednie. Ostatnia wartość w szeregu skumulowanym liczb obserwacji jest równa całkowitej liczbie obserwacji.

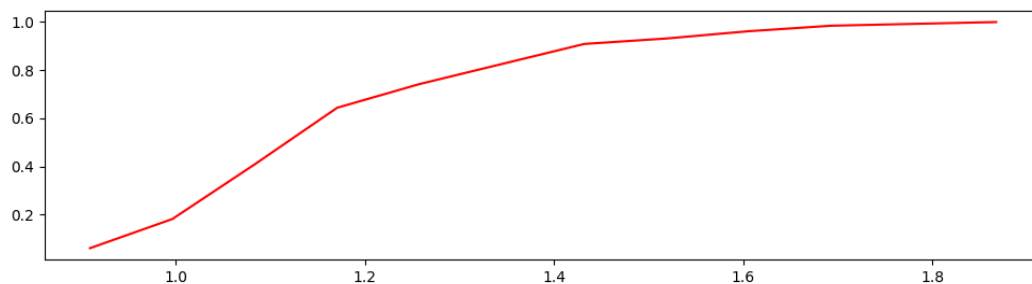
Prawdopodobieństwa empiryczne: [0.06060606060606061, 0.12121212121212122, 0.22727272727272727, 0.23484848484848486, 0.09848484848484848, 0.08333333333333333, 0.08333333333333333, 0.022727272727272728, 0.030303030303030304, 0.022727272727272728, 0.007575757575757576, 0.007575757575757576]



RYSUNEK 3 WYKRES DLA PRAWDOPODOBIENSTW EMPIRYCZNYCH

Analizując wykres dla prawdopodobieństw empirycznych, łatwo zauważyć, że największe prawdopodobieństwa występują w okolicy wartości 1.5.

Dystrybuanta empiryczna: [0.06060606060606061, 0.18181818181818182, 0.4090909090909091, 0.6439393939393939, 0.7424242424242424, 0.8257575757575758, 0.9090909090909091, 0.9318181818181818, 0.9621212121212122, 0.9848484848484849, 0.9924242424242424, 1.0]



RYSUNEK 4 WYKRES DLA DYSTRYBUANT EMPIRYCZNYCH

Obserwując wykres dla dystrybuant empirycznych, można zauważyć, że jest on zbliżony z wykresem dla szeregu skumulowanego i dąży do pełnej próby.

TABELA 1 TABELA WARTOŚCI DLA ŚREDNICH I ODCHYLEŃ STANDARDOWYCH

Średnie	
Średnia z próby	1.2056442951886397
Średnia z szeregu	1.2015301932902978
Błąd oszacowania średniej	0.003412367905492581
Odchylenie standardowe	
Odchylenie standardowe z próby	0.1935719832030477
Odchylenie standardowe z szeregu	0.19686462067619734
Błąd oszacowania odchylenia	0.01700988654797129

Na podstawie przedstawionych informacji można jednoznacznie stwierdzić, że wartości x_{01} oraz h zostały odpowiednio dobrane. Oba błędy oszacowań, zarówno średniej, jak i odchylenia standardowego, są bardzo małe, co świadczy o trafnym doborze wartości x_{01} i h .

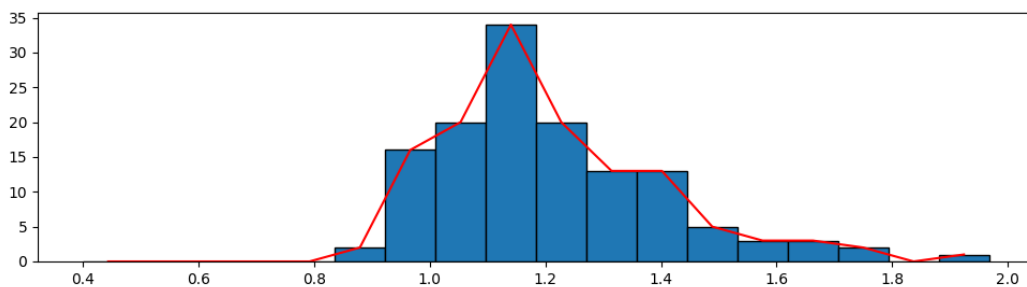
Zadanie 5

Wykonaj polecenia A. - D. dla dwóch nowych szeregów otrzymanych następująco:

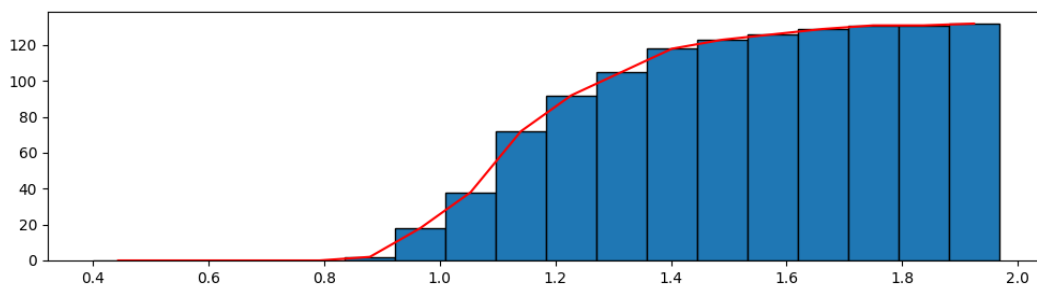
A. Szereg 1 – h zostaje bez zmian, zmieniamy x_{01}

B. Szereg 2 - x_{01} zostaje bez zmian, zmieniamy h

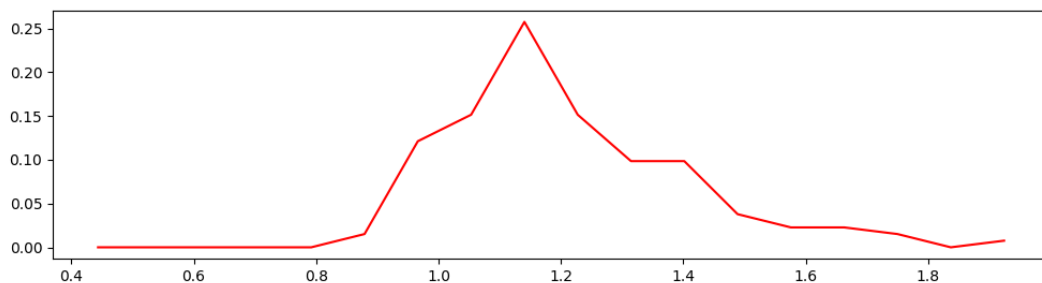
A. Szereg 1 – h zostaje bez zmian, zmieniamy x_{01} na 0.4



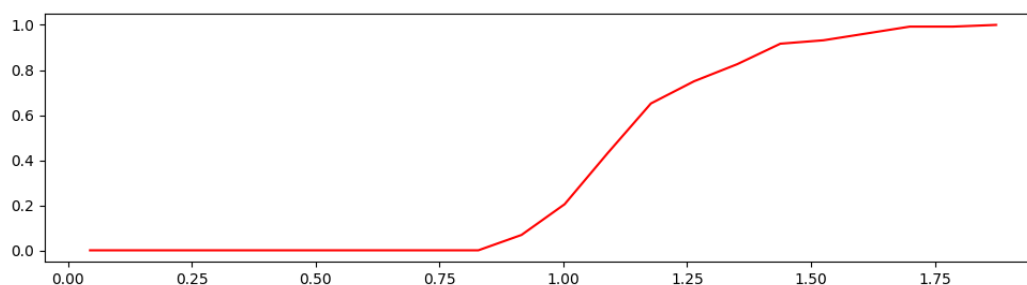
RYSUNEK 5 HISTOGRAM SZEREGU DLA ZMIENIONEGO x_{01}



RYSUNEK 6 HISTOGRAM SZEREGU SKUMULOWANEGO DLA ZMIENIONEGO x_{01}



RYSUNEK 7 WYKRES DLA PRAWDOPODOBIENSTW EMPIRYCZNYCH DLA ZMIENIONEGO x_{01}



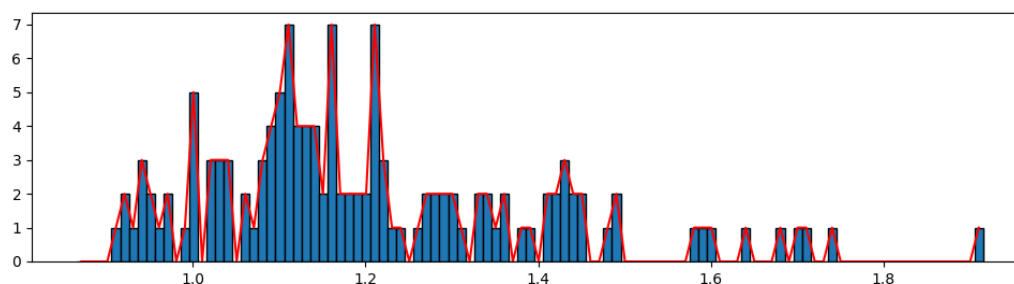
RYSUNEK 8 WYKRES DLA DYSTRYBUANT EMPIRYCZNYCH DLA ZMIENIONEGO x_{01}

TABELA 2 TABELA WARTOŚCI DLA ŚREDNICH I ODCHYLEŃ STANDARDOWYCH DLA ZMIENIONEGO x_{01}

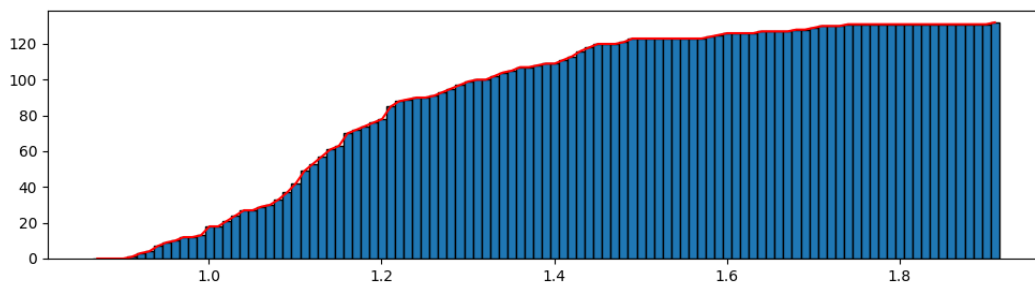
Średnie	
Średnia z próby	1.2056442951886397
Średnia z szeregu	1.2083537182090247
Błąd oszacowania średniej	0.0022472822466771502
Odchylenie standardowe	
Odchylenie standardowe z próby	0.1935719832030477
Odchylenie standardowe z szeregu	0.19686462067619734
Błąd oszacowania odchylenia	0.01700988654797129

Z analizy powyższych wykresów wynika, że zmiana wartości x_{01} na 0.4 spowodowała zmiany w wykresach, które lepiej odzwierciedlają wszystkie dane z próby. Dodatkowo, wykresy wykazują teraz największą liczbę danych w ich centrum. Z tabeli, zawierającej średnie i odchylenia standardowe, wynika, że błędy oszacowań są bardzo małe, co wskazuje na trafność wyboru wartości x_{01} i korzystnie wpływa na dokładność analizy danych.

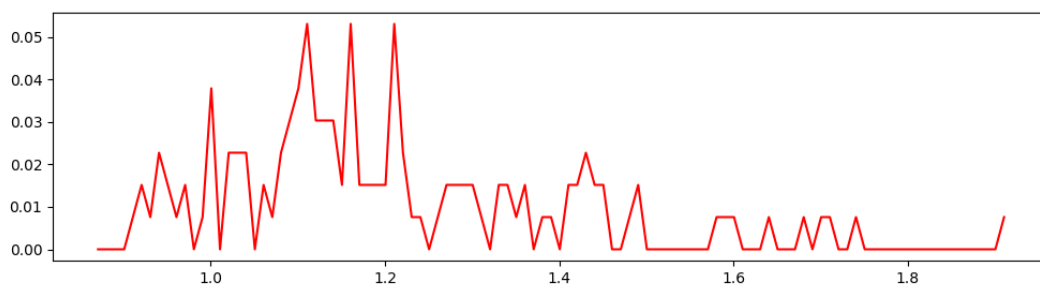
B. Szereg 2 – x_{01} zostaje bez zmian, zmieniamy h na 0.01



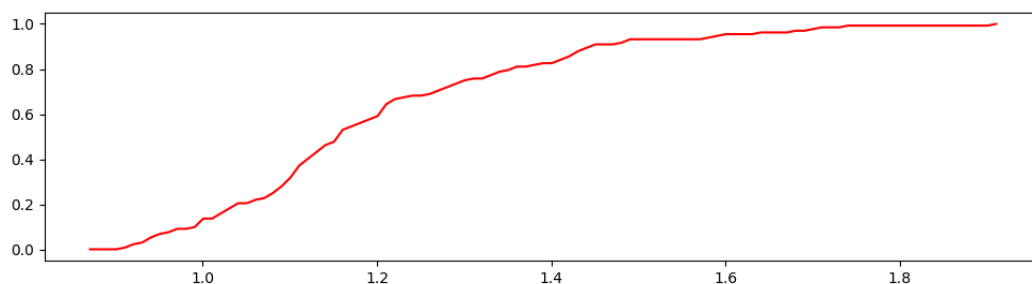
RYSUNEK 9 HISTOGRAM SZEREGU DLA ZMIENIONEGO h



RYSUNEK 10 HISTOGRAM SZEREGU SKUMULOWANEGO DLA ZMIENIONEGO h



RYSUNEK 11 WYKRES DLA PRAWDOPODOBIENSTW EMPIRYCZNYCH DLA ZMIENIONEGO h



RYSUNEK 12 WYKRES DLA DYSTRYBUANT EMPIRYCZNYCH DLA ZMIENIONEGO h

TABELA 3 TABELA WARTOŚCI DLA ŚREDNICH I ODCHYLEŃ STANDARDOWYCH DLA ZMIENIONEGO h

Średnie	
Średnia z próby	1.2056442951886397
Średnia z szeregu	1.2055752172913616
Błąd oszacowania średniej	5.729542084155919e-05
Odchylenie standardowe	
Odchylenie standardowe z próby	0.1935719832030477
Odchylenie standardowe z szeregu	0.19365893250655306
Błąd oszacowania odchylenia	0.0004491833067296569

Z analizy powyższych wykresów wynika, że zmiana wartości na 0.01 spowodowała modyfikacje w wykresach, które prezentują dane w inny sposób, zdecydowanie zwiększając liczbę kolumn. Z tabeli, zawierającej średnie i odchylenia standardowe, można odczytać, że błędy oszacowań nie są duże w porównaniu do przypadku, gdy zmieniliśmy x_{01} .

Zadanie 6

Podsumowując otrzymane wyniki obliczeń można dojść do kilku wniosków, a konkretnie:

1. Wybrane wartości x_{01} oraz h , wynikające z analizy teoretycznej i wartości roboczych, prezentują adekwatną analizę danych wejściowych.
2. Zwiększenie x_{01} skutkować będzie wysokimi błędami oszacowań średniej i odchylenia standardowego, a także nieprawidłowymi wykresami, które nieadekwatnie odzwierciedlą analizę danych wejściowych. Natomiast zmniejszenie x_{01} spowoduje redukcję błędów oszacowań oraz dokładniejsze przedstawienie danych na wykresach, centralizując je.
3. Zwiększenie wartości h skutkować będzie nieprawidłowymi wykresami, jednak błędy oszacowań średniej nie będą duże. Natomiast zmniejszenie h sprawi, że wykresy precyzyjniej odzwierciedlą dane, a wykres dystrybuanty empirycznej jest bardziej precyzyjny, posiadając więcej punktów.
- 4.

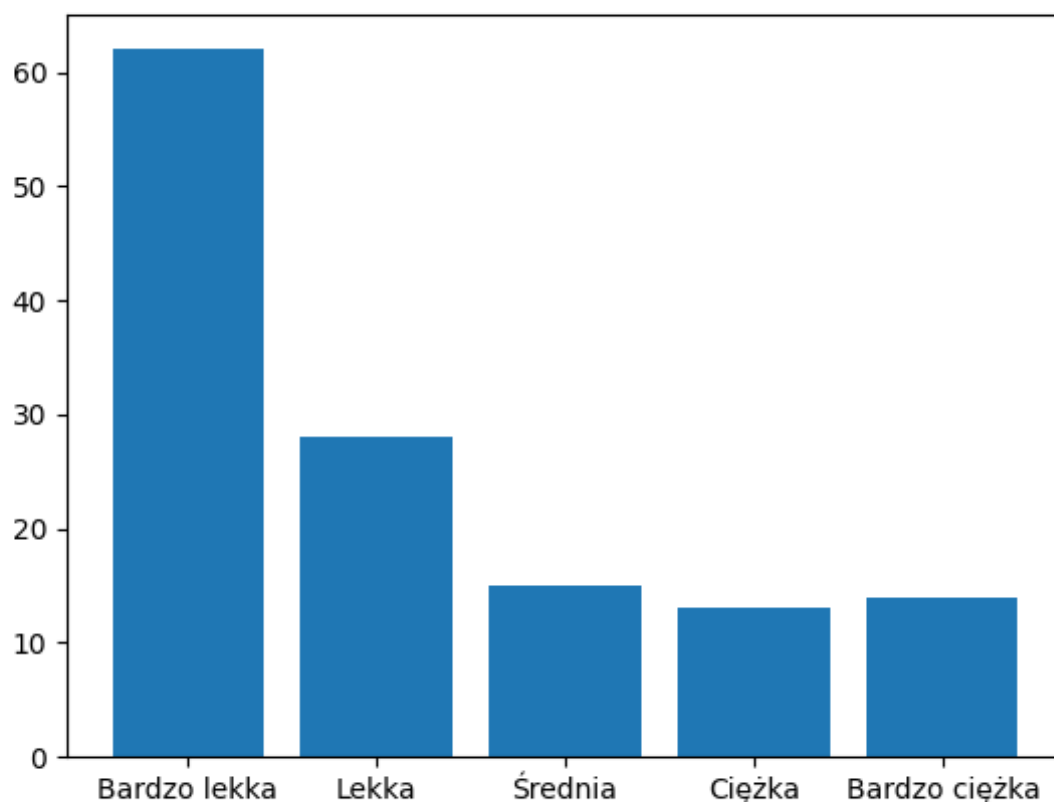
Zadanie 7

Lundman, w swojej klasyfikacji dotyczącej budowy ciała, wprowadził pięć kategorii opisujących masę ciała w stosunku do wzrostu danej osoby. Poniżej przedstawiam zakresy wag dla poszczególnych kategorii:

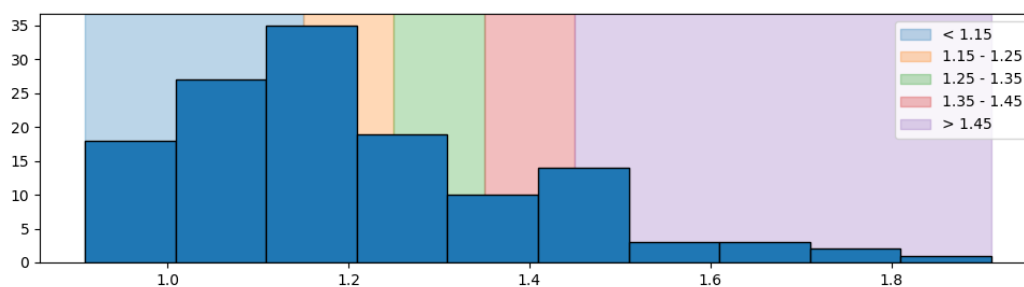
1. Budowa bardzo lekka ($x-1.15$): W tej kategorii znajdują się osoby o bardzo lekkiej budowie ciała w porównaniu do ich wzrostu. Współczynnik masa-wzrost wynosi mniej niż 1.15.
2. Budowa lekka (1.15-1.25): Osoby o budowie lekkiej charakteryzują się stosunkowo niższą masą ciała w porównaniu do wzrostu. Współczynnik masa-wzrost mieści się w zakresie od 1.15 do 1.25.
3. Budowa średnia (1.25-1.35): W tej kategorii mieszczą się osoby o budowie ciała uznawanej za średnią. Współczynnik masa-wzrost mieści się w zakresie od 1.25 do 1.35.
4. Budowa ciężka (1.35-1.45): Osoby o budowie ciężkiej charakteryzują się wyższą masą ciała w porównaniu do wzrostu. Współczynnik masa-wzrost mieści się w zakresie od 1.35 do 1.45.
5. Budowa bardzo ciężka (1.45- x): Ta kategoria obejmuje osoby o bardzo ciężkiej budowie ciała w porównaniu do wzrostu. Współczynnik masa-wzrost wynosi więcej niż 1.45.

TABELA 3 TABELA ILOŚCI DWUNASTOLATKÓW W PRZEDZIAŁACH BUDOWY CIAŁA W WEDŁUG KLASYFIKACJI LUNDMANA

Dane dla dwunastolatków w według klasyfikacji Lundmana	
Budowa bardzo lekka	62
Budowa lekka	28
Budowa średnia	15
Budowa ciężka	13
Budowa bardzo ciężka	14



RYSUNEK 13 WYKRES ILOŚCI DWUNASTOLATKÓW WEDŁUG KLASYFIKACJI LUNDMAN



RYSUNEK 14 HISTOGRAM SZEREGU ROZDZIELCZEGO Z ZAKRESAMI KLASYFIKACJI LUNDMANA

Zadanie 8

Na podstawie podanych danych oraz klasyfikacji według Lundmana można wyciągnąć następujące wnioski:

- Wartości współczynnika Rohrera dla badanych obiektów mieszczą się w przedziale od 0.909 do 1.91, co oznacza, że badane obiekty mają zróżnicowaną budowę ciała.
- Współczynnik Rohrera dla większości badanych obiektów znajduje się w przedziale pomiędzy 1.0 a 1.5, co wskazuje na zróżnicowanie w budowie ciała, ale brak skrajnych wartości.
- Najczęściej występujące kategorie według klasyfikacji Lundmana dla badanych obiektów to "budowa bardzo lekka" i "budowa średnia".
- Wartości współczynnika Rohrera są zróżnicowane, co sugeruje, że badane obiekty mogą pochodzić z różnych populacji lub grup etnicznych, co może mieć wpływ na ich budowę ciała.