



UNIwersytet WArmińsko-MAzurski
W Olsztynie

Analiza wyników biegu po ósemce.

Górski Adam *****

Wydział Matematyki i Informatyki

Projektowanie Systemów Informatycznych i Sieci Komputerowych.

Matematyczne Modelowanie Systemów.

dr *****.

1. Cel badania

W ćwiczeniu numer dwa analizie zostanie poddany zbiór danych udostępniony na pierwszych zajęciach. Wszystkie obliczenia dotyczą parametru 8_t dla grupy wiekowej od 11,00 do 11,99. Do wykonywania obliczeń i prezentacji wyników wykorzystam pakiet Excel.

2. Krótki opis danych

Dane z kolumny 8_t dotyczą grupy wiekowej jedenastoletków. Po wstępnym zapoznaniu się z danymi, gdzie wartość minimalna to 14, maksymalna to 22 mogę przypuszczać, że dane te dotyczą czasu w pewnej konkurencji. Po zapoznaniu się z opisem moje przypuszczenia okazały się prawdziwe i dane z kolumny 8_t dotyczą biegu po „ósemce”. Wartości te zostaną poddane, dokładnej analizie oraz zostanie zbudowany szereg rozdzielczy obliczone prawdopodobieństwo empiryczne, dystrybuanta empiryczna oraz podstawowe cechy takie jak wartość minimalna, maksymalna średnia z szeregu i wiele innych istotnych obliczeń, które wymaga ten zbiór danych.

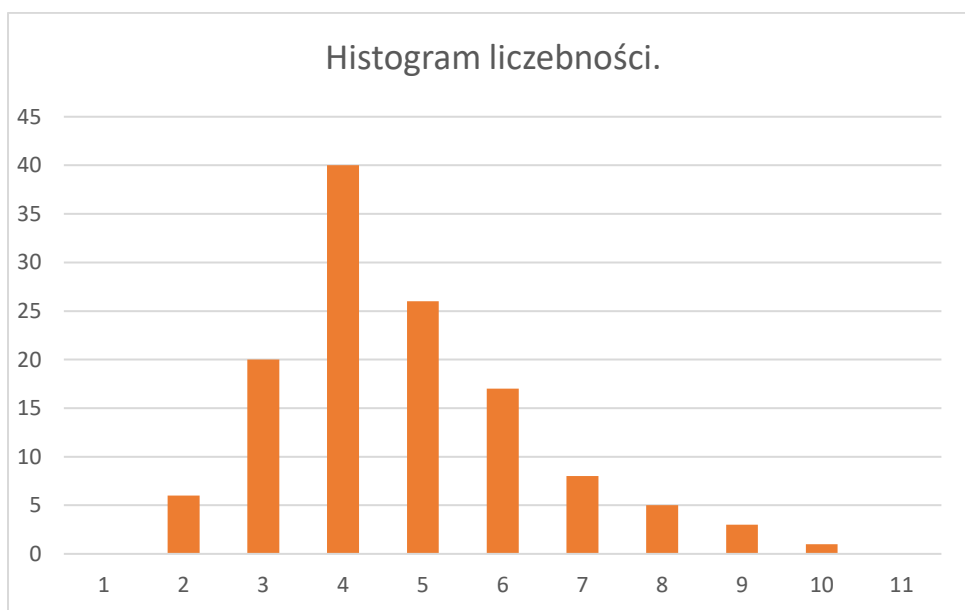
3. Wartości obliczeń wzorcowe

Liczebność	127
Minimalna	14,07
Maksymalna	22,29
Różnica między wartością minimalną, a maksymalną	8,22
Liczba klas	11
Różnica pomiędzy klasami	0,73
Wartość początkowa pierwszego przedziału	13,2
średnia z szeregu	16,16551
odchylenie z szeregu	1,187822
procent typowych wartości wyznaczonych z odchylenia i średniej	68%
współczynnik zmienności	0,073479
procent wartości typowych ² wyznaczonych z odchylenia i średniej	92%
liczba wartości odstających po lewej stronie	0
liczba wartości odstających po prawej stronie	7

W powyższej tabeli zostały przedstawione wybrane wyniki wartości uzyskanych w obliczeniach. Są to wybrane wartości, które uważam, że mogą pomóc w dalszej ocenie wartości danych wybranych do analizy.

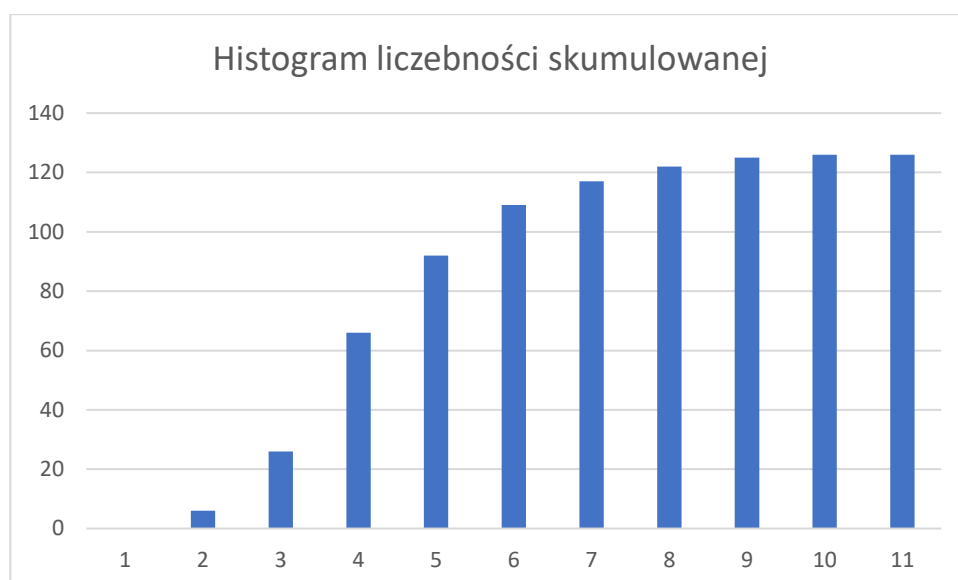
3.1 Wykresy

Akapit ten będzie zawierał szereg wykresów. Na początku zostaną przedstawione wykresy wzorcowe takie które w doskonały sposób przedstawiają wyniki oraz zostały wykonane w oparciu o najlepsze wartości wyjściowe. Następnie w punkcie 4 zmieniając wartości, najpierw „wartość początkową”, następnie „różnicę pomiędzy klasami” zbadam wpływ tych parametrów na uzyskane obliczenia których wynik zostanie przedstawiony na poniższych wykresach.

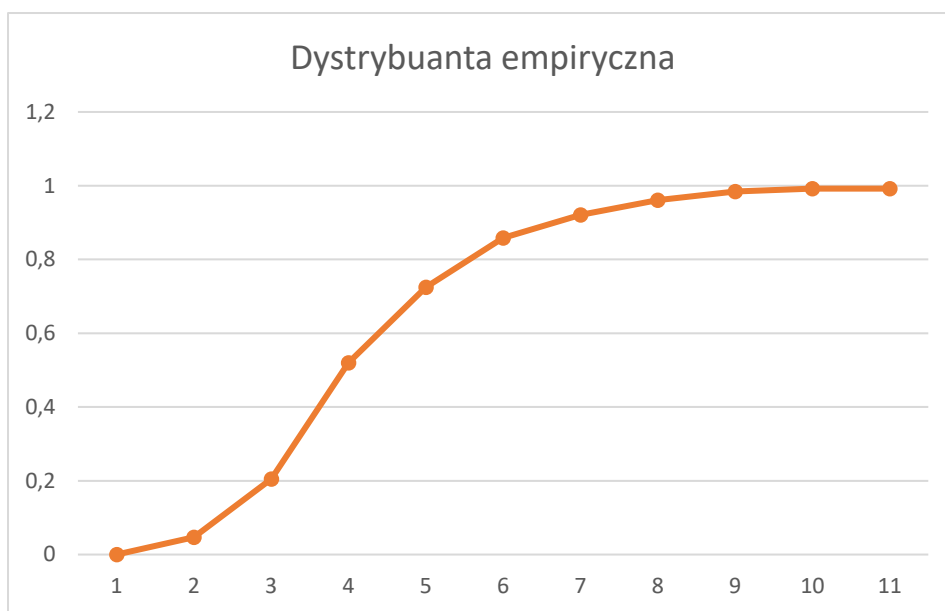


Histogram liczebności przedstawiony powyżej przybiera kształt podobny do wykresu rozkładu normalnego z asymetrią prawostronną. Asymetria prawostronna oznacza, że wartości odstające w dużej mierze będą znajdować się po prawej stronie wykresu. Przedział wartości typowych jeden wynoszący 68% i przedział wartości typowych dwa

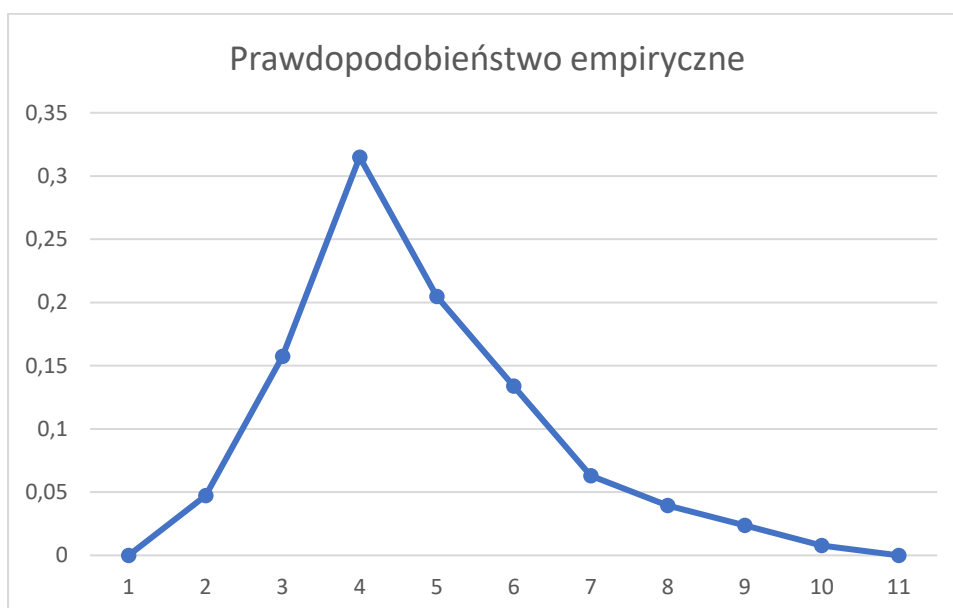
92% wskazuje, że pomiar wartości został przeprowadzony w sposób prawidłowy, a wyniki nie zostały sfałszowane. Średnia wartość wyników powinna być wyższa od mediany oraz dominanty.



Histogram liczebności skumulowanej bardzo dobrze pokazuje jaka jest różnica pomiędzy liczebnością poszczególnych klas, gdy widzimy spory skok wykresu słupkowego oznacza to, że dana klasa jest bardzo liczna. Im liczniejsza klasa tym skok większy, a dalej powinniśmy zauważyć stopniowe wypłaszczenie. Wykres ten bardzo przypomina dystrybuantę rozkładu normalnego. Myślę, że jest to dobre porównanie, ponieważ dystrybuantę definiujemy jako prawdopodobieństwo tego, że zmienna losowa X ma wartości mniejsze bądź równe x . Uważam, że powyższy wykres przedstawia to bardzo dobrze. Więc możemy przyjąć, że ostatnia kolumna to prawdopodobieństwo równe $= 1$ więc prawdopodobieństwo uzyskania takiego czasu lub mniejszego przez jedenastolatka będzie blisko 1.



Dystrybuanta empiryczna pod względem kształtu przypomina wykres wartości skumulowane co oczywiście nie jest bez znaczenia. Jak zauważyłem histogram liczebności skumulowanej bardzo przypomina dystrybuantę rozkładu normalnego tak samo jest w tym przypadku. Dystrybuanta ta, to prawdopodobieństwo tego, że znajdziemy się w poszczególniej grupie wyników w biegu po „ósemce”. Myślę, że gdybyśmy poprzedni wykres znormalizowali do wartości 0-1 to otrzymalibyśmy podobne lub identyczne wyniki.



Ostatnim analizowanym wykresem jest prawdopodobieństwo empiryczne. Jak można zauważyć przypomina on bardzo wykres pierwszy dotyczący liczebności. Jego kształt również jest asymetryczny. Jednak wskazuje na prawdopodobieństwo tego, że nasza wartość uzyskana znajdzie się w wybranym miejscu. Dodatkowo asymetria prawostronna może pokazywać, że więcej jest z danej grupy wiekowej nastolatków, którzy mają dobre wyniki w biegu na „ósemce” z czego wnioskujemy, że są oni wysportowani. W modelowaniu systemów, gdyby taki sam wynik uzyskać dla wyższej grupy wiekowej przełożyłoby się to na trend sportu co oznacza, że warto budować aplikacje pod branżę fitness, ponieważ może skutkować to dobrymi zyskami.

4. Badanie wpływu parametru x_{01} (wartość początkowa przedziału danych) oraz h (różnica pomiędzy klasami) na wyniki analizy danych

4.1 Zmiana wartości x_{01}

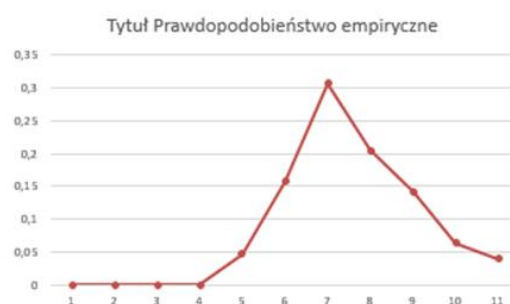
$h=$	0,73
$x_{01}=$	12.8
liczebność próby	127
średnia z szeregu	16,21701
odchylenie z szeregu	1,174673
procent typowych wartości wyznaczonych z odchylenia i średniej	61%
współczynnik zmienności	0,072435
procent wartości typowych ² wyznaczonych z odchylenia i średniej	89%



h=	0,73
x01=	12,2
liczebność próby	127
średnia z szeregu	16,23102
odchylenie z szeregu	1,194365
procent typowych wartości wyznaczonych z odchylenia i średniej	37%
współczynnik zmienności	0,073585
procent wartości typowych ² wyznaczonych z odchylenia i średniej	80%



h=	0,73
x01=	11
liczebność próby	127
średnia z szeregu	16,00862
odchylenie z szeregu	1,160614
procent typowych wartości wyznaczonych z odchylenia i średniej	5%
współczynnik zmienności	0,072499
procent wartości typowych2 wyznaczonych z odchylenia i średniej	51%



Porównanie zmiany wartości			
h=	0,73	0,73	0,73
x01=	12.8	12,2	11
liczebność próby	127	127	127
średnia z szeregu	16,21701	16,23102	16,00862
odchylenie z szeregu	1,174673	1,194365	1,160614
procent typowych wartości wyznaczonych z odchylenia i średniej	61%	37%	5%
współczynnik zmienności	0,072435	0,073585	0,072499
procent wartości typowych ² wyznaczonych z odchylenia i średniej	89%	80%	51%

4.2 Zmiana wartości h

h=	0,6
x01=	13,2
liczebność próby	127
średnia z szeregu	16,19528
odchylenie z szeregu	1,139313
procent typowych wartości wyznaczonych z odchylenia i średniej	53%
współczynnik zmienności	0,070348
procent wartości typowych ² wyznaczonych z odchylenia i średniej	83%



h=	0,5
x01=	13,2
liczebność próby	127
średnia z szeregu	15,86457
odchylenie z szeregu	1,131952
procent typowych wartości wyznaczonych z odchylenia i średniej	31%
współczynnik zmienności	0,071351
procent wartości typowych2 wyznaczonych z odchylenia i średniej	69%



h=	0,3
x01=	13,2
liczebność próby	127
średnia z szeregu	11,28071
odchylenie z szeregu	3,813261
procent typowych wartości wyznaczonych z odchylenia i średniej	5%
współczynnik zmienności	0,338034
procent wartości typowych ² wyznaczonych z odchylenia i średniej	16%



Porównanie zmiany wartości h			
h=	0,6	0,5	0,3
x01=	13,2	13,2	13,2
liczebność próby	127	127	127
średnia z szeregu	16,19528	15,86457	11,28071
odchylenie z szeregu	1,139313	1,131952	3,813261
procent typowych wartości wyznaczonych z odchylenia i średniej	53%	31%	5%
współczynnik zmienności	0,070348	0,071351	0,338034
procent wartości typowych2 wyznaczonych z odchylenia i średniej	83%	69%	16%

5. Podsumowanie punktu 4

W punkcie czwartym dokonałem analizy wpływu wartości h i x_0 na kształt wykresu, a także na wynik wykresu. Każda zmiana tych wartości sprawia, że wyniki które otrzymujemy przyjmują inną wartość i różnią się od pozostałych. Odpowiednio zmieniając wartość h i x_0 możemy uzyskać różny kształt wykresu, z którego czasami ciężko jest wysnuć nieprawidłowości, jeśli ta zmiana jest niewielka. Wyznacznikiem odpowiednio dobranej wartości h i x_0 jest wartość „typowe jeden” i „typowe dwa”. Kierowanie się zasadą „Trzech Sigm” w bardzo dobry sposób ułatwia pracę przy doborze odpowiednich wartości tak aby uzyskać idealne dane wyjściowe. Uważam, że dobra ocena wykresu jest kluczowa, ponieważ może odpowiednio pokazać nam jakie trendy, bądź jak zmienia się poziom sprawności fizycznej u ludzi młodych. Dzięki odpowiednim obliczeniom możemy prognozować, czy należy w przyszłości spodziewać się dużego zainteresowania sportem oraz czy społeczeństwo staje się bardziej aktywne i jaki jest poziom tej aktywności. Jeśli społeczeństwo dąży do zmiany diety być może należy wdrożyć do swej oferty nowy system dotyczący tego typu aktywności. Wykonywanie tego typu obliczeń może też być elementem systemu, w którym dokonujemy oceny naszego użytkownika. Wiemy, kiedy jego wynik jest w gronie najlepszych oraz kiedy musi się poprawić. Podsumowując odpowiednia analiza systemu może sprawić, że oszczędzimy czas, pieniądze i że nasz system będzie odpowiednio przygotowany do warunków panujących na rynku.

