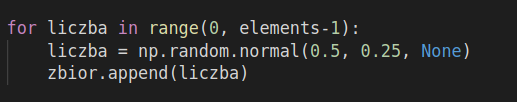
Maja Szymajda, 254313

Sprawozdanie lista 2

Do realizacji listy drugiej użyłam środowiska Python3, napisałam w nim program wykorzystujący następujące biblioteki: matplotlib.pyplot, scipy, statistics, numpy oraz math.

Zadanie 1

Do realizacji zadania 1 wykorzystałam funkcje wbudowana w bibliotekę numpy: np.random.normal(0.5, 0.25, None)

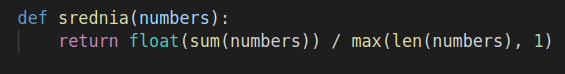
która umożliwiła mi wygenerowanie macierzy 20 i 100 elementowej wypełnionej liczbami z zakresu <0,1) o rozkładzie normalnym. Funkcja tworząca tablice, gdzie w pętli for tworzymy zbiór długości elements (ponieważ tablice numerujemy od 0 to for wykonuje się do elements-1) składający się z liczb wygenerowanych przez funkcje np.random.normal(0.5, 0.25, None).

Przykładowy zbiór 20 elementów:

[0.4585145217392386, 0.39256872999335274, 0.3736153710297714, 0.9398973208341186, 0.5730820765984174, 0.4989734873267324, 0.41683913634768455, 0.8528939715759631, 0.559334160616356, 0.6095519229881184, 0.20229217817076328, 0.8970700600893311, 0.5789265969552638, 0.7251707037005843, 0.7405491561533529, 0.5941585193464644, 0.7526217527154508, 0.7389721033857566, 0.8028997684340614]

Zadanie 2

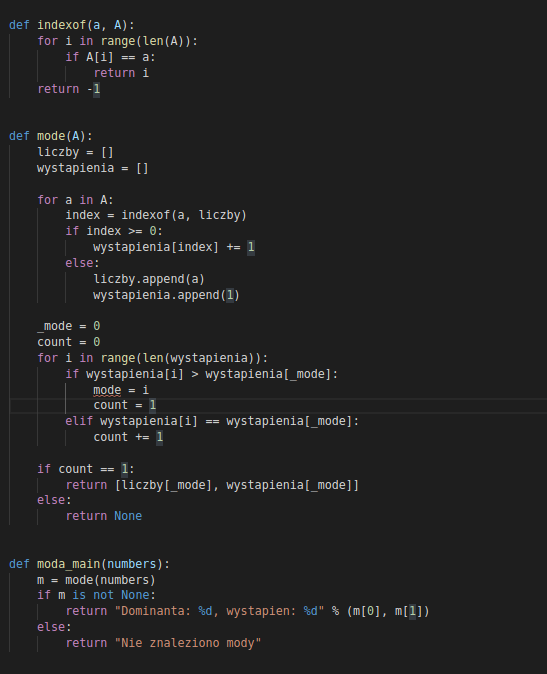
Aby wykonać drugie zadanie musiałam wyliczyć: średnią, medianę, modę, odchylenie standardowe, wariancję, skośność, kurtozę oraz co najmniej jeszcze jedną wybraną przez studenta – średnią harmoniczna.

 Średnia – pierwszą częścią zadania drugiego było wyliczenie średniej arytmetycznej ze zbioru 20 i 100 elementowego. Zrobiłam to za pomocą działania: float(sum(numbers)) / max(len(numbers), 1). Gdzie sum(numbers) to suma wszystkich wygenerowanych liczb a funkcja len(numbers) sprawdza długość tablicy, która później w funkcji max(len(numbers),1) ulega sprawdzeniu. Jeżeli jest mniejsza od 1 to algorytm wybierze większą z nich. Cała funkcja srednia wygląda następująco:

przyjmuje ona jako argument zbiór liczby wygenerowanych przez generator.

Aby obliczyć Medianę czyli wartość środkową można posortować wylosowane liczby i wziąć średnią z dwóch z nich które znajdują się w środku posortowanej listy lub wykorzystać funckje wybudowana w biblioteke numpy: np.median(zbior). Wyliczyła ona wartość środkowa dla zbioru 20 i 100 elementowego.

Do obliczenia mody wykorzystuje funkcje która sprawdza czy jakas liczba powtarza się więcej niż raz i ile razy. Wskazuje ona najczęściej powtarzającą się wartość w zbiorze. W z wziązku z tym ze

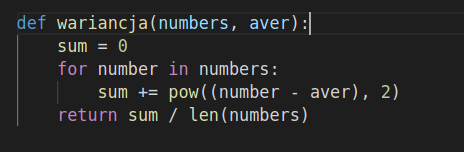
w losowanym przez nas przedziale jest olbrzymia ilość małych liczb to prawdopodobieństwo że jakaś się powtórzy jest bardzo małe. Kod do funkcji:

1. Funkcja indexofia(a,A) – sprawdza czy w zbiorze liczb powtarza się jedna liczba a. Argumentami: A - jest macierz z 20 i 100 elementami, natomiast a – jedna sprawdzana liczba.

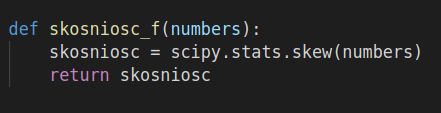
2. Funkcja mode(A) – jej zadaniem jest przejscie po tablicy i zliczenie wystąpień danych liczb oraz wybrania liczby z maksymalną ilością powtórzeń. Jej argumentem jest zbiór prób.

3. Funkcja moda\_main(numbers) – sprawdza czy funkcja mode(numbers) zwrociła liczbe inną niż 0, wtedy zwraca dominante, natomiast jeżeli zwrociła ona 0 to funkcja informuje o braku mody w danym zbiorze.

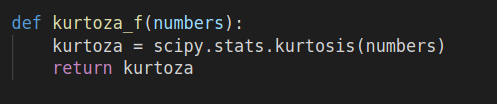
Odchylenie standardowe mówi nam o tym jak bardzo nasze losowe liczby są od siebie oddalone. Im mniejsze jest ono tym bliższe średniej są wartości. Do policzenia ochylenia standardowego użyłam funkcji: statistics.stdev(zbior), która policzyła ochylenie standardowe z podanego zbioru 20 lub 100 elementowego.

 Wariancja to inaczej średnia arytmetyczna kwadratów odchyleń od ich średniej arytmetycznej. Aby obliczyć ją dla obu macierzy 20 i 100 elementowych wykorzystałam funkcje którą nazwałam wariancja, której argumentami jest zbiór wygenerowanych elementów oraz policzona wczesniej średnia arytmetyczna.

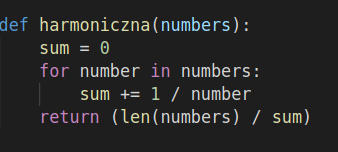
Funkcja w forze sumuje kwadrat różnicy między poszczególnym elementem a średnia arytmetyczna dla całego zbioru, a następnie całą sume dzieli przez ilość elementów.

 Skośność jest miarą symetrii rozkładu, jeżeli rozkład jest idealnie symetryczny wynosi ona 0. Dla rozkładów o lewostronnej asymetrii przyjmuje wartości ujemne a dla prawostronnej dodatnie. Aby obliczyć ją w pythonie wykorzystałam funkcje: scipy.stats.skew(numbers), która wyliczyła mi skośność dla wylosowanego zbioru.

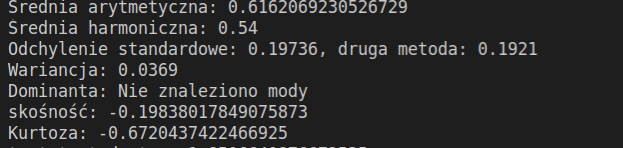
Kurtoza jest jedną z miar koncentracji wyników wokół średniej inaczej mówiąc spłaszczenia rozkładu wartości. Jej dodatnia wartość wskazuje na istnienie dużej ilości wartości bliskiej średniej, ujemna wartość kurtozy odpowiada rozproszonym wynikom wokół średniej. Aby ją obliczyc w Pythonie wykorzystałam funkcje scipy.stats.kurtosis(numbers), która znajduje się w bibliotece scipy.



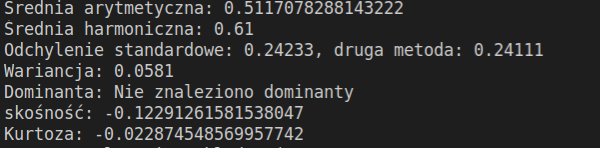
Jako wybraną przeze mnie funkcje obliczyłam średnia harmoniczna stanowi ona odwrotność średniej arytmetycznej odwrotności danych statystycznych. Aby ją obliczyć stworzyłam funkcje harmoniczna która przyjmuje jako argument zbiór wygenerowany w pierwszym zadaniu a następnie sumuje odwrotności tych liczb. By następnie zwrócić ilość liczb podzieloną na wczeniej obliczoną sume odwrotności.



Otrzymane wyniki dla grupy 20 prób:



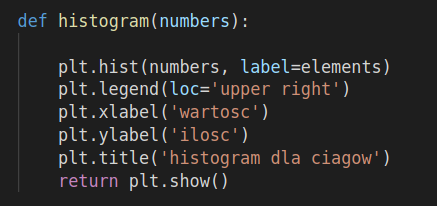
Wyniki dla grupy 100 prób:



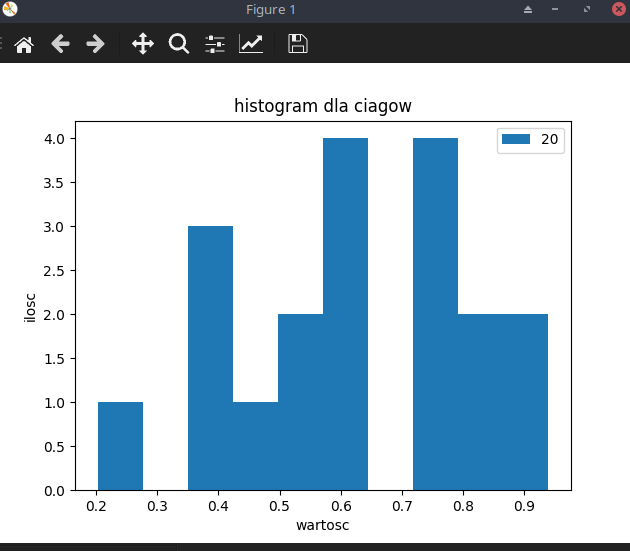
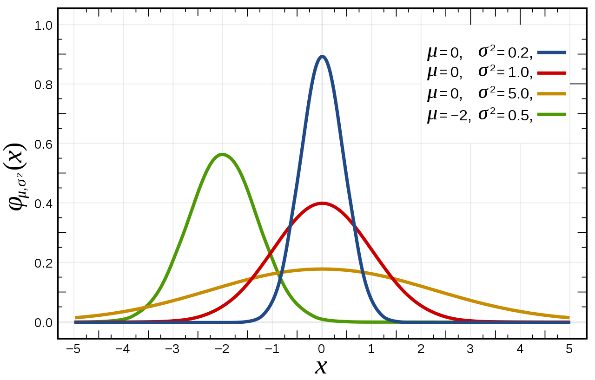
Zadanie 3

Aby wykonać zadanie 3 musiałam wygenerować histogramy dla obu zbiorów 20 i 100 elementowych a następnie porównać je do wykresu gęstości rozkładu normalnego.

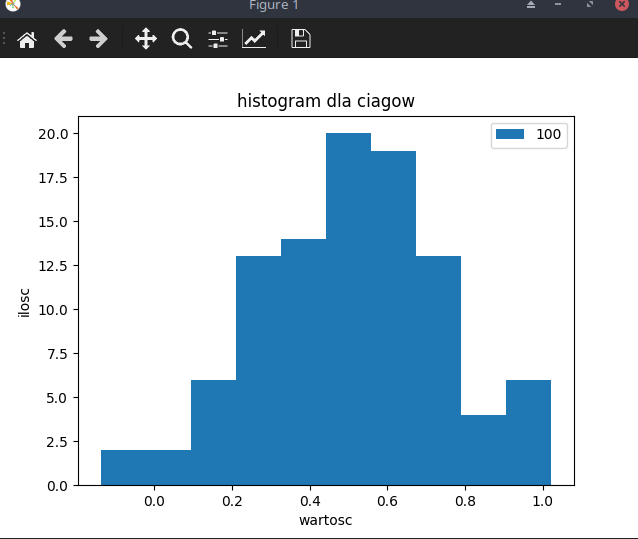
Wykonałam je za pomocą poniżej funkcji histogram:



histogram dla 20 wartosci:



histogram dla 100 wartosci:

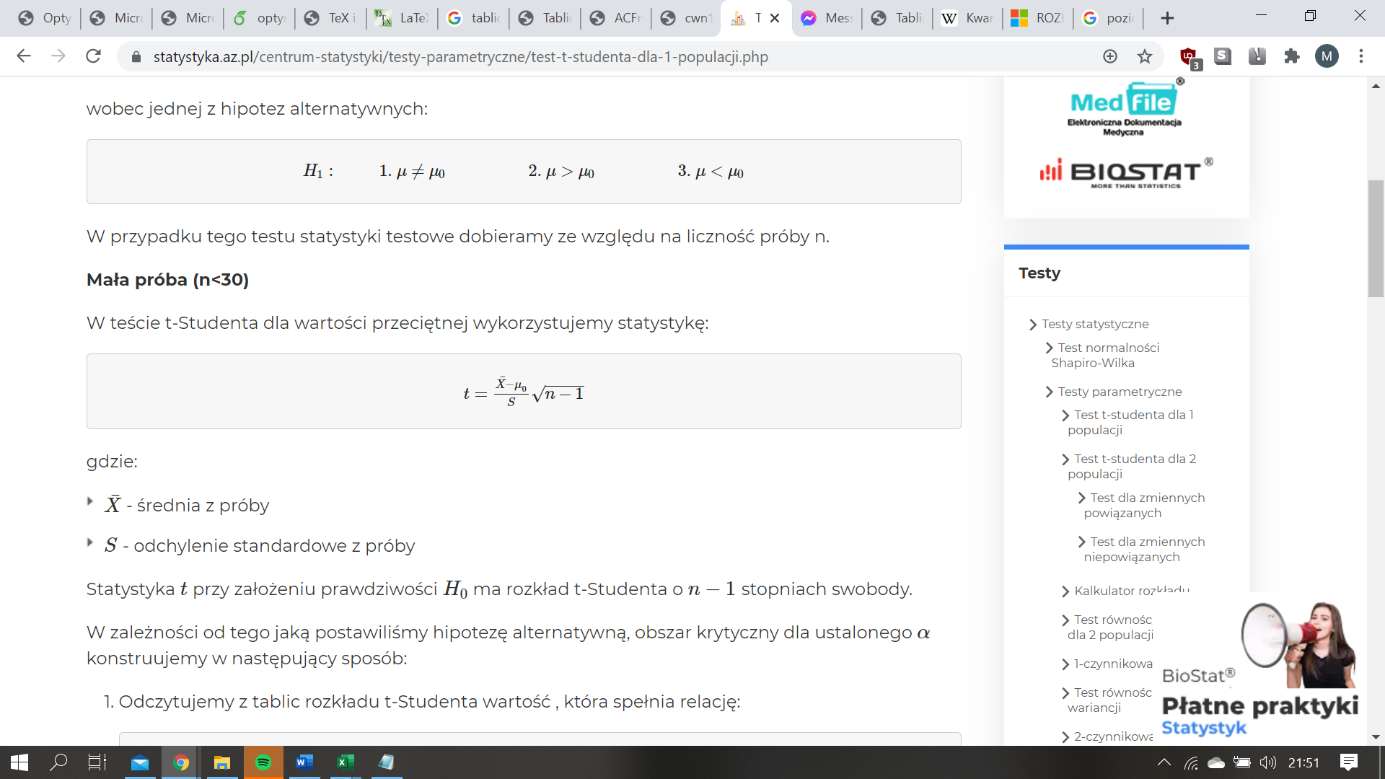


Jak widać wykres dla 20 wartości mocno odbiega od wykresu gęstości dla rozkładu normalnego. Spowodowane to jest małą ilością prób więc mogą wypaść dane ze skajnych obszarów.

Natomiast wykres dla 100 wartości powinien już kształtem przypominac wykres gęstości z powodu większej ilości prób.

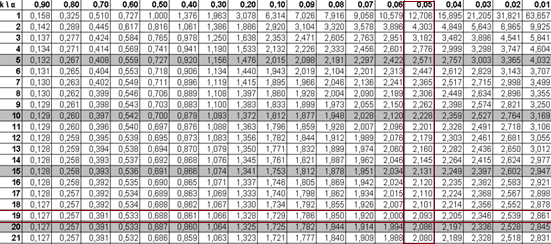
Zadanie 4

W pierwszej części zadania należy porównać średnie są mniejsze od 0,7, dla 20 elementów wykorzystałam do tego test T-studenta. Aby przeprowadzić rozkład T studenta wyznaczam hipotezę: h0 – średnia z prób jest równa 0,7. Hipoteza alternatywna: h1 – średnia z prób nie jest równa 0,7. Jako poziom istotności wybrałam α1 = 0,5, α2 = 0,1, α3=0,05.

Rozkład t studenta wyliczyłam z następującego wzoru:

gdzie X – to średnia z próby, μ0 – wartość hipotezy h0,

S – odchylenie standardowe, n – liczność próby.

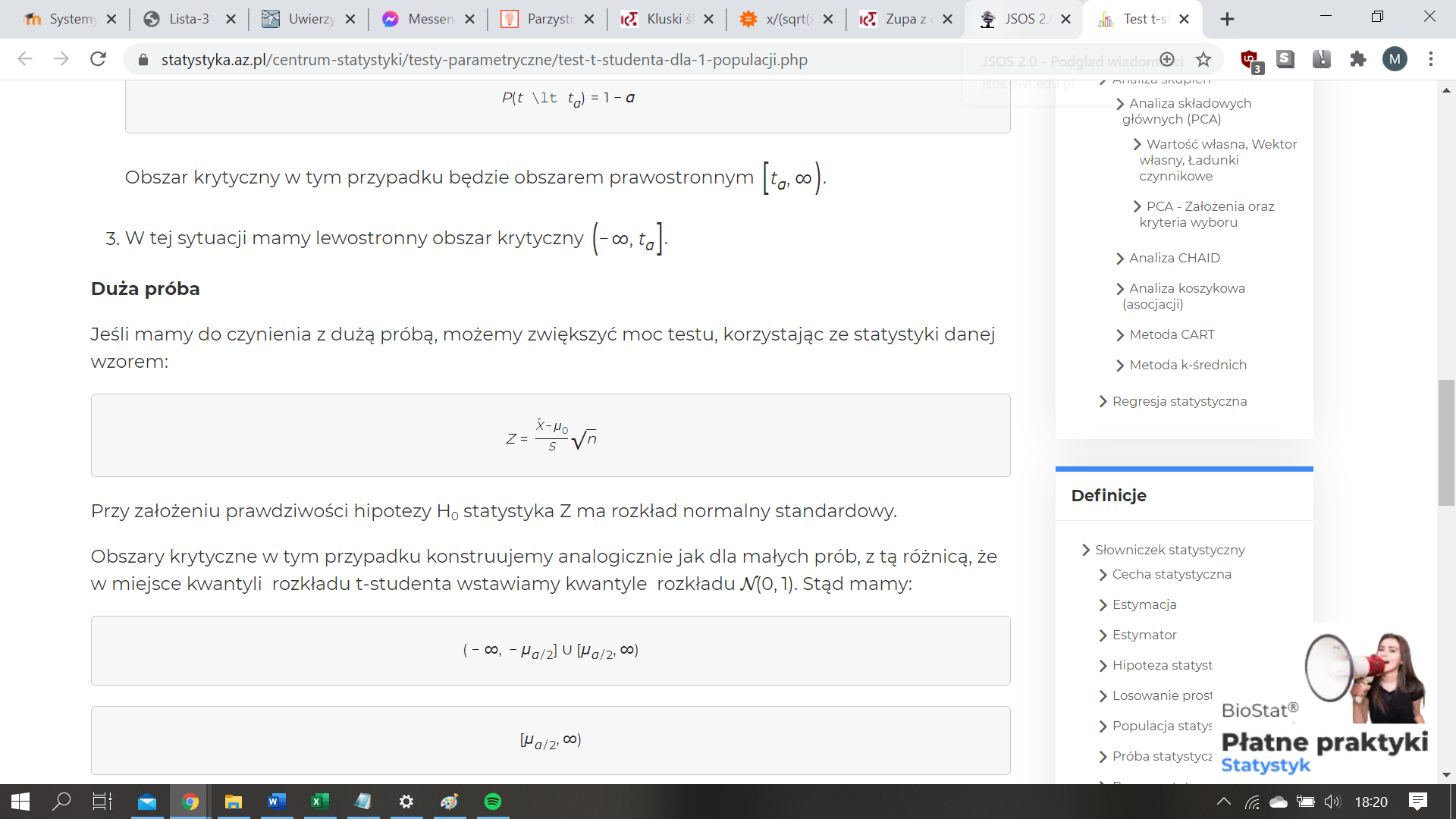




Więc obszar krytyczny dla 20 stopni swobody i poziomie istotności α = 0,05 wynosi

(-∞, -2,093] ∪ [2,093, ∞) wiec dla t = -1,85 nie mamy podstawy do odrzucenia hipotezy h0.

Natomiast dla istotności α = 0,70 obszar krytyczny dla 20 stopni swobody i poziomie wynosi (-∞, -0,39] ∪ [0,39, ∞) wiec dla t = -1,85 więc t należy do obszaru krytycznego, przez co odrzucamy h0 i przyjmujemy hipotezę alternatywna.

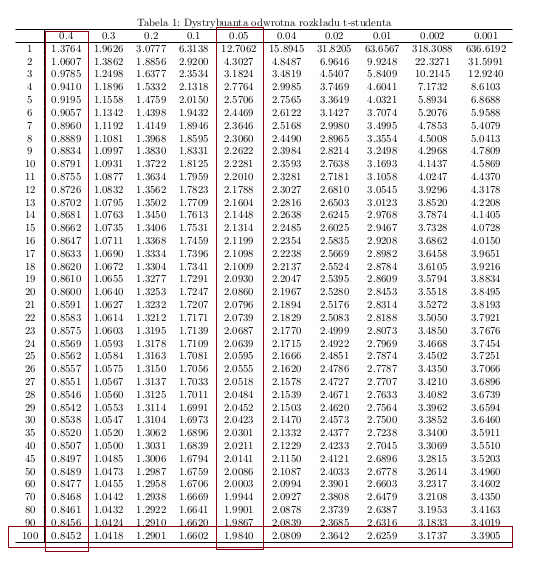
 Natomiast dla ciągu 100 wartości aby sprawdzić średnie z prób wykorzystałam Test Z. Aby przeprowadzić rozkład Z wyznaczam hipotezę: h0 – średnia z prób jest równa 0,5. Hipoteza alternatywna: h1 – średnia z prób nie jest równa 0,5. Jako poziom istotności wybrałam α1 = 0,7, α3=0,05.

Rozkład Z wyliczyłam z następującego wzoru:

gdzie X – to średnia z próby, μ0 – wartość hipotezy h0,

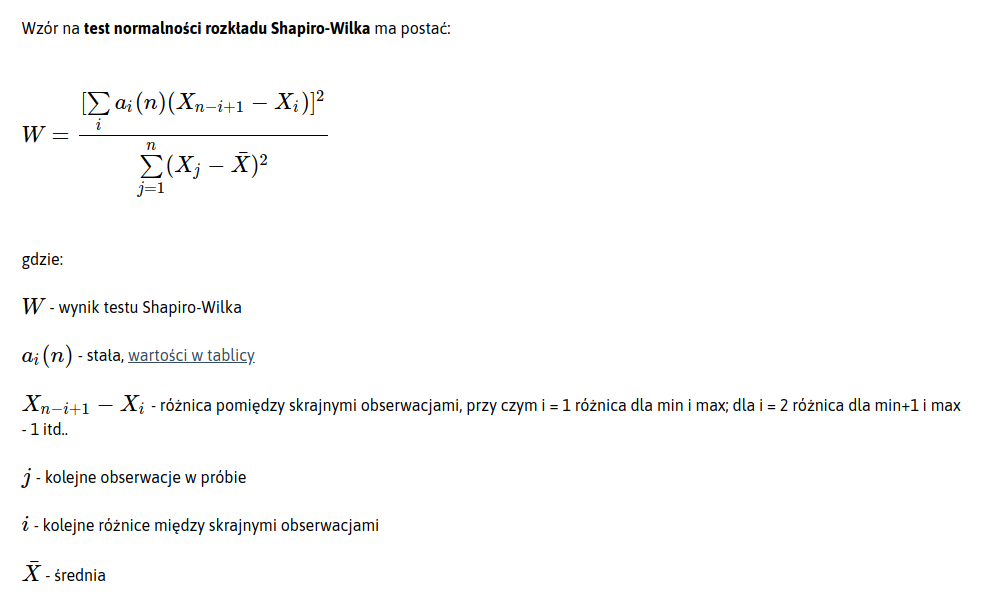
S – odchylenie standardowe, n – liczność próby.

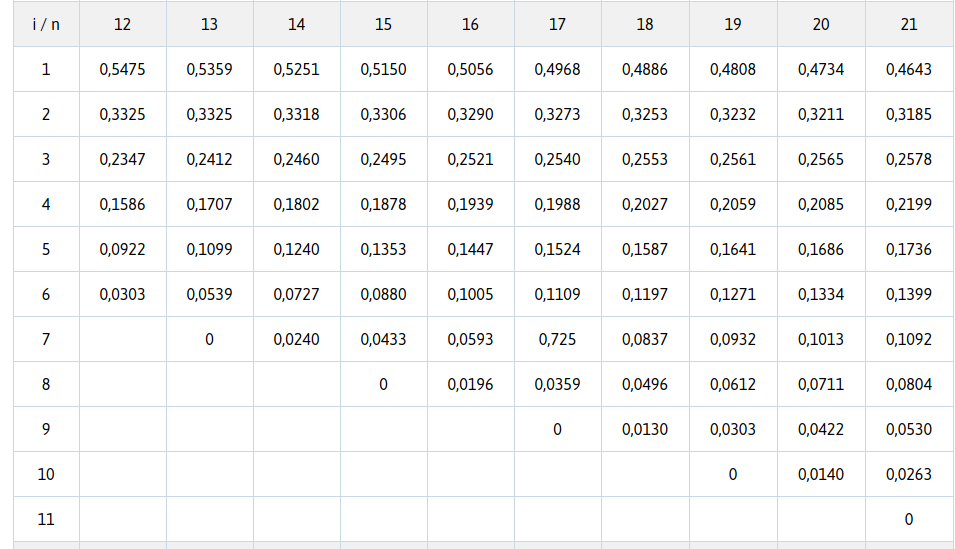


Następnie skorzystałam z tabelki dla k = 100 i α = 0,05 co daje tα = 1,98. Więc obszar krytyczny dla poziomu istotności α = 0,05 wynosi: (-∞, -1,98] ∪ [1,98, ∞) co oznacza dla t = -7.7699 że znajduje się on w obszarze krytycznym, czyli mamy podstawy do odrzucenia hipotezy h0 na korzyść hipotezy alternatywnej.

Wiąze się to z tym że średnia w rozkładzie normalnym zbiliża się do 0.5, więc średnia 0.7 jest mało prawdopodobna w tak duzej próbie.

Aby wykonać porównanie zbioru elementów do rozkładu normalnego wykorzystałam test Shapiro-Wilka. Test ten służący do oceny, czy zebrane przez nas wyniki od badanych prób posiadają rozkład normalny. Jako hipoteze przyjełam h0 : p>0,05 co oznacza że badana próba należy do rozkładu normalnego. Hipoteza alternetywna h1 ; p <0,05 – badana próba nie jest w postaci rozkładu normalnego.

Dla dwudziestu watrości wykonałam go na kartce korzystająć z wrozu:



i tablicy:

Otrzymałam w ten sposób p = 0.88, do sprawdzenia tych obliczeń wykorzystałam funkcje wbudowana w biblioteke scipy: scipy.stats.shapiro(numbers), który zwrocił mi test statystyczny oraz watość p – istotność statystyczna.

Ponieważ p = 0.88 > 0,05 to możemy przyjąć h0 jako prawdziwą.

Tak więc dla 20 elementów

p = 0.8815, więc jest to rozkład naturalny .

Jeśli test Shapiro-Wilka osiąga istotność statystyczną (p < 0,05), świadczy to o rozkładzie oddalonym od krzywej Gaussa. W przypadku tego testu najczęściej chcemy otrzymać wartości nieistotne statystyczne (p > 0,05), ponieważ świadczą one o zgodności rozkładu zmiennej z rozkładem normalnym.

Dla 100 elementów policzyłam tylko i wyłącznie za pomocą funkcji wbudowanej:

Jako hipoteze przyjełam h0 : p>0,05 co oznacza że badana próba należy do rozkładu normalnego. Hipoteza alternetywna h1 ; p <0,05 – badana próba nie jest w postaci rozkładu normalnego.

p = 0,436, więc również jest to rozkład naturalny i nie mamy podstawy do odrzucenia hipotezy h0.