

Statystyka biomedyczna

Laboratorium 5

1. Utworzyć dyskretną funkcję rozkładu prawdopodobieństwa według tabeli:

X	$\Pr(X)$
1	0.1
2	0.15
3	0.05
6	0.2
7	0.1
9	0.4

2. Wylosować $n=5$ wartości z tego rozkładu. Wyznaczyć średnią μ . Eksperyment powtórzyć 10.000 razy. Wyznaczone wartości średniej zapamiętać w tablicy *SampleDst*.
3. Wykreślić histogram rozkładu próbkowego średniej (liczba prążków – 20). Ocenić kształt histogramu. Wyznaczyć wartości kurtozy oraz skośność rozkładu. Powtórzyć obliczenia dla $n=25$ oraz $n=50$. Porównać zarówno kształty wykresów, jak i współczynniki kurtozy i skośności. Kiedy otrzymane histogramy przypominają rozkład normalny?
4. Pobrać z serwera dane *Echocardiogram* i wczytać do środowiska uruchomieniowego Python. Wybrać do dalszej pracy podzbiór danych oznaczających liczbę miesięcy życia po zawale do zgonu (nie bierzemy pod uwagę osób, które żyły w chwili tworzenia zbioru danych) – *monthsToDeath*.
5. W obrębie wyselekcjonowanego podzbioru wyznaczyć średnią liczbę miesięcy życia oraz odchylenie standardowe. Wykreślić histogram tej zmiennej losowej (liczba prążków – 10).
6. Przyjmując podzbiór *monthsToDeath* jako próbę reprezentatywną dla całej populacji, wyznaczyć średnią liczbę miesięcy życia dla dziesięciu ($n=10$) losowo wybranych osób z puli. Powtórzyć eksperyment 10.000 razy. Wyznaczone wartości średniej zapamiętać w tablicy *SampleDst2*.
7. Wykreślić histogram rozkładu próbkowego średniej (liczba prążków – 10). Ocenić kształt histogramu. Wyznaczyć wartości kurtozy oraz skośność rozkładu. Obliczyć błąd standardowy rozkładu *SampleDst2* – zarówno na podstawie danych, jak i posilując się wartością odchylenia standardowego dla całej populacji oraz stosując regułę pierwiastkową. Porównać obydwie wartości.
8. Wyznaczyć prawdopodobieństwo, że w wylosowanej próbce 10-osobowej, średnia liczba długość życia po zawale będzie równa co najmniej 35 miesięcy.
9. Wyznaczyć przedział ufności, w którym mieści się 95% wszystkich średnich długości życia po zawale dla próby 10-osobowej.

10. Dla jak licznej próby osób prawdopodobieństwo tego, że średnia długość życia po zawale w tej próbie nie będzie mniejsza od średniej dla całej populacji bardziej niż o 5 miesięcy, wyniesie 95%?

Przydatne funkcje Pythona:

Tworzenie arbitralnego dyskretnego rozkładu prawdopodobieństwa:
`scipy.stats.rv_discrete(Args)`

Kurtoza:
`scipy.stats.kurtosis(Arg)`

Współczynnik skośności:
`scipy.stats.skew(Arg)`

Ładowanie pliku csv:
`numpy.genfromtxt(Args)`

Próbkowanie zbioru danych:
`random.sample(Args)`

Wyznaczanie prawdopodobieństwa z rozkładu normalnego:
`scipy.stats.norm.cdf(Arg)`