

Statystyka biomedyczna

Laboratorium 3

1. Z repozytorium danych UCI (<https://archive.ics.uci.edu/ml/datasets.html>) pobrać zbiór *Echocardiogram*. Na podstawie informacji dostępnych na stronie repozytorium zapoznać się z zawartością zbioru i charakterem poszczególnych cech.
2. Wyznaczyć średnią liczbę miesięcy przeżycia po zawale oraz szansę na dożycie co najmniej do 12 miesiąca. Których pacjentów należy wykluczyć z tej kwerendy? Dokonać odpowiedniej redukcji wektorów danych.
3. Ile wynosi liczebność pozostałej próby losowej? Wykreślić histogram zmiennej losowej oznaczającej liczbę miesięcy przeżycia po zawale aż do zgonu. ($n = ?$)
4. Na podstawie próby losowej wyznaczyć estymowaną wartość prawdopodobieństwa przeżycia co najmniej 12 miesięcy po zawala ($p = ?$)
5. Wykreślić funkcję masy prawdopodobieństwa rozkładu dwumianowego dla parametrów n i p .
6. Korzystając z rozkładu dwumianowego i wzoru Bernoulliego, znaleźć przedział ufności, w którym z prawdopodobieństwem 95% znajduje się oczekiwana wartość parametru p .
7. Korzystając z próbkowego rozkładu średniej oszacować ten sam przedział ufności, co w pkt. 6. W tym celu wyznaczyć wartość oczekiwaną oraz błąd standardowy. Porównać otrzymane wartości granic przedziałów.

Przydatne funkcje Pythona:

Funkcja masy p-stwa rozkładu dwumianowego dla $k = 0, 1, 2, \dots$:

```
scipy.stats.binom.pmf(k[i], n, p)
```

Funkcja gęstości p-stwa rozkładu normalnego w punkcie:

```
scipy.stats.norm.ppf(0.025)
```

Ładowanie zapisanej tablicy z pliku:

```
monthsSurvived = numpy.load('./monthsSurvive.npy')
```

Tworzenie histogramu:

```
plt.hist(data, bins, align='mid');
```