

Zaliczenie

- Ćwiczenia (zaliczenie = min. 15 punktów)
 - Kolokwium (8/10 czerwca) = maks. 30 punktów
 - Dwa zadania z listy pod linkiem = maks. 1 punkt
<http://www.fuw.edu.pl/~prozanski/WS/upload/20150415-zadania.php>
 - *do ustalenia* = maks. 1 punkt
- Projekt (maks. 2 punkty)
 - Do oddania i ocenienia przed egzaminem

Python

- Proszę pamiętać:
 - ułamki dziesiętne piszemy z kropką (0.3) a nie przecinkiem (0,3)
 - jeśli dzielimy przez siebie dwie liczby, przynajmniej jedna powinna być wartością typu float
 - nie $N/2$ tylko $0.5 * N$ jeśli N jest całkowite
 - Nie N/M tylko $N/\text{float}(M)$ jeśli obie są całkowite
 - nie $x^{**}(1/2)$ tylko $x^{**}0.5$ a jeszcze lepiej `numpy.sqrt(x)`
 - do metody `numpy.concatenate` potrzebny jest dodatkowy nawias
 - `numpy.concatenate([X, Y])` albo `((X, Y))`

Bootstrap

- Startujemy z pojedynczej próby lub z rozkładu
- Powtarzamy wiele razy:

- generujemy z próby próbę „syntetyczną”

- Losując ze zwracaniem

```
Y = X[numpy.random.randint(len(X), size=rozmiar)]
```

lub

- Losując bez zwracania

```
Y = numpy.random.permutation(X)
```

- dla każdej próby (Y) obliczamy wartość statystyki lub zliczamy te, które spełniają warunek

Przedziały ufności

(rozkład normalny ze znaną wariancją)

Importer win musi zbadać średnią zawartość alkoholu w nowej partii win francuskich. Z doświadczenia z poprzednimi gatunkami wina, przyjmuje on, że standardowe odchylenie w populacji wynosi 1,2%. Importer wybrał losową próbę 60 butelek nowego wina i otrzymał średnią z próby 9,3%. Znaleźć przedział ufności 90% dla średniej zawartości alkoholu w nowej partii win.

Przedziały ufności

(rozkład normalny ze znaną wariancją)

Importer win musi zbadać średnią zawartość alkoholu w nowej partii win francuskich. Z doświadczenia z poprzednimi gatunkami wina, przyjmuje on, że standardowe odchylenie w populacji wynosi 1,2%. Importer wybrał losową próbę 60 butelek nowego wina i otrzymał średnią z próby 9,3%. Znaleźć przedział ufności 90% dla średniej zawartości alkoholu w nowej partii win.

- $\bar{x}=9.3$ $\sigma=1.2$ $N=60$ $\alpha=0.1$

Przedziały ufności

(rozkład normalny ze znaną wariancją)

Importer win musi zbadać średnią zawartość alkoholu w nowej partii win francuskich. Z doświadczenia z poprzednimi gatunkami wina, przyjmuje on, że standardowe odchylenie w populacji wynosi 1,2%. Importer wybrał losową próbę 60 butelek nowego wina i otrzymał średnią z próby 9,3%. Znaleźć przedział ufności 90% dla średniej zawartości alkoholu w nowej partii win.

- $\bar{x}=9.3$ $\sigma=1.2$ $N=60$ $\alpha=0.1$
- Znajdujemy kwantyl rozkładu normalnego $z_{\alpha/2}$
`z = scipy.stats.norm.ppf(alpha/2) # uwaga, będzie ujemny`
- Przedział ufności dla średniej z rozkładu normalnego

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

Przedziały ufności

(rozkład normalny z nieznaną wariancją)

Lekarz chce zbadać średni czas trwania kuracji tj. od podania leku do ustąpienia objawów w pewnej chorobie. Losowa próba 15 pacjentów dała średni czas 10,37 dnia i odchylenie standardowe 3,5 dnia. Zakładając normalny rozkład w populacji czasów trwania kuracji znaleźć 95% przedział ufności dla średniego czasu trwania kuracji.

Przedziały ufności

(rozkład normalny z nieznaną wariancją)

Lekarz chce zbadać średni czas trwania kuracji tj. od podania leku do ustąpienia objawów w pewnej chorobie. Losowa próba 15 pacjentów dała średni czas 10,37 dnia i odchylenie standardowe 3,5 dnia. Zakładając normalny rozkład w populacji czasów trwania kuracji znaleźć 95% przedział ufności dla średniego czasu trwania kuracji.

- $\bar{x} = 10.37$ $s = 3.5$ $N = 15$ $\alpha = 0.05$

Przedziały ufności

(rozkład normalny z nieznaną wariancją)

Lekarz chce zbadać średni czas trwania kuracji tj. od podania leku do ustąpienia objawów w pewnej chorobie. Losowa próba 15 pacjentów dała średni czas 10,37 dnia i odchylenie standardowe 3,5 dnia. Zakładając normalny rozkład w populacji czasów trwania kuracji znaleźć 95% przedział ufności dla średniego czasu trwania kuracji.

- $\bar{x}=10.37$ $s=3.5$ $N=15$ $\alpha=0.05$
- Znajdujemy kwantyl rozkładu t o $N-1$ stopniach swobody $t_{\alpha/2}$
`t = scipy.stats.t(N-1).ppf(alpha/2)` # uwaga, będzie ujemny
- Przedział ufności dla średniej z rozkładu t

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{N}}$$

Przedziały ufności

(nieznany rozkład lub dziwna statystyka)

Ogrodnik eksperymentuje z nowym gatunkiem drzew. Posadził 20 sztuk i po dwóch latach zmierzył następujące średnice pni (w cm): 8,5 7,6 9,3 5,5 11,4 6,9 6,5 12,9 8,7 4,8 4,2 8,1 6,5 5,8 6,7 2,4 11,1 7,1 8,8 7,2

- Proszę znaleźć średnią średnicę i 90% przedział ufności dla średniej.
- Proszę znaleźć medianę średnic i 90% przedział ufności dla mediany.

Przedziały ufności

(nieznany rozkład lub dziwna statystyka)

Ogrodnik eksperymentuje z nowym gatunkiem drzew. Posadził 20 sztuk i po dwóch latach zmierzył następujące średnice pni (w cm): 8,5 7,6 9,3 5,5 11,4 6,9 6,5 12,9 8,7 4,8 4,2 8,1 6,5 5,8 6,7 2,4 11,1 7,1 8,8 7,2

- Proszę znaleźć średnią średnicę i 90% przedział ufności dla średniej.
- Proszę znaleźć medianę średnic i 90% przedział ufności dla mediany.

$$\alpha = 0.1$$

Przedziały ufności

(nieznany rozkład lub dziwna statystyka)

- Obliczamy wartości statystyki dla wielu „syntetycznych” prób.

```
X = [...]  
ile_powtorzen = ...  
srednie = numpy.zeros(ile_powtorzen)  
for i in range(ile_powtorzen) :  
    Y = losujemy z X próbę ze zwracaniem  
    srednie[i] = numpy.mean(Y)
```

- Rozdzielamy $\alpha/2$ najczęściej i najrzadziej występujących, czyli znajdujemy odpowiednie kwantyle.

```
xL = scipy.stats.scoreatpercentile(srednie,  
100.0*alpha/2)  
xP = scipy.stats.scoreatpercentile(srednie,  
100.0*(1-alpha/2))
```

Przedziały ufności

(nieznany rozkład lub dziwna statystyka)

- Obliczamy wartości statystyki dla wielu „syntetycznych” prób.

```
X = [...]  
ile_powtorzen = ...  
mediany = numpy.zeros(ile_powtorzen)  
for i in range(ile_powtorzen) :  
    Y = losujemy z X próbę ze zwracaniem  
    srednie[i] = numpy.median(Y)
```

- Rozdzielamy $\alpha/2$ najczęściej i najrzadziej występujących, czyli znajdujemy odpowiednie kwantyle.

```
xL = scipy.stats.scoreatpercentile(mediany,  
100.0*alpha/2)  
xP = scipy.stats.scoreatpercentile(mediany,  
100.0*(1-alpha/2))
```

Przedziały ufności

(zadania dotyczące proporcji)

W próbce 200 osób 7 procent jest bezrobotnych. Określić 95% przedział ufności dla odsetka bezrobotnych w populacji.

- Mamy tylko jedną wartość! (0.07).
Jak z tego zrobić przedział ufności?

Przedziały ufności

(zadania dotyczące proporcji)

W próbce 200 osób 7 procent jest bezrobotnych. Określić 95% przedział ufności dla odsetka bezrobotnych w populacji.

- Mamy tylko jedną wartość! (0.07).
Jak z tego zrobić przedział ufności?

- Tworzymy próbę o długości 200 zawierający 14 jedynek i 186 zer.

```
X = numpy.concatenate([numpy.ones(14), numpy.zeros(186)])
```

- Zauważmy, że średnia z tej próby = 0.07
- Widzimy, że zadanie można rozwiązać poprzez wyznaczenie przedziału ufności dla średniej metodą bootstrap

Przedziały ufności

(zadania dotyczące proporcji)

- Możemy również zauważyć, że:
 - Zamiast tworzyć początkowy wektor i losować z niego, możemy za każdym razem losować bezpośrednio z rozkładu zero-jedynkowego (Bernoulliego)

```
Y = scipy.stats.bernoulli(0.07).rvs(size=200)
srednie[i] = numpy.mean(Y)
```

- Ponieważ interesuje nas tylko średnia z Y , możemy tę średnią obliczyć bezpośrednio z rozkładu dwumianowego

```
srednie[i] = scipy.stats.binom(200, 0.07).rvs() / 200
```

- Skoro tak, to nie musimy w ogóle przeprowadzać symulacji!
Wystarczy wyciągnąć kwantyle z rozkładu dwumianowego

```
scipy.stats.binom(200, 0.07).ppf([alpha/2,
1-alpha/2]) / 200
```