Przedziały ufności

- Poziom istotności = α (zwykle ≤ 0.05)
- Poziom ufności = $1-\alpha$
- Przedział ufności dla parametru μ = taki przedział $[a,b]$, dla którego

$$P(\ \mu \in [a,b]\ ) = 1-\alpha$$

czyli

$$P(\ \mu < a\ ) = \alpha/2$$
$$P(\ \mu > b\ ) = \alpha/2$$

co z kolei oznacza że $P(\ \mu \leq b\ ) = 1-\alpha/2$

Średnia, przypadek 1

- Mamy pomiary $x_1 \ldots x_N$ z rozkładu normalnego o <u>znanej</u> wariancji σ^2
- Zakładamy poziom istotności α
- Przedział ufności dla wartości oczekiwanej (średniej) rozkładu skonstruujemy jako

$$\mu = \bar{x} \mp \frac{z_{\alpha/2}\sigma}{\sqrt{N}}$$

gdzie

$$z_{\alpha/2} = F_N^{-1}\left(\frac{\alpha}{2}\right)$$

jest kwantylem rzędu $\alpha/2$ rozkładu $N(0,1)$

Średnia, przypadek 2

- Mamy pomiary x_1 … x_N z rozkładu normalnego o nieznanej wariancji, którą obliczamy jako

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

- Przedział ufności dla wartości oczekiwanej (średniej) rozkładu skonstruujemy jako

$$\mu = \bar{x} \mp \frac{t_{\alpha/2} s}{\sqrt{N}}$$

gdzie

$$t_{\alpha/2} = F_t^{-1}\left(\frac{\alpha}{2}\right)$$

jest kwantylem rzędu α/2 rozkładu t-Studenta
o N−1 stopniach swobody

Średnia, przypadek 3

- Mamy pomiary $x_1 \dots x_N$ (gdzie N jest duże) z dowolnego rozkładu o <u>znanej</u> wariancji σ^2
- Na podstawie Centralnego Twierdzenia Granicznego, dokładnie tak jak przypadek 1

Wariancja

- Mamy pomiary x_1 … x_N z rozkładu normalnego o <u>nieznanej</u> wariancji, którą obliczamy jako

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - \bar{x} \right)^2$$

- Przedział ufności dla wariancji rozkładu skonstruujemy jako

$$\sigma^2 \in \left[\frac{s^2 (N-1)}{\chi^2_{\alpha/2}} ; \frac{s^2 (N-1)}{\chi^2_{(1-\alpha/2)}} \right]$$

gdzie

$$\chi^2_p = F^{-1}_{\chi^2}(p)$$

jest kwantylem rzędu *p* rozkładu χ^2
o N−1 stopniach swobowy

Czy to jest normalne?

- Jeśli mamy wątpliwości, czy dane w tablicy *X* pochodzą z rozkładu normalnego, możemy użyć testu Shapiro-Wilka:

  ```
  W,p = scipy.stats.shapiro(X)
  ```

  Jeśli otrzymamy $p < \alpha$, nie możemy traktować danych jako pochodzących z rozkładu normalnego
- Co wtedy?

Bootstrap

- Mamy pomiary x_1 … x_N z dowolnego rozkładu
- Chcemy oszacować przedział ufności dla <u>dowolnej</u> statystyki (średnia, wariancja, mediana etc.)
- Powtarzamy wiele (*R*) razy:
  - Losujemy z naszej próby N liczb y_1 … y_N ze zwracaniem
  - Z wylosowanych liczb obliczamy szukaną statystykę i zapisujemy wynik do listy
- Z listy wyników (o długości *R*) wyciągamy kwantyl rzędu $\alpha/2$ i kwantyl rzędu $1-\alpha/2$

Zadanie 1

Producent karmy dla zwierząt chciał przetestować nowy rodzaj karmy. Próbki podawał 12 zwierzakom przez 4 tygodnie. Po tym czasie zanotował następujące przyrosty masy:

15,43 16,92 14,43 12,94 15,92 17,42 18,91 16,92 14,93 14,49 15,92 15,43 kg

Średni przyrost wynosi 15.80 kg. Producent widzi jednak, że w próbie jest dość znaczny rozrzut pomiędzy poszczególnymi zwierzętami 12,94-18,91 i nie jest pewien czy można reklamować nowy produkt podając średni przyrost 15,8 kg. Podejrzewa, że inna grupa zwierząt może mieć zupełnie inną średnią.

- Używając powyższych danych znajdź 95% przedział ufności na średni przyrost masy.
- Wynik zilustruj przy pomocy histogramu.
- Jaki byłby wynik przy założeniu, że masy zwierząt pochodzą z rozkładu normalnego?

Zadanie 2

Producent opon rowerowych chce oszacować średni dystans jaki można przejechać na oponie pewnego rodzaju zanim opona się zużyje. Pobrano losową próbę 32 opon, opona jest używana aż do przetarcia i odległość przejechana na każdej oponie jest rejestrowana. Dane (w tysiącach kilometrów) są następujące:

32, 33, 28, 37, 29, 30, 25, 27, 39, 40, 26, 26, 27, 30, 25, 30, 31, 29, 24, 36, 25, 37, 37, 20, 22, 35, 23, 28, 30, 36, 40, 41

Znaleźć 99% przedział ufności dla średniego przebiegu opon tego rodzaju.

Zadanie 3

Zawartość procentowa aluminium w 18 antycznych naczyniach z Teb była następująca:

11,4 13,4 13,5 13,8 13,9 14,4 14,5 15 15,1 15,8 16 16,3 16,5 16,9 17 17,2 17,5 19,0

Jaka jest mediana procentowej zawartości aluminium i jaki jest 95% przedział ufności?

Zadanie 4

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.

Zadanie 5

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 i 90% przedział ufności dla mediany.
- Wynik zilustrować przy pomocy histogramu.

Zadanie 6

Automat do kawy nalewa kawę do kubków. Jeśli średnia porcja kawy w kubku odbiega od normy, maszynę można wyregulować. Jeśli jednak wariancja porcji kawy jest zbyt duża, maszyna wymaga reperacji. Od czasu do czasu przeprowadzana jest kontrola wariancji porcji kawy. Odbywa się to poprzez wybór losowej próby napełnionych kubków i policzenie wariancji próby. Losowa próba 30 kubków dała wariancję próby $s^2 = 18{,}54$.

Obliczyć 95% przedział ufności dla wariancji populacji σ^2.

Zadanie 7

W badaniach nad cholesterolem u ludzi stwierdzono, że w grupie 135 badanych z wysokim poziomem cholesterolu 10 osób przeszło zawał serca.

Pytanie: Jaki jest 95% przedział ufności dla proporcji 10/135?

Zadanie 8

Biuro podróży chce oszacować średnią ilość pieniędzy wydaną na wakacje przez osoby korzystające z jego usług. Ludzie przeprowadzający analizę chcieliby móc oszacować średni koszt wakacji z dokładnością do 200 zł na poziomie ufności 95%. Z poprzednich doświadczeń tego biura podróży wynika, że odchylenie standardowe w populacji wynosi $\sigma = 400$ zł.

Jaka będzie minimalna wielkość próby?

Zadanie 9

W próbce 20 testowanych baterii stwierdzono średni czas życia 28,85 miesiąca. Określić 95% przedział ufności dla średniej. Wartości dla badanej próbki były następujące:

30 32 31 28 31 29 29 24 30 31 28 28 32 31 24 23 31 27 27 31 miesięcy

Obejrzeć rozkład przy pomocy histfit i zbadać jaki wpływ na przedział ufności ma przyjęcie założenia o normalności rozkładu czasów życia.