Estymacja punktowa i przedziałowa parametrów statystycznych

Estymacja - szacowanie nieznanych wartości parametrów rozkładu pewnej cechy w populacji na podstawie próbki.

Estymacja punktowa - oszacowanie wartości parametru przez wartość estymatora (wynik: θ)

Na podstawie próbki z całej populacji szacujemy parametr θ => Wyciągamy wnioski o parametrze populacji.

Estymacja punktowa jest uzyskiwana za pomocą **estymatora** na podstawie przykładowych danych. Różne próbki będą dawać różne oszacowania, nawet przy użyciu tego samego estymatora.

Estymacja przedziałowa – szacowanie wartości parametru za pomocą pewnego przedziału, który z założonym z góry prawdopodobieństwem pokrywa rzeczywistą wartość parametru θ (wynik: θ leży w przedziale $[\theta_-, \theta_+]$ z prawdopodobieństwem $1-\alpha$)

Przedziałem ufności (θ_1, θ_2) o poziomie ufności $1-\alpha$ nazywamy przedział spełniający warunek:

$$\boldsymbol{P(\theta_1 < \theta < \theta_2) = 1 - \alpha}$$

Poziom ufności $1-\alpha$ – prawdopobieństwo, że rzeczywista wartość szacowanego parametru θ w populacji znajduje się w wyznaczonym przedziale ufności

1. Załaduj pakiet **MASS** i wczytaj zestaw danych **survey**. Zbiór ten przedstawia wyniki ankiety 237 studentów z Uniwersytetu w Adelaide. Ankieta ta opisuje wiele cech – m.in. wiek, płeć, wzrost. Na podstawie wyników tej próby studentów będziemy próbowali wyciągnąć wnioski na temat wieku wszystkich studentów tej uczelni. Podaj punktową ocenę średniej i odchylenia standardowego wieku w populacji. (Przy obliczeniach zwróć uwagę, że nie wszyscy studenci odpowiedzieli na wszystkie pytanie – usuń puste pola)
2. Oblicz przedział ufności dla średniej wieku studentów zakładając przedział ufności na poziomie 95%.
3. Zakładając, że znamy odchylenie standardowe populacji równe 7, oblicz przedział ufności oraz margines błędu dla średniej wieku na poziomie 95%.

   Wyznaczanie przedziału ufności dla średniej, gdy rozkład cech w populacji jest rozkładem normalnym **N(μ,σ) :**

   $$\left(\boldsymbol{\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}\right)$$

   $\boldsymbol{z_{\frac{\alpha}{2}}}$ **–** wartość dystrybuanty standardowego rozkładu normalnego w punkcie $\boldsymbol{1 - \frac{\alpha}{2}}$

   $\boldsymbol{\bar{x}}$ **–** estymator nieznanej średniej

$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ – margines błędu

4. Nie znając odchylenia standardowego wieku populacji, oblicz przedział ufności oraz margines błędu dla średniej wieku na poziomie 95%. Gdy nie znamy odchylenia standardowego populacji, estymujemy je na podstawie wartości z próby. W tym cely korzystamy z tablic **t-Studenta**.

5. Wylosuj 5 wartości z rozkładu normalnego o średniej =10 i odchyleniu standardowym= $\sqrt{3}$. Oblicz średnią z wylosowanej próby 1000 razy. **Przedstaw rozkład estymatora wartości średniej** za pomocą histogramu.
6. **Przedstaw rozkład estymatora wartości średniej** zwiększając wylosowaną próbę do 100 wartości. Porównaj otrzymane wyniki.

7. Wczytaj plik ‘szczury.txt’, który zawiera informacje na temat wag (w gramach) pewnego gatunku szczura dla losowej próby. Załaduj pakiet **mosaic**. Wylosuj kolejno 1000 próbek z wyjściowej próbki. Losowanie odbywa się ze zwracaniem, a wielkości próbek są takie same jak wielkość próbki wyjściowej. Podpowiedź: Użyj funkcji **sample()**

   Metoda bootstrapowa estymacji polega na wielokrotnym losowaniu ze zwracaniem próby. Znajduje ona zastosowanie w sytuacji, gdy nie znamy rozkładu z populacji z której pochodzi próbka.

8. Dla każdej z 1000 próbek wyznacz estymator średniej. Średnia z 1000 uzyskanych estymatorów jest wartością szukanego parametru.
9. Wyznacz histogram średniej estymowanej metodą bootstrapową