ESTYMACJA PARAMETRYCZNA

Przedziały ufności dla średniej μ na poziomie ufności $1-\alpha$

Model 1 $X \sim N(\mu, \sigma)$, σ znane - funkcja `z.test(x)` w R lub ze wzoru $\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

Model 2 $X \sim N(\mu, \sigma)$, σ nieznane - funkcja `t.test(x)` w R lub ze wzoru $\left[\bar{x} - t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x} + t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}}\right]$

Model 3 $X \sim$ rozkład dowolny $(n > 25)$ - funkcja `t.test(x)` w R lub ze wzoru $\left[\bar{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}\right]$

Przedziały ufności dla wariancji σ^2 na poziomie ufności $1-\alpha$

Model 1 $X \sim N(\mu, \sigma)$, σ nieznane - funkcja `sigma.test(x)` w R lub ze wzoru $\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}},\ \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\right]$

Model 2 $X \sim$ rozkład dowolny $(n > 25)$ - własna funkcja w R lub ze wzoru $\left[\frac{s^2(2n-2)}{(\sqrt{2n-3}+z_{1-\alpha/2})^2}, \frac{s^2(2n-2)}{(\sqrt{2n-3}-z_{1-\alpha/2})^2}\right]$

Przedział ufności dla odsetka (procentu) p na poziomie ufności $1-\alpha$

Model $X \sim Bern(p)$, p - nieznane - funkcje `binom.test(k,n)`, `prop.test(k,n)` lub (gdy $np > 5, n(p-1) > 5$) ze wzoru $\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$, gdzie $\hat{p} = \frac{k}{n} = \frac{\text{liczba sukcesów}}{\text{liczność próby}}$

Uwaga: Funkcje z.test(x) i sigma.test(x) są dostępne w pakiecie *TeachingDemos* w R.

---

1. Średnie wynagrodzenie 50 losowo wybranych programistów wyniosło 6000 zł. Wiadomo, że odchylenie standardowe wynagrodzenia programistów wynosi 2100 zł. Wyznacz 95% przedział ufności dla średniego wynagrodzenia programistów, zakładając, że rozkład ich wynagrodzeń jest rozkładem normalnym.

2. Dla wybranego użytkownika zarejestrowano czasy między naciśnięciami klawiszy, gdy wpisywał login i hasło. Pobrano z nich losową próbę 18 pomiarów (w sekundach):

   $$0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27.$$

   Zakładając, że czasy pochodzą z rozkładu normalnego, wyznacz

   a) 99% przedział ufności dla średniego czasu między naciśnięciami klawiszy tego użytkownika,
   b) 95% przedział ufności dla odchylenia standardowego czasu między naciśnięciami klawiszy tego użytkownika.

3. Zmierzono czas świecenia 69 świetlówek i stwierdzono, że dla 14 z nich był on krótszy niż 1000 godzin, dla 15 był w przedziale $[1000, 2000)$, dla 29 świetlówek był dłuższy niż 2000, ale krótszy niż 3000 godzin, zaś dla pozostałych 11 - czas świecenia był dłuższy niż 3000, ale nie dłuższy niż 4000 godzin. Oszacuj przedziałowo średnią i odchylenie standardowe czasu świecenia świetlówek. Przyjmij poziom ufności 0.95.

4. Ramka danych *faithful* zawiera dane dotyczące czasu trwania erupcji gejzera Old Faithful (zmienna *eruptions*) oraz czasu oczekiwania na kolejną erupcję (zmienna *waiting*). Utwórz 99% przedział ufności dla średniego czasu oczekiwania na kolejną erupcję.

5. Ramka danych *Pima.te* z pakietu *MASS* zawiera dane dotyczące zdrowia kilkuset Indianek z plemienia Pima mających co najmniej 21 lat. Zmienna *type* zawiera informację, czy kobieta jest chora na cukrzycę, czy nie.

   a) Utwórz 95% przedział ufności dla odsetka Indianek dotkniętych cukrzycą.
   b) Utwórz 95% przedział ufności dla odsetka Indianek dotkniętych cukrzycą mających co najmniej 35 lat.

6. Zmienna *weight* znajdująca się w ramce danych *chickwts* opisuje wagę kurczaków, natomiast zmienna *feed* rodzaj użtej paszy. Zakładamy, że waga kurczaków ma rozkład normalny. Zbuduj 93% przedział ufności dla wariancji wagi kurczaków karmionych paszą *soybean*.

7. Jak dużą próbę należy pobrać, aby z maksymalnym błędem $\pm 2\%$ oszacować na poziomie ufności 0.99 procent kierowców, którzy nie zapinają pasów bezpieczeństwa? Uwzględnij rezultaty wstępnych badań, z których wynika, że interesująca nas wielkość jest rzędu 16%. Porównaj otrzymaną liczność próby z licznością, jaka byłaby wymagana, gdyby pominąć rezultaty wstępnych badań.

Minimalna liczność próby do oszacowania średniej μ na poziomie ufności $(1-\alpha)$ z max. błędem d

Model 1 $X \sim N(\mu, \sigma)$, σ znane - ze wzoru $n \geq \left(\frac{\sigma}{d} z_{1-\alpha/2}\right)^2$

Model 2 $X \sim N(\mu, \sigma)$, σ nieznane - ze wzoru $n \geq \left(\frac{s}{d} t_{1-\alpha/2}^{n_0-1}\right)^2$, gdzie n_0 - liczność pobranej próby wstępnej

Minimalna liczność próby do oszacowania odsetka p na poziomie ufności $(1-\alpha)$ z max. błędem d

Model 1 Jeśli znany jest szacunkowy procent p_0 - ze wzoru $n \geq \frac{p_0(1-p_0)}{d^2} z_{1-\alpha/2}^2$

Model 2 Jeśli nie jest znany szacunkowy procent p_0 - ze wzoru $n \geq \frac{1}{4d^2} z_{1-\alpha/2}^2$

---

8. Poniższe dane przedstawiają zarejestrowaną przez radar drogowy prędkość 10 losowo wybranych pojazdów, jadących pewną autostradą (km/h):
$$106, 115, 99, 109, 122, 119, 104, 125, 107, 111.$$
Zakładając normalność rozkładu prędkości, wyznacz liczność próby potrzebną do wyestymowania średniej prędkości z dokładnością ± 2 km/h na poziomie ufności 0.95.

9. W celu oszacowania niezawodności pewnego urządzenia dokonano 8 pomiarów czasu bezawaryjnej pracy tego urządzenia i otrzymano następujące wyniki (w godzinach): $1034, 2720, 482, 622, 2624, 420, 342, 703$. Zakładamy, że czas bezawaryjnej pracy tego urządzenia ma rozkład wykładniczy. Oszacuj prawdopodobieństwo, że dane urządzenie nie ulegnie awarii w ciągu 750 godzin pracy.

ESTYMACJA NIEPARAMETRYCZNA

Dystrybuanta empiryczna

Naturalnym estymatorem nieznanej dystrybuanty F zmiennej X jest **dystrybuanta empiryczna** zbudowana na podstawie próby losowej $(X_1, X_2, \ldots, X_n)$ dana wzorem
$$\hat{F}_n(x) = \frac{\#\{X_i : \; X_i \leq x\}}{n}, \; x \in R.$$
Uwaga: Dystrybuantę empiryczną można narysować w R wywołując funkcję `ecdf()`.

Estymatory jądrowe

Innym estymatorem nieznanego rozkładu są **estymatory jądrowe** opisywane wzorem
$$\hat{f}_n(x) = \frac{1}{hn} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$
gdzie $h > 0$ jest zadaną szerokością pasma, natomiast K jest pewną funkcją spełniającą warunek $\int_{-\infty}^{+\infty} K(x)dx = 1$ zwaną jądrem. Często jako K przyjmuje się gęstość rozkładu $N(0,1)$, wtedy przyjmuje się $h = 1.06S/\sqrt[3]{n}$.

---

10. Wygeneruj 4 próby losowe z rozkładu standardowego normalnego: 5, 10, 20 i 100 elementową. Narysuj dla tych prób dystrybuanty empiryczne i porównaj je z odpowiednią dystrybuantą teoretyczną (tw. Gliwienki-Cantellego).

11. Wygeneruj n-elementową $(n = 100)$ próbę losową z rozkładu normalnego standardowego. Utwórz histogram oraz estymator jądrowy dla tej próby. Nałóż na uzyskany obraz wykres gęstości teoretycznej rozkładu normalnego.

12. Wygeneruj $n = 500$ elementową próbę $(Y_1, Y_2, \ldots, Y_n)$ z rozkładu normalnego standardowego.
    a) Dla każdej podpróbki zawierającej i początkowych elementów próbki wyjściowej, tj. dla $X_i = (Y_1, \ldots, Y_i)$, gdzie $i = 1, \ldots, n$ wyznacz średnią $\overline{X}_i$ oraz medianę Med_i. Narysuj na wspólnym wykresie wektory $\{\overline{X}_i : i = 1, \ldots, n\}$ oraz $\{Med_i : i = 1, \ldots, n\}$. Przeanalizuj wpływ liczności próby na zachowanie się średniej i mediany z próby. Czy statystyki te wydają się być sensownymi estymatorami parametru wartości oczekiwanej w tym modelu?
    b) Dla każdej podpróbki zawierającej $i = 2, \ldots, n$ początkowych elementów próbki wyjściowej wyznacz odchylenie standardowe s_i oraz $d_i = IQR(X_i)/1.35$. Przedstaw na wspólnym wykresie wektory $\{s_i : i = 2, \ldots, n\}$ oraz $\{d_i : i = 2, \ldots, n\}$. Przeanalizuj wpływ liczności próby na zachowanie się s_i i d_i. Czy statystyki te wydają się być sensownymi estymatorami odchylenia standardowego w tym modelu?