

9. ESTYMACJA PRZEDZIAŁOWA

Definicja. *Przedziałem ufności* dla parametru $\theta \in \Theta \subset \mathbb{R}$ na poziomie ufności $1 - \alpha$, gdzie $\alpha \in (0, 1)$, nazywamy przedział (θ_1, θ_2) , gdzie $\theta_1 = \theta_1(X_1, X_2, \dots, X_n)$ i $\theta_2 = \theta_2(X_1, X_2, \dots, X_n)$ to mierzalne funkcje próby, takie że $\theta_1 \leq \theta_2$ i

$$P(\theta \in (\theta_1(X_1, X_2, \dots, X_n), \theta_2(X_1, X_2, \dots, X_n))) = 1 - \alpha \quad \text{dla każdego } \theta \in \Theta. \quad (1)$$

Końce przedziału ufności $\theta_1(X_1, X_2, \dots, X_n)$ i $\theta_2(X_1, X_2, \dots, X_n)$ to zmienne losowe. Dla niektórych realizacji próby otrzymamy przedział $(\theta_1(x_1, x_2, \dots, x_n), \theta_2(x_1, x_2, \dots, x_n))$, taki że

$$\theta \in (\theta_1(x_1, x_2, \dots, x_n), \theta_2(x_1, x_2, \dots, x_n)),$$

dla innych realizacji – taki że

$$\theta \notin (\theta_1(x_1, x_2, \dots, x_n), \theta_2(x_1, x_2, \dots, x_n)).$$

Warunek (1) gwarantuje nam, że dla dużej liczby realizacji procent otrzymanych przedziałów takich, że $\theta \in (\theta_1(x_1, x_2, \dots, x_n), \theta_2(x_1, x_2, \dots, x_n))$ będzie w przybliżeniu równy $(1 - \alpha)100\%$.

Oczywiście dobrą precyzję oszacowania parametru θ otrzymamy, gdy tzw. *długość przedziału ufności*, zdefiniowana jako

$$l_n = \theta_2(X_1, X_2, \dots, X_n) - \theta_1(X_1, X_2, \dots, X_n),$$

będzie mała. Zatem, mając do wyboru różne przedziały ufności dla parametru θ na poziomie ufności $1 - \alpha$, będziemy wybierać ten, którego długość jest najmniejsza. Przy takim podejściu mogą się jednak pojawić problemy, bo długość przedziału ufności l_n to zmienna losowa i jeśli tylko nie będzie ona zdegenerowana do liczby, to, podczas porównywania długości dwóch przedziałów ufności l_n^1 i l_n^2 , może się zdarzyć tak, że

$$l_n^1(\omega) < l_n^2(\omega) \quad \text{dla pewnych } \omega \in \Omega$$

i

$$l_n^1(\omega) > l_n^2(\omega) \quad \text{dla innych } \omega \in \Omega,$$

co nie pozwoli nam rozstrzygnąć, który przedział jest krótszy. W tej sytuacji jako kryterium wyboru najlepszego przedziału ufności możemy użyć wartości oczekiwanej długości $E(l_n)$ zamiast samej długości l_n .

Poniżej przedstawimy przykłady jak wyznaczać przedziały ufności dla danych parametrów w zadanych modelach. Najpierw zajmiemy się przedziałami ufności dla wartości średniej, następnie przejdziemy do przedziałów ufności dla wariancji i odchylenia standardowego.

9.1 Przedziały ufności dla wartości średniej

Model I. Niech X_1, X_2, \dots, X_n będzie próbą losową z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$ o nieznannej wartości oczekiwanej $EX_1 = \mu$ i znanej wariancji $Var(X_1) = \sigma^2 > 0$. Wówczas

$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

jako kombinacja liniowa niezależnych zmiennych losowych o rozkładzie normalnym, też ma rozkład normalny. Ponadto z przykładu 7.1 i 7.2 wiemy, że $E(\bar{X}) = \mu$ i $Var(\bar{X}) = \frac{\sigma^2}{n}$ a stąd

$$EU = E\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right) = \frac{\sqrt{n}}{\sigma} E(\bar{X} - \mu) = 0$$

i

$$Var(U) = Var\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right) = \frac{n}{\sigma^2} Var(\bar{X}) = \frac{n}{\sigma^2} \frac{\sigma^2}{n} = 1.$$

Reasumując

$$U \sim \mathcal{N}(0, 1).$$

Zatem znalezienie przedziału ufności dla wartości średniej sprowadza się do znalezienia dwóch stałych a i b , takich że

$$P(a < U < b) = 1 - \alpha. \quad (2)$$

Rzeczywiście wtedy

$$P\left(a < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < b\right) = 1 - \alpha$$

lub równoważnie

$$P\left(\bar{X} - b\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} - a\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

co oznacza, że $\left(\bar{X} - b\frac{\sigma}{\sqrt{n}}, \bar{X} - a\frac{\sigma}{\sqrt{n}}\right)$ jest przedziałem ufności dla parametru μ na poziomie ufności $1 - \alpha$. Długość tego przedziału to $l_n = \frac{\sigma}{\sqrt{n}}(b - a)$. Chcemy znaleźć przedział najkrótszy, więc szukamy a i b spełniających (2) i jednocześnie takich by $b - a$ było jak najmniejsze. Z wykresu gęstości rozkładu $\mathcal{N}(0, 1)$ odczytujemy, że warunki te są spełnione dla następujących a i b :

$$b = u_{1-\frac{\alpha}{2}} \quad \text{i} \quad a = -u_{1-\frac{\alpha}{2}},$$

gdzie u_p to kwantyl rozkładu $\mathcal{N}(0, 1)$ rzędu $p \in (0, 1)$. Ostatecznie poszukiwany przedział ufności ma postać

$$\left(\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right). \quad (3)$$

Przypomnijmy, że w R funkcja

```
> qnorm(p)
```

zwraca przybliżenie (dla $p = 0,5$ jest to wartość dokładna) kwantyla rozkładu $\mathcal{N}(0, 1)$ rzędu p .

Model II. Niech X_1, X_2, \dots, X_n będzie próbą losową z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$ o nieznaney wartości oczekiwanej $EX_1 = \mu$ i nieznaney wariancji $Var(X_1) = \sigma^2 > 0$. Można wykazać, że wówczas

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}, \quad \text{gdzie} \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

jest zmienną losową o rozkładzie t-Studenta z $(n-1)$ -stopniami swobody, co zapisujemy $T \sim t_{[n-1]}$. Przeprowadzając rozumowanie analogiczne do poprzedniego, otrzymujemy następujący przedział ufności dla parametru μ na poziomie ufności $1 - \alpha$:

$$\left(\bar{X} - t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}} \right), \quad (4)$$

gdzie $t_{p,m}$ oznacza kwantyl rzędu $p \in (0, 1)$ rozkładu t-Studenta o m stopniach swobody.

W pakiecie R przybliżenie kwantyla $t_{p,m}$ (dla $p = 0,5$ będzie to wartość dokładna) wyznaczymy używając funkcji

```
> qt(p, df=m)
```

Ponadto w pakiecie R funkcja `t.test()` wyznacza przedział ufności dany w (4). Funkcja ta robi znacznie więcej - przeprowadza test dotyczący wartości średniej, o czym będzie mowa później. Jeśli interesuje nas jedynie przedział ufności to dodajemy `$conf.int`. Dokładniej piszemy komendę

```
> t.test(x, conf.level)$conf.int
```

gdzie

- argument `x` określa wektor z danymi;
- argument `conf.level` określa poziom ufności; pominięcie tego argumentu spowoduje przyjęcie wartości domyślnej `conf.level=0.95`.

Model III. Niech X_1, X_2, \dots, X_n będzie próbą losową z dowolnego rozkładu o nieznaney wartości oczekiwanej $EX_1 = \mu$ i nieznaney wariancji $Var(X_1) = \sigma^2 > 0$. Załóżmy, że n jest duże ($n \geq 100$). Wówczas z centralnego twierdzenia granicznego wynika, że

$$P\left(a < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < b\right) \approx P(a < U < b), \quad \text{gdzie } U \sim \mathcal{N}(0, 1). \quad (5)$$

Ponieważ, zgodnie z przykładem 7.4, $S = S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ jest mocno zgodnym estymatorem parametru σ (tzn. $P(\lim_{n \rightarrow \infty} S_n = \sigma) = 1$), to w (5) można zastąpić σ przez S , otrzymując

$$P\left(a < \frac{\bar{X} - \mu}{S/\sqrt{n}} < b\right) \approx P(a < U < b), \quad \text{gdzie } U \sim \mathcal{N}(0, 1).$$

Powtarzając rozumowanie, przeprowadzone podczas rozważania modelu I, otrzymujemy $b = -a = u_{1-\frac{\alpha}{2}}$ i stąd

$$\left(\bar{X} - u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) \quad (6)$$

to przedział ufności dla parametru μ na poziomie ufności w przybliżeniu równym $1 - \alpha$.

9.2 Przedziały ufności dla wariancji (odchylenia standardowego)

Model. Niech X_1, X_2, \dots, X_n będzie próbą losową z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$ o nieznannej wartości oczekiwanej $EX_1 = \mu$ i nieznannej wariancji $Var(X_1) = \sigma^2 > 0$. Można wykazać, że wówczas

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \text{gdzie } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

ma rozkład chi-kwadrat o $(n-1)$ stopniach swobody, co zapisujemy $\chi^2 \sim \chi_{[n-1]}^2$. Stąd

$$P\left(\chi_{\alpha/2, n-1}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\alpha/2, n-1}^2\right) = 1 - \alpha, \quad (7)$$

gdzie $\chi_{p, m}^2$ oznacza kwantyl rzędu $p \in (0, 1)$ rozkładu chi-kwadrat o m stopniach swobody. Przybliżenie kwantyla $\chi_{p, m}^2$ możemy wyznaczyć w pakiecie R używając funkcji

`> qchisq(p, df=m)`

Po odpowiednich przekształceniach (7) otrzymujemy

$$P\left(\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2}\right) = 1 - \alpha,$$

i

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2}}\right) = 1 - \alpha,$$

co oznacza, że

$$\left(\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}, \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} \right) \quad \text{i} \quad \left(\sqrt{\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}}, \sqrt{\frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2}} \right)$$

są przedziałami ufności na poziomie ufności $1 - \alpha$ dla odpowiednio wariancji (parametr σ^2) i odchylenia standardowego (parametr σ). Niestety nie są to przedziały o najmniejszej długości - kwantyle $\chi_{\alpha/2, n-1}^2$ i $\chi_{1-\alpha/2, n-1}^2$ można zastąpić innymi liczbami i otrzymać przedziały krótsze. Nie będziemy się jednak tym zajmować.

9.3 Planowanie eksperymentu - dobór liczności próby gwarantującej uzyskanie żądanej precyzji estymacji

W podrozdziale tym zajmiemy się wyznaczaniem niezbędnej liczności próby potrzebnej do uzyskania przedziału ufności dla wartości średniej na ustalonym poziomie ufności $1 - \alpha$ i o zadanej precyzji. Będziemy żądać by precyzja była taka, by długość przedziału nie przekraczała założonej wartości $2d$.

Model I. Przedział ufności dla wartości średniej (3), otrzymany dla modelu I w podrozdziale 9.1, ma tę własność, że jego długość $l_n = \frac{2\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}$ jest wielkością nielosową i maleje wraz ze wzrostem liczności próby n . Innymi słowy, zwiększenie liczności próby gwarantuje nam lepszą precyzję oszacowania nieznanego parametru μ . Co więcej, dla ustalonego poziomu ufności $1 - \alpha$ i ustalonej żądanej precyzji $2d$, możemy łatwo wyznaczyć wartość n , która zagwarantuje, że przedział ufności, wyznaczony na podstawie n -elementowej próby losowej, będzie nie dłuższy niż $2d$. Rzeczywiście warunek

$$l_n = \frac{2\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} \leq 2d$$

możemy przekształcić do postaci

$$\sqrt{n} \geq \frac{\sigma u_{1-\frac{\alpha}{2}}}{d}.$$

Stąd dla

$$n \geq \left(\frac{\sigma u_{1-\frac{\alpha}{2}}}{d} \right)^2$$

przedział ufności (3) będzie nie dłuższy niż $2d$.

Zadanie wyznaczenia liczności próby, gwarantującej żądaną dokładność estymacji przedziałowej na ustalonym poziomie ufności, znacznie się komplikuje w przypadku modelu II i III z podrozdziału 9.1. Przeanalizujemy najpierw przypadek modelu III.

Model III. Długość przedziału (6) wynosi $l_n = \frac{2S}{\sqrt{n}}u_{1-\frac{\alpha}{2}}$ i jest wielkością losową - zależy od S . W konsekwencji, wyznaczając n z warunku by przedział ufności dla średniej (6) na poziomie ufności $1 - \alpha$ był nie dłuższy niż $2d$:

$$l_n = \frac{2S}{\sqrt{n}}u_{1-\frac{\alpha}{2}} \leq 2d$$

otrzymujemy

$$n \geq S^2 \left(\frac{u_{1-\frac{\alpha}{2}}}{d} \right)^2, \quad (8)$$

czyli wzór, który niestety nie odpowiada nam na pytanie jak liczną próbę potrzebujemy, bo przed przeprowadzeniem eksperymentu nie wiemy jaką otrzymamy wartość estymatora S^2 . Problem ten można obejść godząc się na rozwiązanie przybliżone. Najpierw pobieramy próbę wstępną $X_1^w, X_2^w, \dots, X_{n_0}^w$ o liczności $n_0 \geq 30$ i obliczamy dla niej wariancję, oznaczając ją S_0^2 :

$$S_0^2 = \frac{1}{n_0 - 1} \sum_{i=1}^{n_0} (X_i^w - \bar{X}_0)^2, \quad \text{gdzie } \bar{X}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} X_i^w. \quad (9)$$

Następnie w (8) zastępujemy S^2 przez S_0^2 , otrzymując wzór przybliżony

$$n \geq S_0^2 \left(\frac{u_{1-\frac{\alpha}{2}}}{d} \right)^2,$$

który, po uwzględnieniu założenia modelu III ($n \geq 100$), przyjmuje postać

$$n \geq \max \left\{ 100, S_0^2 \left(\frac{u_{1-\frac{\alpha}{2}}}{d} \right)^2 \right\}. \quad (10)$$

Ponadto, do konstrukcji poszukiwanego przedziału ufności, możemy wykorzystać obserwacje z próby wstępnej i dołosować do niej próbę o liczności jedynie n_1 , gdzie

$$n_1 \geq \max \left\{ 100, S_0^2 \left(\frac{u_{1-\frac{\alpha}{2}}}{d} \right)^2 \right\} - n_0.$$

Model II. Podejście przybliżone, polegające na zastąpieniu zmiennej losowej S^2 przez wariancję S_0^2 wyznaczoną na podstawie próby wstępnej, możemy zastosować także w przypadku modelu II z podrozdziału 9.1. Korzystając z (4) i powtarzając poprzednie rozumowanie otrzymujemy

$$n \geq S_0^2 \left(\frac{t_{1-\frac{\alpha}{2}, n_0-1}}{d} \right)^2,$$

gdzie S_0^2 jest dane wzorem (9), z tym, że teraz nie wymagamy by $n_0 \geq 30$. Jeśli dla tak wyznaczonego n mamy

- $n \leq n_0$, to liczność próby wstępnej jest wystarczająca by otrzymać zadaną precyzję estymacji przedziałowej, więc szukany przedział ufności konstruujemy na jej podstawie;
- $n > n_0$, to do próby wstępnej dołozowujemy tyle elementów by po ich dołączeniu otrzymać próbę o liczności nie mniejszej niż n .

Wzory wyprowadzone powyżej i inne przydatne wzory zostały zgromadzone w pliku *Wzory ze statystyki*.

Przykład 9.1. Dział kontroli jakości w zakładach chemicznych chce oszacować średnią wagę proszku do prania sprzedawanego w pudełkach o nominalnej wadze 3 kg. Pobrano w tym celu próbkę losową 7 pudełek i otrzymano wyniki (w kg): 2.93, 2.97, 3.05, 2.91, 3.02, 2.87, 2.92. Wiadomo, że rozkład wagi pudełka do prania jest normalny.

(a) Na poziomie ufności 0,95 zbudować przedział ufności dla średniej wagi pudełka proszku do prania.

(b) Na poziomie ufności 0,95 zbudować przedział ufności dla odchylenia standardowego wagi pudełka proszku do prania.

(c) Jak liczną próbkę pudełek proszku należy pobrać, aby z maksymalnym błędem 50 g wyznaczyć przedział ufności na poziomie ufności 0,95 dla średniej wagi pudełka proszku do prania?

Rozwiązanie przykładu 9.1:

Oznaczmy: X -waga proszku do prania. Z treści zadania wiemy, że X ma rozkład normalny, ale parametrów tego rozkładu nie znamy. Zapisujemy to następująco

$$X \sim \mathcal{N}(\mu, \sigma^2), \text{ gdzie } \mu \text{ i } \sigma \text{ są nieznanne.}$$

Stąd widzimy, że będzie nam pasować model oznaczony jako model II w tabeli Przedziały ufności a także model II w tabeli Wyznaczanie niezbędnej ilości pomiarów do próby.

Wpisujemy dane do R:

```
> waga.proszku <- c(2.93, 2.97, 3.05, 2.91, 3.02, 2.87, 2.92)
```

(a) Aby wyznaczać przedział ufności dla średniej wagi pudełka patrzymy na tabelę Przedziały ufności, dokładniej na jej lewą kolumnę (bo interesuje nas przedział ufności dla średniej, a nie dla wariancji) i na wiersz z modelem II. Widzimy, że możemy użyć funkcji `t.test()`.

```
> t.test(x=waga.proszku, conf.level=0.95)$conf.int
```

Otrzymujemy następujący przedział: (2.89; 3.01)

(b) Aby wyznaczyć przedział ufności dla odchylenia standardowego wagi pudełka patrzymy na tabelę Przedziały ufności, dokładniej na jej prawą kolumnę (bo interesuje nas przedział ufności nie dla średniej ale dla odchylenia standardowego) i na wiersz z modelem II. Widzimy tam następujący wzór

$$\frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}.$$

Musimy jeszcze wszystkie jego strony spierwiastkować, aby otrzymać wzór na przedział ufności dla odchylenia standardowego (czyli dla σ):

$$\sqrt{\frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}}.$$

Rachunki przeprowadzamy wykorzystując R. Policzmy najpierw licznik wyrażen znajdujących się pod pierwiastkiem:

```
> licznik <- (7-1)*var(waga.proszku)
```

a następnie mianowniki tych wyrażen (poziom ufności wynosi $1 - \alpha = 0,95$, więc $\alpha = 0,05$):

```
> mianownik.lewy <- qchisq(1-0.05/2,7-1)
```

```
> mianownik.prawy <- qchisq(0.05/2,7-1)
```

Ostatecznie otrzymujemy przedział o następujących końcach

```
> sqrt(licznik/mianownik.lewy)
```

```
> sqrt(licznik/mianownik.prawy)
```

czyli (0.04121;0.1408).

(c) Teraz patrzmy na tabelę Wyznaczanie niezbędnej ilości pomiarów do próby, na model II. Szukamy n a mamy dane $d = 50g = 0,05kg$ i $1 - \alpha = 0,95$ (zatem $\alpha = 0,05$). Korzystamy ze wzoru

$$n \geq \left(t_{1-\alpha/2, n_0-1} \frac{s_0}{d} \right)^2,$$

gdzie liczność próby wstępnej to $n_0 = 7$ (to ta próba podana w treści zadania) a jej odchylenie standardowe s_0 policzymy używając funkcji `sd(waga.proszku)`.

Prawa strona powyższego wzoru jest równa

```
> (qt(1-0.05/2,7-1)*sd(waga.proszku)/0.05)^2
```

czyli 9.79649. Zatem potrzebujemy próbę składającą się z co najmniej 10 pomiarów wagi proszku do prania.

Przykład 9.2. 88 spośród 400 osób, które w 2019 roku nabyły nowy samochód sportowy, stwierdziło, że jakość samochodu jest lepsza niż oczekiwali.

(a) Na poziomie ufności 0,95 zbudować przedział ufności dla badanej frakcji.

(b) Jak liczną próbę należałoby pobrać, aby z maksymalnym błędem $\pm 2\%$ móc oszacować badaną frakcję na poziomie ufności 0,95?

Rozwiązanie przykładu 9.2:

Mamy tutaj do czynienia z rozkładem dwupunktowym: osoba, która w 2019 roku nabyła nowy samochód sportowy może stwierdzić, że jakość zakupionego samochodu jest lepsza niż oczekiwała lub tego nie stwierdzić. Oznaczmy:

$$X = \begin{cases} 1 & \text{jeśli osoba stwierdzi, że jakość zakupionego samochodu jest lepsza niż tego oczekiwała,} \\ 0 & \text{jeśli osoba nie stwierdzi, że jakość zakupionego samochodu jest lepsza niż tego oczekiwała,} \end{cases}$$

zaś p niech będzie prawdopodobieństwem trafienia na osobę, która uzna, że jakość zakupionego samochodu jest lepsza niż tego oczekiwała, tzn. $p = P(X = 1)$. Inaczej mówiąc p jest frakcją osób, które uznają, że jakość zakupionego samochodu jest lepsza niż tego oczekiwali.

(a) Patrzymy na tabelę Przedziały ufności, na wiersz z modelem IV. Zauważamy, że n jest na tyle duże, że możemy użyć `prop.test()`. Rzeczywiście

- liczba sukcesów $n\hat{p} = 88 > 5$,
- liczba porażek $n\hat{q} = 400 - 88 > 5$.

Zatem piszemy

```
> prop.test(x=88,n=400,conf.level=0.95)$conf.int
i otrzymujemy szukany przedział ufności: (0.1810175;0.2644723).
```

(a) Patrzymy na tabelę Wyznaczanie niezbędnej ilości pomiarów do próby, na wiersz z modelem IV. Szacunkowy procent p_0 możemy wyznaczyć na podstawie danych z zadania, więc użyjemy wzoru pierwszego z tam podanych:

$$n \geq u_{1-\alpha/2}^2 \frac{p_0 q_0}{d^2},$$

gdzie p_0 to szacunkowy procent a $q_0 = 1 - p_0$. Oczywiście $p_0 = 88/400$. Ponato $1 - \alpha$ to poziom ufności, zatem $1 - \alpha = 0,95$, co daje $\alpha = 0,05$. Maksymalny błąd to $d = 2\% = 0,02$. Reasumując

```
> p0 <- 88/400
> q0 <- 1-p0
> d <- 0.02
```

i prawa strona powyższej nierówności wynosi

```
> qnorm(1-0.05/2)^2*p0*q0/(d^2)
czyli 1647.986. Potrzebna jest próba o licznosci co najmniej 1648.
```