Wiadomości Statystyczne. The Polish Statistician, 2022, vol. 67, 2, 21–35 DOI: 10.5604/01.3001.0015.7718
Edukacja statystyczna / Statistical education
Zgłoszony/submitted: 6.09.2021, poprawiony/revised: 21.01.2022, zaakceptowany/accepted: 13.02.2022

O różnych kryteriach konstrukcji przedziałów ufności na przykładzie rozkładu normalnego

Stanisław Jaworski[a]

Streszczenie. W artykule opisano różne kryteria konstrukcji przedziałów ufności oparte na kontroli błędu przeszacowania i niedoszacowania parametru, jak również na zadanej z góry precyzji oszacowania tego parametru. Odpowiednie konstrukcje zilustrowano przedziałami ufności dla parametrów rozkładu normalnego. Zamieszczone w pracy przykłady pokazują wpływ zadanych kryteriów na postać przedziałów ufności, a wykonane obliczenia przedstawiają zależność szerokości danego przedziału od rozmiaru próby i odwrotnie – rozmiaru próby od zadanej szerokości tego przedziału. Zagadnienie doboru wielkości próby przy zadanej z góry precyzji oszacowania parametru zaprezentowano z uwzględnieniem rozwiązań, które nie zależą od szacunkowych wartości nieznanego parametru. Celem artykułu jest pokazanie, w jaki sposób można rozszerzyć temat konstrukcji przedziałów ufności w edukacji statystycznej.

Słowa kluczowe: dobór wielkości próby, poziom ufności, precyzja oszacowania, przedziały ufności
JEL: C13

On various criteria for the construction of confidence intervals on the example of normal distribution

Abstract. The paper describes various criteria for the construction of confidence intervals based on the control of the error of overestimation or underestimation of a parameter and on the pre-determined precision of the parameter estimation. The respective constructions are illustrated by confidence intervals for the parameters of the normal distribution. Examples presented in the paper show the influence of some selected criteria on the formula of the appropriate confidence interval. The performed calculations show the dependence of the width of a given interval on the size of the sample, and vice versa – the size of the sample on the pre-determined width of this interval. The problem of selecting the sample size with a pre-determined precision of the estimation was presented taking into account solutions that do not depend on the estimated values of the unknown parameter. The aim of the study is to show how the problem of confidence intervals construction can be presented more comprehensively in statistical education.

Keywords: sample size selection, confidence level, precision of estimation, confidence intervals

[a] Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Wydział Zastosowań Informatyki i Matematyki, Katedra Ekonometrii i Statystyki, Polska / Warsaw University of Life Sciences, Faculty of Applied Informatics and Mathematics, Department of Econometrics and Statistics, Poland. ORCID: https://orcid.org/0000-0002-6169-2886. E-mail: stanislaw_jaworski@sggw.edu.pl.

1. Wprowadzenie

Przedziały ufności, wprowadzone do wnioskowania statystycznego przez Jerzego Spławę-Neymana, należą do najważniejszych narzędzi statystycznych. Zasadnicza praca o przedziałach ufności pochodzi z 1937 r. (Neyman, 1937), aczkolwiek to wcześniejsza publikacja – artykuł zamieszczony w „Journal of the Royal Statistical Society" (Neyman, 1934) – została uznana za jedno z największych dwudziestowiecznych osiągnięć w dziedzinie metodologii statystyki (Kotz i Johnson, 1992). Przełomowy artykuł z 1934 r. powstał na bazie monografii z 1933 r. opracowanej w ramach współpracy Neymana z Instytutem Spraw Społecznych (Krzyśko i in., 2018).

Neyman (1934) zaproponował dwa kryteria konstrukcji przedziału ufności:

1. przedział musi dać się wyznaczyć;
2. przedział powinien być jak najwęższy.

Pierwsze kryterium wyraża pogląd Neymana, że statystyka powinna mieć wymiar praktyczny. Drugie kryterium, wynikające z postulatu, że błąd estymacji powinien być jak najmniejszy, można modyfikować poprzez uwzględnienie z jednej strony ryzyka przeszacowania i niedoszacowania parametru, z drugiej zaś – doboru wielkości próby gwarantującej zadaną precyzję oszacowania. Te modyfikacje prowadzą do kolejnych interesujących kryteriów konstrukcji przedziałów ufności, które są przedmiotem niniejszego artykułu i zostaną omówione, a następnie zilustrowane na przykładzie modelu gaussowskiego. Zamieszczone obliczenia uwzględniają praktyczne możliwości ich realizacji i pokazują wpływ, jaki na konstrukcję przedziału ufności ma kontrola jego szerokości oraz błędu polegającego na przeszacowaniu lub niedoszacowaniu parametru. Wynika z tego, że praktyczne zastosowanie konkretnego przedziału ufności powinno być zawsze poprzedzone pytaniem, czy przedział ten jest odpowiedni dla badanego zagadnienia oraz czy rozmiar próby, którą dysponujemy, jest wystarczający do uzyskania pożądanej precyzji oszacowań. Dla praktyków te pytania są bardzo ważne, dlatego wydaje się, że uwzględnienie omówionych konstrukcji w edukacji statystycznej jest jak najbardziej zasadne.

2. Przedział ufności

Podstawą konstrukcji przedziału ufności jest n-wymiarowy wektor losowy (wektor obserwacji) $\boldsymbol{X}$ o rozkładzie prawdopodobieństwa P_θ. Rozkład ten zależy od nieznanego parametru θ. Wiadomo jedynie, że parametr ten należy do pewnego znanego zbioru Θ, nazywanego *przestrzenią parametrów*. Na podstawie realizacji wektora losowego $\boldsymbol{X}$ można oszacować ten parametr za pomocą przedziału ufności. Klasyczna definicja *przedziału ufności* jest następująca:

Definicja. Przedział $(\theta_L(\boldsymbol{X}), \theta_R(\boldsymbol{X}))$, taki że

$$P_\theta\{\theta \in (\theta_L(\boldsymbol{X}), \theta_R(\boldsymbol{X}))\} \geq \delta, \forall\theta \in \Theta, \quad (1)$$

nazywamy przedziałem ufności dla parametru θ, na poziomie ufności $\delta \in (0, 1)$.

W praktyce przyjmuje się, że δ jest ustaloną wartością, najczęściej równą 0,95 lub 0,99. Podana definicja dotyczy parametru jednowymiarowego. Gdyby parametr θ był wielowymiarowy, podana definicja nie zmieniłaby się, poza tym że słowo *przedział* zostałoby zastąpione słowem *zbiór* lub *obszar*. W podanej definicji prawdopodobieństwo $P_\theta\{\theta \in (\theta_L(\boldsymbol{X}), \theta_R(\boldsymbol{X}))\}$ jest rozumiane następująco:

$$P_\theta\left\{\boldsymbol{X} \in \{\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}: \theta \in (\theta_L(\boldsymbol{x}), \theta_R(\boldsymbol{x}))\}\right\}. \quad (2)$$

Oznacza to, że w przeciwieństwie do wektora $\boldsymbol{X}$ parametr θ nie jest losowy. Dlatego mówimy o prawdopodobieństwie pokrycia parametru θ przez przedział ufności, a nie o prawdopodobieństwie, że parametr θ należy do tego przedziału. Chociaż można uznać, że obie interpretacje są logicznie równoważne, to druga sugeruje, że θ jest zmienną losową, co nie jest prawdą. Przypadek modelu statystycznego, w którym θ jest zmienną losową, prowadzi do pojęcia *bayesowskiego przedziału wiarygodności.*

3. Kryteria konstrukcji przedziału ufności

Wnioskując na podstawie przedziału ufności, gdy wektor losowy $\boldsymbol{X}$ zrealizuje się jako $\boldsymbol{x} \in \mathbb{R}^n$, możemy mieć do czynienia z jedną z trzech niewidocznych dla nas sytuacji:
1. $\theta \in (\theta_L(\boldsymbol{x}), \theta_R(\boldsymbol{x}))$;
2. $\theta \leq \theta_L(\boldsymbol{x})$;
3. $\theta \geq \theta_R(\boldsymbol{x})$.

W każdej z nich wyprowadzamy ten sam wniosek, że prawdziwa i nieznana wartość parametru θ została pokryta przez przedział $(\theta_L(\boldsymbol{x}), \theta_R(\boldsymbol{x}))$. Dlatego jeżeli zajdzie pierwsza sytuacja, wyprowadzony wniosek jest poprawny. W drugiej mówimy o błędzie przeszacowania, a w trzeciej – o błędzie niedoszacowania. Przy poziomie ufności δ ryzyko popełnienia błędu przeszacowania lub niedoszacowania wynosi co najwyżej $1 - \delta$. Jeżeli oznaczymy ryzyko przeszacowania jako α_p, a ryzyko niedoszacowania jako α_n, to zgodnie z ideą konstrukcji przedziału ufności mamy:

$$\alpha_p + \alpha_n \leq 1 - \delta. \quad (3)$$

Formalnie $\alpha_p = P_\theta\{\theta \leq \theta_L(\boldsymbol{X})\}$ i $\alpha_n = P_\theta\{\theta \geq \theta_R(\boldsymbol{X})\}$.

Najpopularniejszym kryterium konstrukcji przedziału ufności jest symetryczny podział ryzyka błędu. Zgodnie z tym kryterium funkcje $\theta_L(\boldsymbol{x})$ oraz $\theta_R(\boldsymbol{x})$ dobiera się tak, aby zachodziła równość $P_\theta\{\theta \leq \theta_L(\boldsymbol{X})\} = P_\theta\{\theta \geq \theta_R(\boldsymbol{X})\}$. W tym przypadku ryzyko przeszacowania i ryzyko niedoszacowania są takie same: $\alpha_p = \alpha_n$. W przypadku gdy wektor losowy $\boldsymbol{X}$ ma rozkład typu ciągłego, można przyjąć $\alpha_p = \alpha_n = (1 - \delta)/2$.

Zdarzają się sytuacje, w których z powodów merytorycznych bardziej wskazany jest niesymetryczny podział ryzyka błędu. Na przykład nie chcąc popełnić błędu przeszacowania, należy przyjąć $\alpha_p = 0$. Konsekwencją takiego założenia jest przyjęcie $\theta_L \equiv \inf \Theta$. Jeżeli parametr θ jest średnią w rozkładzie normalnym, wówczas $\theta_L \equiv -\infty$, a jeżeli oznacza prawdopodobieństwo sukcesu w rozkładzie dwumianowym, to $\theta_L \equiv 0$. Innymi słowy, nie ma niebezpieczeństwa przeszacowania, ponieważ lewy koniec przedziału zawsze znajdzie się poniżej nieznanej wartości parametru, bez względu na obserwacje otrzymane w doświadczeniu. Uzyskujemy przedział, w którym wyznaczany jest tylko jego prawy koniec. Taki przedział ufności nazywany jest *jednostronnym przedziałem ufności* (dokładniej: *prawostronnym*). Przedział lewostronny uzyskuje się dla $\alpha_n = 0$. Wtedy $\theta_R \equiv \sup \Theta$, co oznacza brak możliwości niedoszacowania parametru.

Bardzo istotnym elementem we wnioskowaniu opartym na przedziałach ufności jest, wskazana przez Neymana (1934), szerokość przedziału ufności. Dwa wyżej wymienione kryteria kładły nacisk jedynie na ryzyko błędów, co nie zapewniało uzyskiwania najwęższych przedziałów. Minimalizacja szerokości przedziału ufności polega na takim doborze funkcji $\theta_L(\boldsymbol{x})$ i $\theta_R(\boldsymbol{x})$, aby dla wszystkich możliwych obserwacji $\boldsymbol{x}$ szerokość przedziału, tzn. różnica $\theta_R(\boldsymbol{x}) - \theta_L(\boldsymbol{x})$, była najmniejsza. Gdy nie jest to możliwe, można przyjąć nieco słabsze kryterium minimalizacji przeciętnej szerokości przedziału ufności, czyli taki dobór funkcji $\theta_L(\boldsymbol{x})$ i $\theta_R(\boldsymbol{x})$, dla których wielkość

$$E_\theta(\theta_R(\boldsymbol{X}) - \theta_L(\boldsymbol{X})) \tag{4}$$

jest najmniejsza dla każdego $\theta \in \Theta$.

Na podstawie szerokości przedziału ufności można przyjąć kryterium, którego spełnienie zależy nie tyle od końców przedziału ufności, ile od rozmiaru próby. Nawet najwęższy przedział ufności może w praktyce nie być dostatecznie wąski. Dlatego też rozważa się dwa podejścia polegające na konstrukcji przedziałów ufności o zadanej szerokości, w szczególności takiej, że:

1. średnia szerokość przedziału ufności ma nie przekraczać zadanej wartości $d > 0$, tzn.

$$E_\theta\big(\theta_R(\boldsymbol{X}) - \theta_L(\boldsymbol{X})\big) \leq d, \forall \theta \in \Theta; \tag{5}$$

2. z dużym prawdopodobieństwem szerokość przedziału ufności ma nie przekraczać zadanej wartości $d > 0$, czyli

$$P_\theta\{\theta_R(\boldsymbol{X}) - \theta_L(\boldsymbol{X}) \leq d\} \geq \gamma, \forall\theta \in \Theta, \tag{6}$$

   gdzie γ jest daną wartością bliską 1.

4. Konstrukcja przedziału ufności

Zasadnicza konstrukcja przedziału ufności polega na znalezieniu takiego zbioru A_θ, dla którego zachodzi $P_\theta\{\boldsymbol{X} \in A_\theta\} \geq \delta$, dla każdej wartości parametru $\theta \in \Theta$. Następnie dla dowolnej realizacji $\boldsymbol{x}$ wektora losowego $\boldsymbol{X}$ definiuje się następujący zbiór:

$$C(\boldsymbol{x}) = \{\theta \in \Theta: \boldsymbol{x} \in A_\theta\}. \tag{7}$$

Zbiór losowy $C(\boldsymbol{X})$ jest zbiorem ufności na poziomie ufności δ. W praktyce dla jednowymiarowego parametru θ przyjmuje on zazwyczaj postać przedziału.

Przy konstrukcji przedziału ufności przydaje się funkcja centralna $t(\boldsymbol{x}, \theta)$ (Krzyśko, 1996). Zgodnie z definicją tej funkcji $t(\boldsymbol{X}, \theta)$ jest zmienną losową o rozkładzie niezależnym od parametru $\theta \in \Theta$ oraz dla każdej realizacji $\boldsymbol{x}$ wektora losowego $\boldsymbol{X}$ odwzorowanie $\theta \rightarrow t(\boldsymbol{x}, \theta)$ jest monotoniczną funkcją parametru $\theta \in \Theta$. Stąd

1. można wyznaczyć takie dwie liczby $t_1, t_2 \in \mathbb{R}$, że

$$P_\theta\{t_1 < t(\boldsymbol{X}, \theta) < t_2\} \geq \delta \tag{8}$$

   dla dowolnego $\theta \in \Theta$;
2. można dobrać takie dwie funkcje $\theta_L(\boldsymbol{x})$ i $\theta_R(\boldsymbol{x})$, dla których zachodzi równoważność

$$t_1 < t(\boldsymbol{x}, \theta) < t_2 \Leftrightarrow \theta \in (\theta_L(\boldsymbol{x}), \theta_R(\boldsymbol{x})). \tag{9}$$

Wtedy przedział $(\theta_L(\boldsymbol{X}), \theta_R(\boldsymbol{X}))$ jest przedziałem ufności dla parametru $\theta \in \Theta$, ponieważ

$$P_\theta\{\theta \in (\theta_L(\boldsymbol{x}), \theta_R(\boldsymbol{x}))\} = P_\theta\{t_1 < t(\boldsymbol{X}, \theta) < t_2\} \geq \delta \tag{10}$$

dla dowolnego $\theta \in \Theta$. Jeżeli znamy funkcję centralną, to konstrukcja przedziału ufności jest dość prosta. Trzeba jednak wiedzieć, jak taką funkcję skonstruować.

W tym przypadku wymagana jest dobra znajomość rozkładów prawdopodobieństwa zmiennych losowych.

Poza omówionymi wyżej kryteriami są jeszcze inne, związane z takimi pojęciami, jak *jednostajnie najdokładniejsze zbiory ufności*, *nieobciążone zbiory ufności* oraz *ekwiwariantne zbiory ufności*. Ich konstrukcja wymaga znajomości teorii weryfikacji hipotez. Wiele informacji na temat tych koncepcji można znaleźć w książce Lehmanna (1968) lub Bartoszewicza (1996).

5. Przykłady

Zakładamy, że próba $X_1, X_2, \dots, X_n$ składa się z niezależnych zmiennych losowych o tym samym rozkładzie normalnym $N(\mu, \sigma^2)$. Przy powyższym założeniu, przyjmując oznaczenie

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \tag{11}$$

można udowodnić, że:

1. zmienna losowa $\frac{(n-1)S^2}{\sigma^2}$ ma rozkład chi-kwadrat z $n-1$ stopniami swobody;
2. zmienna losowa $\frac{\bar{X}-\mu}{S}\sqrt{n}$ ma rozkład t-Studenta o $n-1$ stopniach swobody.

Z podanych faktów wynika, że przy oznaczeniu $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ funkcje

$$t_1(\boldsymbol{X}, \mu) = \frac{\bar{X}-\mu}{S}\sqrt{n} \text{ oraz } t_2(\boldsymbol{X}, \sigma^2) = \frac{(n-1)S^2}{\sigma^2} \tag{12}$$

są funkcjami centralnymi. Pierwsza jest monotoniczna ze względu na parametr μ, a druga – ze względu na σ^2. Obie też mają rozkłady niezależne od parametrów μ i σ^2. Wektor $\theta = (\mu, \sigma^2)$ można zatem traktować jako parametr wielowymiarowy. Zwróćmy uwagę na to, że do tego miejsca parametr był jednowymiarowy. Przejście na dwuwymiarowość nie zmienia istoty rzeczy, co zostanie pokazane w przykładach. Niemniej, żeby uniknąć nieporozumień, zostanie przytoczona definicja przedziału ufności dla przypadku, gdy $\theta = (\theta_1, \theta_2)$.

Definicja. Przedział $(\theta_L(\boldsymbol{X}), \theta_R(\boldsymbol{X}))$, taki że

$$P_\theta\{\theta_1 \in (\theta_L(\boldsymbol{X}), \theta_R(\boldsymbol{X}))\} \geq \delta, \forall\theta \in \Theta, \tag{13}$$

nazywamy przedziałem ufności dla parametru θ_1, na poziomie ufności $\delta \in (0, 1)$.

Analogicznie definiuje się przedział ufności dla θ_2. W obu przypadkach przestrzeń parametrów Θ jest podzbiorem przestrzeni $\mathbb{R}^2$. W sytuacji, w której przedział ufności jest wyznaczany dla parametru θ_1, parametr θ_2 nazywa się *parametrem zaburzającym* lub *zakłócającym*. I na odwrót, gdy wyznaczamy przedział ufności dla θ_2, parametrem zakłócającym jest θ_1. O znaczeniu takich parametrów w estymacji lub weryfikacji hipotez statystycznych można przeczytać np. w pracach Lehmanna (1968, 1991).

Korzystając z funkcji centralnej $t_1(\boldsymbol{X},\mu)$, budujemy przedział ufności dla parametru μ. Niech t_ν oznacza kwantyl rzędu ν rozkładu t-Studenta o $n-1$ stopniach swobody. Dla ustalonego $\delta_1 \in [0,1]$ mamy

$$P_{\mu,\sigma^2}\left\{t_{\delta_1} < t_1(\boldsymbol{X},\mu) < t_{\delta+\delta_1}\right\} = \delta, \forall(\mu \in R, \sigma^2 \in R_+). \tag{14}$$

Ponieważ

$$t_{\delta_1} < t_1(\boldsymbol{X},\mu) < t_{\delta+\delta_1} \Leftrightarrow \mu \in \left(\bar{X} - t_{\delta+\delta_1}\frac{S}{\sqrt{n}}, \bar{X} - t_{\delta_1}\frac{S}{\sqrt{n}}\right), \tag{15}$$

przedział ufności dla wartości oczekiwanej μ na poziomie ufności δ ma postać

$$\left(\bar{X} - t_{\delta+\delta_1}\frac{S}{\sqrt{n}}, \bar{X} - t_{\delta_1}\frac{S}{\sqrt{n}}\right). \tag{16}$$

Stąd wynika również, że dla prawdopodobieństwa przeszacowania α_p zachodzi

$$\alpha_p = P_{\mu,\sigma^2}\left\{\mu \leq \bar{X} - t_{\delta+\delta_1}\frac{S}{\sqrt{n}}\right\} = 1 - (\delta + \delta_1), \tag{17}$$

a dla niedoszacowania α_n –

$$\alpha_n = P_{\mu,\sigma^2}\left\{\mu \geq \bar{X} - t_{\delta_1}\frac{S}{\sqrt{n}}\right\} = \delta_1. \tag{18}$$

Zauważmy, że w rozważanym przypadku, w którym σ^2 jest parametrem zaburzającym, mamy

$$A_\mu = \left\{\boldsymbol{x}: t_{\delta_1} < t_1(\boldsymbol{x},\mu) < t_{\delta+\delta_1}\right\} \tag{19}$$

oraz

$$C(\boldsymbol{x}) = \{\mu : \boldsymbol{x} \in A_\mu\} = \left(\bar{x} - t_{\delta+\delta_1}\frac{S}{\sqrt{n}}, \bar{x} - t_{\delta_1}\frac{S}{\sqrt{n}}\right). \tag{20}$$

Ponieważ rozkład t-Studenta jest symetryczny względem 0, mamy $t_{\delta_1} = -t_{1-\delta_1}$ i przedział ufności dla wartości średniej można zapisać w postaci

$$\left(\bar{X} - t_{\delta+\delta_1}\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\delta_1}\frac{S}{\sqrt{n}}\right). \tag{21}$$

Przy symetrycznym podziale ryzyka przyjmujemy $\delta_1 = \frac{1-\delta}{2}$, ponieważ wtedy

$$\alpha_n = \alpha_p = \tfrac{1-\delta}{2}. \tag{22}$$

Zastępując δ przez $1 - \alpha$, otrzymamy znaną z podręczników postać

$$\left(\bar{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right). \tag{23}$$

Jeżeli przyjmiemy prawdopodobieństwo przeszacowania $1 - (\delta + \delta_1)$ lub niedoszacowania δ_1 równe 0, otrzymamy jednostronny przedział ufności

$$\left(-\infty, \bar{X} + t_\delta\frac{S}{\sqrt{n}}\right) \text{lub} \left(\bar{X} - t_\delta\frac{S}{\sqrt{n}}, \infty\right). \tag{24}$$

Jednostronne przedziały nie są z pewnością najwęższe, ponieważ ich szerokość jest nieograniczona. Aby znaleźć najwęższy przedział ufności, za δ_1 należy przyjąć taką wartość z przedziału $(0, 1 - \delta)$, dla której szerokość przedziału

$$\left(t_{\delta+\delta_1} - t_{\delta_1}\right)\frac{S}{\sqrt{n}} \tag{25}$$

jest najmniejsza. Osiągamy to poprzez minimalizację funkcji

$$l(\delta_1) = t_{\delta+\delta_1} - t_{\delta_1} = F^{-1}(\delta + \delta_1) - F^{-1}(\delta_1), \tag{26}$$

gdzie F oznacza dystrybuantę rozkładu t-Studenta o $n - 1$ stopniach swobody.

Dystrybuanta F oraz jej odwrotność F^{-1} są funkcjami różniczkowalnymi, a zatem

$$\frac{dl(\delta_1)}{d\delta_1} = \frac{1}{f(F^{-1}(\delta+\delta_1))} - \frac{1}{f(F^{-1}(\delta_1))}, \tag{27}$$

gdzie f jest funkcją gęstości rozkładu t-Studenta.

Pochodna funkcji $l(\delta_1)$ zeruje się, gdy

$$F^{-1}(\delta+\delta_1) = -F^{-1}(\delta_1). \tag{28}$$

Z własności kwantyli rozkładu t-Studenta wynika, że jest to możliwe tylko wtedy, gdy $\delta_1 = \frac{1-\delta}{2}$. Oznacza to, że najwęższy przedział ufności dla wartości średniej jest jednocześnie przedziałem o symetrycznym podziale ryzyka błędu. Jest również przedziałem o najmniejszej średniej szerokości tego przedziału, która w tym przypadku wynosi

$$l_n\left(\frac{1-\delta}{2}\right) E_{\mu,\sigma^2}\left(\frac{S}{\sqrt{n}}\right) = \sqrt{2}\,\frac{l_n\left(\frac{1-\delta}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\sqrt{n(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}\,\sigma, \tag{29}$$

gdzie indeks n dopisany do wcześniej zdefiniowanej funkcji l służy podkreśleniu, że funkcja ta zależy od rozmiaru próby poprzez stopnie swobody rozkładu t-Studenta.

Należy zauważyć, że szerokość przedziału ufności jest proporcjonalna do odchylenia standardowego S z próby, a średnia szerokość owego przedziału – do parametru σ. Stąd też, aby kontrolować szerokość przedziału ufności, należy rozważyć szerokość względną (szerokość przedziału podzieloną przez odchylenie standardowe σ). W szczególności szukamy takiej minimalnej wielkości próby, by dla zadanego $d > 0$ zachodził jeden z warunków:

1. średnia względna szerokość przedziału nie przekracza d:

$$\sqrt{2}\,\frac{l_n\left(\frac{1-\delta}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\sqrt{n(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)} \le d, \tag{30}$$

2. z bardzo dużym prawdopodobieństwem względna szerokość przedziału nie przekracza d:

$$P\left\{l_n\left(\frac{1-\delta}{2}\right)\frac{S/\sigma}{\sqrt{n}} \leq d\right\} = \gamma, \tag{31}$$

dla ustalonego $\gamma \in [0, 1]$ bliskiego 1.

Ponieważ zmienna losowa $\sqrt{n-1}\, S/\sigma$ ma rozkład chi z $n-1$ stopniami swobody, drugi warunek można zapisać w postaci

$$l_n\left(\frac{1-\delta}{2}\right)\frac{G_n^{-1}(\gamma)}{\sqrt{(n-1)n}} \leq d, \tag{32}$$

gdzie $G_n^{-1}(\gamma)$ oznacza kwantyl rzędu γ rozkładu chi z $n-1$ stopniami swobody.

Oba warunki można rozwiązać numerycznie ze względu na n. W tabl. 1 podano rozmiary próby dla zadanej dokładności d. W drugiej kolumnie znajdują się rozmiary dla szerokości oczekiwanej, a w trzeciej – dla szerokości, która jest realizowana z prawdopodobieństwem $\gamma = 0{,}9$. W obu przypadkach przyjęto poziom ufności $\delta = 0{,}95$.

Tabl. 1. Rozmiary próby dla zadanej dokładności d

d	Rozmiar próby dla szerokości	
	oczekiwanej	prawdopodobnej
0,45	78	93
0,40	98	116
0,35	128	148
0,30	173	196
0,25	248	276

Źródło: opracowanie własne.

Na przykład dla próby rozmiaru 98 oczekiwana względna szerokość przedziału ufności nie przekracza 40% odchylenia standardowego. Przy tej samej precyzji w drugim przypadku rozmiar próby jest większy. Możemy wtedy powiedzieć, że przy próbie rozmiaru 116 z prawdopodobieństwem 0,9 uzyskujemy przedział ufności o względnej szerokości nieprzekraczającej 40% odchylenia standardowego.

Podobnie można skonstruować przedziały ufności dla wariancji σ^2, przy czym w tym przypadku to μ jest parametrem zaburzającym. Korzystając z funkcji centralnej $t_2(\boldsymbol{x}, \sigma^2)$, otrzymujemy

$$P_{\mu,\sigma^2}\{\chi^2_{\delta_1} \leq t_2(\boldsymbol{X},\sigma^2) \leq \chi^2_{\delta+\delta_1}\} = \delta, \forall(\mu \in R, \sigma^2 \in R_+), \tag{33}$$

gdzie χ^2_ν oznacza kwantyl rzędu ν rozkładu chi-kwadrat z $n-1$ stopniami swobody oraz $\delta_1 \in [0, 1-\delta]$ jest ustaloną dowolnie wartością.

Z równoważności

$$\chi^2_{\delta_1} \leq t_2(\boldsymbol{X},\sigma^2) \leq \chi^2_{\delta+\delta_1} \Leftrightarrow \sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{\delta+\delta_1}}, \frac{(n-1)S^2}{\chi^2_{\delta_1}}\right), \tag{34}$$

wynika, że przedział ufności dla wariancji σ^2 na poziomie ufności δ ma postać

$$\left(\frac{(n-1)S^2}{\chi^2_{\delta+\delta_1}}, \frac{(n-1)S^2}{\chi^2_{\delta_1}}\right). \tag{35}$$

Stąd prawdopodobieństwo przeszacowania

$$\alpha_p = P_{\mu,\sigma^2}\left\{\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\delta+\delta_1}}\right\} = 1-(\delta+\delta_1), \tag{36}$$

a niedoszacowania

$$\alpha_n = P_{\mu,\sigma^2}\left\{\sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{\delta}}\right\} = \delta_1. \tag{37}$$

Analogicznie do przypadku oszacowania wartości średniej, dla $\delta_1 = \frac{1-\delta}{2}$ otrzymamy przedział ufności o symetrycznym podziale ryzyka błędu. W przypadku braku ryzyka przeszacowania otrzymamy przedział

$$\left(0, \frac{(n-1)S^2}{\chi^2_{1-\delta}}\right), \tag{38}$$

a przy braku ryzyka niedoszacowania – przedział

$$\left(\frac{(n-1)S^2}{\chi^2_{1+\delta}}, \infty\right). \tag{39}$$

Najwęższy przedział ufności otrzymuje się dla δ_1 wyznaczonego z równania

$$\frac{1}{f(F^{-1}(\delta+\delta_1))}-\frac{1}{f(F^{-1}(\delta_1))}=0, \tag{40}$$

gdzie F i f oznaczają odpowiednio dystrybuantę i funkcję gęstości rozkładu chi--kwadrat z $n-1$ stopniami swobody (dla $n \geq 2$).

Ponieważ rozkład chi-kwadrat nie jest symetryczny, warunek ten można rozwiązać jedynie numerycznie. W tabl. 2 zawarto szerokości przedziałów ufności o symetrycznym podziale ryzyka błędu oraz przedziałów najwęższych. W ostatniej kolumnie podano prawdopodobieństwa niedoszacowania.

Tabl. 2. Szerokości przedziałów ufności o symetrycznym podziale ryzyka błędu oraz przedziałów najwęższych

Rozmiar próby	Szerokość przedziału		δ_1
	o symetrycznym podziale błędu	najwęższego	
10	0,259	0,221	0,048
20	0,075	0,069	0,044
30	0,038	0,036	0,042
40	0,024	0,023	0,040
50	0,017	0,016	0,039
100	0,006	0,006	0,035

Źródło: opracowanie własne.

Zagadnienie dotyczące przedziałów ufności dla wariancji w odniesieniu do ich zadanej szerokości można rozpatrywać w ujęciu względnym, tzn. w jednostkach szacowanej wariancji. Na przykład warunek, że średnia względna szerokość przedziału dla wariancji ma nie przekroczyć wartości d, zapisujemy następująco:

$$E_{\mu,\sigma^2}\left\{\frac{(n-1)S^2}{\sigma^2}\left(\frac{1}{\chi^2_{\delta_1}}-\frac{1}{\chi^2_{\delta+\delta_1}}\right)\right\}=\left(\frac{1}{\chi^2_{\delta_1}}-\frac{1}{\chi^2_{\delta+\delta_1}}\right)(n-1)\leq d. \tag{41}$$

Natomiast warunek, że z bardzo dużym prawdopodobieństwem (dla $\gamma \in (0,1)$ bliskiego 1) względna szerokość przedziału nie przekracza d, zapisujemy w postaci $P_{\mu,\sigma^2}\left\{\frac{(n-1)S^2}{\sigma^2}\left(\frac{1}{\chi^2_{\delta_1}}-\frac{1}{\chi^2_{\delta+\delta_1}}\right)\right\}\geq\gamma$ lub w postaci równoważnej:

$$\left(\frac{1}{\chi^2_{\delta_1}}-\frac{1}{\chi^2_{\delta+\delta_1}}\right)G_n^{-1}(\gamma)\leq d, \tag{42}$$

gdzie $G_n^{-1}(\gamma)$ oznacza kwantyl rzędu γ rozkładu chi-kwadrat z $n-1$ stopniami swobody.

Oba podane warunki można rozwiązać numerycznie ze względu na rozmiar próby n. Celem jest znalezienie takiego najmniejszego n, dla którego zachodzi pierwszy albo drugi warunek. Należy przy tym pamiętać, że kwantyle $\chi^2_{\delta_1}$ i $\chi^2_{\delta+\delta_1}$ również zależą od rozmiaru próby. Prawdopodobieństwo niedoszacowania δ_1 można uprzednio ustalić lub przyjąć takie, które daje najwęższy przedział ufności. W tym drugim przypadku prawdopodobieństwo to również zależy od rozmiaru próby. Na przykład dla $\delta_1 = \frac{1-\delta}{2}, \delta = 0{,}95$ i $\gamma = 0{,}9$ w tabl. 3 podano wielkości próby dla zadanej szerokości względnej przedziału ufności.

Tabl. 3. Rozmiary próby dla oczekiwanej oraz prawdopodobnej szerokości przedziału ufności

d	Rozmiar próby dla szerokości	
	oczekiwanej	prawdopodobnej
0,45	161	203
0,40	201	248
0,35	260	314
0,30	350	415
0,25	501	578

Źródło: opracowanie własne.

Na przykład dla próby rozmiaru 201 oczekiwana względna szerokość przedziału ufności nie przekracza 40% szacowanej wariancji. Natomiast aby względna szerokość z prawdopodobieństwem 0,9 również nie przekroczyła 40% szacowanej wariancji, rozmiar próby musi być większy i wynosić co najmniej 248. W obu przypadkach prawdopodobieństwo niedoszacowania jest równe prawdopodobieństwu przeszacowania i wynosi 2,5%.

6. Wymiar edukacyjny

Podręczniki oraz skrypty do statystyki dla studentów zawierają wiedzę albo o wysokim stopniu ogólności, jak prace Lehmanna (1968) lub Bartoszewicza (1996), albo przeciwnie – o niewielkim wycinku wnioskowania statystycznego. W obu przypadkach brakuje odniesień do – przedstawionych w niniejszym artykule – koncepcji podziału ryzyka błędu przeszacowania lub niedoszacowania parametru oraz koncepcji optymalnej liczebności próby.

Pierwsza z wymienionych koncepcji jest użyteczna, gdy niesymetryczny podział ryzyka błędu jest lepszy od symetrycznego – wówczas gdy strata wynikająca z niedoszacowania danego parametru jest inna niż wynikająca z przeszacowania. Na przykład lepiej nie doszacować niż przeszacować wartość zagrożoną, która zabezpiecza banki przed utratą płynności finansowej.

Druga koncepcja poszerza, ujęte w programach nauczania, zagadnienie doboru wielkości próby. Standardowe podejście wymaga oszacowania odchylenia standardowego przed ustaleniem rozmiaru próby, co wymusza pobranie próby wstępnej albo korzystanie z wyników innych badań. Działania te mogą okazać się niepraktyczne, a otrzymane oszacowania – nieprecyzyjne. Ponadto dotyczą jedynie przypadku, w którym podział ryzyka błędu pomiędzy przeszacowanie a niedoszacowanie jest symetryczny. Są to wady, których nie ma zaprezentowany w artykule dobór wielkości próby.

Wyznaczenie przedziałów ufności o niesymetrycznym podziale ryzyka błędu oraz dobór optymalnych wielkości nie są skomplikowane. Wartości kwantyli znanych rozkładów prawdopodobieństwa oraz funkcji gamma można wyznaczać w arkuszach kalkulacyjnych oraz w ogólnie dostępnych pakietach statystycznych. Iloraz funkcji gamma w nierówności (30) może powodować problemy numeryczne. Jeżeli jednak iloraz ten będzie odpowiednio liczony (np. przy wykorzystaniu własności rekurencyjnej funkcji gamma), problemy te nie wystąpią.

Należy zwrócić uwagę, że przedstawione przykłady dotyczą szczególnego przypadku rozkładu normalnego, który jest rozkładem typu ciągłego. Konstrukcja przedziałów ufności dla parametrów rozkładu dyskretnego okazuje się pod pewnymi względami trudniejsza. Z tego powodu w edukacji statystycznej popularne stały się asymptotyczne przedziały ufności, np. dla wskaźnika struktury (rozumianego jako odsetek obiektów całej populacji mających interesującą nas własność). Przedziały te nie są jednak przedziałami ufności w myśl podanej definicji i można je z powodzeniem zastąpić przedziałem dokładnym (Zieliński, 2009), a nawet, dokładając pewien mechanizm losowy, poprawić jego właściwości (Zieliński, 2017).

7. Podsumowanie

W artykule przedstawiono przykłady konstrukcji przedziałów ufności na podstawie kryteriów wynikających z kontroli ryzyka przeszacowania i niedoszacowania parametru oraz zagadnienia doboru wielkości próby gwarantującej zadaną precyzję oszacowania. Podane kryteria uwzględniają ewentualny wymóg asymetrii tego ryzyka oraz, w przypadku najkorzystniejszego wyboru optymalnego rozmiaru próby, dają rozwiązania niezależne od szacunkowych wartości nieznanych parametrów. W ten sposób podane konstrukcje mogą, w ramach typowej edukacji statystycznej, poszerzyć zakres zagadnień związanych z konstrukcją przedziałów ufności, a także przyczynić się do ich lepszego zrozumienia.

Bibliografia

Bartoszewicz, J. (1996). *Wykłady ze statystyki matematycznej*. Państwowe Wydawnictwo Naukowe.

Kotz, S., Johnson, N. L. (red.). (1992). *Breakthroughs in Statistics: Volume 2. Methodology and Distribution*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-4380-9.

Krzyśko, M. (1996). *Statystyka matematyczna*. Państwowe Wydawnictwo Naukowe.

Krzyśko, M., Adamczewski, W., Berger, J., Gołata, E., Kruszka, K., Łazowska, B. (2018). *Statystycy polscy. Biogramy*. Główny Urząd Statystyczny. https://bws.stat.gov.pl/BWS/Archiwum/gus_bws_67_Statystycy_polscy_Biogramy.pdf.

Lehmann, E. L. (1968). *Weryfikacja hipotez statystycznych*. Państwowe Wydawnictwo Naukowe.

Lehmann, E. L. (1991). *Teoria estymacji punktowej*. Państwowe Wydawnictwo Naukowe.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, *97*(4), 558–625. https://doi.org/10.2307/2342192.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London: Series A. Mathematical and Physical Sciences*, *236*(767), 333–380. https://doi.org/10.1098/rsta.1937.0005.

Zieliński, R. (2009). Przedział ufności dla frakcji. *Matematyka Stosowana*, *9*(50), 76–90.

Zieliński, W. (2017). The shortest Clopper-Pearson randomized confidence interval for binomial probability. *REVSTAT – Statistical Journal*, *15*(1), 141–153.