

Przedział ufności

95% przedział ufności dla parametru populacji to przedział, dla którego mamy 95% pewności, że zawiera badany parametr.

Typowa sytuacja: nie znamy średniej populacji μ . Estymujemy ją przy pomocy średniej próbkowej $\bar{y}$. Konstruujemy taki przedział o środku $\bar{y}$, który z prawdopodobieństwem 95% zawiera μ .

Konstrukcja przedziału ufności

Założmy, że badana zmienna ma rozkład $N(\mu, \sigma)$, przy czym σ jest znane (założenie mało realistyczne, ale dobre na początek).

Wówczas średnia $\bar{Y}$ z n obserwacji ma rozkład $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Dla standardowego rozkładu normalnego wyliczamy kwantyle rzędu 0.025 oraz 0.975:

$$P(Z > 1.96) = 0.025 \quad P(Z < -1.96) = 0.025$$

Oznaczamy: $Z_{0.025} = 1.96$ Szukane kwantyle dla rozkładu $\bar{Y}$ wynoszą

$$\mu \pm Z_{0.025} \frac{\sigma}{\sqrt{n}}$$

Inaczej:

$$P(\bar{Y} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + 1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

Konstrukcja przedziału ufności

Założmy, że badana zmienna ma rozkład $N(\mu, \sigma)$, przy czym σ nie jest znane. Wówczas średnia $\bar{Y}$ z n obserwacji ma rozkład $N(\mu, \frac{\sigma}{\sqrt{n}})$, ale nie znamy σ i musimy je oszacować przy pomocy odchylenia standardowego próbki (które jest zbliżone do σ , ale zazwyczaj nie jest mu równe).

- Estymujemy σ przy pomocy s (próbkowe odchylenie standardowe)
- Definiujemy *standardowy błąd średniej* (*standard error*) jako

$$SE = \frac{s}{\sqrt{n}}$$

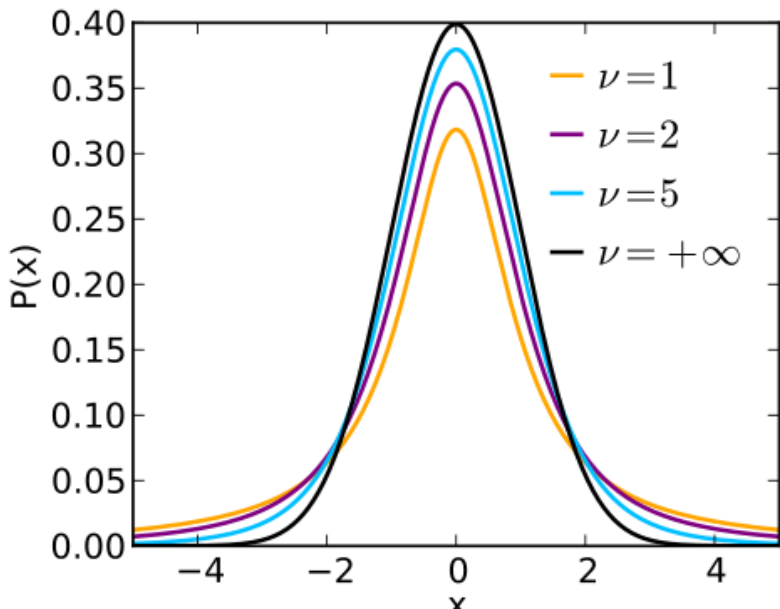
- SE jest estymatorem odchylenia standardowego $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$ średniej $\bar{Y}$

Za nieznaną σ musimy zapłacić pewną cenę:

- 1 estymacja σ wprowadza dodatkową niepewność (nie możemy brać kwantyli z rozkładu normalnego),
- 2 przedziały ufności są szersze niż w przypadku, gdy znamy σ .

Aby rozwiązać problem (1) posługujemy się *rozkładem Studenta / rozkładem t Studenta (t distribution)*. Jest to rodzina ciągłych rozkładów, o gęstościach przypominających standardowy rozkład normalny, ale mających „cięższe ogony”. Zależą od parametru df (degrees of freedom), czyli liczby stopni swobody.

Rozkład t Studenta



Estymując σ za pomocą s , do konstrukcji przedziału ufności bierzemy kwantyle z rozkładu Studenta T (zamiast ze standardowego rozkładu normalnego Z) z $df = n - 1$ stopniami swobody (gdzie n to rozmiar próby).

Przykład 1

Dla jakiej wartości t zachodzi $P(T > t) = 0.025$, gdzie T jest zmienną losową o rozkładzie Studenta z 8 stopniami swobody?

Z Tablicy rozkładu odczytujemy: 2.306. Zwróćmy uwagę, że analogiczna wartość dla zmiennej Z o rozkładzie normalnym, tzn. wartość z dla której zachodzi $P(Z > z) = 0.025$ wynosi 2 (dokładnie 1.96), zgodnie z regułą 68%–95%–99.7%.

Przykład 2

Znajdź dwie symetryczne wartości takie, że między nimi zawiera się 95% masy rozkładu Studenta z 11 stopniami swobody.

Z Tablicy rozkładu odczytujemy: ± 2.201

Przedział ufności na poziomie C w przypadku, gdy σ nie jest znane:

$$\bar{y} \pm t^* SE = \bar{y} \pm t^* \frac{s}{\sqrt{n}}$$

gdzie t^* to taka wartość, że $P(-t^* < T < t^*) = C$ dla zmiennej T o rozkładzie Studenta z $n - 1$ stopniami swobody (gdzie n to rozmiar próby).

Przykład 3

Dla $n = 5$ obserwacji zmierzono $\bar{y} = 31.72$ oraz $s = 8.729$.
Wyznacz 95% oraz 90% przedział ufności dla μ .

Posługujemy się rozkładem Studenta z $5 - 1 = 4$ stopniami swobody. Z tablic odczytujemy:

$$P(T > t^*) = 0.025 \text{ daje } t^* = 2.776$$

$$P(T > t^*) = 0.050 \text{ daje } t^* = 2.132$$

Stąd dostajemy 95% PU: $31.72 \pm 2.776 \cdot \frac{8.729}{\sqrt{5}} = 31.72 \pm 10.84$
oraz 90% PU: $31.72 \pm 2.132 \cdot \frac{8.729}{\sqrt{5}} = 31.72 \pm 8.32$

Gdybyśmy w poprzednim przykładzie znali odchylenie standardowe $\sigma = 8.729$ (a nie szacowali przy pomocy s), to posługivalibyśmy się rozkładem normalnym, czyli

$$P(Z > z^*) = 0.025 \text{ daje } z^* = 1.96$$

$$P(Z > z^*) = 0.050 \text{ daje } z^* = 1.65$$

Stąd dostajemy 95% PU: $31.72 \pm 1.96 \cdot \frac{8.729}{\sqrt{5}} = 31.72 \pm 7.65$

oraz 90% PU: $31.72 \pm 1.65 \cdot \frac{8.729}{\sqrt{5}} = 31.72 \pm 6.44$

Przykład 4

Dla $n = 30$ obserwacji zmierzono $\bar{y} = 31.72$ oraz $s = 8.729$.
Wyznacz 95% oraz 90% przedział ufności dla μ .

Dla rozkładu Studenta z $30 - 1 = 29$ stopniami swobody odczytujemy:

$$P(T > t^*) = 0.025 \text{ daje } t^* = 2.045$$

$$P(T > t^*) = 0.050 \text{ daje } t^* = 1.699$$

$$95\% \text{ PU: } 31.72 \pm 2.045 \cdot \frac{8.729}{\sqrt{30}} = 31.72 \pm 3.26$$

$$90\% \text{ PU: } 31.72 \pm 1.699 \cdot \frac{8.729}{\sqrt{30}} = 31.72 \pm 2.71$$

Podczas, gdy w sytuacji znanego odchylenia standardowego $\sigma = 8.729$ dostajemy:

$$95\% \text{ PU: } 31.72 \pm 1.96 \cdot \frac{8.729}{\sqrt{30}} = 31.72 \pm 3.12$$

$$90\% \text{ PU: } 31.72 \pm 1.65 \cdot \frac{8.729}{\sqrt{30}} = 31.72 \pm 2.63$$

Przykład 5

Zmierzono zawartość witaminy C w kilku próbkach mieszanki sojowo-kukurydzianej. Uzyskano wyniki (w mg/100 g):

26 31 23 22 11 22 14 31

Wyznacz 95% przedział ufności dla zawartości witaminy C.

Dla rozkładu Studenta z 7 stopniami swobody ($n = 8$):

$$P(T > t^*) = 0.025 \text{ daje } t^* = 2.365$$

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_8}{8} = 22.5$$

$$s = \sqrt{\frac{((y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_8 - \bar{y})^2)}{8-1}} = 23.1$$

Stąd 95% przedział ufności to:

$$22.5 \pm 2.365 \cdot \frac{23.1}{\sqrt{8}} = 22.5 \pm 19.3$$