## Estymacja przedziałowa

W estymacji przedziałowej podajemy przedziały ufności dla nieznanych parametrów. *Przedziałem ufności* (ang. confidence interval) dla parametru $\theta$ na poziomie ufności $(1-\alpha)$ nazywamy przedział $(\theta_1, \theta_2)$ spełniający następujące warunki:

♣jego końce $\theta_1=\theta_1(x_1,x_2,...x_n)$; $\theta_2=\theta_2(x_1,x_2,...,x_n)$ są funkcjami próby i nie zależą od szacowanego parametru $\theta$

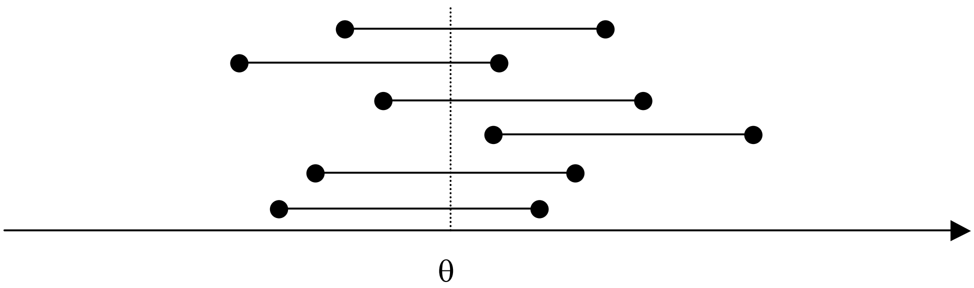
♣prawdopodobieństwo pokrycia przez ten przedział nieznanego parametru jest równe $(1-\alpha)$, co zapisujemy w postaci

$$P(\theta_1(x_1,x_2,...x_n)<\theta<\theta_2(x_1,x_2,...,x_n))=1-\alpha$$

gdzie $\alpha$ jest ustalonym z góry prawdopodobieństwem.

Stosuje się następującą terminologię: $\alpha$ poziom istotności

$1-\alpha$ poziom ufności (ang. confidence level)




Na rysunku powyżej poziom ufności $1-\alpha$ jest równy 5/6.

Do estymacji przedziałowej (dla małych prób, n<30, losowanych z populacji o rozkładzie normalnym)

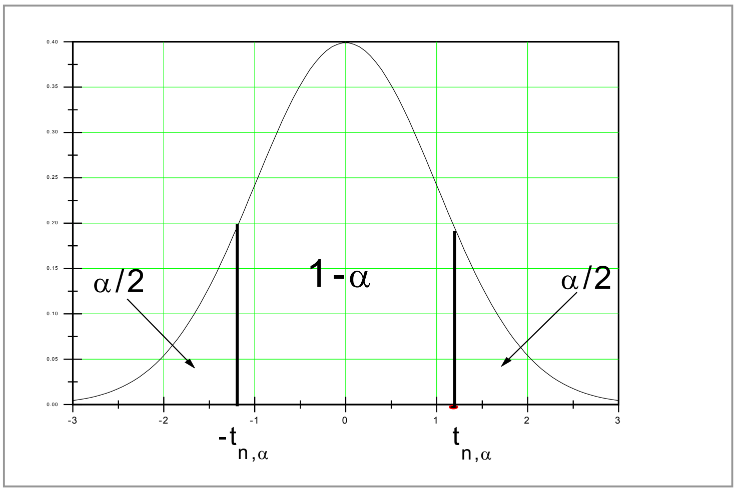
wartości oczekiwanej - stosuje się rozkład Studenta

wariancji i odchylenia standardowego - stosuje się rozkład chi kwadrat

*Estymacja przedziałowa wartości oczekiwanej $\mu$*

Ponieważ statystyka $t=\dfrac{\bar{x}-\mu}{s}\sqrt{n-1}$, gdzie $\bar{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$, $s=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$, n – liczebność próbki, ma rozkład Studenta, więc

$$P(|t|<t_{n,\alpha})=1-\alpha$$




Przekształcając powyższe równanie

$$P\left(\left|\frac{\bar{x}-\mu}{s}\sqrt{n-1}\right|<t_{n,\alpha}\right)=1-\alpha$$

$$P\left(-t_{n,\alpha}<\frac{\bar{x}-\mu}{s}\sqrt{n-1}<t_{n,\alpha}\right)=1-\alpha$$

$$P\left(-\frac{s}{\sqrt{n-1}}t_{n,\alpha} < \overline{x} - \mu < \frac{s}{\sqrt{n-1}}t_{n,\alpha}\right) = 1-\alpha$$

$$P\left(-\overline{x} - \frac{s}{\sqrt{n-1}}t_{n,\alpha} < -\mu < \frac{s}{\sqrt{n-1}}t_{n,\alpha} - \overline{x}\right) = 1-\alpha$$

$$P\left(\overline{x} + \frac{s}{\sqrt{n-1}}t_{n,\alpha} > \mu > \overline{x} - \frac{s}{\sqrt{n-1}}t_{n,\alpha}\right) = 1-\alpha$$

otrzymamy ostatecznie

$$\boxed{P\left(\overline{x} - t_{n,\alpha}\hat{S}_{\overline{x}} < \mu < \overline{x} + t_{n,\alpha}\hat{S}_{\overline{x}}\right) = 1-\alpha} \qquad (\square)$$

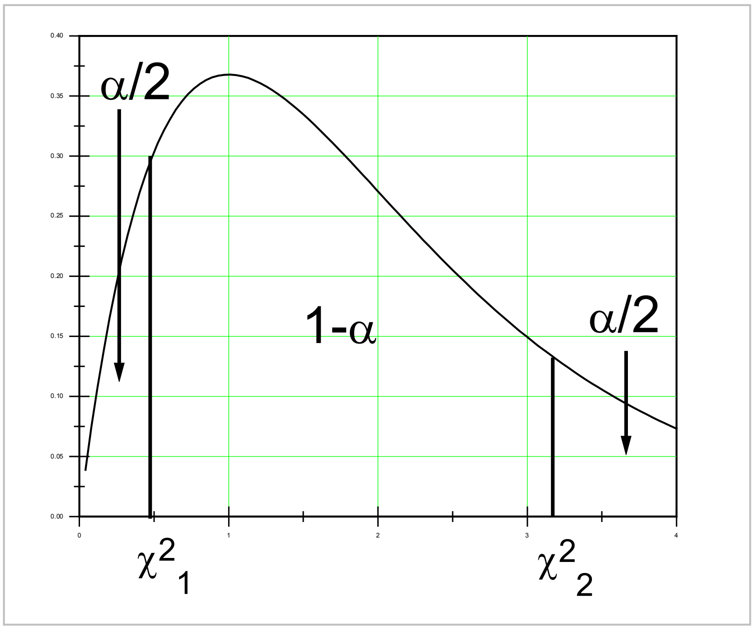
gdzie $\hat{S}_{\overline{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n(n-1)}}$ odchyleniem standardowym średniej arytmetycznej.

Równanie (□) czytamy następująco: (1-α)100% przedziałem ufności dla nieznanej wartości oczekiwanej jest przedział określony podwójną nierównością: $\overline{x} - t_{n,\alpha}\hat{S}_{\overline{x}} < \mu < \overline{x} + t_{n,\alpha}\hat{S}_{\overline{x}}$. Wartości krytyczne $t_{n,\alpha}$ rozkładu Studenta odczytujemy z tablic dla liczby stopni swobody $r=n-1$.

*Estymacja przedziałowa wariancji $\sigma^2$ i odchylenia standardowego $\sigma$*

Ponieważ statystyka $\chi^2 = \frac{nS_x^2}{\sigma^2}$, gdzie $S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$, ma rozkład $\chi^2$ o $r=n-1$ stopniach swobody, to

$$P\left(\chi_1^2 < \chi^2 < \chi_2^2\right) = 1-\alpha$$




Oznacza to, że

$$P\left(\chi_1^2 < \frac{nS_x^2}{\sigma^2} < \chi_2^2\right) = 1-\alpha$$

Po prostym przekształceniu otrzymamy końcowy rezultat

$$P\left(\frac{nS_x^2}{\chi_2^2} < \sigma^2 < \frac{nS_x^2}{\chi_1^2}\right) = 1-\alpha$$

Wartości $\chi_1^2$ i $\chi_2^2$ odczytujemy z tablic dystrybuanty rozkładu $\chi^2$.Wartość $\chi_1^2$ odczytujemy w wierszu odpowiadającym liczbie stopni swobody $r$ i w kolumnie odpowiadającej prawdobodobieństwu ½ α, wartość $\chi_2^2$ odczytujemy dla prawdopodobieństwa 1 - ½ α. Gdy nie znamy wariancji dla populacji, to liczba stopni swobody $r=n-2$.

Dla odchylenia standardowego przedział ufności otrzymamy przez spierwiastkowanie nierówności stojącej pod znakiem prawdopodobieństwa w powyższym wyrażeniu dla wariancji

$$P\left(\sqrt{\frac{nS_x^2}{\chi_2^2}} < \sigma < \sqrt{\frac{nS_x^2}{\chi_1^2}}\right) = 1-\alpha$$

Przykład . Przedziały ufności dla wartości oczekiwanej i odchylenia standardowego.
Ze zbioru 2000 liczb (znajdujących się w pliku gauss.dat) losujemy 10 liczb (początkowo losujemy pozycje tych liczb, a potem wybieramy liczby na tych pozycjach). W programie Mathematica użyto instrukcji:

**gau=Import["gauss.dat"]; gau=Flatten[gau]; ga=Part[gau, Table[Random[Integer,{1,2000}], {10}]]**

Oto przykład wylosowanych liczb (tablica ga):

80, 82, 100, 114, 90, 106, 86, 100, 100, 102.

Dla tej próby: wartość średnia jest równa 96.0000, odchylenie standardowe (pojedynczego pomiaru) 11.0353

| Poziom ufności | Przedział ufności dla wartości oczekiwanej | Przedział ufności dla odchylenia standardowego |
|---|---|---|
| 0.999 | 79.32 – 112.68 | 6.08 – 33.58 |
| 0.99 | 84.66 – 107.34 | 6.82 – 25.13 |
| 0.95 | 88.11 – 103.89 | 7.59 – 20.15 |
| 0.90 | 89.60 – 102.40 | 8.05 – 18.16 |
| 0.80 | 91.17 – 100.83 | 8.64 – 16.22 |
| 0.70 | 92.16 – 99.84 | 9.08 – 15.08 |
| 0.60 | 92.92 – 99.08 | 9.46 – 14.37 |

Poniżej przedstawiono obliczone przedziały ufności dla wartości oczekiwanej (poziom ufności 0.8) w kolejnych 12 próbkach (liczność próbki n=10), wylosowanych z tej samej tablicy 2000 liczb:

| Nr próbki | Średnia arytmetyczna próbki | Przedział ufności dla wartości oczekiwanej | Szerokość przedziału ufności |
|---|---|---|---|
| 1 | 98.60 | 95.93 – 101.27 | 5.34 |
| 2 | 100.59 | 97.68 – 103.51 | 5.83 |
| 3 | 99.99 | 97.26 – 102.73 | 5.47 |
| 4 | 98.40 | 95.21 – 101.59 | 6.38 |
| 5 | 100.60 | 97.05 – 104.15 | 7.10 |
| 6 | 101.00 | 96.73 – 105.27 | 8.54 |
| 7 | 101.40 | 98.79 – 104.01 | 5.22 |
| 8 | 102.60 | 99.10 – 106.10 | 7.00 |
| 9 | 99.50 | 95.67 – 103.33 | 7.66 |
| 10 | 100.60 | 97.51 – 103.69 | 6.18 |
| 11 | 100.00 | 96.29 – 103.71 | 7.42 |
| 12 | 104.40 | 101.93 – 106.87 | 4.94 |

Jak widać to z powyższej tabeli, szerokość przedziału ufności dla wartości oczekiwanej zmienia się od próby do próby, pomimo tego, że liczebność próbek jest taka sama i taki sam jest poziom ufności.