

Estymacja parametrów rozkładu cechy

Estymujemy parametr θ rozkładu cechy X

Próba: X_1, X_2, \dots, X_n

Estymator (punktowy) jest funkcją próby

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

przybliżającą wartość parametru θ

Przedział ufności (estymator przedziałowy)

jest przedziałem o końcach zależnych od próby, który z pewnym z góry zadany prawdopodobieństwem pokrywa nieznaną wartość parametru θ

$$P\{\theta \in (\underline{\theta}(X_1, \dots, X_n), \bar{\theta}(X_1, \dots, X_n))\} = 1 - \alpha$$

Poziom ufności: prawdopodobieństwo $1 - \alpha$

Co wpływa na długość d przedziału ufności?

1. Liczność próby ($n \nearrow \implies d \searrow$)
2. Poziom ufności ($1 - \alpha \nearrow \implies d \nearrow$)
3. Wariancja cechy ($\sigma^2 \searrow \implies d \searrow$)

Rozkład normalny

Estymacja parametrów

Próba (prosta): X_1, X_2, \dots, X_n

Estymator średniej μ — średnia arytmetyczna

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + \dots + X_n}{n}$$

Estymator wariancji σ^2 — wariancja próbkowa

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

Suma kwadratów odchyłeń od średniej

$$\text{var} X = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

Estymator odchylenia standardowego σ

$$S = \sqrt{S^2}$$

Szereg rozdzielczy (dane skumulowane)

Przedział klasowy	Liczebność
$x_0 - x_1$	n_1
$x_1 - x_2$	n_2
\vdots	\vdots
$x_{k-1} - x_k$	n_k
	n

Średnia z próby ($\dot{x}_i = (x_{i-1} + x_i)/2$)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k \dot{x}_i n_i$$

Suma kwadratów odchyleń od średniej

$$\text{var}X = \sum_{i=1}^k (\dot{x}_i - \bar{X})^2 n_i$$

Liczność próby

Jeżeli X_1, \dots, X_n są niezależnymi zmiennymi losowymi takimi, że $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$, to

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

.....

Przykład.

Jakie jest prawdopodobieństwo, że średnia \bar{X} „trafi” bliżej μ niż 0.1σ ?

$$P\{|\bar{X} - \mu| < 0.1\sigma\} = P\{\bar{X} \in (\mu - 0.1\sigma, \mu + 0.1\sigma)\}$$

$$F\left(\frac{(\mu + 0.1\sigma) - \mu}{\sigma/\sqrt{n}}\right) - F\left(\frac{(\mu - 0.1\sigma) - \mu}{\sigma/\sqrt{n}}\right) =$$

$$F(0.1\sqrt{n}) - F(-0.1\sqrt{n}) = 2F(0.1\sqrt{n}) - 1$$

n	1	5	10	50
P	0.07966	0.17694	0.24817	0.52050
n	100	500	1000	1500
P	0.68269	0.97465	0.99843	0.99989

Przedział ufności dla średniej

Wariancja σ^2 jest nieznana

Poziom ufności: $1 - \alpha$

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

$t(\alpha; n - 1)$: wartość krytyczna rozkładu t (Studenta) z ν stopniami swobody

Długość przedziału: $d = 2t(\alpha; n - 1) \frac{S}{\sqrt{n}}$

Przedziały jednostronne

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

$$\left(\bar{X} - t(2\alpha; n - 1) \frac{S}{\sqrt{n}}, \quad \infty \right)$$

Przykład.

Na podstawie próby 1.1, 1.2, 0.8, 0.9, 1.2, 1.3, 1.0, 0.7, 0.8, 1.0 oszacować wartość średnią rozkładu obserwowanej cechy.

$$\bar{x} = \frac{1.1 + 1.2 + \dots + 1.0}{10} = 1.0$$

$$\text{var}X = (1.1 - 1.0)^2 + \dots + (1.0 - 1.0)^2 = 0.36$$

$$s^2 = \frac{0.36}{10 - 1} = 0.04, \quad s = \sqrt{s^2} = 0.2$$

Poziom ufności $1 - \alpha = 0.95$, czyli $\alpha = 0.05$.

$$t(0.05; 9) = 2.2622$$

$$t(0.05; 9) \frac{s}{\sqrt{n}} = 2.2622 \frac{0.2}{\sqrt{10}} = 0.14$$

$$(1 - 0.14, 1 + 0.14) = (0.86, 1.14)$$

Wniosek. Średnia wartość cechy jest jakąś liczbą z przedziału (0.86, 1.14). Zaufanie do tego wniosku wynosi 95%.

Przykład.

Oszacować przeciętną ilość punktów uzyskiwanych na klasówce.

$$n = 300 \quad \sum x_i = 176.566 \quad \sum x_i^2 = 107.845302$$

Populacja:

Słuchacze podstawowego kursu statystyki

Cecha X :

ilość punktów zdobytych na klasówce

Założenie:

cecha X ma rozkład normalny $N(\mu, \sigma^2)$

Zadanie: oszacować parametr μ

Technika statystyczna:

przedział ufności dla średniej
poziom ufności $1 - \alpha = 0.95$

Obliczenia

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{176.566}{300} = 0.589$$

$$\begin{aligned} \text{var} X &= \sum x_i^2 - \frac{1}{n} \left(\sum x_i \right)^2 \\ &= 107.845302 - \frac{176.566^2}{300} = 3.92679 \end{aligned}$$

$$s^2 = \frac{3.92679}{300 - 1} = 0.01313, \quad s = \sqrt{s^2} = 0.11460$$

$$t(0.05; 299) \approx 1.96$$

$$t(0.05; 299) \frac{s}{\sqrt{n}} = 1.96 \frac{0.11460}{\sqrt{300}} = 0.01297$$

$$(0.589 - 0.013, 0.589 + 0.013) = (0.576, 0.602)$$

Odpowiedź: $\mu \in (0.576, 0.602)$

Wniosek. Przeciętna liczba punktów zdobywana na klasówce jest liczbą z przedziału $(0.576, 0.602)$. Zaufanie do tego wniosku wynosi 95%.

Przedział ufności dla wariancji

Średnia μ jest nieznaną

Poziom ufności: $1 - \alpha$

$$\left(\frac{\text{var}X}{\chi^2\left(\frac{\alpha}{2}; n-1\right)}, \frac{\text{var}X}{\chi^2\left(1-\frac{\alpha}{2}; n-1\right)} \right)$$

$\chi^2(\alpha; n-1)$ jest stabilizowaną wartością krytyczną rozkładu chi-kwadrat z ν stopniami swobody.

Przedziały jednostronne

$$\left(0, \frac{\text{var}X}{\chi^2(\alpha; n-1)} \right)$$
$$\left(\frac{\text{var}X}{\chi^2(1-\alpha; n-1)}, \infty \right)$$

Przykład.

Na podstawie próby 1.1, 1.2, 0.8, 0.9, 1.2, 1.3, 1.0, 0.7, 0.8, 1.0 oszacować zróżnicowanie rozkładu obserwowanej cechy.

$$\bar{x} = \frac{1.1 + 1.2 + \dots + 1.0}{10} = 1.0$$

$$\text{var}X = (1.1 - 1.0)^2 + \dots + (1.0 - 1.0)^2 = 0.36$$

$$s^2 = \frac{0.36}{10 - 1} = 0.04, \quad s = \sqrt{s^2} = 0.2$$

Poziom ufności $1 - \alpha = 0.95$, czyli $\alpha = 0.05$.

$$\chi^2 \left(\frac{\alpha}{2}; n - 1 \right) = \chi^2 (0.025; 9) = 19.0228$$

$$\chi^2 \left(1 - \frac{\alpha}{2}; n - 1 \right) = \chi^2 (0.975; 9) = 2.7004$$

$$\left(\frac{0.36}{19.0228}, \frac{0.36}{2.7004} \right) = (0.019, 0.133)$$

Wniosek. Wariancja cechy jest jakąś liczbą z przedziału $(0.019, 0.133)$. Zaufanie do tego wniosku wynosi 95%.

Przedział ufności dla odchylenia standardowego

Średnia μ jest nieznana

Poziom ufności: $1 - \alpha$

$$\left(\sqrt{\frac{\text{var}X}{\chi^2(\frac{\alpha}{2}; n - 1)}}, \sqrt{\frac{\text{var}X}{\chi^2(1 - \frac{\alpha}{2}; n - 1)}} \right)$$

Przedziały jednostronne

$$\left(0, \sqrt{\frac{\text{var}X}{\chi^2(\alpha; n - 1)}} \right)$$

$$\left(\sqrt{\frac{\text{var}X}{\chi^2(1 - \alpha; n - 1)}}, \infty \right)$$

.....

Przykład (cd).

Przedział ufności dla odchylenia standardowego:

$$(\sqrt{0.019}, \sqrt{0.133}) = (0.136, 0.365)$$

Przykład.

Oszacować zróżnicowanie ilości punktów uzyskiwanych na klasówce.

$$n = 300 \quad \sum x_i = 176.566 \quad \sum x_i^2 = 107.845302$$

Populacja:

Słuchacze podstawowego kursu statystyki

Cecha X :

ilość punktów zdobytych na klasówce

Założenie:

cecha X ma rozkład normalny $N(\mu, \sigma^2)$

Zadanie: oszacować parametr σ

Technika statystyczna:

przedział ufności dla odchylenia standardowego
poziom ufności 0.95

Obliczenia

$$\bar{x} = 0.589 \quad \text{var} X = 3.92679$$

$$\chi^2 \left(\frac{\alpha}{2}; n - 1 \right) = \chi^2 (0.025; 299) = 348.79420$$

$$\chi^2 \left(1 - \frac{\alpha}{2}; n - 1 \right) = \chi^2 (0.975; 299) = 252.99251$$

$$\left(\sqrt{\frac{3.92679}{348.79420}}, \sqrt{\frac{3.92679}{252.99251}} \right) = (0.10610, 0.12458)$$

Odpowiedź: $\sigma \in (0.10610, 0.12458)$

Wniosek. Odchylenie standardowe liczby punktów zdobywanych na klasówce jest liczbą z przedziału $(0.106, 0.125)$. Zaufanie do tego wniosku wynosi 95%.

Rozkład dwupunktowy

Estymacja parametru

p — frakcja, wskaźnik struktury

Próba: X_1, \dots, X_n ($X_i = 0$ lub $= 1$)

$k = \sum_{i=1}^n X_i$ — ilość jedynek (sukcesów)

Estymator punktowy:

$$\hat{p} = \frac{k}{n}$$

Przedział ufności na poziomie ufności $1 - \alpha$

$$\left(p_1 \left(1 - \frac{\alpha}{2}; k, n - k \right), 1 - p_1 \left(1 - \frac{\alpha}{2}; n - k, k \right) \right)$$

Jednostronne przedziały ufności

$$(p_1 (1 - \alpha; k, n - k), 1)$$

$$(0, 1 - p_1 (1 - \alpha; n - k, k))$$

Przykład.

Wśród 20 zbadanych detali znaleziono dwa braki.
Oceń na tej podstawie wadliwość produkcji.

Cecha X — jakość detalu (dobry, zły).

Sukces — detal wybrakowany

Pytanie: $p = ?$

$$n = 20, k = 2 \implies \hat{p} = 2/20 = 0.1$$

Poziom ufności $1 - \alpha = 0.9$, czyli $\alpha = 0.1$

$$p_1 \left(1 - \frac{\alpha}{2}; k, n - k \right) = p_1(0.95; 2, 18) = 0.0123$$

$$p_1 \left(1 - \frac{\alpha}{2}; n - k, k \right) = p_1(0.95; 18, 2) = 0.6830$$

$$(0.0123, 1 - 0.6830) = (0.0123, 0.3170)$$

Wniosek. Wadliwość produkcji wyraża się liczbą z przedziału (1.23%, 31.70%). Zaufanie do wniosku wynosi 90%.

Przybliżony przedział ufności

$$\left(\hat{p} - u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

u_α jest kwantylem rzędu α rozkładu $N(0, 1)$.

.....
Przykład. (cd)

$$n = 200, k = 20 \implies \hat{p} = 20/200 = 0.1$$

Poziom ufności $1 - \alpha = 0.9$, czyli $\alpha = 0.1$

$$u_{1-\alpha/2} = u_{0.95} = 1.6449$$

$$0.1 - 1.6449 \sqrt{\frac{0.1(1-0.1)}{200}} = 0.0651$$

$$0.1 + 1.6449 \sqrt{\frac{0.1(1-0.1)}{200}} = 0.1349$$

Wniosek. Wadliwość produkcji wyraża się liczbą z przedziału (6.51%, 13.49%). Zaufanie do wniosku wynosi 90%.

Przykład.

Oszacować odsetek ocen dostatecznych otrzymywanych na klasówce.

$$n = 300 \quad k = 88$$

Populacja:

Słuchacze podstawowego kursu statystyki

Cecha X :

ocena dostateczna z klasówki

Założenie:

cecha X ma rozkład $D(p)$

Zadanie: oszacować parametr p

Technika statystyczna:

przybliżony przedział ufności dla prawdopodobieństwa

poziom ufności 0.95

Obliczenia

$$\hat{p} = \frac{88}{300} = 0.29$$

$$u_{1-\alpha/2} = u_{0.975} = 1.96$$

$$0.29 - 1.96\sqrt{\frac{0.29(1 - 0.29)}{300}} = 0.2387$$

$$0.29 + 1.96\sqrt{\frac{0.29(1 - 0.29)}{300}} = 0.3413$$

Odpowiedź: $p \in (0.2387, 0.3413)$

Wniosek.

Odsetek ocen dostatecznych zdobywanych na klasówce jest liczbą z przedziału (23.87%, 34.13%). Zaufanie do tego wniosku wynosi 95%.