Elementy statystyki

STA - Wykład 4

dr hab. Waldemar Wołyński
Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza

Przedziały ufności

Przedziały ufności

Niech $\theta \in \Theta$ oznacza parametr modelu statystycznego.

Definicja

Przedział (L, R) *określony parą statystyk L i R takich, że* $P_\theta(L \leq R) = 1$ *dla każdego* $\theta \in \Theta$*, nazywamy* **przedziałem ufności** *dla parametru* θ*, na* **poziomie ufności** $1 - \alpha$ *(*$0 < \alpha < 1$*), gdy dla każdego* $\theta \in \Theta$

$$P_\theta(L < \theta < R) \geq 1 - \alpha.$$

Typowe wartości poziomu ufności to: 0,9; 0,95; 0,99 (zazwyczaj podawane w %).

Trochę historii ...

Jerzy Spława-Neyman
(1894 - 1981)

Przedziały ufności

1937

Konstrukcja przedziałów ufności

Definicja

Funkcję $Q(\boldsymbol{X}, \theta)$ *nazywamy* ***funkcją centralną*** *dla parametru* θ*, gdy*

1. *rozkład prawdopodobieństwa zmiennej losowej Q jest absolutnie ciągły i nie zależy od parametru* θ*,*
2. *funkcja* $Q(\boldsymbol{X}, \theta)$ *jest ciągła i ściśle monotoniczna względem* θ*.*

Uwaga! Warunek pierwszy można osłabić, żądają tylko aby rozkład graniczny zmiennej losowej Q był absolutnie ciągły. Wtedy uzyskany przedział można stosować jedynie dla dużych prób.

Konstrukcja przedziałów ufności

1. Obieramy funkcję centralną $Q(\boldsymbol{X}, \theta)$.
2. Wyznaczamy stałe a i b tak, aby

$$\forall \theta \in \Theta : \; P_\theta(a < Q < b) = 1 - \alpha.$$

3. Rozwiązujemy nierówność

$$a < Q(\boldsymbol{X}, \theta) < b$$

względem θ otrzymując szukany przedział

$$(L(\boldsymbol{X}), R(\boldsymbol{X})).$$

Uwaga! Stałe a i b można dobrać na wiele sposobów. Zazwyczaj dobieramy je tak, aby

$$P_\theta(Q \leq a) = P_\theta(Q \geq b) = \frac{\alpha}{2}.$$

Rachunek prawdopodobieństwa

Definicja

Niech $X \sim N(0,1)$ *oraz* $Y \sim \chi^2(n)$ *będą niezależnymi zmiennymi losowymi.*
Mówimy, że zmienna losowa

$$\frac{X}{\sqrt{\frac{1}{n}Y}}$$

ma rozkład ***t-Studenta*** *z n stopniami swobody (ozn.* $t(n)$*).*

Fakt

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \ x \in \mathbf{R}.$$

Fakt

W modelu jednej próby prostej z rozkładu normalnego zmienna losowa

$$\frac{\bar{X} - \mu}{S}\sqrt{n} \sim t(n-1).$$

Trochę historii ...

William S. Gosset
(1876 - 1937)

Student

Przedział ufności dla parametru μ w modelu normalnym

Fakt

W modelu jednej próby prostej z rozkładu normalnego $(1-\alpha)\cdot 100\%$ *przedział ufności dla parametru* μ *ma postać:*

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

gdzie $t(p,n)=F_t^{-1}(p)$ *oznacza kwantyl rzędu* p *z rozkładu* $t(n)$.

Przedział ufności dla prognozy

W modelach prostej i wielokrotnej regresji liniowej najczęściej wykorzystywany jest przedział ufności dla prognozy Y_p. Jego konstrukcja wymaga dodatkowego założenia normalności rozkładu błędów!!!

Fakt

W modelu prostej regresji liniowej $(1-\alpha)\cdot 100\%$ *przedział ufności dla prognozy ma postać:*

$$\left(\hat{Y}_p - t(1-\frac{\alpha}{2}, n-2)\sqrt{S^2+S_p^2}, \hat{Y}_p + t(1-\frac{\alpha}{2}, n-2)\sqrt{S^2+S_p^2}\right).$$

W modelu wielokrotnej regresji liniowej $(1-\alpha)\cdot 100\%$ *przedział ufności dla prognozy ma postać:*

$$\left(\hat{Y}_p - t(1-\frac{\alpha}{2}, n-m-1)\sqrt{S^2+S_p^2}, \hat{Y}_p + t(1-\frac{\alpha}{2}, n-m-1)\sqrt{S^2+S_p^2}\right).$$

R

Funkcje związane z przedziałami ufności:

enorm (EnvStats) – pozwala wyznaczyć przedziały ufności dla parametrów w modelu normalnym,

predict – pozwala wyznaczyć przedział ufności dla prognozy w modelach prostej i wielokrotnej regresji liniowej.