

Przedział ufności dla frakcji w populacji

Pewna zmienna X przyjmuje jedynie wartości 1 (sukces) lub 0 (porażka) i jej średnia dla całej populacji wynosi p . Chcemy skonstruować przedział ufności dla p .

- Losujemy n -elementową próbę i liczymy liczbę sukcesów (zmienna Y).
- Zmienna Y ma rozkład dwumianowy z parametrami (n, p) .
- Rozkład Y przybliżamy (dla dużych n) rozkładem normalnym $N(np, \sqrt{np(1-p)})$.
- Zmienna $\hat{p} = \frac{Y}{n}$ ma wartość oczekiwaną $\mu = p$ oraz odchylenie standardowe $\sigma = \sqrt{\frac{p(1-p)}{n}}$.
- Zatem $\hat{p}$ ma w przybliżeniu rozkład $N(p, \sqrt{\frac{p(1-p)}{n}})$.

Klasyczny przybliżony przedział ufności

Klasyczny przybliżony przedział ufności dla p otrzymujemy:

- biorąc $\hat{p} = \frac{Y}{n}$ za środek przedziału,
- zastępując p przez $\hat{p}$ we wzorze na odchylenie standardowe i otrzymując $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Otrzymujemy przedział ufności na poziomie istotności α :

$$\hat{p} \pm Z_{\alpha/2} SE_{\hat{p}} = \hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Klasyczny przedział ufności źle się zachowuje, gdy liczba sukcesów

Y jest bliska 0 lub bliska n (może się nawet zdarzyć, że klasyczny PU zawiera ujemne wartości).

Przybliżony przedział ufności Agrestiego-Coula

Przybliżony przedział ufności Agrestiego-Coula dla p otrzymujemy biorąc:

- biorąc $\tilde{p} = \frac{Y + 0.5(Z_{\alpha/2})^2}{n + (Z_{\alpha/2})^2}$ za środek przedziału (pewna modyfikacja $\hat{p}$),
- zastępując p przez $\tilde{p}$ we wzorze na odchylenie standardowe i modyfikując je do postaci $SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n + (Z_{\alpha/2})^2}}$

Otrzymujemy przedział ufności na poziomie istotności α :

$$\tilde{p} \pm Z_{\alpha/2} SE_{\tilde{p}}$$

Przedziały ufności – porównanie

Dla 95% przedziału ufności (tzn. $\alpha = 0.05$) mamy ($Z_{0.025} = 1.96$):

Klasyczny przedział ufności

$$\hat{p} = \frac{Y}{n} \quad \text{oraz} \quad SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

PU ma postać $\hat{p} \pm 1.96 \cdot SE_{\hat{p}}$

Przedziału ufności Agrestiego-Coula

$$\tilde{p} = \frac{Y + 0.5 \cdot (1.96)^2}{n + (1.96)^2} = \frac{Y + 1.92}{n + 3.84} \approx \frac{Y + 2}{n + 4}$$

$$SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + (1.96)^2}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 3.84}} \approx \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

PU ma postać $\tilde{p} \pm 1.96 \cdot SE_{\tilde{p}}$

Przykład 1

Złapano 125 myszy i 6 z nich ma brzuszki nakrapiane na biało. Jaka część myszek w tej populacji ma brzuszki nakrapiane na biało? Przyjmij poziom istotności $\alpha = 0.05$.

$$\tilde{p} \approx \frac{8}{129} = 0.062$$

$$SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} = 0.021$$

Otrzymujemy przedział ufności $0.062 \pm 1.96 \cdot 0.021 = 0.062 \pm 0.041$ czyli $p \in (0.021, 0.103)$.

Przykład 2

Złapano 125 myszy i 6 z nich ma brzuszki nakrapiane na biało. Jaka część myszek w tej populacji ma brzuszki nakrapiane na biało? Przyjmij poziom istotności $\alpha = 0.1$.

$$\tilde{p} \approx \frac{6 + 0.5 \cdot (1.65)^2}{125 + (1.65)^2} = 0.058$$

$$SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + (1.65)^2}} = 0.021$$

Otrzymujemy przedział ufności $0.058 \pm 1.65 \cdot 0.021 = 0.058 \pm 0.035$ czyli $p \in (0.023, 0.093)$.

Wybór wielkości próby

Chcemy, aby 95% PU miał szerokość nie większą od pewnej zadanej liczby. Jak ustalić rozmiar próby?

Szerokość przedziału ufności wynosi

$$2 \cdot 1.96 \cdot \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

czyli zależy nie tylko od n (które chcemy wyznaczyć), ale również od $\tilde{p}$, którego nie znamy. Rozwiązania są dwa:

- 1 jeśli znamy przybliżoną wartość p , to możemy jej użyć jako $\tilde{p}$,
- 2 jeśli nie mamy żadnych wstępnych informacji, użyjemy $p = 0.5$.

Przykład 2

Chcemy ustalić jaka część myszek w pewnej populacji ma brzuszek nakrapiane na biało. Przypuszczamy, że jest to ok. 5%. Chcemy uzyskać 95% PU o szerokości ok. 0.02. Jaki powinien być rozmiar próby?

Chcemy:

$$SE_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} \approx 0.005$$

skąd

$$n \approx \frac{\tilde{p}(1 - \tilde{p})}{(0.005)^2} - 4$$

Przyjmując $\tilde{p} \approx 0.05$, otrzymujemy $n \approx 190$.