Przedział ufności dla odsetka

Rozważamy następujący problem. Odsetek/frakcja będzie procentem populacji, który spełnia zadany warunek (np. popieranie konkretnego kandydata w wyborach prezydenckich). Populacja generalna o liczebności n jest opisana przez rozkład Bernoulliego, z parametrem $p = \mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0)$. Do konstrukcji przedziału ufności na poziomie ufności $1-\alpha$ dla odsetka elementów wyróżnionych w populacji można wykorzystać fakt, że $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ ma przy dostatecznie dużym n rozkład $N(p, \frac{p(p-1)}{n})$. Ze standaryzacji:

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow_{n\to\infty} N(0,1)$$

$$\lim_{n\to\infty} \mathbb{P}\left[-U_{\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq U_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

Korzystamy z symetryczności rozkładu $N(0,1)$. Przedział ufności ma postać:

$$\hat{p} - U_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + U_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$$

Granice przedziału w powyższym wzorze są zależne od wartości p, która nie jest znana. Dla dużej próby prostej o rozmiarze n_1 z populacji, w której $\sum_{i=1}^{n} X = k$, zastosujemy podstawienie $p = \frac{k}{n_1}$. Oznacza to, że w praktyce przedział ufności będzie mieć postać:

$$\hat{p} - U_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}} \leq p \leq \hat{p} + U_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}}$$

Zadanie 1: Przykładowe zadanie – fragment zadania 15 z zestawu uzupełniającego

Władze województwa zamierzają sprawdzić skuteczność programu przeciwdziałania przemocy w rodzinie. Na terenie powiatów umieszczone zostały plakaty informujące o możliwości telefonicznego wsparcia psychologicznego. Przeprowadzono badanie ankietowe w powiatach, sprawdzające zasięg oddziaływania programu. (...) W każdym z powiatów wylosowano niezależnie 200 osób, którym zadano pytanie, czy widziały plakaty. W powiecie 7 twierdząco odpowiedziało 49 osób. Zbudować przybliżony przedział ufności dla odsetka osób poddanych programowi w powiecie 7 na poziomie ufności 0.9.

Rozwiązanie: Każda osoba ma prawdopodobieństwo zobaczenia plakatu p. Zmienna losowa przyjmuje wartość 1, gdy dana osoba zobaczyła plakat, a 0 gdy nie. W związku z tym $\hat{p} = \frac{k}{n_1}$ gdzie n_1 to liczebność próby, a k to liczba osób, które twierdzą, że zobaczyły plakaty. Poziom ufności to $0.9 = 1-\alpha$, czyli $\alpha = 0.1$. W związku z tym przedział ufności wygląda w następujący sposób:

$$\hat{p} - U_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}} \leq p \leq \hat{p} + U_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}}$$

Po podstawieniu liczb ($U_{1-\frac{\alpha}{2}} = U_{0.95} = 1.64$ – kwantyl rozkładu normalnego standardowego na poziomie 0.95):

$$\frac{49}{200} - 1.64\sqrt{\frac{\frac{49}{200}(1-\frac{49}{200})}{200}} \leq p \leq \frac{49}{200} + 1.64\sqrt{\frac{\frac{49}{200}(1-\frac{49}{200})}{200}}$$

Co daje przedział [0.19512, 0.29487].