Notatka o średniej i wariancji

Rozpatrzmy *wartość średnią próby* jako średnią arytmetyczną

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) \tag{1}$$

Będąc funkcją zmiennych losowych, wartość średnia $\bar{x}$ jest również zmienną losową. Jej wartość oczekiwana

$$E(\bar{x}) = \frac{1}{n}\{E(x_1) + E(x_2) + \ldots + E(x_n)\} = \hat{x} \tag{2}$$

Obliczmy wariancję $\bar{x}$

$$\begin{aligned}\sigma^2(\bar{x}) &= E\{(\bar{x} - E(\bar{x}))^2\} = E\{(\bar{x} - \hat{x})^2\} \\ &= E\Big\{\Big(\frac{x_1 + x_2 + \ldots + x_n}{n} - \hat{x}\Big)^2\Big\} \\ &= E\Big\{\Big(\frac{x_1 + x_2 + \ldots + x_n - n\hat{x}}{n}\Big)^2\Big\} \\ &= \frac{1}{n^2}E\Big\{[(x_1 - \hat{x}) + (x_2 - \hat{x}) + \ldots + (x_n - \hat{x})]^2\Big\}\end{aligned} \tag{3}$$

Z faktu, że wszystkie elementy próby x_i są niezależne wynika, że wszystkie wyrazy typu

$$E[(x_i - \hat{x})(x_j - \hat{x})] \tag{4}$$

(czyli kowariancje) znikają, co daje

$$\sigma^2(\bar{x}) = \frac{1}{n^2}\sum_{i=1}^{n} E[(x_i - \hat{x})^2] = \frac{1}{n^2} n\sigma^2(x) = \frac{1}{n}\sigma^2(x) \tag{5}$$

Obecnie musimy znaleźć estymator dla wariancji populacji. W pierwszym przybliżeniu zdefiniujemy *wariancję próby* jako średnią arytmetyczną odchyleń kwadratowych od wartości średniej próby

$$s'^2 = \frac{1}{n}\Big\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\Big\} \tag{6}$$

Wartość ta posiada wartość oczekiwaną

$$
\begin{aligned}
E(s'^2) = \frac{1}{n}E\Big\{\sum_{i=1}^{n}(x_i-\bar{x})^2\Big\} = \frac{1}{n}E\Big\{\sum_{i=1}^{n}(x_i-\hat{x}+\hat{x}-\bar{x})^2\Big\} \\
= \frac{1}{n}E\Big\{\sum_{i=1}^{n}[(x_i-\hat{x})-(\bar{x}-\hat{x})]^2\Big\} \\
= \frac{1}{n}E\Big\{\sum_{i=1}^{n}(x_i-\hat{x})^2 - 2(\bar{x}-\hat{x})\sum_{i=1}^{n}(x_i-\hat{x}) + \sum_{i=1}^{n}(\bar{x}-\hat{x})^2\Big\} \\
= \frac{1}{n}E\Big\{\sum_{i=1}^{n}(x_i-\hat{x})^2 - 2n(\bar{x}-\hat{x})^2 + n(\bar{x}-\hat{x})^2\Big\} \\
= \frac{1}{n}E\Big\{\sum_{i=1}^{n}(x_i-\hat{x})^2 - n(\bar{x}-\hat{x})^2\Big\} \\
= \frac{1}{n}\Big\{n\sigma^2(x) - n\Big(\frac{1}{n}\sigma^2(x)\Big)\Big\} = \frac{1}{n}\Big(n\sigma^2(x)-\sigma^2(x)\Big) \\
= \frac{n-1}{n}\sigma^2(x)
\end{aligned}
\tag{7}
$$

i otrzymujemy

$$
\sigma^2(x) = \frac{1}{n-1}E\Big\{\sum_{i=1}^{n}(x_i-\bar{x})^2\Big\} \tag{8}
$$

Skorzystaliśmy z własności

$$
\sum_{i=1}^{n} x_i = n\bar{x} \tag{9}
$$

oraz z tego, że

$$
\sum_{i=1}^{n}(x_i-\hat{x}) = n\bar{x} - n\hat{x} = n(\bar{x}-\hat{x}) \tag{10}
$$

Możemy zatem wprowadzić nieobciążony estymator wariancji jako

$$
s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{11}
$$

Literatura

[1] Siegmund Brandt, (1976) *Metody statystyczne i obliczeniowe analizy danych* PWN Warszawa

Pawel Jan Piskorz (paweljs@gmail.com)