



# Estymatory

Statystyka i analiza danych 2019

---

Jurek Błaszczński

28 marca 2019

Estymator  $\hat{\theta}$  parametru  $\theta$  to funkcja próby  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ , która szacuje nieznaną wartość parametru rozkładu  $\theta$ .

Estymator  $\hat{\theta}$  jest więc (jako funkcja próby) **zmienną losową**.

Oznacza to, że ma swój rozkład, wartość oczekiwaną, wariancję, itp. (jak każda zmienna losowa).

**$\theta$  nie jest zmienną losową.**

Jest parametrem (nieznanym), który chcemy oszacować (estymować).

Estymator  $\hat{\theta}$  parametru  $\theta$  jest **nieobciążony**, jeśli

$$\mathbb{E}[\hat{\theta}] = \theta,$$

czyli obciążenie  $\theta - \mathbb{E}[\hat{\theta}]$  jest równe 0.

## Przykład 1

Pokażmy, że estymator  $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  wartości oczekiwanej  $\mu$  jest nieobciążony.

## Przykład 1

Pokażmy, że estymator  $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  wartości oczekiwanej  $\mu$  jest nieobciążony.

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}[X_i]}_{\mu} = \mu.$$

## Przykład 2

Pokażmy, że estymator  $\hat{\sigma}^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest obciążony.

## Przykład 2

Pokażmy, że estymator  $\hat{\sigma}^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest obciążony.

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2 - \bar{X}^2 \quad (1)$$

## Przykład 2

Pokażmy, że estymator  $\hat{\sigma}^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest obciążony.

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2 - \bar{X}^2 \quad (1)$$

$$\mathbb{E}[S_n^2] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}^2] = \quad (2)$$

$$= \frac{1}{n} n \mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] = \mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] \quad (3)$$



## Przykład 2

Pokażmy, że estymator  $\hat{\sigma}^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest obciążony.

$$\mathbb{E}[S_n^2] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}^2] = \quad (4)$$

$$= \frac{1}{n} n \mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] = \mathbb{E}[X^2] - \mathbb{E}[\bar{X}^2] \quad (5)$$

ponieważ  $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$\mathbb{E}[S_n^2] = \text{Var}(X) + (\mathbb{E}[X])^2 - \text{Var}(\bar{X}) - (\mathbb{E}[\bar{X}])^2 \quad (6)$$

## Przykład 2

Pokażmy, że estymator  $\hat{\sigma}^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest obciążony.

$$\mathbb{E}[S_n^2] = \text{Var}(X) + (\mathbb{E}[X])^2 - \text{Var}(\bar{X}) - (\mathbb{E}[\bar{X}])^2 = \quad (7)$$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \quad (8)$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \quad (9)$$

### Przykład 3

Pokażmy, że estymator  $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest nieobciążony.

## Estymator nieobciążony

### Przykład 3

Pokażmy, że estymator  $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest nieobciążony.

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}^2 \quad (10)$$

## Przykład 3

Pokażmy, że estymator  $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest nieobciążony.

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}^2 \quad (10)$$

$$\mathbb{E}[S_n^2] = \mathbb{E}\left[\frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}^2\right] = \quad (11)$$

$$= \frac{1}{n-1} \left( \sum_{i=1}^n \mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2] \right) = \quad (12)$$

$$= \frac{1}{n-1} \left( n\mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2] \right) \quad (13)$$

## Przykład 3

Pokażmy, że estymator  $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest nieobciążony.

$$\mathbb{E}[S_n^2] = \frac{1}{n-1} \left( n\mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2] \right) \quad (14)$$

ponieważ  $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$\mathbb{E}[S_n^2] = \frac{n}{n-1} \left( \text{Var}(X) + (\mathbb{E}[X])^2 - \text{Var}(\bar{X}) - (\mathbb{E}[\bar{X}])^2 \right) \quad (15)$$

## Przykład 3

Pokażmy, że estymator  $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  wariancji  $\sigma^2$  jest nieobciążony.

$$\mathbb{E}[S_n^2] = \frac{n}{n-1} \left( \text{Var}(X) + (\mathbb{E}[X])^2 - \text{Var}(\bar{X}) - (\mathbb{E}[\bar{X}])^2 \right) = \quad (16)$$

$$= \frac{n}{n-1} \left( \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \right) = \frac{n}{n-1} \left( \sigma^2 - \frac{\sigma^2}{n} \right) = \quad (17)$$

$$= \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} = \frac{n\sigma^2 - \sigma^2}{n-1} = \frac{\sigma^2(n-1)}{(n-1)} = \sigma^2 \quad (18)$$

Zamiast punktowego (pojedynczego) estymatora  $\hat{\theta}$  tworzy się **przedział ufności**  $[\hat{\theta}_L, \hat{\theta}_P]$  tak, aby prawdziwa wartość parametru  $\theta$  znajdowała się w stworzonym przedziale z prawdopodobieństwem  $1 - \alpha$  (zwanym **poziomem ufności**):

$$\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P) = 1 - \alpha.$$

### Co tu jest losowe?

Przedział ufności tworzy się zwykle poprzez rozpięcie przedziału wokół estymatora punktowego  $\hat{\theta}$ :

$$\hat{\theta}_L = \hat{\theta} - \Delta, \quad \hat{\theta}_P = \hat{\theta} + \Delta.$$



Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

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

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

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

Tworzymy przedział ufności  $[\bar{X} - \Delta, \bar{X} + \Delta]$  taki, że:

$$P(\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta) = 1 - \alpha$$

## Estymator przedziałowy

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

$$P(\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta) = P(-\Delta \leq \bar{X} - \mu \leq \Delta)$$

## Estymator przedziałowy

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

$$P(\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta) = P(-\Delta \leq \bar{X} - \mu \leq \Delta)$$

Doprowadźmy do tego aby wyrażenie “w środku” nierówności miało rozkład  $N(0,1)$

$$P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\Delta}{\sigma}\sqrt{n}\right)$$

## Estymator przedziałowy

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

$$P(\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta) = P(-\Delta \leq \bar{X} - \mu \leq \Delta)$$

Doprowadźmy do tego aby wyrażenie “w środku” nierówności miało rozkład  $N(0,1)$

$$P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\Delta}{\sigma}\sqrt{n}\right)$$

Wprowadźmy  $z_\alpha = \frac{\Delta}{\sigma}\sqrt{n}$

$$P\left(-z_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\right)$$

## Estymator przedziałowy

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

$$P\left(-z_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\right) = \Phi(z_\alpha) - (1 - \Phi(z_\alpha)) = 2\Phi(z_\alpha) - 1$$

## Estymator przedziałowy

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

$$P\left(-z_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\right) = \Phi(z_\alpha) - (1 - \Phi(z_\alpha)) = 2\Phi(z_\alpha) - 1$$

Możemy ustalić ile wynosi  $z_\alpha$

$$2\Phi(z_\alpha) - 1 = 1 - \alpha$$

$$\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$$

## Estymator przedziałowy

Zbudujmy estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znaną wariancją  $\sigma^2$ .

$$P\left(-z_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\right) = \Phi(z_\alpha) - (1 - \Phi(z_\alpha)) = 2\Phi(z_\alpha) - 1$$

Możemy ustalić ile wynosi  $z_\alpha$

$$2\Phi(z_\alpha) - 1 = 1 - \alpha$$

$$\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$$

Stąd  $z_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$ .  $z_\alpha = \frac{\Delta}{\sigma} \sqrt{n}$ , czyli  $\Delta = z_\alpha \frac{\sigma}{\sqrt{n}}$ .



# Efektywność estymatorów

Weźmy dwa **nieobciążone** estymatory  $\hat{\theta}_1$  i  $\hat{\theta}_2$  parametru  $\theta$ .

Mówimy, że  $\hat{\theta}_1$  jest efektywniejszy<sup>1</sup>  $\hat{\theta}_2$ , jeśli  $\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2)$  przy każdej wartości parametru  $\theta$ .

Estymator nieobciążony o minimalnej wariancji, czyli efektywniejszy od wszystkich nieobciążonych estymatorów, nazywamy estymatorem **efektywnym**.

---

<sup>1</sup>ściślej, powinno być: „co najmniej tak efektywny jak”.

Estymator  $\hat{\theta}$  parametru  $\theta$  jest zgodny, jeśli gdy  $n \rightarrow \infty$ , to  $\hat{\theta} \rightarrow \theta$ , czyli estymator zbiega<sup>2</sup> do prawdziwej wartości parametru  $\theta$ .

---

<sup>2</sup>Jeśli zbieżność jest prawie na pewno (z prawdopodobieństwem 1), to estymator nazywa się silnie zgodnym, gdy zbieżność jest według prawdopodobieństwa, to estymator nazywa się słabo zgodnym.