Mediana, współczynnik skośności i średnia ucinana

W wyniku pomiaru zmiennej losowej X otrzymujemy pewien zbiór obserwacji (**szereg statystyczny**). W kolejnym kroku możemy uporządkować obserwacje według pewnego kryterium (np. niemalejąco). Rozważmy następujący przykład: grupa $N = 11$ studentów pisała test zaliczeniowy ze statystyki. Studenci otrzymali następujące oceny z testu:

$$5, 3, 4, 2, 3, 3, 5, 5, 5, 4, 5$$

Porządkujemy obserwacje w sposób niemalejący

$$2, 3, 3, 3, 4, \mathbf{4}, 5, 5, 5, 5, 5$$

Uporządkowany szereg statystyczny może służyć do wyznaczenia **mediany** M.

a) dla nieparzystej liczby obserwacji mediana M stanowi wartość rzeczywistą poniżej i powyżej której znajduje się jednakowa liczba obserwacji.

$$M = x_{\frac{N+1}{2}}$$

b) dla parzystej liczby obserwacji (wartość rzeczywista lub abstrakcyjna) mediana M stanowi średnią arytmetyczną dwóch centralnych wartości zmiennej losowej w uporządkowanym szeregu statystycznym.

$$M = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$$

Mediana M jest znacznie bardziej odporna na wartości odstające niż średnia $\langle x \rangle$ z tej samej próby. Nawet olbrzymie zmiany wartości skrajnych występujących w próbie (x_{min} i/lub x_{max}) nie wpływają na wartość liczbową mediany!

Dla uporządkowanego szeregu statystycznego A

$$2, 3, 3, 3, 4, 4, 5, 5, 5, 5, 5$$

$$M_A = 4 \ \wedge \ \langle x \rangle_A = 4$$

Dla uporządkowanego szeregu statystycznego B

$$0, 3, 3, 3, 4, 4, 5, 5, 5, 5, 29$$

$$M_B = M_A = 4 \ \wedge \ \langle x \rangle_B = 6 \ \rightarrow \ \langle x \rangle_B = \frac{3}{2} \langle x \rangle_A$$

Mediana pojawia się w definicji **wskaźnika skośności** A_M

$$A_M = \frac{M - \langle x \rangle}{\Delta x}$$

Współczynnik skośności A_M podobnie jak moment centralny trzeciego rzędu opisuje asymetrię badanego rozkładu ale w porównaniu do momentu centralnego trzeciego rzędu jest prostszy w wyznaczeniu. Dla symetrycznego rozkładu zmiennej losowej: $\langle x \rangle = M \ \rightarrow \ A_M = 0$. Dla rozkładu zmiennej losowej o lewostronnej asymetrii: $M < \langle x \rangle \ \rightarrow \ A_M < 0$. Dla rozkładu zmiennej losowej o lewostronnej asymetrii: $M > \langle x \rangle \rightarrow A_M > 0$.

Zadanie 1. Dla zbioru danych (2, 4, 6, 8) proszę wyznaczyć: $\langle x \rangle$, M, Δx oraz A_M.

$$\langle x \rangle = M = 5 \quad \text{oraz} \quad \Delta x = \sqrt{5}$$

$$A_M = \frac{M - \langle x \rangle}{\Delta x} = \frac{5 - 5}{\sqrt{5}} = 0$$

Zadanie 2. Proszę wyznaczyć współczynnik skośności A_M dla następujących zbiorów danych:

(1, 1, 1, 3, 4)

(1, 1, 2, 3, 3)

(0, 1, 3, 3, 3)

Średnia ucinana (**trymowana**, ang. *to trim* - przycinać) **z parametrem k -** przy obliczaniu średniej ucinanej tworzymy uporządkowany szereg statystyczny. Następnie odrzucamy po k obserwacji ($k = 0,1,2,3, \ldots$) występujących na początku i końcu uporządkowanego (rosnąco lub niemalejąco) szeregu statystycznego.

$$\langle x \rangle_k = \frac{1}{N - 2k} \sum_{i=1+k}^{i=N-k} x_i$$

Średnia ucinana $\langle x \rangle_k$ jest stosowana gdy podejrzewamy, że wartości najmniejsze oraz największe występujące w próbie są skutkiem błędnego przetworzenia danych i/lub wadliwej pracy przyrządów pomiarowych.

Uwaga! Średnia ucinana stanowi uogólnienie definicji średniej arytmetycznej, tzn. $\langle x \rangle_0 = \langle x \rangle$.

Zadanie 3. Studenci otrzymali następujące oceny z testu:

$$5, 3, 4, 2, 3, 3, 5, 5, 5, 4, 5$$

Uporządkujmy obserwacje w sposób niemalejący

$$2, 3, 3, 3, 4, 4, 5, 5, 5, 5, 5$$

W ten sposób został utworzony uporządkowany szereg statystyczny, na podstawie którego wyznaczymy średnią ucinaną $\langle x \rangle_1$.

$$\langle x \rangle_1 = \frac{1}{N-2k} \sum_{i=k+1}^{i=N-k} x_i = \frac{1}{11-2} \sum_{i=1+1}^{i=11-1} x_i = \frac{1}{9} \sum_{i=2}^{i=10} x_i = 4\frac{1}{10}$$

Jak jest wada średniej ucinanej $\langle x \rangle_k$ w porównaniu do klasycznej średniej arytmetycznej? Ile wynosi $\langle x \rangle_2$? Proszę wykazać, że w przypadku nieparzystej liczby obserwacji N mediana stanowi średnią ucinaną z parametrem

$$k = \frac{N-1}{2}$$

Błąd pomiaru i pojęcia z nim związane

Błąd pomiaru B **-** odstępstwo wyniku pojedynczego pomiaru x_i od wartości rzeczywistej x_R

$$B = x_i - x_R$$

Bezwzględny błąd pomiaru BB **-** moduł różnicy między wartością pomiaru x_i a wartością rzeczywistą x_R

$$BB = |x_i - x_R| = |x_R - x_i| \geq 0$$

Względny błąd pomiaru WB **-** iloraz bezwzględnego błędu pomiaru $|x - x_R|$ i wartości rzeczywistej x_R

$$WB = \frac{BB}{x_R} = \frac{|x - x_R|}{x_R}$$

Uwaga! Względny błąd pomiaru jest wartością bezwymiarową (niemianowaną) i zazwyczaj jest wyrażany w procentach (**procentowy względny błąd pomiaru**)

$$\%WB = \frac{|x - x_R|}{x_R} \cdot 100\%$$

Uwaga! Na podstawie pojedynczego pomiaru nie możemy wnioskować o jego dokładności. Aby wyznaczyć wartość mierzonej wielkości fizycznej dokonuje się zazwyczaj serii pomiarów. Z serii pomiarów: $x_1, x_2, x_3, \ldots$ wartością najbardziej zbliżoną do wartości rzeczywistej x_R jest średnia arytmetyczna $\langle x \rangle$. Jest to podstawowe twierdzenie teorii błędów i tzw. **pierwszy postulat Gaussa**. Średnia arytmetyczna $\langle x \rangle$ wyznaczona na podstawie serii pomiarów jest położona zazwyczaj blisko wartości rzeczywistej x_R, tzn. $\langle x \rangle \approx x_R$. Przybliżenie jest tym lepsze im dłuższa jest seria pomiarowa.