Rozstęp (empiryczny obszar zmienności, ang. *range*) - jest najprostszą miarą zmienności. Rozstęp ($R \geq 0$) jest różnicą między wartością maksymalną x_{max} a wartością minimalną x_{min} w analizowanej próbie (jego obliczenie wymaga wykonania tylko jednego odejmowania). Rozstęp opiera się jedynie na dwóch skrajnych wartościach zmiennej, które często różnią się istotnie od pozostałych wartości, dlatego stanowi miarę o małej wartości poznawczej. Znając rozstęp, wiemy jaka jest różnica między krańcowymi wartościami cechy, nie mamy natomiast żadnej informacji o strukturze pozostałych $N-2$ wartości!

$$R = x_{max} - x_{min} = |x_{min} - x_{max}|$$

Typowy przedział zmienności

$$\langle x \rangle - \Delta x < \ x_{\mathrm{typ}} < \ \langle x \rangle + \Delta x$$

Szerokość typowego przedziału zmienności wynosi $2\Delta x$. Możemy wyznaczyć również względną szerokość typowego przedziału zmienności – np. względem modułu (wartości bezwzględnej) wartości średniej $\langle x \rangle \neq 0$, tzn.

$$\frac{\langle x \rangle + \Delta x - [\langle x \rangle - \Delta x]}{|\langle x \rangle|} = \frac{2\Delta x}{|\langle x \rangle|} = 2\,\frac{\langle x^2 \rangle - \langle x \rangle^2}{|\langle x \rangle|}$$

Odchylenie przeciętne d - średnia arytmetyczną modułów (wartości bezwzględnych) odchyleń obserwacji x_i od średniej arytmetycznej $\langle x \rangle$ wyznaczonej dla analizowanej zbiorowości statystycznej ($d \geq 0$).

$$d = \sum_{i=1}^{n} |x_i - \langle x \rangle|\, p_i = \sum_{i=1}^{n} |\langle x \rangle - x_i|\, p_i$$

Odchylenie standardowe Δx oraz odchylenie przeciętne d wyznaczone dla tej samej zbiorowości statystycznej zawsze spełniają relację: $\Delta x \geq d$. Odchylenie przeciętne d ma pewną przewagę nad odchyleniem standardowym Δx - jest znacznie mniej wrażliwe na obserwacje odstające w próbie: w definicji d odchylenia $x_i - \langle x \rangle$ nie są podnoszone do kwadratu!

Względne odchylenie standardowe Δx_r (ang. *relative standard deviation*) - odchylenie standardowe Δx podzielone przez moduł średniej arytmetycznej $\langle x \rangle$ z analizowanej próby (zbiorowości statystycznej). Jak sugeruje nazwa Δx_r jest miarą względną oraz bezwymiarową (niemianowaną). Moduł, który pojawia się w mianowniku zapewnia, że $\Delta x_r \geq 0$ gdy $\langle x \rangle < 0$. Względne odchylenie standardowe Δx_r określa jaki ułamek średniej arytmetycznej stanowi odchylenie standardowe Δx.

$$\Delta x_r = \frac{\Delta x}{|\langle x \rangle|} = \frac{\sqrt{\langle x^2 \rangle - \langle x \rangle^2}}{|\langle x \rangle|}$$

Uwaga! Względne odchylenie standardowe Δx_r nazywane jest również **współczynnikiem zmienności** V.

Podobnie definiujemy **względne odchylenie przeciętne**

$$d_r = \frac{d}{|\langle x \rangle|} \cdot 100\%$$

Uwaga! Względne odchylenie standardowe Δx_r możemy podawać również w procentach

$$\Delta x_r = \frac{\Delta x}{|\langle x \rangle|} \cdot 100\%$$

Moment centralny trzeciego rzędu ($k = 3$)

$$\sum_{i=1}^{n} [x_i - \langle x \rangle]^3 \, p_i = \langle x^3 \rangle - 3\langle x \rangle \langle x^2 \rangle + 2\langle x \rangle^3$$

Trzeci moment centralny znajduje zastosowanie przy wyznaczaniu **asymetrii** rozkładu zmiennej losowej - przyjmuje wartość zero dla rozkładu symetrycznego, wartości ujemne dla rozkładów o lewostronnej asymetrii (wydłużone lewe ramię rozkładu) i wartości dodatnie dla rozkładów o prawostronnej asymetrii (wydłużone prawe ramię rozkładu).

Moment centralny czwartego rzędu ($k = 4$)

$$\sum_{i=1}^{n} [x_i - \langle x \rangle]^4 \, p_i = \langle x^4 \rangle - 4\langle x \rangle \langle x^3 \rangle + 6\langle x \rangle^2 \langle x^2 \rangle - 3\langle x \rangle^4 \geq 0$$

Moment centralny czwartego rzędu znajduje zastosowanie przy wyznaczaniu **koncentracji** wartości x_i zmiennej X wokół wartości średniej $\langle x \rangle$.

Momenty centralne wysokich rzędów - to momenty wykraczające poza centralny moment czwartego rzędu. Im wyższy moment centralny, tym trudniej go oszacować. Momenty centralne wysokich rzędów są również subtelne w interpretacji. Moment centralny piątego rzędu nazywany jest **hiperskośnością** (ang. *hyper-skewness*). Moment centralny szóstego rzędu nazywany jest **hiperkurtozą** (ang. *hyper-kurtosis*).

Jak wiemy **moment *k*-tego rzędu** stanowi fundamentalne pojęcie statystyki opisowej

$$\sum_{i=1}^{n} (x_i - c)^k \, p_i$$

Możemy również zdefiniować inne momenty np. **odwrócony moment *k*-tego rzędu**

$$\sum_{i=1}^{n} \left(\frac{1}{x_i} - c \right)^k p_i$$

Na przykład, odwrócony moment zwykły pierwszego rzędu ($c = 0 \;\wedge\; k = 1$) ma postać

$$\sum_{i=1}^{n} \left(\frac{1}{x_i} - c \right)^k p_i = \sum_{i=1}^{n} \left(\frac{1}{x_i} - 0 \right)^1 p_i = \sum_{i=1}^{n} \frac{p_i}{x_i}$$

Logarytmiczny moment *k*-tego rzędu

$$\sum_{i=1}^{n} (lnx_i - c)^k \, p_i$$

Zadanie 1. Rozkład dyskretnej zmiennej losowej X przedstawia się następująco

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Przyjmijmy, że powyższy rozkład posiada wartość średnią $\langle x \rangle$.

a) Jak zmieni się wartość średnia $\langle x \rangle$ jeżeli do każdej wartości x_i zmiennej X dodamy stałą c?

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} = \begin{pmatrix} x_1 + c & x_2 + c & \dots & x_n + c \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$\langle y \rangle = \sum_{i=1}^{n} y_i p_i = \sum_{i=1}^{n} (x_i + c) p_i = \sum_{i=1}^{n} (x_i p_i + c p_i) = \sum_{i=1}^{n} x_i p_i + c \sum_{i=1}^{n} p_i = \langle x \rangle + c$$

b) Jak zmieni się średnia $\langle x \rangle$ jeżeli każdą wartość x_i zmiennej X pomnożymy przez stałą c?

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} = \begin{pmatrix} cx_1 & cx_2 & \dots & cx_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$\langle y \rangle = \sum_{i=1}^{n} y_i p_i = \sum_{i=1}^{n} (c x_i) p_i = c \sum_{i=1}^{n} x_i p_i = c \langle x \rangle$$

c) Jak zmieni się wariancja jeżeli do każdej wartości x_i zmiennej losowej X dodamy stałą c?

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} = \begin{pmatrix} x_1 + c & x_2 + c & \dots & x_n + c \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$\langle y \rangle = \langle x \rangle + c$$

$$(\Delta y)^2 = \sum_{i=1}^{n} [y_i - \langle y \rangle]^2 \, p_i = \sum_{i=1}^{n} [(x_i + c) - \langle x \rangle - c]^2 \, p_i = \sum_{i=1}^{n} [x_i - \langle x \rangle]^2 \, p_i = (\Delta x)^2$$

Wariancja nie ulegnie zmianie! Żaden wskaźnik rozproszenia nie powinien zmieniać swej wartości, gdy do wszystkich elementów próby zostanie dodana ta sama liczba (dodatnia lub ujemna): mówimy wówczas o przesunięciu elementów próby o zadaną wartość. Postulat ten wynika stąd, że takie przesunięcie nie zmienia kształtu rozkładu i w szczególności nie zmienia stopnia rozproszenia próby!

d) Jak zmieni się wariancja jeżeli każdą wartość x_i zmiennej X pomnożymy przez stałą c?

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} = \begin{pmatrix} cx_1 & cx_2 & \dots & cx_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$\langle y \rangle = c\langle x \rangle$$

$$(\Delta y)^2 = \sum_{i=1}^{n} [y_i - \langle y \rangle]^2 \, p_i = \sum_{i=1}^{n} [cx_i - c\langle x \rangle]^2 \, p_i$$

$$\downarrow$$

$$(\Delta y)^2 = c^2 \sum_{i=1}^{n} {x_i}^2 \, p_i - 2c^2 \langle x \rangle \sum_{i=1}^{n} x_i \, p_i + c^2 \langle x \rangle^2 \sum_{i=1}^{n} p_i$$

$$\downarrow$$

$$(\Delta y)^2 = c^2 \langle x^2 \rangle - 2c^2 \langle x \rangle^2 + c^2 \langle x \rangle^2 = c^2 \langle x^2 \rangle - c^2 \langle x \rangle^2 = c^2 (\Delta x)^2$$

Jeżeli każdą wartość x_i zmiennej losowej X pomnożymy przez stałą c to wariancja $(\Delta y)^2$ będzie równa wariancji wyjściowej $(\Delta x)^2$ pomnożonej przez kwadrat stałej c. Zauważmy, że wariancja $(\Delta y)^2$ będzie zawsze nieujemna (niezależnie od znaku stałej c).

Standaryzacja zmiennej losowej

Zmienna losowa X jest podana za pomocą dwuwierszowej tablicy

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Przyjmijmy, że średnia arytmetyczna dla powyższego rozkładu prawdopodobieństwa wynosi $\langle x \rangle$, a odchylenie standardowe Δx. **Wartością standaryzowaną** odpowiadającą obserwacji x_i jest wartość u_i otrzymana w wyniku przekształcenia ($\Delta x \neq 0$):

$$u_i = \frac{x_i - \langle x \rangle}{\Delta x}$$

Wartości u_i odpowiadające obserwacjom $x_i < \langle x \rangle$ są ujemne, natomiast te, które odpowiadają wartościom $x_i > \langle x \rangle$ są dodatnie. Oczywiście $u_i = 0$ dla $x_i = \langle x \rangle$.

Standaryzowana zmienna U, która została otrzymana w oparciu o zmienną losową X ma postać

$$U = \begin{pmatrix} u_1 & u_2 & \dots & u_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Wyznaczmy ile wynosi średnia arytmetyczna $\langle u \rangle$ dla standaryzowanej zmiennej losowej U

$$\langle u \rangle = \sum_{i=1}^{n} u_i \, p_i \quad \wedge \quad u_i = \frac{x_i - \langle x \rangle}{\Delta x}$$

$$\langle u \rangle = \sum_{i=1}^{n} u_i \, p_i = \sum_{i=1}^{n} \left[\frac{x_i - \langle x \rangle}{\Delta x} \right] p_i = \sum_{i=1}^{n} \left[\frac{x_i p_i}{\Delta x} - \frac{\langle x \rangle p_i}{\Delta x} \right] = \frac{1}{\Delta x} \sum_{i=1}^{n} x_i p_i - \frac{\langle x \rangle}{\Delta x} \sum_{i=1}^{n} p_i$$

Ponieważ

$$\sum_{i=1}^{n} x_i p_i = \langle x \rangle$$

$$\sum_{i=1}^{n} p_i = 1$$

$$\langle u \rangle = \frac{\langle x \rangle}{\Delta x} - \frac{\langle x \rangle}{\Delta x} = 0$$

Średnia arytmetyczna $\langle u \rangle$ standaryzowanej zmiennej losowej U jest zawsze równa zero! Wartości standaryzowane u_i mogą wskazywać **nietypowe wartości** jakie przyjmuje zmienna losowa X. W praktyce, za nietypowe uznaje się takie wartości x_i, którym odpowiadają wartości standaryzowane $|u_i| > 3$.

Wyznaczmy moment zwykły drugiego rzędu dla standaryzowanej zmiennej losowej U.

$$\langle u^2 \rangle = \sum_{i=1}^{n} {u_i}^2 \, p_i$$

$$\langle u^2 \rangle = \sum_{i=1}^{n} {u_i}^2 \, p_i = \sum_{i=1}^{n} \left[\frac{x_i - \langle x \rangle}{\Delta x} \right]^2 p_i = \sum_{i=1}^{n} \left[\frac{{x_i}^2 - 2x_i \langle x \rangle + \langle x \rangle^2}{(\Delta x)^2} \right] p_i$$

$$\langle u^2 \rangle = \sum_{i=1}^{n} \left[\frac{{x_i}^2 p_i - 2x_i \langle x \rangle p_i + \langle x \rangle^2 p_i}{(\Delta x)^2} \right] = \sum_{i=1}^{n} \frac{{x_i}^2 p_i}{(\Delta x)^2} - \sum_{i=1}^{n} \frac{2x_i \langle x \rangle p_i}{(\Delta x)^2} + \sum_{i=1}^{n} \frac{\langle x \rangle^2 p_i}{(\Delta x)^2}$$

$$\langle u^2 \rangle = \frac{1}{(\Delta x)^2} \sum_{i=1}^{n} {x_i}^2 p_i - \frac{2\langle x \rangle}{(\Delta x)^2} \sum_{i=1}^{n} x_i p_i + \frac{\langle x \rangle^2}{(\Delta x)^2} \sum_{i=1}^{n} p_i$$

Ponieważ

$$\sum_{i=1}^{n} {x_i}^2 p_i = \langle x^2 \rangle$$

$$\sum_{i=1}^{n} x_i p_i = \langle x \rangle$$

$$\sum_{i=1}^{n} p_i = 1$$

$$\langle u^2 \rangle = \frac{\langle x^2 \rangle}{(\Delta x)^2} - \frac{2\langle x \rangle^2}{(\Delta x)^2} + \frac{\langle x \rangle^2}{(\Delta x)^2} = \frac{\langle x^2 \rangle - \langle x \rangle^2}{(\Delta x)^2} = \frac{(\Delta x)^2}{(\Delta x)^2} = 1$$

W związku z tym:

$$(\Delta u)^2 = \langle u^2 \rangle - \langle u \rangle^2 = 1 - 0^2 = 1$$

Wariancja $(\Delta u)^2$ standaryzowanej zmiennej losowej U jest zawsze równa jedności! Procedura standaryzacji zmienia jedynie wartości liczbowe jaki przyjmuje zmienna losowa X. Odpowiadające im wartości p_i pozostają niezmienione!

Zadanie 2. Dla zbioru danych: 2, 4, 6, 8 ($N = 4$) wyznaczmy wartość średnią, odchylenie standardowe, szerokość typowego obszaru zmienności, współczynnik zmienności, odchylenie przeciętne, rozstęp oraz przeprowadźmy procedurę standaryzacji.

$$\langle x \rangle = 5$$

$$\langle x^2 \rangle = 30$$

$$(\Delta x)^2 = \langle x^2 \rangle - \langle x \rangle^2 = 30 - 25 = 5 \quad \rightarrow \quad \Delta x = \sqrt{5}$$

$$\langle x \rangle - \Delta x : \langle x \rangle + \Delta x \;\equiv\; 5 - \sqrt{5} : 5 + \sqrt{5}$$

$$\Delta x_r = \frac{\Delta x}{|\langle x \rangle|} = \frac{\sqrt{5}}{5}$$

$$d = \frac{1}{N}\sum_{i=1}^{4}|x_i - \langle x \rangle| = \frac{|2-5| + |4-5| + |6-5| + |8-5|}{4} = 2$$

Wartości jakie przyjmuje zmienna standaryzowana U:

$$u_1 = \frac{x_1 - \langle x \rangle}{\Delta x} = \frac{2-5}{\sqrt{5}} = -\frac{3}{\sqrt{5}}$$

$$u_2 = \frac{x_2 - \langle x \rangle}{\Delta x} = \frac{4-5}{\sqrt{5}} = -\frac{1}{\sqrt{5}}$$

$$u_3 = \frac{x_3 - \langle x \rangle}{\Delta x} = \frac{6-5}{\sqrt{5}} = \frac{1}{\sqrt{5}}$$

$$u_4 = \frac{x_4 - \langle x \rangle}{\Delta x} = \frac{8-5}{\sqrt{5}} = \frac{3}{\sqrt{5}}$$