Podstawy statystyki matematycznej (opisowej)

Dyskretna (skokowa) zmienna losowa X - masa prawdopodobieństwa skupiona jest w izolowanych punktach $(x_1, x_2, \dots, x_n)$, które stanowią zbiór skończony. Dyskretna zmienna losowa X może być zdefiniowana za pomocą dwuwierszowej tablicy, która jest nazywana **rozkładem zmiennej losowej** X.

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

W powyższej tablicy $x_1, x_2, \dots, x_n$ są możliwymi wartościami zmiennej losowej X, a p_1, p_2, $\dots, p_n$ są odpowiadającymi im prawdopodobieństwami. Liczby $x_1, x_2, \dots, x_n$ mogą być dowolne, natomiast prawdopodobieństwa p_1, $p_2, \dots, p_n$ powinny spełniać trzy zasadnicze warunki

a) wszystkie p_i są większe lub równe zeru

$$p_i \geq 0$$

b) suma wszystkich p_i jest równa 1 (suma prawdopodobieństw jest unormowana do jedności)

$$p_1 + p_2 + \cdots + p_n = \sum_{i=1}^{n} p_i = 1$$

c) wszystkie p_i są liczbami rzeczywistymi

Powszechnie dziś przyjmowane **aksjomaty prawdopodobieństwa** została podane w 1933 roku przez radzieckiego matematyka Andrieja Kołmogorowa. Wprost z aksjomatów Kołmogorowa wynikają następujące własności: **prawdopodobieństwo jest miarą skończoną** co oznacza, że prawdopodobieństwa p_1, $p_2, \dots, p_n$ są określone skończonymi liczbami rzeczywistymi.

Prawdopodobieństwo P zdarzenia niemożliwego $\emptyset$ jest równe $P(\emptyset) = 0$, a prawdopodobieństwo **zdarzenia pewnego** jest równe jedności.

Zapis $P(X = x_i) = p_i$ czytamy: prawdopodobieństwo, że zmienna losowa X przyjmuje wartość x_i wynosi p_i. Prawdopodobieństwo otrzymania i-tego wyniku w próbie składającej się z N obserwacji (pomiarów), gdzie n_i to liczba wystąpień i-tego wariantu x_i cechy mierzalnej ($n_i \leq N$)

$$0 \leq p_i = \frac{n_i}{N} \leq 1$$

Moment *k*-tego rzędu - podstawowe pojęcie opisowej statystyki matematycznej

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

gdzie n to liczba różnych wariantów x_i cechy mierzalnej.

Dla $c = 0$ definiujemy **moment zwykły *k*-tego rzędu**

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

Moment zwykły zerowego rzędu ($k = 0$) **-** jest zawsze równy jedności. W związku z tym ma niewielką wartość poznawczą.

$$\langle x^0 \rangle = \sum_{i=1}^{n} {x_i}^0 \, p_i = \sum_{i=1}^{n} p_i = 1$$

Moment zwykły pierwszego rzędu ($k = 1$) - nazywamy **średnią arytmetyczną**

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

Kilka faktów dotyczących średniej arytmetycznej $\langle x \rangle$ (momentu zwykłego pierwszego rzędu)

a) jest **wielkością abstrakcyjną -** jej wartość zazwyczaj nie występuje w próbie na podstawie której jest wyznaczana,
b) jest wielkością stałą: $\langle x \rangle = constans$
c) jest wielkością mianowaną - jest wyrażona w takich samych jednostkach miary jak badana zmienna losowa (cm, g, s, itp.)
d) jest miarą prawidłową tylko w przypadku zbiorowości jednorodnych o niewielkim zróżnicowaniu wartości x_i jakie przyjmuje zmienna losowa - w miarę wzrostu asymetrii i dyspersji danych $\langle x \rangle$ traci na znaczeniu,
e) na poziom wartości średniej duży wpływa mają wartości skrajne w badanej zbiorowości statystycznej: $x_{\min}$ (wartość minimalna) i $x_{\max}$ (wartość maksymalna),
f) spełnia nierówność $x_{\min} < \langle x \rangle < x_{\max}$,
g) jest nazywana **nadzieją matematyczną**, **wartością oczekiwaną** lub po prostu **średnią,**
h) jest jedna,
i) może być ujemna, dodatnia lub równa zero w zależności od struktury badanej zbiorowości,
j) jest oznaczana za pomocą symboli $\langle x \rangle$ lub $\bar{x}$.

Zadanie 1. Wykonano serię pomiarów pewnej bezwymiarowej wielkości fizycznej. Otrzymano następujące wyniki

0, 2, 4, 6, 8, 10

Proszę obliczyć średnią arytmetyczną $\langle x \rangle$.

Seria danych składa się z $N = 6$ obserwacji, z których każda występuje w próbie tylko raz.

$$\langle x \rangle = \sum_{i=1}^{n} x_i \, p_i = \frac{1}{6}(0 + 2 + 4 + 6 + 8 + 10) = \frac{30}{6} = 5$$

Zadanie 2. Wykonano serię pomiarów pewnej bezwymiarowej wielkości fizycznej. Otrzymano następujące wyniki

2, 4, 5, 5, 6, 8

Proszę obliczyć średnią arytmetyczną $\langle x \rangle$.

Seria danych składa się z $N = 6$ obserwacji. Liczba różnych wariantów zmiennej losowej wynosi $n = 5$.

$$\langle x \rangle = \sum_{i=1}^{n} x_i \, p_i = \left(2 \cdot \frac{1}{6}\right) + \left(4 \cdot \frac{1}{6}\right) + \left(5 \cdot \frac{2}{6}\right) + \left(6 \cdot \frac{1}{6}\right) + \left(8 \cdot \frac{1}{6}\right) = \frac{30}{6} = 5$$

Zadanie 3. Wykonano serię pomiarów pewnej bezwymiarowej wielkości fizycznej. Otrzymano następujące wyniki

5, 5, 5, 5, 5, 5

Proszę obliczyć średnią arytmetyczną $\langle x \rangle$.

$$\langle x \rangle = \sum_{i=1}^{n} x_i \, p_i = 5 \cdot 1 = 5$$

Moment zwykły drugiego rzędu ($k = 2$) - znajduje zastosowanie przy wyznaczaniu **warianacji** oraz **odchylenia standardowego**.

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

W powyższej definicji pojawia się suma iloczynów wartości x_i zmiennej losowej podniesionych do kwadratu i odpowiadających im prawdopodobieństw p_i. Każda liczba podniesiona do kwadratu i pomnożona przez liczbę nieujemną ($p_i \geq 0$) jest liczbą nieujemną. Zachodzi więc relacja: $\langle x^2 \rangle \geq 0$. Moment zwykły drugiego rzędu jest średnią arytmetyczną z kwadratów wartości x_i zmiennej losowej X. W analogiczny sposób możemy zdefiniować moment zwykły k-tego rzędu, gdzie $k = 3, 4, 5, \ldots$

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

Zadanie 4. Wykonano serię pomiarów pewnej bezwymiarowej wielkości fizycznej. Otrzymano wyniki

0, 2, 4, 6, 8, 10

Proszę obliczyć ile wynosi moment zwykły drugiego rzędu $\langle x^2 \rangle$.

Mamy $N = 6$ obserwacji, z których każda występuje w próbie tylko raz

$$\langle x^2 \rangle = \sum_{i=1}^{n} {x_i}^2 \, p_i = \frac{1}{6}(0^2 + 2^2 + 4^2 + 6^2 + 8^2 + 10^2) = \frac{220}{6}$$

Zadanie 5. Wykonano serię pomiarów pewnej bezwymiarowej wielkości fizycznej. Otrzymano wyniki

2, 4, 5, 5, 6, 8

Proszę obliczyć ile wynosi moment zwykły drugiego rzędu $\langle x^2 \rangle$.

$$\langle x^2 \rangle = \sum_{i=1}^{n} {x_i}^2 \, p_i = \left(2^2 \cdot \frac{1}{6}\right) + \left(4^2 \cdot \frac{1}{6}\right) + \left(5^2 \cdot \frac{2}{6}\right) + \left(6^2 \cdot \frac{1}{6}\right) + \left(8^2 \cdot \frac{1}{6}\right) = \frac{170}{6}$$

Powracamy teraz do pierwotnej definicji **momentu *k*-tego rzędu**

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

Dla $c = \langle x \rangle$ definiujemy **moment centralny *k*-tego rzędu**

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

Moment centralny zerowego rzędu ($k = 0$) **-** jest zawsze równy jedności

$$\sum_{i=1}^{n} [x_i - \langle x \rangle]^0 \, p_i = 1$$

Moment centralny pierwszego rzędu ($k = 1$) **-** jest zawsze równy zero

$$\sum_{i=1}^{n} [x_i - \langle x \rangle]^1 \, p_i = \sum_{i=1}^{n} x_i p_i - \langle x \rangle p_i = \sum_{i=1}^{n} x_i p_i - \sum_{i=1}^{n} \langle x \rangle p_i = \langle x \rangle - \langle x \rangle \sum_{i=1}^{n} p_i = 0$$

Jak widzimy, suma odchyleń wartości x_i od wartości średniej $\langle x \rangle$ jest zawsze równa zero. Moment centralny zerowego i moment centralny pierwszego rzędu mają **niewielką wartość poznawczą** ponieważ dla dowolnego rozkładu zmiennej losowej (zwanej dalej **zmienną**) przyjmują zawsze te same wartości (odpowiednio jeden i zero).

Moment centralny drugiego rzędu ($k = 2$) nazywamy **wariancją** i oznaczamy $(\Delta x)^2$

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

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

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

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

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

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

A więc

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