Miary zróżnicowania

Miary zmienności (dyspersji) charakteryzują zbiorowość statystyczną, uwzględniając różnice między poszczególnymi jednostkami wchodzącymi w jej skład.
Dokonują one charakterystyki stopnia zróżnicowania zbiorowości ze względu na wyróżnioną cechę zmienną.

Miary zróżnicowania

Obserwowane jednostki statystyczne, mimo takiej samej średniej mogą się znacząco różnić pomiędzy sobą, np. przeciętne wynagrodzenie w dwóch przedsiębiorstwach może być takie samo bądź zbliżone, jednak pracownicy tych przedsiębiorstw mogą otrzymywać różne wynagrodzenia.

Przykład

Przeciętne wynagrodzenie w firmie A: 2,3; 2,5; 2,7;

przeciętne wynagrodzenie w firmie B: 1,9; 2,5 3,1.

Średnia wynosi 2,5.

Wartość średnia, nie pozwala na ocenę zróżnicowania lub inaczej zmienności badanej cechy (dyspersji, rozproszenia, rozrzutu).

Miary zróżnicowania

Tymczasem większa lub mniejsza zmienność w różnych sytuacjach może być mniej lub bardziej korzystna.

W przypadku sprzedaży artykułów typu koszule, garnitury, obuwie itp. małe zróżnicowanie oferty będzie niekorzystne, gdyż ograniczać będzie możliwości sprzedażowe, w takiej sytuacji zróżnicowanie powinno być dostosowane do zróżnicowania wzrostu i wagi społeczeństwa.

Miary zróżnicowania

Z kolei oceniając np. rozmiar kół pasowych, stanowiących część zamienną silnika, zróżnicowanie rozmiaru może sięgać co najwyżej setnych części milimetra. Nadmierne zróżnicowanie może powodować szybkie zużywanie się takich części.

Ocena zmienności stanowi ważne uzupełnienie oceny poziomu przeciętnego i wzbogaca wiedzę o strukturze zbiorowości.

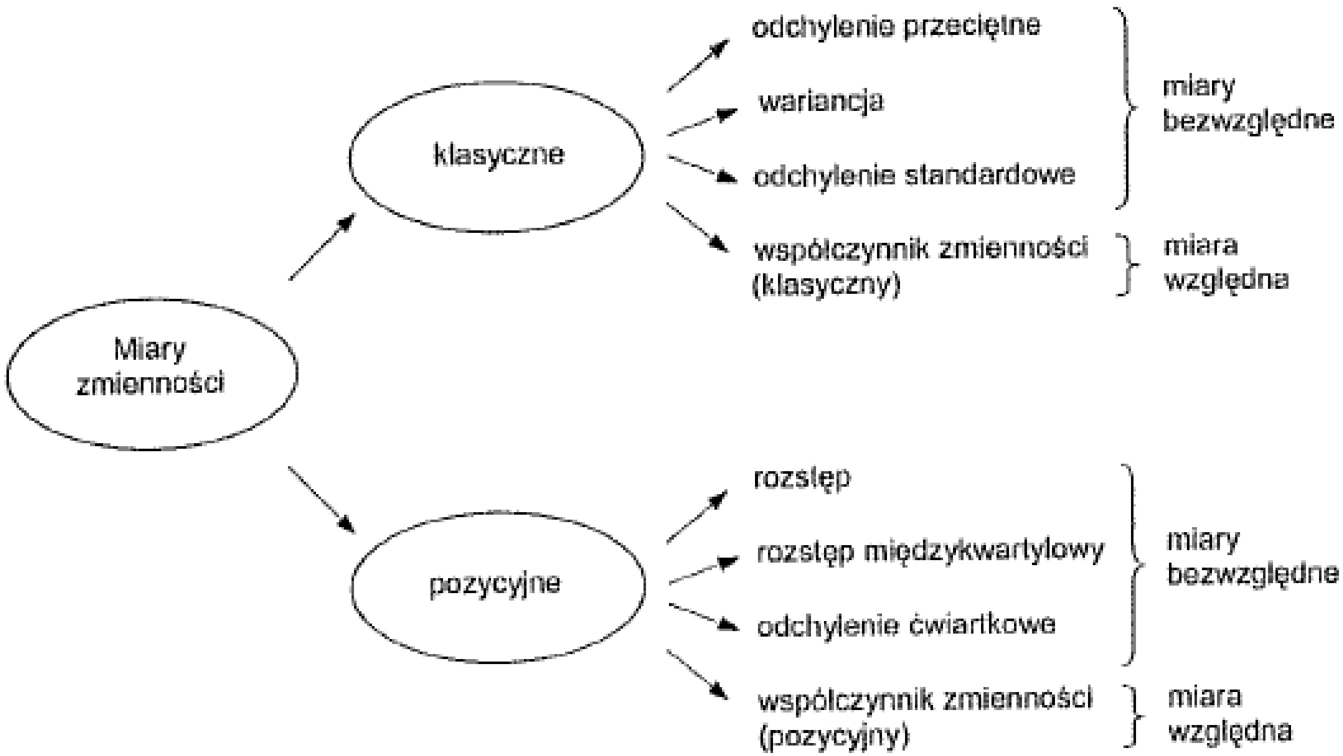
Miary zróżnicowania

Miary zróżnicowania można podzielić na miary klasyczne i miary pozycyjne.

Miary klasyczne to miary oparte na średniej. Za ich pomocą określa się zróżnicowanie wartości przyjmowanych przez obserwowane jednostki statystyczne w stosunku do wartości średniej.

Pozycyjne miary zróżnicowania pozwalają na ocenę różnicy pomiędzy wybranymi wartościami.

Klasyfikacja miar zróżnicowania



Miary zróżnicowania

Obszar zmienności - miara służąca do wstępnej oceny rozproszenia wyników.

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

gdzie:

x_{max} – największa wartość zmiennej;

x_{min} – najmniejsza wartość zmiennej.

Interpretacja: Wartości cech zawierają się w przedziale od … do….

Obszar zmienności

Przykład:

Na podstawie informacji o poziomie wynagrodzenia 12 pracowników pewnego oddziału banku ustal obszar zmienności:

2,0 2,0 2,3 2,4 2,5 2,5 2,6 3,0 3,0 3,1 3,2 9,0

$R = 9{,}0 – 2{,}0 = 7{,}0$

Wynagrodzenie ostatniego pracownika wyraźnie odstaje od pozostałych, gdyby go wykluczyć ze zbiorowości to obszar zmienności wyniósłby :

$R = 3{,}2 – 2{,}0 = 1{,}2$

Miary zróżnicowania

Miary szeroko wykorzystywane we wnioskowaniu statystycznym to wariancja i odchylenie standardowe.

Wariancja jest średnią arytmetyczną kwadratów odchyleń zmiennej *X* od jej średniej, nazywa się ją drugim momentem centralnym.

Dla szeregu wyliczającego wzór przedstawia się następująco:

$$S^2(x) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

W szeregach jednowariantowych i przedziałowych:

$$S^2(x) = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 n_i$$

$$S^2(x) = \frac{1}{n}\sum_{i=1}^{k}\left(\hat{x}_i - \bar{x}\right)^2 n_i$$

Miary zróżnicowania

Odchylenie standardowe jest pierwiastkiem kwadratowym z wariancji:

$$S(x) = \sqrt{S^2(x)}$$

Odchylenie standardowe *informuje o ile średnie zróżnicowanie obserwowanych jednostek odchyla się od wartości średniej.*

Miano odchylenia standardowego jest w takich samych jednostkach jak wartości dokonywanych pomiarów.

Odchylenie standardowe

- ***Przykład***

Wyznacz odchylenie standardowe dla szeregu przedziałowego na przykładzie informacji o wielkości działek gruntu (w arach) przeznaczonych pod budownictwo jednorodzinne.

$x_{id} - x_{ig}$	x_i	n_i	$x_i n_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert n_i$	$(x_i - \bar{x})^2 n_i$
5-10	7,5	18	135	-6,02	108,36	652,3272
10-15	12,5	20	250	-1,02	20,4	20,8080
15-20	17,5	7	122,5	3,98	27,86	110,8828
20-25	22,5	5	112,5	8,98	44,9	403,2020
25-30	27,5	4	110	13,98	55,92	781,7616
	-	**54**	**730**	-	**257,44**	**1968,9816**

$$\bar{x} = \frac{730}{54} = 13,52 \qquad S(x) = \sqrt{\frac{1}{54} \cdot 1968,9816} = 6,04$$

Wielkość badanych działek budowlanych odchyla się średnio o 6,04 a od wartości średniej 13,52 a.

Miary zróżnicowania

Odchylenie standardowe można wykorzystać do określenia **typowego obszaru zmienności** analizowanej cechy:

$$\bar{x} - S(x) < x_{typ} < \bar{x} + S(x)$$

Jeżeli rozkład cechy w zbiorowości jest rozkładem normalnym, to w **granicach**

typowego obszaru zmienności mieści się około 2/3 jednostek badanej zbiorowości.

Obserwacje o wartościach wykraczających poza typowy obszar zmienności nazywa się nietypowymi.

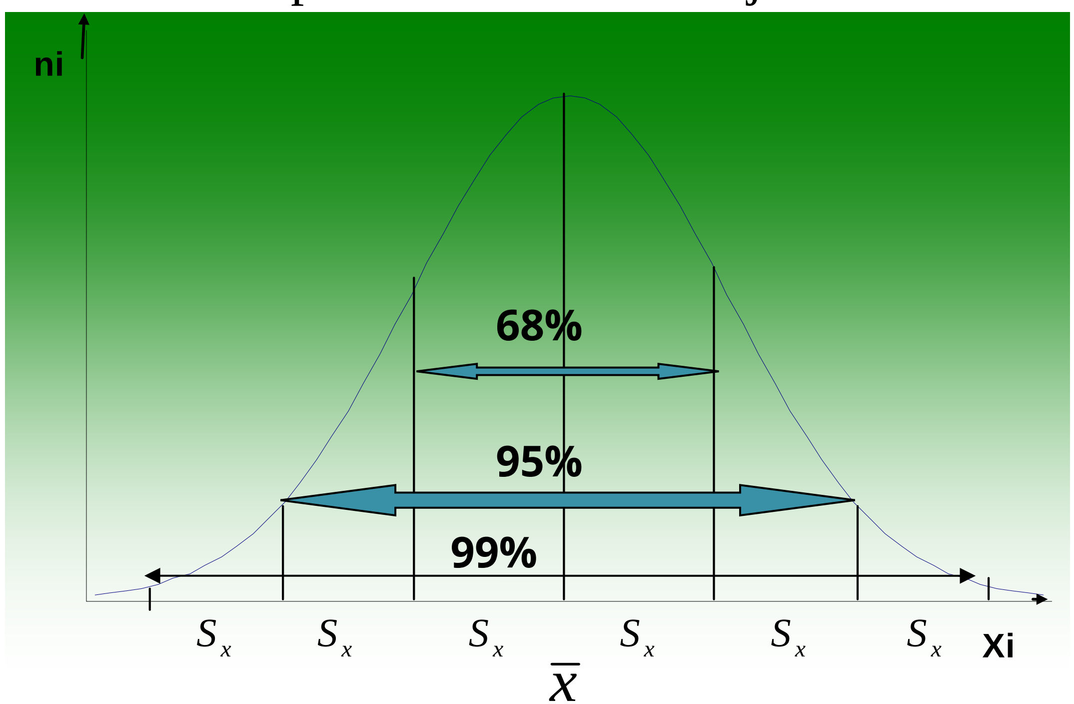
Reguła trzech sigm (Twierdzenie Czebyszewa)

- jeżeli rozkład populacji jest w przybliżeniu normalny, to:

$\bar{x} \square S_x$ zawiera około 68 % obserwacji

$\bar{x} \square 2S_x$ zawiera 95% obserwacji

$\bar{x} \square 3S_x$ zawiera ponad 99% obserwacji



Twierdzenie Czebyszewa

Ideę typowego obszaru zmienności można rozszerzyć na regułę trzech sigm. W myśl tej reguły, określanej jako twierdzenie Czebyszewa przedział:

1. zawiera około 68,27% wszystkich obserwacji;
2. zawiera około 95,45% wszystkich obserwacji (odstające);
3. zawiera około 99,73% wszystkich obserwacji (izolowane).

Nadmierna liczba obserwacji odstających o więcej niż 2 odchylenia standardowe, albo i 3 odchylenia standardowe może sugerować błędy pomiarowe.

Miary zróżnicowania

Odchylenie przeciętne jest średnią arytmetyczną z bezwzględnych wartości odchyleń zmiennej *X* od jej średniej.

Wskazuje na średnie zróżnicowanie wartości obserwowanych jednostek statystycznych od wartości przeciętnej.

Uwzględniając grupowanie danych w **szeregach jednowariantowych**:

$$d(x) = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$$

Uwzględniając grupowanie danych w **szeregach przedziałowych**:

$$d(x) = \frac{1}{n}\sum_{i=1}^{k}\left|\hat{x}_i - \bar{x}\right| n_i$$

Miary zróżnicowania

Przykład:

Wyznacz odchylenie przeciętne dla szeregu wyliczającego przedstawiającego zatrudnienie w sześciu przedsiębiorstwach.

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
8	-11,33	11,33
9	-10,33	10,33
19	-0,33	0,33
21	1,67	1,67
25	5,67	5,67
34	14,67	14,67
116	**-**	**44**

$$\bar{x} = \frac{116}{6} = 19,33$$

$$d(x) = \frac{1}{n}\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert \qquad d(x) = \frac{1}{6} \cdot 44 = 7,33$$

Miary zróżnicowania

Jeżeli wartości odchylenia przeciętnego i odchylenia standardowego odniesie się do poziomu przeciętnego, to w ten sposób ocenić można poziom zróżnicowania względnego, w ten sposób uzyskuje się współczynniki zmienności.

Współczynnik zmienności oparty na odchyleniu przeciętnym wyraża się wzorem:

$$V_{d(x)} = \frac{d(x)}{\bar{x}}$$

a **współczynnik zmienności oparty na odchyleniu standardowym**:

$$V_{s(x)} = \frac{s(x)}{\bar{x}}$$

Współczynnik zmienności

Zbiorowości o charakterze społeczno-ekonomicznym o współczynniku zmienności:

- poniżej 35% określa się jako **względnie jednorodne**
- współczynniki z przedziału 35-65% wskazują na **przeciętne zróżnicowanie**
- o współczynniku zmienności powyżej 65% jako **silnie zróżnicowane**

Miary zróżnicowania

- ***<u>Przykład</u>***

Zbadano osiem gospodarstw domowych ze względu na spożycie chleba (kg/dziennie) oraz spożycie mleka (l/dziennie).

spożycie chleba (*X*): 0,4; 0,7; 0,3; 0,7; 0,6; 0,2; 0,5; 0,8.

spożycie mleka (*Y*): 1,2; 0,4; 0,9; 1,0; 0,5; 0,9; 1,1; 0,5.

Dla zmiennych tych otrzymano:

[illegible]

[illegible]

Badane gospodarstwa są nieznacznie silniej zróżnicowanie ze względu na spożycie chleba niż spożycie mleka.

Pozycyjne miary zróżnicowania

Uzupełnieniem klasycznych miar zróżnicowania lub podstawą opisu zbiorowości w sytuacji, gdy nie można określić wartości miar klasycznych, są pozycyjne miary zróżnicowania (np. w szeregach otwartych).

Rozstęp ćwiartkowy wyznacza się ze wzoru:

$$R_Q = Q_3 - Q_1$$

Rozstęp ćwiartkowy jako różnica pomiędzy trzecim a pierwszym kwartylem **wskazuje na obszar zmienności 50% środkowych obserwacji (**zakres ograniczony do II i III ćwiartki obserwacji).

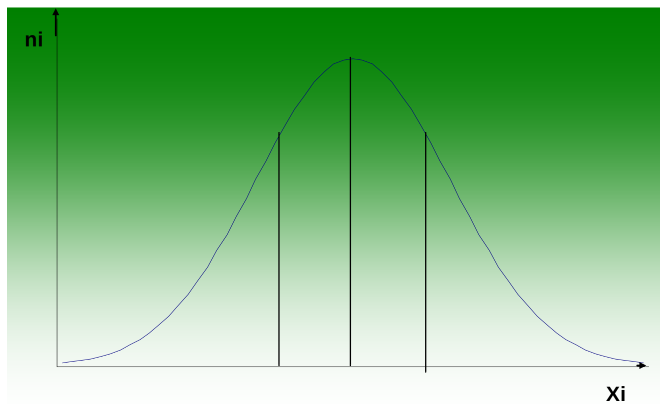
W porównaniu do klasycznego obszaru zmienności jest to miara niewrażliwa na obserwacje skrajne.

Pozycyjne miary zróżnicowania

Odchylenie ćwiartkowe:

- 1/2 obszaru zmienności 50 % środkowych jednostek zbiorowości
- przeciętne odchylenie od mediany dla 50% środkowych jednostek zbiorowości

$$Q = \frac{Q_3 - Q_1}{2}$$



Podobną informację uzyskuje się z wyznaczenia odchylenia ćwiartkowego, które interpretuje się jako **średnie zróżnicowanie elementów z II i III ćwiartki szeregu.**

Pozycyjne miary zróżnicowania

Pozycyjny współczynnik zmienności wyznacza się ze wzoru:

$$V_Q = \frac{Q}{Me} \cdot 100\%$$

Odchylenie ćwiartkowe odniesione do wartości mediany wyznacza pozycyjny współczynnik zmienności, który pozwala na ocenę względnego zróżnicowania elementów z II i III ćwiartki.

Pozycyjne miary zróżnicowania

- ***Przykład:***

Szereg przedstawia wydatki na energię elektryczną w wylosowanej próbie gospodarstw domowych:

wydatki (w zł)	liczba gosp.dom.	liczebności skumulowane
x_{id}-x_{ig}	n_i	$n_{i(sk)}$
-50	21	21
50-100	35	56
100-150	52	108
150-200	24	132
200-250	8	140
250-	7	147
-	147	-

Q1=72,5; Me=116,83; Q3=154,69.

Wyznacz rozstęp ćwiartkowy, odchylenie ćwiartkowe oraz pozycyjny współczynnik zmienności.

Pozycyjne miary zróżnicowania

- ***<u>Rozwiązanie:</u>***

Q1=72,5;   Me=116,83;   Q3=154,69.

Rozstęp ćwiartkowy

$$R_Q = Q_3 - Q_1 \qquad\qquad R_Q = 154,69 - 72,5 = 82,19$$

Różnica pomiędzy wydatkami w grupie 50% środkowych gospodarstw domowych nie przekracza 82,19 zł.

Odchylenie ćwiartkowe

$$Q = \frac{Q_3 - Q_1}{2} \qquad\qquad Q = \frac{82,19}{2} = 41,1$$

Średnie zróżnicowanie wydatków w gospodarstwach domowych należących do II i III ćwiartki wynosi 41,1 zł.

Pozycyjny współczynnik zmienności:

$$V_Q = \frac{Q}{Me} \cdot 100\% \qquad\qquad V_Q = \frac{41,1}{116,83} \cdot 100\%$$

Gospodarstwa domowe ograniczone do II i III ćwiartki stanowią zbiorowość przeciętnie zróżnicowaną ze względu na ponoszone wydatki na energię elektryczną.

W praktyce badawczej zdarza się, że pobiera się z populacji generalnej jednocześnie kilka próbek, które charakteryzuje się osobno ze względu na wartość przeciętną i odchylenie standardowe (wariancję).

Średnią arytmetyczną dla próbek połączonych wyznacza się ze wzoru:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r}\bar{x}_i n_i ,$$

gdzie:
r – liczba próbek;
$\bar{x}_i$ – wartość średnia i-tej próbki, $i = 1, 2, \ldots, r$;
n_i – liczebność i-tej próbki, $i = 1, 2, \ldots, r$;
n – liczebność próbek połączonych $n = \sum_{i=1}^{r} n_i$.

Wzór na średnią arytmetyczną z próbek połączonych analogiczny jest do wzoru na średnią ważoną. Wagami dla średnich z poszczególnych próbek są ich liczebności.

Wariancję dla próbek połączonych wyznacza się ze wzoru:

$$S^2 = \overline{s_i^2} + s^2(\bar{x}_i) = \frac{1}{n}\sum_{i=1}^{r} s_i^2 n_i + \frac{1}{n}\sum_{i=1}^{r}(\bar{x}_i - \bar{x})^2 n_i ,$$

wariancja = wariancja wewnątrzgrupowa + wariancja międzygrupowa

gdzie:
s_i^2 – wartość wariancji i-tej próbki, $i = 1, 2, \ldots, r$;
pozostałe oznaczania jak wyżej.

Składnikami wzoru na wariancję są wariancja wewnętrzna (wewnątrzgrupowa) oraz wariancja zewnętrzna (międzygrupowa). Wariancja wewnątrzgrupowa pozwala na ocenę przeciętnego poziomu zróżnicowania w pojedynczej grupie, a wariancja międzygrupowa pozwala na ocenę zróżnicowania pomiędzy grupami. Całkowite zróżnicowanie obserwacji w próbach połączonych jest sumą zróżnicowania wewnątrzgrupowego i międzygrupowego.

Przykład. W czterech oddziałach przedsiębiorstwa stwierdzono następujące parametry dotyczące wynagrodzeń: (Tabela Obliczenia pomocnicze dla wyznaczenia średniej i wariancji z prób połączonych)

r	n_i	$\bar{x}_i$	s_i^2	$\bar{x}_i \cdot n_i$	$s_i^2 \cdot n_i$	$(\bar{x}_i - \bar{x})^2 \cdot n_i$
1	25	3,4	0,27	85,0	6,75	0,1444
2	40	2,9	0,37	116,0	14,80	7,1910
3	37	3,5	0,34	129,5	12,58	1,1461
4	18	3,8	0,29	68,4	5,22	4,0784
-	120	-	-	398,9	39,35	12,5599

$$\bar{x} = \frac{1}{120} \cdot 398{,}9 = 3{,}324\,,$$

$$S^2 = \frac{1}{120} \cdot 39{,}35 + \frac{1}{120} \cdot 12{,}5599 = 0{,}328 + 0{,}105 = 0{,}433\,,$$

$$S = \sqrt{0{,}433} = 0{,}658\,.$$

Z uzyskanych rezultatów wynika, że przeciętny poziom wynagrodzeń w badanym przedsiębiorstwie to 3,324 tys. zł z odchyleniem standardowym 0,658 tys. zł. Ze składowych wariancji ogólnej wynika, że zróżnicowanie wewnątrzgrupowe (0,328) jest silniejsze niż zróżnicowanie międzygrupowe (0,105).

Klasyczne miary zróżnicowania mogą być wyznaczane jedynie w szeregach wyliczających oraz rozdzielczych domkniętych. Dla danych umieszczonych w takich szeregach możliwe jest wyznaczenie wartości średniej, i tym samym odchylenia od niej. W sytuacji, gdy szereg rozdzielczy jest typu otwartego i nie można wyznaczyć wartości średniej, to automatycznie nie można wyznaczyć żadnej z miar klasycznych. Rozwiązaniem w takiej sytuacji może być skorzystanie z pozycyjnych miar zróżnicowania: rozstępu ćwiartkowego, odchylenia ćwiartkowego i pozycyjnego współczynnika zmienności.