MIARY POŁOŻENIA I ROZPROSZENIA WYNIKÓW SERII POMIAROWYCH

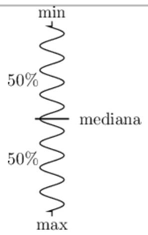
	MIARY KLASYCZNE Miary opisujące rozkład badanej cechy w zbiorowości, które obliczamy na podstawie wszystkich zaobserwowanych wartości cechy	**MIARY POZYCYJNE** Miary opisujące rozkład badanej cechy statystycznej, które obliczamy na podstawie tylko niektórych wartości cechy, zajmujących szczególną pozycję w szeregu statystycznym
MIARY POŁOŻENIA Opisują średni lub typowy poziom wartości cechy. Określają tą wartość cechy, wokół której skupiają się wszystkie pozostałe wartości badanej cechy. Wśród nich można wyróżnić miary tendencji centralnej wskazujące położenie centralnych (przeciętnych) wartości cechy w rozkładzie	średnia arytmetyczna średnia geometryczna średnia harmoniczna inne średnie	dominanta mediana kwantyle
MIARY ZRÓŻNICOWANIA (rozproszenia, rozrzutu, dyspersji) miary opisujące jak bardzo zróżnicowane są wartości cechy w zbiorowości	odchylnie przeciętne odchylenie standardowe wariancja klasyczny współczynnik zmienności	Rozstęp (max-min) Rozpiętość (max-min+1) Rozstęp ćwiartkowy (Q_3-Q_1) odchylenie ćwiartkowe (Q_3-Q_1)/2

Miary położenia (tendencji centralnej) to tzw. miary przeciętne charakteryzujące średni lub typowy poziom wartości cechy.

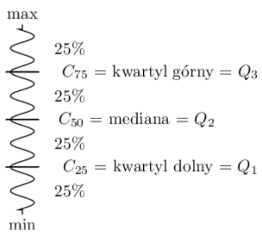
Średnia arytmetyczna: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, gdzie X_i to kolejne wartości zmiennej a n - liczność próby.

Moda/Dominanta (M_o) jest to wartość, która występuje najczęściej wśród uzyskanych pomiarów.

Mediana (M_e, Q_2) W **uporządkowanym** zbiorze danych mediana jest wartością dzielącą ten zbiór na dwie równe części. Połowa wszystkich obserwacji znajduje się poniżej, a połowa powyżej mediany.



Kwartyle (Q_1, Q_2, Q_3) dzielą uporządkowany szereg na 4 równe części. Drugi kwartyl jest równy medianie.



Miary rozproszenia - Znajomość miar tendencji centralnej nie wystarcza do scharakteryzowania struktury zbiorowości statystycznej. Badana grupa może charakteryzować się różnym stopniem zmienności w zakresie badanej cechy. Potrzebne są zatem formuły pozwalające wyznaczyć wartości, które charakteryzują rozrzut danych.

Rozstęp: $R = X_{\max} - X \min$, gdzie X_i to wartości badanej zmiennej.

Rozstęp kwartylowy: $R_Q = Q_3 - Q_1$, gdzie Q_1 i Q_3 to dolny i górny kwartyl.

Wariancja (z próby) - mierzy stopień rozproszenia pomiarów wokół średniej arytmetycznej: $s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$,

gdzie X_i to kolejne wartości zmiennej a $\overline{x}$ to średnia arytmetyczna tych wartości, n - liczność próby.

Odchylenie standardowe mierzy stopień rozproszenia pomiarów wokół średniej arytmetycznej: $s = \sqrt{s^2}$.

Im wyższa wartość odchylenia standardowego lub wariancji, tym bardziej zróżnicowana grupa pod względem badanej cechy. **Uwaga!** Odchylenie standardowe z próby jest pewnym przybliżeniem (estymatorem) odchylenia standardowego z populacji. Populacyjna wartość odchylenia standardowego mieści się w pewnym przedziale zawierającym odchylenie standardowe z próby. Przedział ten nazywany jest przedziałem ufności dla odchylenia standardowego.

Odchylenie przeciętne: $d = \frac{\sum_{i=1}^{n}(X_i - \overline{X})}{n}$

Odchylenie ćwiartkowe (Q) jest miarą rozproszenia wartości cechy od mediany: $Q = \frac{Q_3 - Q_1}{2}$

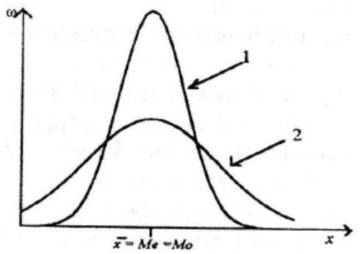
Współczynnik zmienności podobnie jak odchylenie standardowe pozwala ocenić stopień jednorodności badanej zbiorowości: $V_s = \frac{s}{\overline{X}}$ lub $V_d = \frac{d}{\overline{X}}$, gdzie *s* to odchylenie standardowe, *d* – odchylenie przeciętne, $\overline{x}$ - średnia arytmetyczna. Jest to wielkość niemianowana. Pozwala on na ocenę zróżnicowania kilku zbiorowości pod względem tej samej cechy oraz tej samej zbiorowości pod względem kilku różnych cech (wyrażonych w różnych jednostkach). Przyjmuje się, że jeżeli współczynnik *V* nie przekracza 10%, to cechy wykazują zróżnicowanie statystycznie nieistotne.

Współczynnik zmienności pozycyjny: $V_Q = \frac{Q}{M_e}$.

Błąd średniej arytmetycznej: $s_r = \frac{s}{\sqrt{n}}$.

Uwaga! Na podstawie błędu standardowego średniej arytmetycznej z próby można określić przedział ufności dla parametru populacji.

Ocena rozproszenia na podstawie obserwacji diagramów



Na rysunku pokazano dwa diagramy częstości (1) i (2). Dla uproszczenia miary położenia (średnia, mediana i modalna) są sobie równe i identyczne dla obu zbiorowości.

- Mniejsze rozproszenie wokół średniej występuje w zbiorowości (1). Diagram jest smuklejszy i wyższy.
- Większe rozproszenie wokół średniej występuje w zbiorowości (2). Diagram jest bardziej rozłożysty i niższy.

Odchylenie standardowe w zbiorowości (1) jest mniejsze niż w zbiorowości (2)

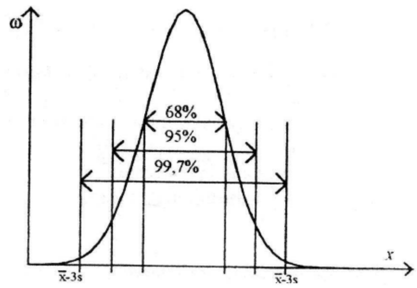
$$s_1 < s_2$$

Przedział TYPOWYCH wartości cechy (miary klasyczne)

$$\overline{x} - s < x_{typ} < \overline{x} + s$$

Przedział taki ma tą własność, że około70% jednostek badanej zbiorowości charakteryzuje się wartością cechy należącą do tego przedziału.

Reguła „3 sigma”

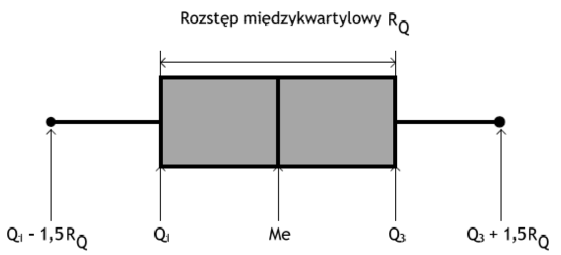


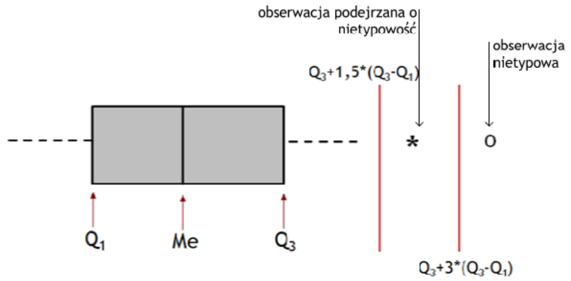
Przedział TYPOWYCH wartości cechy (miary pozycyjne)

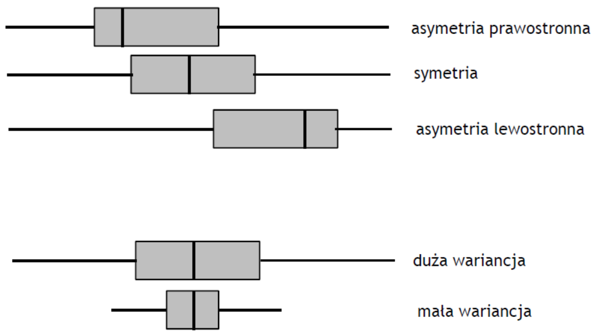
Definiujemy go podobnie jak w przypadku miar klasycznych (rolę średniej przejmuje tutaj mediana, a rolę odchylenia standardowego – odchylenie ćwiartkowe)

$$M_e - Q < x_{typ} < M_e + Q$$

Wykres RAMKA-WĄSY



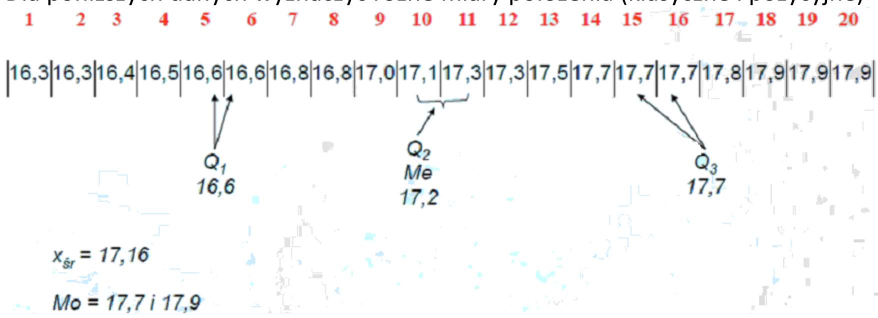




PRZYKŁADY i ZADANIA

Przykład 1.

Dla poniższych danych wyznaczyć różne miary położenia (klasyczne i pozycyjne)



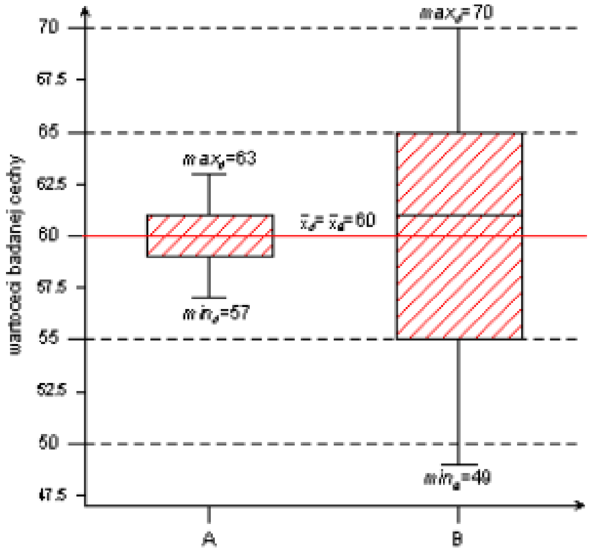
Przykład 2. Policzyć odchylenia standardowe dla serii pomiarowych uzyskanych z wykorzystaniem przyrządów kontrolno pomiarowych o różnej rozdzielczości

	Przyrząd 1	Przyrząd 2	Przyrząd 3
Uzyskane wyniki	17	16,8	16,83
	17	17,1	17,14
	17	16,9	16,88
	17	17,4	17,43
	17	17,3	17,27
	17	17,2	17,24
	17	17,0	16,96
s	0	0,22	0,223

Przykład 3. Dla obu serii danych, uzyskanych 2 różnymi metodami A i B, wykreślić i skomentować wykresy ramka-wąsy

Metoda	Dane									
A	60	59	58	61	60	61	57	62	59	63
B	53	60	67	49	65	62	56	70	63	55

Obie serie pomiarowe cechuje taka sama ilość pobranych próbek/uzyskanych wyników (n=10). Średnie arytmetyczne są sobie równe i wynoszą 60. Czy istnieje jakaś różnica pomiędzy obiema seriami?

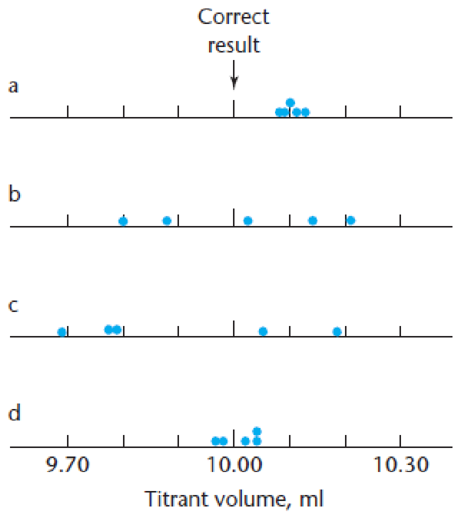


obszar	n_i	$\bar{x}$	M_e	min	max	Q_1	Q_3	R
A	10	60	60	57	63	59	61	6
B	10	60	61	49	70	55	65	21

Wyniki uzyskane obiema metodami różnią się PRECYZJĄ, co obrazuje rozrzut danych wokół wartości średniej. Dla serii A wartość *min* jest znacznie wyższa niż dla serii B, a wartość *max* jest znacznie niższa niż dla serii B.

Przykład 4. Czterech studentów (A, B, C i D) miareczkowało tą samą analizę. Ich wyniki są w poniższej tabeli. Udowodnić stwierdzenia podane w komentarzach; wartość uznana za prawdziwą to 10,00 ml.

Student	Results (ml)					Comment
A	10.08	10.11	10.09	10.10	10.12	Precise, biased
B	9.88	10.14	10.02	9.80	10.21	Imprecise, unbiased
C	10.19	9.79	9.69	10.05	9.78	Imprecise, biased
D	10.04	9.98	10.02	9.97	10.04	Precise, unbiased



In summary, precision describes random error, bias describes systematic error, and the accuracy, i.e. closeness to the true value of a single measurement or a mean value, incorporates both types of error.

Zadanie 1.

A standard sample of pooled human blood serum contains 42.0 g of albumin per litre. Five laboratories (A–E) each do six determinations (on the same day) of the albumin concentration, with the following results (g l^{-1} throughout):

A	42.5	41.6	42.1	41.9	41.1	42.2
B	39.8	43.6	42.1	40.1	43.9	41.9
C	43.5	42.8	43.8	43.1	42.7	43.3
D	35.0	43.0	37.1	40.5	36.8	42.2
E	42.2	41.6	42.0	41.8	42.6	39.0

Comment on the bias, precision and accuracy of each of these sets of results.

Przykład 5. Obliczanie przedziału ufności

Table 2.1 Results of 50 determinations of nitrate ion concentration, in μg ml^{-1}

0.51	0.51	0.51	0.50	0.51	0.49	0.52	0.53	0.50	0.47
0.51	0.52	0.53	0.48	0.49	0.50	0.52	0.49	0.49	0.50
0.49	0.48	0.46	0.49	0.49	0.48	0.49	0.49	0.51	0.47
0.51	0.51	0.51	0.48	0.50	0.47	0.50	0.51	0.49	0.48
0.51	0.50	0.50	0.53	0.52	0.52	0.50	0.50	0.51	0.51

For large samples, the confidence limits of the mean are given by

$$\bar{x} \pm zs/\sqrt{n}$$

where the value of z depends on the degree of confidence required.

For 95% confidence limits, $z = 1.96$
For 99% confidence limits, $z = 2.58$
For 99.7% confidence limits, $z = 2.97$

For small samples, the confidence limits of the mean are given by

$$\bar{x} \pm t_{n-1}s/\sqrt{n}$$

Calculate the 95% and 99% confidence limits of the mean for the nitrate ion concentration measurements in Table 2.1.

We have $\bar{x} = 0.500$, $s = 0.0165$ and $n = 50$. Using equation (2.8) gives the 95% confidence limits as:

$$\bar{x} \pm 1.96s/\sqrt{n} = 0.500 \pm 1.96 \times 0.0165/\sqrt{50} = 0.500 \pm 0.0046 \text{ μg ml}^{-1}$$

and the 99% confidence limits as:

$$\bar{x} \pm 2.58s/\sqrt{n} = 0.500 \pm 2.58 \times 0.01651/\sqrt{50} = 0.500 \pm 0.0060 \text{ μg ml}^{-1}$$

Degrees of freedom	Values of t for confidence interval of	
	95%	99%
2	4.30	9.92
5	2.57	4.03
10	2.23	3.17
20	2.09	2.85
50	2.01	2.68
100	1.98	2.63

Przykład 6.

The sodium ion content of a urine specimen was determined by using an ion-selective electrode. The following values were obtained: 102, 97, 99, 98, 101, 106 mM. What are the 95% and 99% confidence limits for the sodium ion concentration?

The mean and standard deviation of these values are 100.5 mM and 3.27 mM respectively. There are six measurements and therefore 5 degrees of freedom. From Table A.2 the value of t_5 for calculating the 95% confidence limits is 2.57 and from equation (2.9) the 95% confidence limits of the mean are given by:

$$100.5 \pm 2.57 \times 3.27/\sqrt{6} = 100.5 \pm 3.4 \text{ mM}$$

Similarly the 99% confidence limits are given by:

$$100.5 \pm 4.03 \times 3.27/\sqrt{6} = 100.5 \pm 5.4 \text{ mM}$$

Przykład 7.

The absorbance scale of a spectrometer is tested at a particular wavelength with a standard solution which has an absorbance given as 0.470. Ten measurements of the absorbance with the spectrometer give $\bar{x} = 0.461$, and $s = 0.003$. Find the 95% confidence interval for the mean absorbance as measured by the spectrometer, and hence decide whether a systematic error is present.

The 95% confidence limits for the absorbance as measured by the spectrometer are [equation (2.9)]:

$$\bar{x} \pm t_{n-1}s/\sqrt{n} = 0.461 \pm 2.26 \times 0.003/\sqrt{10} = 0.461 \pm 0.002$$

(The value of t_9 was obtained from Table A.2.)

Since the confidence interval does not include the known absorbance of 0.470, it is likely that a systematic error has occurred.

Zadanie 2.

The reproducibility of a method for the determination of selenium in foods was investigated by taking nine samples from a single batch of brown rice and determining the selenium concentration in each. The following results were obtained:

0.07 0.07 0.08 0.07 0.07 0.08 0.08 0.09 0.08 $\mu g\ g^{-1}$

(Moreno-Domínguez, T., García-Moreno, C. and Mariné-Font, A. 1983. *Analyst* 108: 505)

Calculate the mean, standard deviation and relative standard deviation of these results.

Zadanie 3.

Ten replicate analyses of the concentration of mercury in a sample of commercial gas condensate gave the following results:

23.3 22.5 21.9 21.5 19.9 21.3 21.7 23.8 22.6 24.7 $ng\ ml^{-1}$

(Shafawi, A., Ebdon, L., Foulkes, M., Stockwell, P. and Corns, W. 1999. *Analyst* 124: 185)

Calculate the mean, standard deviation, relative standard deviation and 99% confidence limits of the mean.

Zadanie 4.

In an evaluation of a method for the determination of fluorene in sea-water, a synthetic sample of sea-water was spiked with 50 $ng\ ml^{-1}$ of fluorene. Ten replicate determinations of the fluorene concentration in the sample had a mean of 49.5 $ng\ ml^{-1}$ with a standard deviation of 1.5 $ng\ ml^{-1}$.

(Gonsález, M. A. and López, M. H. 1998. *Analyst* 123: 2217)

Calculate the 95% confidence limits of the mean. Is the spiked value of 50 $ng\ ml^{-1}$ within the 95% confidence limits?

Zadanie 5.

A 0.1 M solution of acid was used to titrate 10 ml of 0.1 M solution of alkali and the following volumes of acid were recorded:

9.88 10.18 10.23 10.39 10.21 ml

Calculate the 95% confidence limits of the mean and use them to decide whether there is any evidence of systematic error.