MIANO ROZTWORU TITRANTA

Analiza statystyczna wyników oznaczeń

Estymatory punktowe

Średnia arytmetyczna $\left(\bar{x}\right)$ – jest to suma wyników w serii podzielona przez ich liczbę:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

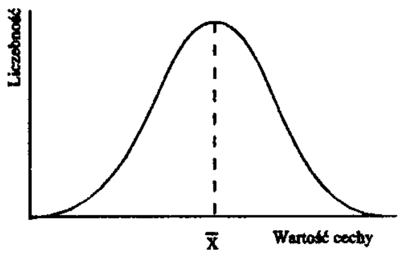
gdzie:
x_i - wynik poszczególnego oznaczenia
n - liczba pomiarów w serii

Mediana (M) – jest to wartość środkowa w uporządkowanym szeregu wartości pomiarów np. gdy dokonano dziewięciu pomiarów, to wartością środkową (medianą) będzie piąta wartość.

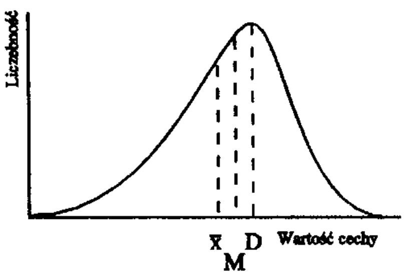
Gdy liczba pomiarów jest parzysta mediana jest równa średniej z dwóch wyników środkowych. Wyznaczanie mediany sprowadza się do uszeregowania wyników według rosnącej wartości i wyborze wartości środkowej tego ciągu. Mediana nie zależy od wyników skrajnych. Nie jest wymagane jak w przypadku wartości średniej stosowanie procedur odrzucenia wyników wątpliwych.

Dominanta (D) – określa, która z wartości w danej zbiorowości występuje najczęściej (dominuje). Można ją wyznaczyć metodą graficzną. Jeśli na osi odciętych odłożymy wartość cechy (wynik pomiaru) a na osi rzędnych liczbę pomiarów przyjmujących określoną wartość cechy, to dominanta jest wartością odciętej odpowiadającą maksimum krzywej liczebności. Dominanta podobnie jak mediana nie zależy od wyników skrajnych.

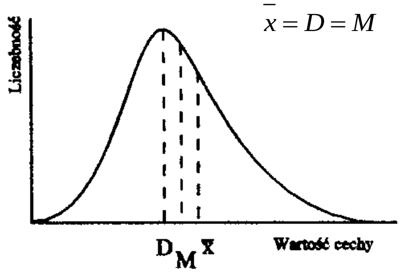
Rozkłady liczebności mogą być różne – symetryczne i niesymetryczne. Zależnie od rozkładu cechy zachodzą trzy podstawowe relacje pomiędzy średnią arytmetyczną $\bar{x}$, dominantą D i medianą M przedstawione na rysunkach:



Rys. 1. Symetryczny rozkład cechy $\bar{x} = D = M$



Rys. 2. Lewostronnie skośny cechy $\bar{x} < D < M$



Rys. 3. Prawostronnie skośny rozkład rozkład cechy $\bar{x} > D > M$

Zastosowanie powyższych miar centralnej tendencji ma jednak sens tylko przy większej ilości danych pomiarowych.

Estymacja przedziałowa

Metoda estymacji punktowej, dążąca do przyporządkowania parametrom wartości liczbowej stanowiącej jego statystyczną ocenę ma swoje wady, ponieważ ocena ta nie musi się pokrywać z prawdziwą wartością parametru. W takim przypadku niezbędne jest podanie informacji o wielkości oczekiwanych odstępstw od wartości rzeczywistej. Problem ten sprowadza się do pytania: jak wielki musi być przedział Δx dookoła wartości stanowiącej wynik oznaczenia *x*, aby prawdopodobieństwo tego, że wartość oczekiwana populacji mieści się w tym przedziale miało zadaną wartość. Warunek ten można opisać nierównością:

$$x - \Delta x < E(x) < x + \Delta x$$

gdzie: $E(x)$ – estymacja punktowa wartości *x*

Tak określony przedział nosi nazwę przedziału ufności a zadana wartość prawdopodobieństwa – poziomu ufności. W celu wyznaczenia przedziału ufności średniej arytmetycznej korzystamy z testu W.S. Gosseta (tzw. testu t-Studenta) Gosset wprowadził nową zmienną losową, która ma rozkład prawdopodobieństwa **t**, symetryczny, niezależny od liczby oznaczeń, podobny do rozkładu normalnego. W przypadku liczby pomiarów n>30 rozkład prawdopodobieństwa t i rozkład Gaussa praktycznie się pokrywają. Gdy liczba pomiarów jest mniejsza, niepewność oszacowania odchylenia standardowego jest kompensowana większą wartością parametru **t** (Tabela 6.1)
Przedział ufności średniej arytmetycznej wyznacza się z zależności:

$$\mu = \bar{x} \pm t \cdot \bar{s}$$

gdzie: μ - wartość rzeczywista
t - wartość parametru funkcji Studenta
$\bar{s}$ - odchylenie standardowe średniej z próby
$\bar{x}$ - wartość średnia z próby

$$\bar{s} = \frac{s}{\sqrt{n}}$$

s – odchylenie standardowe z próby. Liczbowo odchylenie standardowe z próby jest równe pierwiastkowi kwadratowemu z wariancji.

Wariancja (*V*) jest miarą efektywności oszacowania średniej arytmetycznej z próby (miara precyzji):

$$V = \frac{\sum \left(x_i - \bar{x}\right)^2}{n - 1}$$

$$s = \sqrt{V} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Odchylenie standardowe z próby (s) stanowi ocenę błędu przypadkowego, jakim obarczone jest oznaczenie. To odchylenie określa rozrzut wyników wokół wartości średniej. Wartość s jest miarą powtarzalności wyników, gdy analiza jest wykonywana przez tego samego analityka. Im wartość s jest mniejsza tym mniejszy jest rozrzut wyników a analiza wykonywana precyzyjniej. Odchylenie standardowe ma wymiar wielkości mierzonej.

Szerokość przedziału ufności $\left(2 \cdot t\bar{s}\right)$ wyrażana jest w takich jednostkach, w jakich podane są wyniki pomiarów. Można operować także tzw. względną szerokością przedziału (τ) wyrażoną w procentach:

$$\tau = \frac{\left(\bar{x} + t\bar{s}\right) - \left(\bar{x} - t\bar{s}\right)}{\bar{x}} \cdot 100\,\% = \frac{2t\bar{s}}{\bar{x}} \cdot 100\,\%$$

Współczynnik zmienności (względne odchylenie standardowe) (RSD) podawany w procentach jest przykładem błędu względnego tj. błędu oszacowania dzielonego przez oszacowaną absolutną wartość mierzonej wielkości:

$$RSD\% = \frac{s}{\bar{x}} \cdot 100\,\%$$

Przedział ufności może być wykorzystywany do sprawdzenia czy pomiary obarczone są błędem systematycznym. Jeśli przedział ufności nie zawiera znanej (prawdziwej) wartości wielkości mierzonej wskazuje to na obecność błędu systematycznego (dodatniego lub ujemnego).

Tabela 6.1. Wartości współczynnika t oraz parametru $t/\sqrt{n}$ w zależności od liczby oznaczeń (n), poziomu ufności (p) lub poziomu istotności (α) i prawdopodobieństwa (w %).

Liczba oznaczeń	Stopnie swobody K = n-1	Poziom ufności p, poziom istotności α, prawdopodobieństwo w %					
		p = 0,90; α = 0,1; 90°/o		p = 0,95; α = 0,05; 95°/o		p = 0,99; α = 0,01; 99°/o	
		t	$t/\sqrt{n}$	t	$t/\sqrt{n}$	t	$t/\sqrt{n}$
2	1	6,314	4,478	12,706	9,00	63,656	45,146
3	2	2,920	1,688	4,303	2,48	9,92	5,734
4	3	2,353	1,177	3,182	1,59	5,841	2,921
5	4	2,132	0,953	2,776	1,24	4,604	2,06
6	5	2,015	0,901	2,571	1,05	4,032	1,646
7	6	1,943	0,703	2,447	0,92	3,707	1,399
8	7	1,895	0,670	2,365	0,84	3,499	1,236
9	8	1,860	0,620	2,306	0,77	3,355	1,118
10	9	1,833	0,580	2,262	0,72	3,250	1,028 ·
11	10	1,80	0,54	2,20	0,66	3,11	0,94
12	11	1,78	0,51	2,18	0,63	3,06	0,88
13	12	1,77	0,49	2,16	0,59	3,01	0,83
14	13	1,76	0,47	2,14	0,57	2,98	0,80
15	14	1,75	0,45	2,13	0,55	2,95	0,76
16	15	1,75	0,44	2,12	0,53	2,92	0,73
17	16	1,74	0,42	2,11	0,51	2,90	0,70
18	17	1,73	0,40	2,10	0,49	2,88	0,68
19	18	1,73	0,40	2,09	0,48	2,86	0,66
20	19	1,72	0,38	2,09	0,47	2,85	0,64
22	21	1,72	0,37	2,07	0,44	1,82	0,60
24	23	1,71	0,35	2,06	0,42	2,80	0,57
26	25	1,71	0,34	2,06	0,41	2,78	0,56
28	27	1,70	0,33	2,05	0,39	2,76	0,53
30	29	1,67	0,31	2,04	0,38	2,75	0,51

Przykład 1.
Oznaczając stężenie roztworu HCl otrzymano wyniki zamieszczone w poniższej tabeli:

Numer pomiaru	Stężenie HCl [mol/dm^3]	Numer pomiaru	Stężenie HCl [mol/dm^3]	Numer pomiaru	Stężenie HCl [mol/dm^3]
1	0,9987	11	0,9990	21	0,9992
2	0,9973	12	0,9989	22	0,9984
3	0,9986	13	0,9978	23	0,9981
4	0,9980	14	0,9971	24	0,9987
5	0,9975	15	0,9982	25	0,9987
6	0,9982	16	0,9983	26	0,9983
7	0,9984	17	0,9988	27	0,9982
8	0,9982	18	0,9975	28	0,9991
9	0,9981	19	0,9980	29	0,9981
10	0,9980	20	0,9994	30	0,9969

Przeprowadzić statystyczną ocenę serii pomiarowej.
Obliczyć wartość średniej arytmetycznej $(\bar{x})$, wariancji (V), odchylenia standardowego (s) oraz względnego odchylenia standardowego (RSD). Wyznaczyć wartości dominanty (D) i mediany (M) oraz przedział ufności średniej arytmetycznej (dla zadanego poziomu istotności).

Rozwiązanie

Wartość średniej arytmetycznej wynosi:

$$\bar{x}_{30} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{30}\big(0{,}9987 + 0{,}9973 + 0{,}9986 + 0{,}9980 + 0{,}9975 + 0{,}9982 + 0{,}9984 + 0{,}9982 +$$
$$+ 0{,}9981 + 0{,}9980 + 0{,}9990 + 0{,}9989 + 0{,}9978 + 0{,}9971 + 0{,}9982 + 0{,}9983 +$$
$$+ 0{,}9988 + 0{,}9975 + 0{,}9980 + 0{,}9994 + 0{,}9992 + 0{,}9984 + 0{,}9981 + 0{,}9987 +$$
$$+ 0{,}9987 + 0{,}9983 + 0{,}9982 + 0{,}9991 + 0{,}9981 + 0{,}9969\big) = 0{,}9983$$

W celu wyznaczenia mediany (M) uszeregowujemy wyniki według wzrastającej wartości:

Numer pomiaru	Stężenie HCl [mol/dm^3]	Numer pomiaru	Stężenie HCl [mol/dm^3]	Numer pomiaru	Stężenie HCl [mol/dm^3]
1	0,9969	11	0,9981	21	0,9986
2	0,9971	12	0,9981	22	0,9987
3	0,9973	13	0,9982	23	0,9987
4	0,9975	14	0,9982	24	0,9987
5	0,9975	15	0,9982	25	0,9988

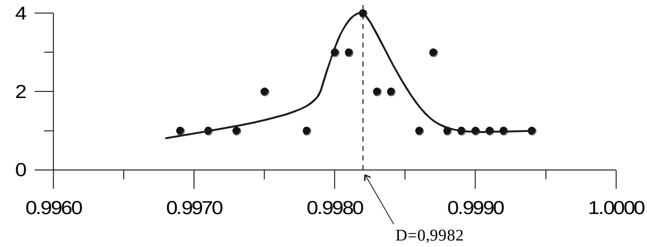
6	0,9978	16	0,9982	26	0,9989
7	0,9980	17	0,9983	27	0,9990
8	0,9980	18	0,9983	28	0,9991
9	0,9980	19	0,9984	29	0,9992
10	0,9981	20	0,9984	30	0,9994

Wartość mediany wynosi:

$M = 0{,}9982$

Wartość dominanty wynosi:

$D = 0{,}9982$



Wartości $\overline{x}$, D i M są identyczne – rozkład jest symetryczny.
Oceniamy następnie precyzję wykonania oznaczenia. W tym celu obliczamy: wariancję w ocenianej serii pomiarowej:

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

$$= \frac{1}{29}\Big[(0{,}9969 - 0{,}9982)^2 + (0{,}9971 - 0{,}9982)^2 + (0{,}9973 - 0{,}9982)^2 + (0{,}9975 - 0{,}9982)^2 +$$

$$+ (0{,}9978 - 0{,}9982)^2 + 3\cdot(0{,}9980 - 0{,}9982)^2 + 3\cdot(0{,}9981 - 0{,}9982)^2 +$$

$$+ 4\cdot(0{,}9982 - 0{,}9982)^2 + 2\cdot(0{,}9983 - 0{,}9982)^2 + 2\cdot(0{,}9984 - 0{,}9982)^2 +$$

$$+ (0{,}9986 - 0{,}9982)^2 + 3\cdot(0{,}9987 - 0{,}9982)^2 + (0{,}9988 - 0{,}9982)^2 + (0{,}9989 - 0{,}9982)^2 +$$

$$+ (0{,}9990 - 0{,}9982)^2 + (0{,}9991 - 0{,}9982)^2 + (0{,}9992 - 0{,}9982)^2 + (0{,}9994 - 0{,}9982)^2\Big]$$

$$= 3{,}67\cdot 10^{-7}$$

Odchylenie standardowe:

$$s = \sqrt{V} = \sqrt{3{,}67\cdot 10^{-7}} = 6{,}06\cdot 10^{-4}$$

Względne odchylenie standardowe:

$$RSD = \frac{s}{\overline{x}}\cdot 100\,\% = \frac{6{,}06\cdot 10^{-4}}{0.9983}\cdot 100\,\% = 6{,}07\cdot 10^{-2}\%$$

Przedział ufności:

$$\overline{x} - \frac{t}{\sqrt{n}}\cdot s < \mu < \overline{x} + \frac{t}{\sqrt{n}}\cdot s$$

dla poziomu ufności P = 95% (α = 0,05) i k = n-1 = 29

$$t = 2{,}04 \qquad \frac{t}{\sqrt{n}} = \frac{2{,}04}{\sqrt{30}} = 0{,}37$$

$$0{,}9982 - 0{,}37 \cdot 6{,}08 \cdot 10^{-4} < \mu < 0{,}9982 + 0{,}37 \cdot 6{,}08 \cdot 10^{-4}$$

$$0{,}9980 < \mu < 0{,}9984$$

Wartość prawdziwa stężenia HCl zawiera się z prawdopodobieństwem 95% w przedziale 0,9982 ± 0,0002

Wyniki

$\bar{x}$	D	M	V	RSD	Przedział ufności
0,9982	0,9982	0,9982	$3{,}7 \cdot 10^{-4}$	$6{,}09 \cdot 10^{-2}$%	$0{,}9980 < \mu < 0{,}9984$ (dla α = 0,05)

Na podstawie uzyskanych powyżej wyników obliczeń przyjmujemy, że c_{HCl} = 0,9982.

Nastawianie miana $AgNO_3$

Ocena statystyczna serii pomiarowej:

Numer pomiaru	Stężenie$AgNO_3$ [mol/dm^3]	Numer pomiaru	Stężenie$AgNO_3$ [mol/dm^3]	Numer pomiaru	Stężenie$AgNO_3$ [mol/dm^3]
1		11		21	
2		12		22	
3		13		23	
4		14		24	
5		15		25	
6		16		26	
7		17		27	
8		18		28	
9		19		29	
10		20		30	

- *Wartość średniej arytmetycznej wynosi:* $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i =$ …………
- *Wartość mediany wynosi:* $M =$ ....
- *Wartość dominanty wynosi:* $D =$ ....
- *Wariancja w ocenianej serii pomiarowej wynosi:* $V = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} =$ ....
- *Odchylenie standardowe:* $s = \sqrt{V} =$ ....
- *Względne odchylenie standardowe:* $RSD = \frac{s}{\bar{x}} =$ ....

dla poziomu ufności P = 95% (α = 0,05) i k = n-1 = …….. $t =$ …… $\frac{t}{\sqrt{n}} =$ ……

- *Przedział ufności* ($\bar{x} - \frac{t}{\sqrt{n}} \cdot s < \mu < \bar{x} + \frac{t}{\sqrt{n}} \cdot s$) .... < μ < ....

Wnioski:

Wartość prawdziwa stężenia $AgNO_3$ zawiera się z prawdopodobieństwem 95% w przedziale: .... ± ....

Nastawianie miana titranta

Ocena statystyczna serii pomiarowej:

Numer pomiaru	Stężenie titranta [mol/dm^3]	Numer pomiaru	Stężenie titranta [mol/dm^3]	Numer pomiaru	Stężenie titranta[mol/dm^3]
1		11		21	
2		12		22	
3		13		23	
4		14		24	
5		15		25	
6		16		26	
7		17		27	
8		18		28	
9		19		29	
10		20		30	

- *Wartość średniej arytmetycznej wynosi:* $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i =$ ....
- *Wartość mediany wynosi:* $M =$ ....
- *Wartość dominanty wynosi:* $D =$ ....
- *Wariancja w ocenianej serii pomiarowej wynosi:* $V = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} =$ ....
- *Odchylenie standardowe:* $s = \sqrt{V} =$ ....
- *Względne odchylenie standardowe:* $RSD = \frac{s}{\bar{x}} =$ ....

dla poziomu ufności P = 95% (α = 0,05) i k = n-1 = …….. $t =$ …… $\frac{t}{\sqrt{n}} =$ ……

- *Przedział ufności* $(\bar{x} - \frac{t}{\sqrt{n}} \cdot s < \mu < \bar{x} + \frac{t}{\sqrt{n}} \cdot s)$ .... < μ < ....

Wnioski:

Wartość prawdziwa stężenia titranta zawiera się z prawdopodobieństwem 95% w przedziale: .... ± ....