

Eliminacja odstających danych.

Uwaga.

Odstające dane należy odrzucać z bardzo dużą ostrożnością.

Kryterium Grafa.

Obliczamy $\bar{x}, \hat{s}$ dla wszystkich danych ($n > 25$).

Odrzucamy dane gdy leżą poza przedziałem

$$\langle \bar{x} - 4\hat{s}; \bar{x} + 4\hat{s} \rangle$$

Kryterium Grubbs'a.

(zakładamy normalność danych).

Dane porządkujemy.

x_n – wynik największy,

x_1 – wynik najmniejszy,

Obliczamy

$$u_1 = \frac{x_n - \bar{x}}{\hat{s}}$$

$$u_2 = \frac{\bar{x} - x_1}{\hat{s}}$$

W tablicy odczytujemy k_1, k_2 :
Tablica dla $p=0,95$.

n	5	10	15	20	25
k_1	2,08	2,441	2,617	2,732	2,815
k_2	1,869	2,294	2,493	2,623	2,717

Jeśli $u_1 > k_1$ to x_n odrzucamy
Jeśli $u_2 > k_2$ to x_1 odrzucamy
dla przyjętego poziomu p .

Uwaga.

Dla $n > 25$ k_I możemy wyznaczyć za pomocą kwantyla u_α rozkładu $N(0;1)$:

$$k_I = u_{1-\frac{1-p}{n}} \sqrt{\frac{n-1}{n}}$$

Zadanie.

Wyznacz k_I dla $p = 0,95$ i $n=25$, $n=50$.

Uwaga.

Dla $n > 5$ i $p = 0,95$ k_2 możemy wyznaczyć za pomocą aproksymacji:

$$k_2 = \begin{cases} 1,31 + 0,435 \ln(n - 2,7) & \text{dla } 5 \leq n < 35 \\ 1,962 + 0,281 \ln(n - 15) & \text{dla } 35 \leq n \leq 500 \end{cases}$$

Zadanie.

Wyznacz k_2 dla $p = 0,95$ i $n=25$, $n=50$.

Kryterium Q-Dixona.

(zakładamy normalność danych).

Dane porządkujemy.

x_w – wynik wątpliwy,

x_s – wynik sąsiadujący z wątpliwym,

R – rozstęp. $R = x_{max} - x_{min}$

Obliczamy

$$Q = \frac{|x_w - x_s|}{R}$$

W tablicy odczytujemy Q_{kryt} :

Tablica dla $p=0,9$, $p=0,95$, $p=0,99$.

p n	3	4	5	6	7	8	9	10
0,9	0,89	0,68	0,56	0,48	0,43	0,4	0,370	0,349
0,95	0,94	0,77	0,64	0,56	0,51	0,48	0,437	0,412
0,99	0,99	0,89	0,76	0,7	0,64	0,58	0,555	0,527

Jeśli $Q > Q_{\text{kryt}}$ to x_w odrzucamy dla przyjętego poziomu p .

Kryterium Chauventa. I wariant

(zakładamy normalność danych).

Raczej nie stosować gdy jest więcej niż jeden wynik wątpliwy.

x_w – wynik wątpliwy,

Wyznaczamy $\bar{x}, \hat{s}$ dla wszystkich danych

Obliczamy

$$U = \frac{|x_w - \bar{x}|}{\hat{s}}$$

Wyznaczamy $p = 2 - 2\Phi(U)$.

Jeśli $np < 0,5$ to x_w odrzucamy.

Kryterium Chauventa. II wariant (zakładamy normalność danych).

x_w – wynik wątpliwy,

Wyznaczamy $\bar{x}, \hat{s}$ dla wszystkich danych

Obliczamy

$$U = \frac{|x_w - \bar{x}|}{\hat{s}}$$

W tablicy odczytujemy k :

n	4	5	6	10	15	25	50	100	300
k	1,54	1,65	1,73	1,96	2,13	2,33	2,57	2,81	3,14

Jeśli $U > k$ to x_w odrzucamy.

Kryterium Chauventa. III wariant (zakładamy normalność danych).

x_w – wynik wątpliwy,

Wyznaczamy $\bar{x}, \hat{s}$ dla wszystkich danych

Dla $n=10$, to jeśli $|x_w - \bar{x}| > 2\hat{s}$ to x_w odrzucamy.

Dla $n=40$, to jeśli $|x_w - \bar{x}| > 2,5\hat{s}$ to x_w odrzucamy.

Dla $n>250$, to jeśli $|x_w - \bar{x}| > 3\hat{s}$ to x_w odrzucamy.