Teste pentru depistarea erorilor grosiere

- Exista doua categorii de teste, functie de volumul selectiei:
  - teste pentru selectii de volum ridicat (Chauvenet, 3σ);
  - teste pentru selectii de volum scazut (Grubbs-Smirnov).

Testul Chauvenet consta în verificarea încadrarii datelor experimentale într-un interval de încredere $[m_v - \Delta_l; m_v + \Delta_l]$ în care sunt admise erorile aleatorii (s-a presupus ca erorile sistematice au fost eliminate). Datele afectate de erori mai mari limitelor $\pm \Delta_l$ se considera afectate de erori grosiere si se elimina din sirul de determinari.

Testul presupune urmatoarele etape(pasi):

1. ordonarea datelor $V_1, V_2, ..., V_n$ în sens crescator (seria variationala), având ca extreme V_{min} si V_{max};

2. calculul mediei m_v si a estimatiei erorii standard σ^;

3. calculul intervalului de valori în care trebuie sa se încadreze datele:

$$V_l \in [m_V - z_l \cdot \hat{\sigma}; m_V + z_l \cdot \hat{\sigma}]$$

Teste pentru depistarea erorilor grosiere

în care z_l se deduce din relatia (considerata la limita):

$$1-2\Phi(z_l)\geq\frac{1}{2n}$$

4. toate datele care sunt în afara intervalului dat de relatia de la pasul 3 se elimina fiind considerate ca afectate de erori grosiere;

5. se reia testul cu noile date ramase de la pasul 2, pâna când nici un rezultat experimental nu iese în afara intervalului dat de relatia de la pasul 3.

❖ ***Testul 3σ*** consta în verificarea încadrarii datelor în intervalul:

$$V_i\in[m_V-3\hat{\sigma};m_V+3\hat{\sigma}]$$

! Etapele sunt similare testului Chauvenet, cu observatia ca, în acest caz, se subintelege ca nivelul de încredere adoptat este $\eta = 0{,}9973$.

Teste pentru depistarea erorilor grosiere

❖ ***Testul Grubbs - Smirnov*** consta în verificarea încadrarii valorilor extreme ale datelor experimentale si presupune urmatoarele etape:

1. ordonarea datelor V_1, V_2, ..., V_n în sens crescator (seria variationala având ca extreme $V_{\min}$ si $V_{\max}$);

2. calculul mediei m_v si a estimatiei σ^;

3. daca valoarea suspectata de a fi afectata de erori grosiere este $V_{\max}$ ($V_{\min}$) atunci se calculeaza:

$$v = \frac{V_{\max} - m_V}{\hat{\sigma}}; \quad \text{sau} \quad v = \frac{m_V - V_{\min}}{\hat{\sigma}}$$

4. se compara v cu valoarea $v_{n,\alpha}$ din tabelul Grubbs-Smirnov, în care n este volumul selectiei de date, iar α reprezinta riscul ales (la masurarile uzuale riscul α se ia, de obicei, 0,05 sau 0,01, în timp ce pentru masurarile de foarte mare precizie se adopta $\alpha = 0{,}005$ sau $\alpha = 0{,}001$); daca $v \geq v_{n,\alpha}$ atunci $V_{\max}$ ($V_{\min}$) este afectata de o eroare grosiera si se elimina din sirul de determinari, în caz contrar se mentine în sirul de rezultate;

5. se reia testul cu n-1 date experimentale de la pasul 2 si se continua pâna când $v < v_{n\text{-}i,\alpha}$

Erori aleatorii; repartitii empirice (exemple)

Exemplul 1.

Pentru determinarea valorii curentului într-un circuit s-au efectuat 22 măsurări, de egală precizie, obţinându-se următoarele rezultate:

Nr.det.	1	2	3	4	5	6	7	8	9	10	11
I_i[mA]	205	197	211	209	202	205	204	206	195	213	199
Nr.det.	12	13	14	15	16	17	18	19	20	21	22
I_i[mA]	227	201	205	204	207	203	208	203	207	215	206

a) Să se traseze histograma şi poligonul de frecvenţe;

b) Să se calculeze m_v, Me, Mo, μ, σ^, θ^, E, Δ_{lim}, p(Δ);

c) Să se determine intervalul în care este situată valoarea reală a curentului adoptând acelaşi nivel de încredere η = 0,9642 pentru determinările individuale şi pentru medie.

Rezolvare

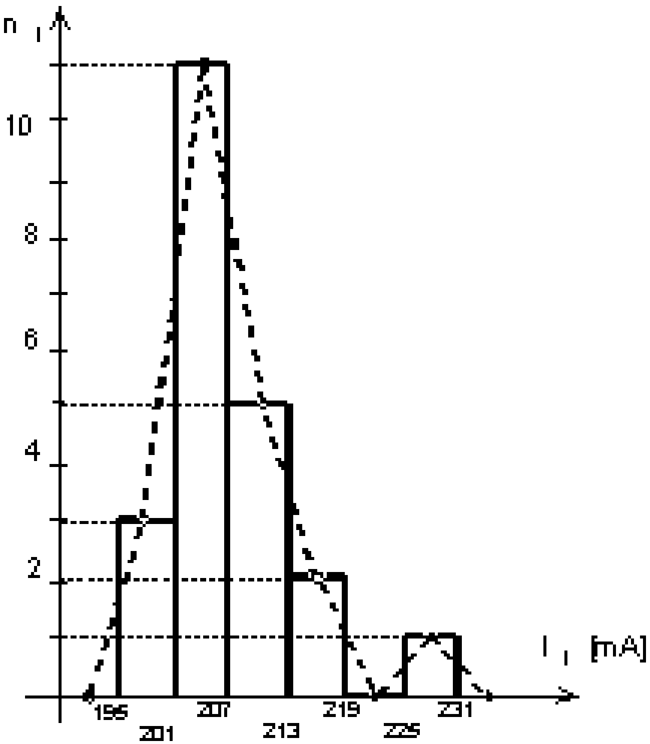
Se calculează intervalul de grupare Δ cu relaţia:

$$\Delta = \frac{V_{\max} - V_{\min}}{1 + 3{,}22 \lg n} = \frac{227 - 195}{1 + 3{,}22 \lg 22} = 6{,}02 \cong 6 \text{ mA}$$

Erori aleatorii; repartitii empirice (exemple)

Histograma este trasată în fig….. Prin unirea mijloacelor de grupare se obţine poligonul de frecvenţe (curba punctată din fig….).

Se observă că cele mai multe date se grupează în intervalul [201 mA; 207 mA], unde ne aşteptăm să se găsească valoarea medie, iar alura poligonului de frecvenţe corespunde unei repartiţii empirice normale, cu excepţia valorii de 227 mA, care este posibil să fie afectată de o eroare grosieră. Cum numărul de date este mic acestea nu trebuie să depăşească pragul de siguranţă $\Delta_{\lim} = 3\ \sigma$^. In consecinţă, se calculează:

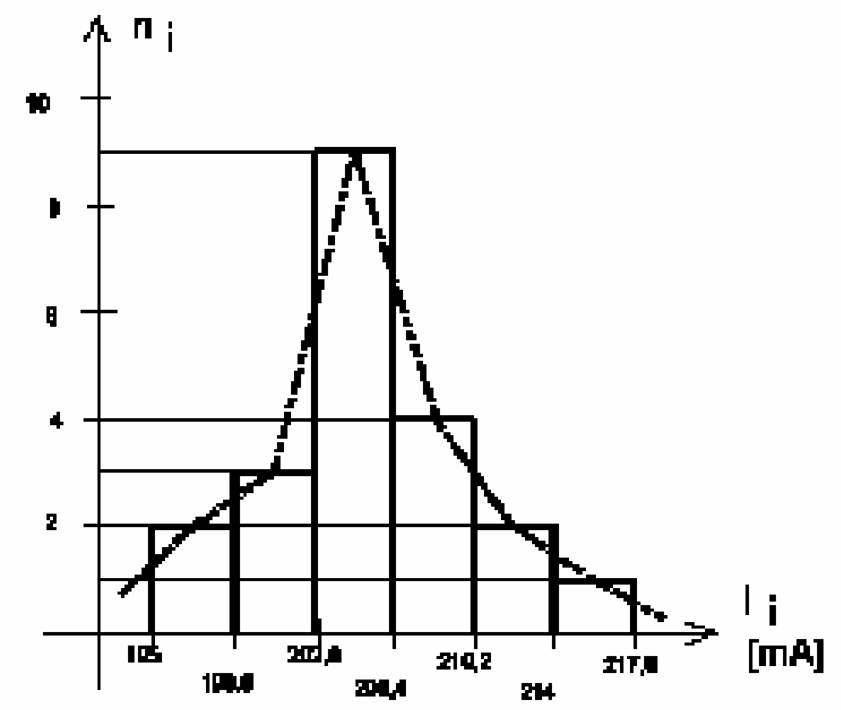


$$m_V = \frac{1}{n}\sum_{i=1}^{n} V_i = \frac{1}{22}\sum_{i=1}^{22} I_i = 206 \text{ mA}$$

$$\hat{\sigma} = \sqrt{\frac{1}{\text{n - 1}}\sum_{i=1}^{n} \Delta V_i^2} = \sqrt{\frac{1}{22-1}\sum_{i=1}^{22} (m_V - I_i)^2} = 6{,}66 \text{ mA}$$

Erori aleatorii; repartitii empirice (exemple)

Valorile individuale I_i trebuie să fie cuprinse în intervalul [m_v - 3 σ^; m_v + 3 σ^] adică [186,02 mA; 225,98 mA]; într-adevăr valoarea I_{12} = 227 mA se află în afara acestui interval, deci este afectată de erori grosiere, în consecinţă o eliminăm din şirul de rezultate, calculele în continuare făcându-se pentru 21 de valori.



Dacă se trasează histograma, respectiv poligonul de frecvenţe, cu cele 21 de valori, rezultă Δ = 3,8 mA, astfel că se obţin reprezentările din fig.… mult mai semnificative faţă de cele din fig.. anterioara.

In continuare se calculează:

Erori aleatorii; repartitii empirice (exemple)

$$m_V = \frac{1}{21}\sum_{i=1}^{21} I_i = 205 \text{ mA}$$

$$\hat{\sigma} = \sqrt{\frac{1}{21-1}\sum_{i=1}^{21}(m_V - I_i)^2} = 4{,}85 \text{ mA}$$

Se observă că intervalul [m_v - 3 σ^; m_v + 3 σ^] este [190,45 mA; 219,55 mA] în care sunt incluse toate cele 21 de date. In consecinţă se pot pune în evidenţă ceilalţi indicatori ceruţi de problemă.

b) Valoarea medie şi estimaţia erorii standard au fost calculate la punctul a). In continuare se construieşte seria variaţională (şirul de date ordonat în sens crescător) ataşată rezultatelor

k	1	2	3	4	5	6	7	8	9	10	11
I_k[mA]	195	197	199	201	202	203	203	204	204	205	205
k	12	13	14	15	16	17	18	19	20	21	
I_k[mA]	205	206	206	207	207	208	209	211	213	215	

Erori aleatorii; repartitii empirice (exemple)

din care rezultă

$$M_e = 205 \text{ mA}$$

$$M_o = 205 \text{ mA}$$

valori care verifică relaţia $Mo = 3 \cdot Me - 2 \cdot mv \cdot$

$$\mu = \sqrt{\frac{n-1}{n}}\hat{\sigma} = \sqrt{\frac{21-1}{21}} \cdot 4{,}85 = 4{,}73 \text{ mA}$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left|V_i - m_V\right| = \frac{1}{21}\sum_{i=1}^{21}\left|I_i - 205\right| = 3{,}52 \text{ mA}$$

$$E = \frac{2}{3} \cdot \hat{\sigma} = 3{,}23 \text{ mA}$$

$$\Delta_{\lim} = 3 \cdot \hat{\sigma} = 14{,}55 \text{ mA}$$

Erori aleatorii; repartitii empirice (exemple)

$$p(\Delta)=\frac{1}{\hat{\sigma}\sqrt{2\pi}}\mathrm{e}^{-\frac{\Delta^2}{2\hat{\sigma}^2}}=\frac{1}{12{,}15}\mathrm{e}^{-\frac{\Delta^2}{47{,}045}}\left[\frac{1}{\mathrm{mA}}\right]$$

$$p(I)=\frac{1}{12{,}15}\mathrm{e}^{-\frac{(I-205)^2}{47{,}045}}\left[\frac{1}{\mathrm{mA}}\right]$$

c) Se ştie că legătura între nivel şi interval de încredere este

$$\eta=2\cdot\Phi\left(\frac{\Delta_\eta}{\hat{\sigma}}\right)$$

unde

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$ este funcţia lui Laplace, care este tabelată.

Erori aleatorii; repartitii empirice (exemple)

Pentru η = 0,9642 rezultă Φ(z) = 0,4821 iar din tabele (v. ANEXA C din carte!) se găseşte $z = 2{,}1$. Deci

$$\Delta_\eta = z \cdot \hat{\sigma} = 2{,}1 \cdot 4{,}85 = 10{,}185 \text{ mA}$$

Aşadar

$$I_i \in [m_V - \Delta_\eta ; m_V + \Delta_\eta] = [194{,}815 \text{ mA; } 215{,}185 \text{ mA}] \quad \text{cu} \quad \eta = 0{,}9642$$

iar valoarea reală I a curentului din circuitul respectiv aparţine intervalului

$$I \in [I_i - 10{,}185 \text{ mA}; I_i + 10{,}185 \text{ mA}] \quad \text{cu} \quad \eta = 0{,}9642 \quad \text{pentru orice} \quad i = 1, 2, ..., 21.$$

Dacă se adoptă acelaşi nivel de încredere pentru valoarea medie, atunci

$$\Delta_\eta^{'} = \frac{\Delta_\eta}{\sqrt{n}} = \frac{10{,}185}{\sqrt{21}} = 2{,}22 \text{ mA}$$

deci valoarea reală a curentului I va aparţine intervalului:

$$I \in [202{,}78 \text{ mA}; 207{,}22 \text{ mA}] \quad \text{cu} \quad \eta = 0{,}9642$$

Erori aleatorii; repartitii empirice (exemple)

Exemplul 2.

Pentru aflarea valorii căderii de tensiune de la bornele unui consumator s-au efectuat 15 măsurări obţinându-se următoarele rezultate

Nr.det.	1	2	3	4	5	6	7	8
U_i[V]	205	204	207	202	208	204	203	209
Nr.det.	9	10	11	12	13	14	15	
U_i[V]	199	203	206	201	206	211	207	

Să se precizeze în ce interval se situează valoarea reală a tensiunii dacă se adoptă un nivel de încredere η = 0,95; dar dacă η = 0,9642 ca la exemplul anterior pct.c).

Rezolvare

Întrucât numărul de date este n < 20 se foloseşte repartiţia Student. Se calculează:

$$m_V = \frac{1}{n}\sum_{i=1}^{n} V_i = \frac{1}{15}\sum_{i=1}^{15} U_i = 205\ \mathrm{V}$$

Erori aleatorii; repartitii empirice (exemple)

$$\hat{s} = \frac{\hat{\sigma}}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\Delta V_i^2} = \sqrt{\frac{1}{15(15-1)}\sum_{i=1}^{n}(205-U_i)^2} = 0{,}82\ \text{V}$$

Din tabelele prezentate în ANEXA D (carte) rezultă pentru n = 15 şi η = 0,95 valoarea t = 2,131; deci valoarea reală U a tensiunii este cuprinsă în intervalul

$$U \in [m_V - t\cdot\hat{s};\ m_V - t\cdot\hat{s}] = [203{,}25\ \text{V};\ 206{,}75\ \text{V}] \quad \text{cu} \quad \eta = 0{,}95$$

Dacă η = 0,9642, întrucât în ANEXA D nu este precizată valoarea lui t pentru acest nivel de încredere, se procedează la interpolare liniară

- pentru η' = 0,95 t' = 2,131
- pentru η'' = 0,98 t'' = 2,602

atunci

$$t = t' + \frac{t''-t'}{\eta''-\eta'}(\eta-\eta') = 2{,}131 + \frac{2{,}602-2{,}131}{0{,}98-0{,}95}(0{,}9642-0{,}95) = 2{,}354$$

astfel că $U \in [203{,}07\ \text{V};\ 205{,}93\ \text{V}] \quad cu \quad \eta = 0{,}9642$

Erori aleatorii; medii ponderate (exemple)

De remarcat faptul că, dacă s-ar adopta repartiţia normală, ar rezulta, din ANEXA C (carte), $z = 1{,}96$ pentru $\eta = 0{,}95$ şi $z = 2{,}1$ pentru $\eta = 0{,}9642$, astfel că s-ar obţine intervale mai mici de încadrare a valorii reale (acest aspect este evident prin simpla comparare a valorilor z cu cele ale valorilor t pentru cele două nivele de încredere).

Exemplul 3.

Asupra rezistenţei de izolaţie s-au efectuat simultan, din 3 locuri diferite, seturi de măsurări cu aparatură de aceeaşi precizie. S-au obţinut următoarele rezultate:

R_i [MΩ]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Poz.1	50	48,5	49	48,7	51	50,3	50,7	51,2	50	50,4	49,7	49	48,8	49,9	50,2
Poz.2	51	49	49,5	50,5	50,2	50,3	51	51,3	51,5	50	50,1	49,8	49,7	49,2	50,1
Poz.3	49,5	49,9	48,8	50,2	50,4	50	50,3	51	49	49,7	48,7	50,3	50,8	50,2	50

Să se precizeze care este valoarea cea mai probabilă a rezistenţei de izolaţie măsurate şi în ce interval se situează valoarea reală a acesteia.

Erori aleatorii; medii ponderate (exemple)

Rezolvare

Cu rezultatele obţinute de la fiecare punct de măsurare se pot calcula valorile medii şi abaterile medii pătratice parţiale:

$$m_{V_1} = \frac{1}{15}\sum_{i=1}^{15} R_{1_i} = 49{,}83\ \mathrm{M\Omega}\ ; \qquad \hat{\sigma}_1 = \sqrt{\frac{1}{15-1}\sum_{i=1}^{15}(49{,}83 - R_{1_i})^2} = 0{,}896\ \mathrm{M\Omega}$$

$$m_{V_2} = \frac{1}{15}\sum_{i=1}^{15} R_{2_i} = 50{,}21\ \mathrm{M\Omega}\ ; \qquad \hat{\sigma}_2 = \sqrt{\frac{1}{15-1}\sum_{i=1}^{15}(50{,}21 - R_{2_i})^2} = 0{,}741\ \mathrm{M\Omega}$$

$$m_{V_3} = \frac{1}{15}\sum_{i=1}^{15} R_{3_i} = 49{,}92\ \mathrm{M\Omega}\ ; \qquad \hat{\sigma}_3 = \sqrt{\frac{1}{15-1}\sum_{i=1}^{15}(49{,}92 - R_{3_i})^2} = 0{,}763\ \mathrm{M\Omega}$$

Valoarea cea mai probabilă se obţine ca medie ponderată a mediilor parţiale. Cum $p_i = c/\sigma\hat{}_i{}^2$, c - constantă arbitrar aleasă, se alege, de exemplu, c = 10, rezultând $p_1 = 12{,}45$; $p_2 = 18{,}21$; $p_3 = 17{,}18$. Aşadar:

Erori aleatorii; medii ponderate (exemple)

$$m_p = \frac{\sum_{i=1}^{3} p_i m_{V_i}}{\sum_{i=1}^{3} p_i} = \frac{12{,}45 \cdot 49{,}83 + 18{,}21 \cdot 50{,}21 + 17{,}18 \cdot 49{,}92}{12{,}45 + 18{,}21 + 17{,}18} = 50\ \mathrm{M\Omega}$$

$$\hat{s}_p = \sqrt{\frac{1}{(n-1)\sum_{i=1}^{n} p_i} \sum_{i=1}^{n} p_i (m_p - m_{V_i})^2} = 0{,}115\ \mathrm{M\Omega}$$

Valoarea reală a rezistenţei de izolaţie R se află situată în intervalul

$$R \in [m_p - 3 \cdot \hat{s}_p;\ m_p + 3 \cdot \hat{s}_p] = [49{,}655\ \mathrm{M\Omega};\ 50{,}345\ \mathrm{M\Omega}]$$

Erori aleatorii; medii ponderate (exemple)

Exemplul 4.

Prin măsurarea repetată, la momente de timp diferite, a puterii la bornele unui consumator invariant în timp, cu wattmetre de aceeaşi clasă de precizie, s-au obţinut următoarele valori medii parţiale:

- 220 W provenită din 25 determinări
- 218 W provenită din 20 determinări
- 225 W provenită din 15 determinări
- 222 W provenită din 30 determinări
- 221 W provenită din 20 determinări

Să se găsească valoarea cea mai probabilă a puterii la bornele consumatorului.

Rezolvare

Întrucât mediile parţiale provin din măsurări de egală precizie, dar dintr-un număr diferit de determinări, valoarea cea mai probabilă este media ponderată a mediilor parţiale, la care ponderările vor fi egale cu numărul de determinări:

$$m_p = \frac{\sum_{i=1}^{k} p_i m_{V_i}}{\sum_{i=1}^{k} p_i} = \frac{25 \cdot 220 + 20 \cdot 218 + 15 \cdot 225 + 30 \cdot 222 + 20 \cdot 221}{25 + 20 + 15 + 30 + 20} = 221 \text{ W}$$

Teste pentru depistarea datelor afectate de erori grosiere (exemplificare)

Prin măsurarea repetată a tensiunii unei surse s-au obţinut valorile: 304,5; 305,2; 303,3; 304,9; 304,8; 305,0; 304,6; 305,1; 304,7; 304,9 [mV]. Să se verifice dacă valorile obţinute prin măsurări sunt sau nu afectate de erori grosiere.

Rezolvare

Şirul de date fiind de volum mic se aplică testul Grubbs - Smirnov. După ordonarea în sens crescător a datelor, adică obţinerea seriei variaţionale
303,3; 304,5; 304,6; 304,7; 304,8; 304,9; 304,9; 305,0; 305,1; 305,2 [mV]
se constată vizual că valoarea 303,3 mV diferă semnificativ de restul determinărilor.

Aplicarea testului Grubbs - Smirnov presupune etapele:

1. Se calculează

$$m_v = \frac{1}{n}\sum_{i=1}^{n} V_i = \frac{1}{10}\sum_{i=1}^{10} U_i = 304{,}7\ mV$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\Delta V_i^2} = \sqrt{\frac{1}{9}\sum_{i=1}^{10}(U_i - m_v)^2} = 0{,}54\ mV$$

Teste pentru depistarea datelor afectate de erori grosiere (exemplificare)

2. Având în vedere că valoarea suspectată este cea mai mică din seria variaţională, se calculează

$$v = \frac{m_v - V_{\min}}{\hat{\sigma}} = \frac{304,7 - 303,3}{0,54} = 2,59$$

3. Din tabelul prezentat în ANEXA E, adoptând un risc α = 0,05 (valoare medie pentru α, corespunzătoare unui nivel de încredere η = 0,95), rezultă $v_{n,\alpha} = v_{10;0,05} = 2,176$. Cum

$v > v_{n,\alpha}$ avem toate motivele (fiind un număr mic de determinări) să excludem valoarea 303,3 mV din şirul de determinări.

4. Testul se continuă cu recalcularea mediei şi estimaţiei erorii medii pătratice folosind numai datele rămase în şir:

$$m_v = \frac{1}{9}\sum_{i=1}^{9} U_i = 304,86 \ mV \qquad \hat{\sigma} = \sqrt{\frac{1}{8}\sum_{i=1}^{9}(U_i - m_v)^2} = 0,23 \, mV$$

Teste pentru depistarea datelor afectate de erori grosiere (exemplificare)

iar
$$v_M = \frac{V_{\max} - m_v}{\hat{\sigma}} = \frac{305{,}2 - 304{,}86}{0{,}23} = 1{,}478$$

$$v_m = \frac{m_v - V_{\min}}{\hat{\sigma}} = \frac{304{,}86 - 304{,}5}{0{,}23} = 1{,}565$$

valori inferioare lui $v_{n,\alpha} = v_{9;0,05} = 2{,}11$. Aşadar testul se opreşte aici, cele 9 valori rămase fiind afectate numai de erori aleatorii.

Nota 1. Deşi testul Chauvenet dă rezultate concludente pentru date cu volum de selecţie mare, totuşi, şi în acest caz, conduce la aceeaşi concluzie. Intr-adevăr, pentru n = 10, rezultă

$$\phi(z_l) = \frac{1 - \frac{1}{2n}}{2} = 0{,}475$$

iar din ANEXA C se găseşte $z_l = 1{,}96$, astfel că intervalul de încadrare al datelor trebuie să fie

$$U_i \in [\, m_v - z_l \bullet \hat{\sigma}\,;\, m_v + z_l \bullet \hat{\sigma}\,] = [\, 303{,}64\ mV\,;\, 305{,}76\ mV\,]$$

deci valoarea 303,3 mV trebuie eliminată din şirul de determinări.

Teste pentru depistarea datelor afectate de erori grosiere (exemplificare)

Cu cele 9 valori rămase se obţine

$$\phi(z_l) = \frac{1 - \frac{1}{2n}}{2} = 0,4722$$

care, din ANEXA C, conduce la valoarea z_l = 1,914, astfel că

$$U_i \in [304,86 - 1,914 \bullet 0,23; 304,86 + 1,914 \bullet 0,23] = [304,42\ mV\ ; 305,3\ mV]$$

deci toate datele şirului de determinări rămase verifică ultima relaţie.

Nota 2. Nu acelaşi lucru (ca la nota 1) se poate afirma despre testul 3σ^ care, fiind mult acoperitor, conduce la un interval de încadrare

$$U_i \in [\ m_v - 3\hat{\sigma}\ ; m_v + 3\hat{\sigma}\ ] = [\ 303,08\ mV\ ; 306,32\ mV\ ]$$

în care sunt cuprinse toate cele 10 determinări (totuşi, având în vedere apropierea valorii 303,3 mV mult mai evidentă de limita inferioară de încadrare, comparativ cu apropierea valorii maxime din şir în raport cu limita superioară de încadrare, nu se poate lua decizia păstrării valorii 303,3 mV numai pe baza testului 3σ^).