

Eugenij T. VOLODARSKY<sup>1</sup>, Zygmunt WARSZA<sup>2</sup>, Larisa A. KOSHEVAYA<sup>3</sup>

<sup>1</sup>NARODOWY UNIWERSYTET TECHNICZNY UKRAINY - POLITECHNIKA KIJOWSKA

<sup>2</sup>PRZEMYSŁOWY INSTYTUT AUTOMATYKI I POMIARÓW (PIAP), Warszawa

<sup>3</sup>NARODOWY UNIWERSYTET LOTNICTWA, Kijów, Ukraina

## Odporna ocena dokładności metod pomiarowych

Prof. dr nauk tech. Evgeniy T. VOLODARSKY

Narodowy Techniczny Uniwersytet Ukrainy - Politechnika Kijowska „KPI” Kiev, Katedra Automatyzacji badań eksperymentalnych; Przewodniczący Podkomisji „Metrologia” w Państwowej Agencji Akredytacyjnej Ukrainy. Senior member of IEEE. Zainteresowania naukowe: systemy oprogramowania pomiarowego do monitorowania i testowania. Autor ponad 200 publikacji.



e-mail: vet-1@ukr.net

Doc. dr inż. Zygmunt Lech WARSZA

Studia 1959, doktorat 1967 i praca 1960-70 na wydz. Elektrycznym Politechniki Warszawskiej oraz w Instytucie Elektrotechniki 1958-63. Docent od 1970. Zorganizował i prowadził: Wydział Transportu Pol. Świętokrzyskiej 1970-76, Ośrodek Aparatury Pomiarowej IMGW 1978-81, Zakład Techniki Pomiarowej Instytutu Chemii Przemysłowej 1983-91. Obecnie: główny specjalista w Przemysłowym Instytucie Automatyki i Pomiarów PIAP Warszawa. Autor około 130 publikacji, 2 monografie, 11 patentów.



e-mail: zlw@op.pl

Prof. dr nauk tech. Larisa A. KOSHEVAJA

Zatrudniona w Narodowym Uniwersytecie Lotnictwa, Kiev, Ukraina; w Katedrze Biocybernetyki i Medycyny Lotniczej. Jej zainteresowania naukowe obejmują: metrologiczne wspomaganie pomiarów medycznych i biologicznych oraz statystyczne przetwarzanie danych. Jest Autorką ponad 70 publikacji.



e-mail: arnis@ukrpost.net

### Streszczenie

Przedstawiono dwie metody oceny niepewności próbki danych doświadczalnych o niewielkiej liczebności, odporne na zawarte w niej tzw. odstające obserwacje pomiarowe, tj. o wartościach znacznie różniących się od pozostałych. Umożliwiają one wyznaczenie w sposób wiarygodny statystycznych parametrów wyniku pomiaru na podstawie całości danych eksperymentalnych. Rozważania ilustruje przykład liczbowy wykorzystujący dane z porównań międzylaboratoryjnych. Porównano otrzymane w nim wyniki obliczone metodą o przeskalowanym odchyleniu medianowym  $MAD_S$  i metodą iteracyjną dwukryterialną. Podano wnioski i bibliografię.

**Słowa kluczowe:** dane odstające, niepewność pomiaru, odchylenie standardowe, mediana, średnia odporna, przedział międzykwartyłowy.

### Robust evaluation of the accuracy of measurement methods

#### Abstract

Presented are two methods of assessing the value and uncertainty of the measurand from the sample of experimental data which are resist to contained therein small number of outliers, i.e. values of measurement data significantly different from the others. This allows to set a credible statistical parameters of the measurement result on the basis of all experimental data. The considerations are illustrated by the numerical example of inter-laboratory measurement data. Compared are results obtained by method applied the rescaled median absolute deviation  $MAD_S$  and by the iterative two-criteria method. Given are conclusions and bibliography.

**Keywords:** outliers, uncertainty of measurements, standard deviation, median, robust mean value, inter-quartile mid-range.

### 1. Wprowadzenie

W praktyce pomiarowej zdarza się iż wartości jednej lub kilku obserwacji w próbce istotnie odstają od pozostałych danych. Nawet niewielka ich liczba może znacząco zmienić klasyczne parametry statystyczne próbki o małej liczbie elementów, a w wielu przypadkach nie ma możliwości powtórzenia, lub uzupełnienia pomiarów. Wartość wyniku pomiarów i niepewność statystyczna  $u_A$ , obliczone w dotychczas stosowany sposób

według zasad międzynarodowego przewodnika GUM, mogą okazać się niewiarygodne, a nawet sprzeczne ze zdrowym rozsądkiem. Aby uniknąć takiej sytuacji, tradycyjnie stosuje się testy statystyczne. Pozwalają one zidentyfikować, a następnie usunąć nietypowe, tj. odstające wartości danych (ang. *outliers*). Postępowanie takie jest skuteczne tylko dla dużych próbek, gdyż dla małych testy te tracą wrażliwość na wartości odstające. Ponadto, gdy próbki o małej liczbie danych, uzyskuje się z badań eksperymentalnych np. o dużym koszcie, wykonywanych metodami niszczącymi nieliczne posiadane obiekty, lub niemożliwych do powtórzenia, należy starać się wykorzystać wszystkie otrzymane dane. Usunięcie dowolnej obserwacji z próbki zmniejsza wiarygodność jej oceny statystycznej. Stosunek odchylenia standardowego  $s(u_A)$  do wartości oczekiwanej niepewności  $u_A$  oszacowanej w eksperymencie dla próbki o  $n$  obserwacjach wynosi około  $1/\sqrt{2(n-1)}$ , [1, 2] (dodatek E.1). Na przykład dla  $n=4$  stosunek ten równa się 42%, a dla  $n=3$  - rośnie do 52%. Usunięcie tylko jednej obserwacji z tak mało-licznej próbki powoduje wzrost odchylenia standardowego  $s(u_A)$  aż o około 24%. Z drugiej strony rozrzut parametru MAD, bardziej odpornego niż  $s = u_A$ , jest około 1,5 razy większy.

Klasyczne metody statystyczne przetwarzania danych opierają się na założeniu modelowania prawdopodobieństwa ich rozrzutu rozkładem normalnym. Jednak dla szeregu przypadków w praktyce liczba danych nie jest wystarczająca do budowy w pełni odpowiadających im modeli parametrycznych. W 60-tych latach ubiegłego wieku statystycy, w oparciu o wyniki szczegółowych badań ustalili, że dane eksperymentalne przetwarzane przy założeniu rozkładu normalnego zawierają zwykle średnio w około 10% (od 1 do 20%) dane odstające od tego rozkładu. Dawniej, ze względu na zmusne obliczenia, opracowywano tylko dane o wysokiej jakości, a teraz, przy stosowaniu komputerów - prawie każde.

Już Poincare wskazywał na nieuzasadnioną, ale panującą powszechnie głęboką wiarę w uniwersalność rozkładu normalnego: matematycy sądzą, że obserwuje się go w eksperymencie, a eksperymenciarze - że matematycy mogą udowodnić teoretycznie, że z centralnego twierdzenia granicznego wynika, iż musi zachodzić rozkład normalny. W nielicznych tylko przypadkach np. gdy wynik pomiaru wyznacza się z dużej liczby zliczeń, normalność rozkładu dla samego mezurandu można traktować jako ściśle prawo fizyki, a przy małej ich liczbie, że spełniony jest rozkład Poissona. Wszystkie założenia teoretyczne opierają się na możliwości prowadzenia eksperymentu (uzyskiwania obserwacji) w tych samych stałych warunkach. W makro-rzeczywistości zwykle występuje przestrzenno-czasowa zmienność obiektu badań, systemu pomiarowego i warunków otoczenia. Dlatego J. Tukey [5] wyznawał nawet pogląd, że normalność rozkładu danych uzyskanych w eksperymencie jest mitem i takiego rozkładu nie ma i nigdy się nie uzyska. Jest to jak widać zagadnienie dla filozofów. Natomiast w praktyce pomiarowej, szczególnie przy małej liczbie danych powodem powstania odstających wartości

niektórych wyników mogą być błędy wynikłe z nieprzestrzegania zasad prowadzenia doświadczenia, niewykryte uszkodzenia przyrządów pomiarowych, błędy i omyłki w przetwarzaniu wyników obserwacji, wpływy zewnętrzne o dużej intensywności i wiele innych.

Przyjęty kształt rozkładu danych doświadczalnych jest tylko proponowanym umownym modelem, któremu one powinny podlegać. Sam ten wybór może być nieidealny i część wartości obserwacji podlega innemu rozkładowi niż przyjęty. Parametryczne podejście przy założeniu, że rozkład jest znany (i powinien być normalny) tak głęboko zakorzeniło się w praktyce statystycznego przetwarzania danych, że rezygnacja z niego zwykle nie wydaje się zasadna.

Celem pracy jest przedstawienie oceny dokładności próbki metodą iteracyjną opartą na statystyce odpornościowej (robust statistics). Dzięki małej wrażliwości na dane odstające (ang. *outliers*) umożliwia ona poprawę wiarygodności wyników przy próbkach również o małej liczności. Podstawy tej statystyki opracowano kilkadziesiąt lat temu [4 - 10]. Rozwija się ją dalej dla różnych zadań przetwarzania danych, w tym do kalibracji pomiarów wieloparametrowych [6 - 10, 12, 14, 15] w wyspecjalizowanych technikach pomiarowych, np. w Polsce w chemometrii na Uniwersytecie Śląskim [11 - 13]. Niektóre metody odpornościowe są już oprogramowane, m. in. w MatLab, ale nie trafiły jeszcze do podstawowych międzynarodowych przepisów metrologicznych, takich jak GUM. Podstawowe informacje o tej statystyce przedstawił polskiemu środowisku metrologicznemu A. Zięba z AGH na sympozjum Podstawowych Problemów Pomiarów w 2007 r. oraz ujął je w książce przekazanej do wydania w PWN [16].

Poniżej omówi się szczegółowo iteracyjną metodę przetwarzania danych należącą do typu o angielskim symbolu IRLS (*iteratively reweighted least squares*). Metodę zilustruje się przykładem jej zastosowania do atestacji dokładności procedur badawczych i kontrolnych dokonywanej poprzez porównanie wyników w badaniach międzylaboratoryjnych.

## 2. Metoda odporna dwukryterialna

Statystyka odpornościowa polega na zastosowaniu metod, które zapewniają mniejszy niż modele klasyczne wpływ danych odstających i innych anomalii na wynik pomiaru. Znajdują one szerokie zastosowanie przy szacowaniu odchylenia standardowego  $\sigma$ . Pod pojęciem odporność rozumie się niewrażliwość na nieprawidłowości i niejednorodności w próbie wywołane przez różne, w ogólnym przypadku nieznanne przyczyny. W szczególności rozpatrzy się metodę, w której zakłada się, że część środkowa rozkładu gęstości prawdopodobieństwa (*probability density function*, PDF) pochodząca od małych odchylenia od estymatora wartości mierzandru nie odbiega zbytnio od przyjętego modelu teoretycznego

Aby zapewnić odporność wybiera się model z outlierami, czyli danymi odstającymi, którego cechą charakterystyczną są bardziej rozciągnięte "ogony" niż dla normalnego rozkładu gęstości prawdopodobieństwa. Skorzystanie z tej prawidłowości pozwala zachować wygodne do analizy założenie o hipotetycznej jednorodności populacji, na którym opierają się wszelkie statystyczne oszacowania i częściowo uwzględnia informację zawartą w odchyleniach odstających. Z danych próbki pomiarowej oblicza się estymator wartości mierzandru, zwykle jako ich wartość średnią  $\mu$ . Wskutek występowania wpływów losowo zmiennych oddziaływań można jedynie ocenić prawdopodobieństwo, z jakim ta wartość może znajdować się w danym przedziale, czyli określić jej niepewność  $u$ .

Otrzymywane w praktyce wyniki zwykle rozkładają się niesymetrycznie względem średniej. Im próbka jest mniejsza, tym asymetria ta może być większa.

Ocena niepewności według GUM opiera się na sumie kwadratów odchylenia wartości obserwacji od estymatora wartości mierzandru i osiąga minimum dla średniej. I to właśnie jest źródłem wrażliwości tej oceny na odstające dane, które znacznie

zwiększają jej wartość. Tylko centralna część rozkładu danych doświadczalnych odpowiada założeniu o normalnym rozkładzie populacji. Dla danych z tej części należy przeprowadzać uśrednianie, tj. stosować metodę najmniejszych kwadratów (MNK). Bardziej odporne na duże odchylenia jest podane przez Laplace'a kryterium modułowe MNM. Dlatego przy tworzeniu metod odpornych warto dokonywać "symbiozy" obu kryteriów. Dla centralnej części wartości danych można stosować metodę najmniejszych kwadratów, a poza granicami tego przedziału, wykorzystać kryterium modułowe w celu zmniejszenia wpływu danych odstających (outlierów), ale ich nie usuwać. Dla zmniejszenia czułości na występowanie odstających wartości danych w [5] proponuje się więc stosować funkcjonal

$$\sum_{i=1}^n \rho(x_i - \mu) \quad (1)$$

gdzie:  $\rho(x_i - \mu)$  - funkcja o postaci zależnej od parametru  $c$ .

Dla obserwacji o odchyleniach mniejszych od  $c\sigma$  (gdzie  $\sigma$  to odchylenie standardowe,  $c$  - współczynnik) jest to funkcja kwadratowa, a dla większych minimalizuje się moduł odchylenia  $|\varepsilon| = |x_i - \mu|$ . Te obszary czułości metody ilustruje rys. 1.

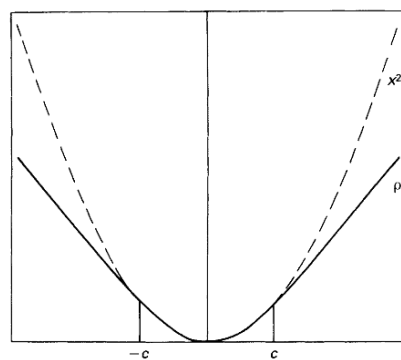
Wybrana w taki sposób funkcja  $\rho(\varepsilon)$  pozwala na mniej "ostre" podejście do danych odstających na wartość  $|\varepsilon| > c\sigma$  od środka grupowania rozkładu danych. Stała  $c$  określa stopień odporności. Wartość tej stałej zależy od stopnia "zanieczyszczenia" rozkładu. Tak więc dla "zanieczyszczenia" 1%,  $c=2$ , a dla "zanieczyszczenia" 5%  $c=1,4$ . Zwykle wybiera się  $c=1,5$ .

Zgodnie z tak wybranym kryterium konieczne jest zmodyfikowanie pozyskanych danych, a mianowicie:

$$x_i^* = \begin{cases} x_i & \text{przy } |x_i - \hat{\mu}| \leq c\sigma, \\ \hat{\mu} - c\sigma & \text{przy } x_i < \hat{\mu} - c\sigma, \\ \hat{\mu} + c\sigma & \text{przy } x_i > \hat{\mu} + c\sigma \end{cases} \quad (2)$$

gdzie:  $\hat{\mu} = \text{med}\{x_i\}$  danych  $x_i$  uszeregowanych rosnąco.

Opisana w (2) „przeróbka” danych nazywa się *windsoryzacją*.



Rys. 1. Względna funkcja czułości  $\rho$  metody odpornej. Uwaga:  $\rho(\varepsilon) = \varepsilon^2$  dla  $|\varepsilon| \leq \sigma$  i  $\rho(\varepsilon) < \varepsilon^2$  dla dużych  $|\varepsilon|$

Fig. 1. A robust method relative sensitivity function  $\rho$ . Note:  $\rho(\varepsilon) = \varepsilon^2$  for  $|\varepsilon| \leq \sigma$  and  $\rho(\varepsilon) < \varepsilon^2$  for large  $|\varepsilon|$

Jako odporną na dane odstające ocenę środka grupowania rozkładu  $\hat{\mu}$  próbki wstępnie preferuje się medianę  $\text{med}\{x_i\}$ . Badania wykazały [9], że najlepszy jest środek rozstępu między trzecim ( $p=3/4$ ) i pierwszym ( $p=1/4$ ) kwartylami próbki (*inter-quartile mid-range*). Ich odległość od tego środka nazywa się *odchyleniem ćwiartkowym* [16].

Długość odcinka pomiędzy kwartylami wynika jednoznacznie z rozkładu gęstości prawdopodobieństwa przyjętego dla rozrzutu danych próbki. Powierzchnia pod krzywą rozkładu dla tego przedziału wynosi 50%. Wprowadza się bezwzględne odchylenie

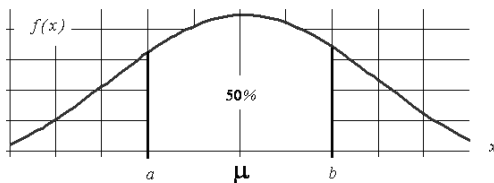
medianowe MAD (*median absolute deviation*), które jest oszacowaniem zakresu przejścia z rozkładu pełnego do obciętego

$$MAD_n = \text{med}\{|x_i - M_n|\}$$

gdzie  $M_n = \text{med}\{x_i\}$ ,  $n$  - liczba elementów w próbie.

Dla ustalenia zależności pomiędzy parametrami rozkładu PDF "obciętego" i przewidywanego dla całej populacji, tj. do spełnienia warunku skalowania, trzeba przeliczyć odchylenie standardowe  $\sigma$  przy użyciu współczynnika korygującego, który wyznacza się z początkowego rozkładu gęstości prawdopodobieństwa:

$$f(x) = \frac{d}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Rys. 2. Definicja rozstępu między-kwartylowego,  $a, b$  - rzędne pierwszego i trzeciego kwartyla

Fig. 2. Definition of inter-quartile mid-range,  $a, b$  - ordinates of first and third quartile

W rozkładzie "nieobciętym"  $d=1$ , zaś w "obciętym" w kwartylach

$$d = \frac{0,5}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

gdzie:  $\Phi(z)$  - znormalizowana skumulowana funkcja (CPDF) rozkładu normalnego.

Z tabeli rozkładu normalnego wynika

$$d = \frac{1}{0,6745} = 1,4826 \approx 1,483.$$

Tak więc zmiana skali dla przejścia od rozkładu nieobciętego do obciętego wynosi: 1,483. Ocena odchylenia standardowego  $s^*$  próbki o liczbie obserwacji  $n$ , wyznaczona na podstawie znormalizowanego rozstępu pomiędzy kwartylami

$$s^* = 1,483 \cdot MAD_n \quad (3)$$

Różni się ona od odchylenia standardowego próbki

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} \quad (4)$$

Huber [10] wykazał, że  $s^*$  stanowi nieobciążoną ocenę odchylenia standardowego  $\sigma$  i będzie bardziej stabilne nawet, jeśli próbka zawiera do 50% odstających wartości obserwacji (outlierów). Aby uzyskać stabilność przy występowaniu takich danych, należy dla wybranej funkcji  $\rho$  wyznaczyć medianę  $\text{med}$  oryginalnego zestawu danych. Wartość tę przyjmuje się jako ocenę środka grupowania danych bardziej odporną na wartości odstające niż ich średnia. Następnie z (3) oblicza się  $s^*$  i po wyborze  $c=1,5$  znajduje się granicę przejścia  $\varphi = cs^*$  z metody najmniejszych kwadratów MNK do metody najmniejszego modułu MNM.

Stosując warunki (2), otrzymuje się zmodyfikowany szereg danych  $x_i^*$  o wartości średniej

$$\bar{x}^* = \frac{1}{n} \sum_{i=1}^n x_i^* \quad (5)$$

Po podstawieniu  $x_i^*$  i  $\bar{x}^*$ , wyznacza się z (3) odchylenie standardowe  $s(x^*)$ . Obliczoną wartość  $s(x^*)$  wykorzystuje się z kolei do obliczenia "nowego" stabilnego odchylenia standardowego

$$s^* = 1,134 \cdot s(x^*)$$

Współczynnik 1,134 odpowiada  $c=1,5$  dla rozkładu normalnego. Otrzymaną wartość  $s^*$  wykorzystuje się do obliczenia nowej granicy  $\varphi$  i kontynuuje się procedurę w sposób taki, jak omówiono powyżej. Na podstawie wartości  $\bar{x}_i^*$  i  $\bar{x}_{i-1}^*$  obliczonych w bieżącym i poprzednim kroku iteracji określa się zbieżność algorytmu. Procedurę powtarza się, aż zmiany  $\bar{x}_i^*$  i  $s_i^*$  pomiędzy kolejnymi krokami staną się minimalne.

Zastosowanie metody przedstawiono na przykładzie w p. 4.2.

### 3. Istota porównań międzylaboratoryjnych

W celu sprawdzenia lub walidacji [14] danej metody pomiarowej prowadzi się wspólny eksperyment, w którym uczestniczą wybrane laboratoria o odpowiednim przygotowaniu merytorycznym. Rozkłady wyników poprawnych badań laboratoryjnych powinny z założenia asymptotycznie zbliżać się do rozkładu normalnego, ale tak nie jest i właśnie takie porównanie ujawnia nagminne występowanie wartości odstających. Wstępnie przyjmuje się, że pomiary wykonywane przez wszystkie te laboratoria mają jednakową powtarzalność. Jednak zdarza się w praktyce, że powtarzalność wyników w niektórych z nich jest z obiektywnych powodów gorsza niż dla pozostałych. Przy stosowaniu klasycznych testów statystycznych oddziela się z danych wyniki odstające [10], a pozostałe służą do wyznaczenia wariancji jako oceny odtwarzalności. Bardzo często niektóre z tych wyników są bliskie wartości granicznej oddzielającej odchylenia quasi-odstające i odstające. Nieuwzględnienie wszystkich wyników wpływa znacząco na wiarygodność oszacowania wariancji. Aby tego uniknąć, preferuje się stosowanie metod odpornych.

Międzylaboratoryjne badania atestacyjne (*key comparisons*), polegają na określaniu poprawności (*trueness*) badanej metody [3] i także jej powtarzalności (*repeatability*) szacowanej według średniej dyspersji wyników otrzymanych przez laboratoria. Stanowi ona obiektywną ocenę rzeczywistych technicznych i organizacyjnych możliwości danego laboratorium. Według [3], dotychczas przy ocenie wskaźników precyzji - powtarzalności i odtwarzalności (*reproducibility*), usuwa się odstające wartości wyników obserwacji za pomocą testu Grubbsa. Usunięcie „najgorszych” danych uzyskanych przez niektóre laboratoria idealizuje rzeczywiste warunki badań i zwiększa niepewność wspólnego wyniku. Nieuwzględnienie odstających wyników zmniejsza wiarygodność statystycznych ocen dokładności, tj. powtarzalności i odtwarzalności metody pomiarowej badanej w porównaniach międzylaboratoryjnych. Jest to szczególnie istotne przy kosztownych testach i próbach niszczących, gdy liczba badanych obiektów i danych w każdej z próbek jest zwykle ograniczona. W przykładzie dwoma metodami odpornymi na wartości odstające wyznaczy się i porówna wyniki i niepewności uzyskane przez laboratoria.

### 4. Przykład

Dane liczbowe tego przykładu zaczerpnięto z [3]. W dziewięciu laboratoriach przeprowadzono wspólny eksperyment polega-

jący na pomiarach porównawczych pewnej metody badawczej w celu oszacowania jej dokładności. Założono wstępnie, że wiarygodność pomiarów wszystkich laboratoriów jest jednakowa. Z pomiarów w laboratoriach otrzymano  $n = 9$  następujących wartości średnich  $x_i$ :

$$\begin{array}{ccccc} \underline{24,140} & 20,155 & 19,500 & 20,300 & 20,705 \\ \underline{17,570} & 20,100 & 20,940 & 21,185 & \end{array}$$

Dwa podkreślone wyniki ( $x_1$  oraz  $x_6$ ) istotnie odstają od pozostałych. Przy modelu tradycyjnym (*kontaminacji*) zakłada się, że poprawne obserwacje pochodzą tylko z rozkładu normalnego. Konsekwencją jest zastosowanie testu, np. Grubbsa, do znalezienia wartości odstających i ich odrzucenie. Z pozostałych wyznacza się wynik wspólny dla całego eksperymentu jako wartość średnią  $\bar{x} = 20,41$  i standardowe odchylenie  $s = 0,50$ . Tak obliczone  $s$  tylko z 7-miu wartości będzie jako ocena mało wiarygodna statystycznie.

Natomiast jeśli założy się, że wszystkie 9 obserwacji pochodzi z *pojedynczego rozkładu niegaussowskiego*, to logiczną konsekwencją jest zastosowanie algorytmu, który przy obliczaniu wyniku nie usuwa żadnej z wartości danych otrzymanych w pomiarach. Dotyczy to dwu rozpatrywanych tu metod odpornych: o przeskalowanym odchyleniu medianowym  $MAD_s$  i iteracyjnej dwukryterialnej.

#### 4.1. Obliczenie przeskalowanego odchylenia medianowego $MAD_s$

Najprostszy sposób postępowania to:

- wyznaczenie ze wszystkich  $n = 9$  danych  $x_i$  mediany  $med$  i uznanie jej za wartość wyniku pomiaru,
- obliczenie odchylenia medianowego (*median absolute deviation*)  $MAD = 0,64$  (również ze wszystkich 9. odchyleń danych).

Za standardową niepewność pomiaru  $s(x)$  można znać przeskalowane odchylenie medianowe  $MAD_s$

$$s(x) \equiv MAD_s = \kappa \cdot MAD \quad (6)$$

Wartość asymptotyczna  $\kappa_\infty = 1,483$  (odpowiadająca  $d = 1,483$  dla populacji generalnej) jest stosunkiem  $s(x)/MAD$  dla rozkładu normalnego w granicy  $n \rightarrow \infty$ .

Według Zięby [16] taką odporną procedurę proponowali Randa (2000), Burke (2001), Müller (2001) i nieco inną Cox (2002). W rozdziale 5 przygotowywanej do druku książki [16] Zięba, za opublikowanym w Internecie preprintem Randy z 2005 r. [17], podał jej modyfikację, tj. by zamiast wartości asymptotycznej  $\kappa_\infty = 1,483$ , która dla próbek o skończonej liczbie elementów  $n$  systematycznie zaniża wartość oceny wyniku pomiarów, stosować wartości  $\kappa(n)$  z tabeli 1. Dla próbki o  $n$  danych  $s(x_n) > s(x_\infty)$ . Stąd dla analizowanych danych otrzymuje się  $med = 20,3$ ;  $MAD = 0,64$  oraz

$$\begin{aligned} s(x_\infty) &\equiv 1,483 \cdot 0,64 = 0,95 \\ s(x_n) &\equiv \kappa(n) \cdot MAD = 1,633 \cdot 0,64 = 1,045. \end{aligned}$$

Tab. 1. Wartości mnożnika  $\kappa$  dla prób losowych o liczebności  $n$ , [17]  
Tab. 1. Values of coefficient  $\kappa$  for random samples of  $n$  elements [17]

$n$	$\kappa(n)$	$n$	$\kappa(n)$	$n$	$\kappa(n)$
2	1,773	10	1,626	50	1,507
3	2,206	11	1,602	100	1,494
4	2,019	12	1,596	1000	1,484
5	1,800	13	1,581	2000	1,483
6	1,764	14	1,577	.	.
7	1,686	15	1,566	.	.
8	1,671	20	1,544	$\infty$	1,483
9	1,633	25	1,530		

#### 4.2. Obliczenia oceny metodą dwukryterialną

W kolumnie „0” tabeli 2 podano w porządku rosnącym wyniki pomiarów poszczególnych  $i = (1, \dots, 9)$  laboratoriów jako dane  $x_{i(0)}^*$  stanu początkowego iteracji. Kolumny 1 – 4 zawierają kolejno dane następnych kroków iteracji.

Tab. 2. Przykład obliczeń odpornego wyniku pomiarów  
Tab. 2. Example of calculation of the robust measurement result

Nr iteracji $j$	0	1	2	3	4
$\varphi_j$		1,424	1,478	1,514	1,539
$x_j - \varphi_j$		18,876	18,909	18,893	18,872
$x_j + \varphi_j$		21,724	21,865	21,921	21,950
$x_1^*(j)$	17,570	18,876	18,909	18,893	18,872
$x_2^*(j)$	19,500				
$x_3^*(j)$	20,100				
$x_4^*(j)$	20,155				
$x_5^*(j)$	20,300				
$x_6^*(j)$	20,705				
$x_7^*(j)$	20,940				
$x_8^*(j)$	21,185				
$x_9^*(j)$	24,140	21,724	21,865	21,921	21,950
<b>Średnia <math>\bar{x}_j^*</math></b>	20,511	20,387	20,407	20,411	20,412
<b>SD <math>s_j^*</math></b>	1,727	0,869	0,890	0,905	0,916
<b>nowe <math>\bar{x}_{j+1}^*</math></b>	20,300	20,387	20,407	20,411	20,412
<b>nowe <math>s_{j+1}^*</math></b>	0,949	0,985	1,009	1,026	1,039

Dla danych początkowych oblicza się: wartość średnią

$$\bar{x}_0^* = \frac{1}{n} \sum_{i=1}^9 x_{i(0)}^* = 20,511$$

i z (4) odchylenie standardowe SD próbki

$$s_0^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^9 (x_{i(0)}^* - \bar{x}_0^*)^2} = 1,727$$

gdzie:  $x_{i(0)}^*$  - dane z kolumny "0" tabeli 2.

Jako początkowe (dla surowych danych  $x_{i(0)}^*$ ) oszacowanie położenia centrum grupowania populacji przyjmuje się medianę  $x_0^* \equiv med(x_{i(0)}^*) = x_{5(0)}^* = 20,300$  i z (3)

$$s_1^* = 1,483 \cdot med \left\{ \left| x_{i(0)}^* - x_0^* \right| \right\}$$

gdzie: seria różnic  $\left| x_{i(0)}^* - x_0^* \right|$  przyjmuje wartości:

$$2,73 \quad 0,80 \quad 0,20 \quad 0,145 \quad 0 \quad 0,405 \quad \underline{0,640} \quad 0,889 \quad 3,84.$$

Otrzymuje się  $s_1^* = 1,483 \cdot 0,640 = 0,949$  i następnie

$$\varphi_1 = 1,5 \cdot s_1^* = 1,5 \cdot 0,949 = 1,424$$

które określa granice dla odchylen przy przejściu od metody o najmniejszej sumie kwadratów MNK do metody najmniejszego modułu MNM:

$$x_0^* - \varphi_1 = 20,300 - 1,424 = 18,876$$

$$x_0^* + \varphi_1 = 20,300 + 1,424 = 21,724$$

Ponieważ  $x_{1(0)}^* < (x_0^* - \varphi_1)$ ,  $x_{9(0)}^* > (x_0^* + \varphi_1)$ , to wartości  $x_{1(0)}^*$ ,  $x_{9(0)}^*$  wykraczają poza te progi i w kolumnie 1 danym  $x_{1(1)}^*$ ,  $x_{9(1)}^*$  przypisuje się wartości progów, tj.  $x_{1(1)}^* = 18,876$ ,  $x_{9(1)}^* = 21,724$ , a pozostałe dane  $x_{2(1)}^* = x_{2(0)}^*$ , ...,  $x_{8(1)}^* = x_{8(0)}^*$  nie zmieniają się.

Po podobnych jak poprzednio obliczeniach otrzymuje się: wartość średnią  $\bar{x}_1^* = \frac{1}{n} \sum_{i=1}^9 x_{i(1)}^* = 30,387$ ; i odchylenie standardowe

$s_1^* = 0,869$ . Do ich obliczenia użyto wartości  $x_{i(1)}^*$  pobrane z kolumny "1". Centrum grupowania rozkładu będzie teraz  $x_1^* = \bar{x}_1^* = 30,387$ .

Następnie wyznacza się nową wartość

$$s_2^* = 1,134 \sqrt{\frac{1}{n-1} \sum_{i=1}^9 (x_{i(1)}^* - x_1^*)^2} = 0,985$$

gdzie:  $med \left\{ \left| x_{i(1)}^* - x_1^* \right| \right\} = 0,664$ .

Górna i dolna granice wynoszą odpowiednio

$$x_1^* - \varphi_2 = 18,909 \quad \text{i} \quad x_1^* + \varphi_2 = 21,865,$$

gdzie:  $\varphi_2 = 1,5 \cdot s_2^* = 1,5 \cdot 0,985 = 1,478$ .

W drugim etapie iteracji otrzymuje się wartości zmodyfikowane:

$$x_{1(2)}^* = x_1^* - \varphi_2 = 18,909$$

$$x_{9(2)}^* = x_1^* + \varphi_2 = 21,865.$$

Wprowadza się zmienione dane  $x_{1(2)}^*$  i  $x_{9(2)}^*$  do kolumny 2

i wraz z niezmodyfikowanymi danymi  $x_{i(0)}^*$  ( $i = 2, \dots, 8$ ) stosuje się je w drugim etapie iteracji. Wyniki obliczeń umieszcza się w kolejnej kolumnie 3 i dalej realizuje się ten sam algorytm jak poprzednio. Wartości średnie dla kroku czwartego i piątego, tj.  $\bar{x}_4^* = 20,411$  i  $\bar{x}_5^* = 20,412$  praktycznie nie różnią się. Przy tym odporne oceny odchylenia standardowego  $s_i^*$  wartości średniej pomiarów uzyskanych w laboratoriach są też zbliżone, tj.  $s_4^* = 1,026$  i  $s_5^* = 1,039$ .

### 4.3. Porównanie wyników obliczeń

Przy tradycyjnym podejściu, tj. zidentyfikowaniu wartości odstających (outlierów) za pomocą kryterium Grabbasa i wyeliminowaniu z dalszego przetworzenia wyników krańcowych

$x_{1(0)}^* = 17,570$  i  $x_{9(0)}^* = 24,140$ , dla  $n=7$  otrzymuje się średnią wartość  $m=20,4$ . W tym przypadku bierze się pod uwagę wyniki tylko z  $n=7$  laboratoriów i oszacowane z nich odchylenie standardowe próbek  $s = 0,501$ . Niepewność wspólnych badań będzie mniej wiarygodna, gdyż według dodatku E.1 do GUM [1, 2] ma ona większe własne odchylenie standardowe niż odchylenie  $s^*$  wyznaczone metodami odpornymi dla  $n=9$ .

Wartości centrum grupowania danych otrzymane w punktach 4.1 i 4.2 dla obu metod odpornych są zbliżone dlatego, że oceniano dane i dokładność wyników tego samego badania. Ich niepewności różnią się tylko o 9%, ale ponieważ dla próbek o 9 danych dokładność oceny jest około  $1/\sqrt{2(n-1)} = 25\%$ , to zbliżoność obu metod odpornych jest w praktyce niemal całkowita. Korzystając z metody dwukryterialnej uzyskuje się  $s^* = 1,039 > s$ , dobrze odpowiadające rzeczywistej dyspersji wartości otrzymanych wyników badań. Po usunięciu dwu krańcowych wyników początkowych  $x_{1(0)}^*$  i  $x_{9(0)}^*$  otrzymano mniejszą wartość  $s$ . Odbiega to od rzeczywistości, gdyż idealizuje się pomiar, zakładając, że dane eksperymentalne należą do populacji generalnej o rozkładzie normalnym.

Dla potwierdzenia tej tezy można rozpatrzeć prosty przykład o symulowanych danych pomiarowych pobranych z populacji o rozkładzie normalnym o wartości średniej  $m=75,238$  i odchyleniu standardowym  $\sigma = 13,475$ . Próbką zawierała następujące wartości:

$$75,3 \quad 76,0 \quad 76,3 \quad 102,1$$

Przy zastosowaniu kryterium Grabbasa największa wartość 102,1 została by uznana jako odstająca, chociaż w rzeczywistości  $m+2\sigma=102,201$ . Z odchylenia medianowego  $MAD_s$  otrzyma się estymator  $s(x) = 1,0095$  daleko odbiegający od rzeczywistej wartości odchylenia standardowego  $\sigma = 13,475$ .

Natomiast stosując omówiony algorytm iteracyjny otrzyma się wartość  $s_{28}^* = 14,882$  dużo bliższą rzeczywistemu  $\sigma$  i bardziej wiarygodnie opisującą możliwy rozrzut wartości obserwacji pomiarowych. Warto zaznaczyć, iż otrzymano ją dopiero aż po 28 iteracjach. Nie stanowi to jednak problemu dla współczesnej techniki obliczeniowej.

## 5. Podsumowanie

Metoda przeskalowanego odchylenia medianowego z p. 4.1 jest bardzo prosta, ale nie daje poprawnych rezultatów, gdy wartość odstająca daleko odbiega od grupy pozostałych danych. Uwidacznia się to szczególnie dla mało licznych próbek.

Metoda dwukryterialna z p. 4.2 ma bardziej skomplikowany, lecz łatwy do zautomatyzowania algorytm. Dzięki wprowadzeniu progu  $\pm c\sigma$  zmniejszającego czułość na dane odstające, ukierunkowana jest głównie na wyznaczenie odpornej oceny niepewności.

Przeprowadzona w p. 4.2 analiza wykazała przydatność zastosowania rozważanej odpornej metody dwukryterialnej do wyznaczania parametrów statystycznych próbek o małej liczbie danych, gdy pobrane są one z populacji generalnej o założonym rozkładzie normalnym, ale zawierają wyniki istotnie odstające od pozostałych. Pozwala ona ocenić obiektywnie wartość wyniku i dokładność metody badań.

Występujące w przykładzie różnice pomiędzy statystycznymi ocenami dokładności metody pomiaru wyznaczonymi przez poszczególne laboratoria i niepewnością wyniku całego międzylaboratoryjnego eksperymentu można wykorzystać jako podstawę obiektywnej analizy rzeczywistego stanu organizacji procesu badań i sposobu prowadzenia działalności w tych dwu laboratoriach, które miały wyniki odstające.

*Autorzy serdecznie dziękują D. A. Palyanichko z Politechniki Kijowskiej za pomoc w obliczeniach.*

## 6. Literatura

- [1] Guide to the Expression of Uncertainty in Measurement (GUM), revised and corrected version of GUM 1995, BIPM\_JCGM 100:2008.
- [2] Wyrażanie Niepewności Pomiaru. Przewodnik. Tłumaczenie GUM z 1995 r. z komentarzem J. Jaworskiego, Wydawnictwo Głównego Urzędu Miar Alfabero Warszawa 1999, 2002.
- [3] ISO 5725:1994 - Accuracy (trueness and precision) of measurement methods and results.
- [4] Sarhan Ahmed E., Greenberg Bernard G., editors: Contributions to order statistics, John Wiley & Sons, 1962, pp. 482.
- [5] Tukey John W.: Exploratory Data Analysis. Addison-Wesley, 1978.
- [6] Domański Cz., Pruska K.: Nielokalne metody statystyczne. Polskie Wydawnictwo Ekonomiczne SA., 2000.
- [7] Frosch Moller S., von Frese J.: Robust methods for multivariate data analysis. J. Chemometrics 2005; 19: 549–563.
- [8] Doksum, Kjell A. Mathematical Statistics: Basic and Selected Topics. 1. Second updated edition, Pearson Prentice-Hall 2007.
- [9] Olive David J.: Applied Robust Statistics - Southern Illinois University Department of Mathematics. June 23, 2008.
- [10] Huber Peter J., Ronchetti Elvezio M.: Robust Statistics. 2nd edition. Wiley - 2011 pp. 380.
- [11] Daszykowski M. i inni: Robust statistics in data analysis - A review basic concepts. Chemometrics and Intelligent Laboratory Systems 85 (2007) 203–219.
- [12] Stanimirova I. i inni: How to construct a multiple regression model for data with missing elements and outlying objects. Analytica Chimica Acta 581 (2007) 324–332.
- [13] Stanimirova I., Walczak B.: Classification of data with missing elements and outliers, Talanta 76 (2008) 602–609.
- [14] Piotrowski J., Kostyrko K.: Wzorcowanie Aparatury Pomiarowej. Wydawnictwo Naukowe PWN Warszawa 2002.
- [15] Buchcik D.: Procedura kalibracji przy wykorzystaniu metody najmniejszej mediany w przypadku modeli wielowymiarowych. PAK nr 5 2008, str.294 – 297.
- [16] Zięba A.: Opracowanie danych w naukach ścisłych i technicznych z uwzględnieniem zaleceń konwencji GUM oceny niepewności pomiaru. PWN (w trakcie wydawania).
- [17] Randa J.: Proposal for KCRV and Degree of Equivalence for GTRF Key Comparisons. NIST, (2000) and update (2005). Preprint dostępny w internecie.
- [18] [http://pl.wikipedia.org/wiki/Statystyka\\_odpornościowa](http://pl.wikipedia.org/wiki/Statystyka_odpornościowa)

*otrzymano / received: 16.01.2012*

*przyjęto do druku / accepted: 02.03.2012*

*artykuł recenzowany / revised paper*

## INFORMACJE

### Informacje dla Autorów

Redakcja przyjmuje do publikacji tylko prace oryginalne, nie publikowane wcześniej w innych czasopismach. Redakcja nie zwraca materiałów nie zamówionych oraz zastrzega sobie prawo redagowania i skracania tekstów oraz streszczeń.

Artykuły naukowe publikowane w czasopiśmie PAK są formatowane jednolicie zgodnie z ustaloną formatką zamieszczoną na stronie redakcyjnej [www.pak.info.pl](http://www.pak.info.pl). Dlatego artykuły przekazywane redakcji należy przygotowywać w edytorze Microsoft Word 2003 (w formacie DOC) z zachowaniem:

- wielkości czcionek,
- odstępów między wierszami tekstu,
- odstępów przed i po rysunkach, wzorach i tabelach,
- oznaczeń we wzorach, tabelach i na rysunkach zgodnych z oznaczeniami w tekście,
- układu poszczególnych elementów na stronie.

Osobno należy przygotować w pliku w formacie DOC notki biograficzne autorów o objętości nie przekraczającej 450 znaków, zawierające podstawowe dane charakteryzujące działalność naukową, tytuły naukowe i zawodowe, miejsce pracy i zajmowane stanowiska, informacje o uprawianej dziedzinie, adres e-mail oraz aktualne zdjęcie autora o rozmiarze 3,8 x 2,7 cm zapisane w skali odcieni szarości lub dołączone w osobnym pliku (w formacie TIF).

Wszystkie materiały:

- artykuł (w formacie DOC),
- notki biograficzne autorów (w formacie DOC),
- zdjęcia i rysunki (w formacie TIF lub CDR),

prosimy przesyłać w formie plików oraz dodatkowo jako wydruki na białym papierze (lub w formacie PDF) na adres e-mail: [wydawnictwo@pak.info.pl](mailto:wydawnictwo@pak.info.pl) lub pocztą zwykłą, na adres:

Redakcja Czasopisma Pomiaru Automatyka Kontrola,  
Sekretarz Redaktora Naczelnego mgr Agnieszka Skórkowska,  
ul. Akademicka 10, p.21A, 44-100 Gliwice.

Wszystkie artykuły naukowe są dopuszczane do publikacji w czasopiśmie PAK po otrzymaniu pozytywnej recenzji. Autorzy materiałów nadesłanych do publikacji są odpowiedzialni za przestrzeganie prawa autorskiego. Zarówno treść pracy, jak i wykorzystane w niej ilustracje oraz tabele powinny stanowić dorobek własny Autora lub muszą być opisane zgodnie z zasadami cytowania, z powołaniem się na źródło cytatu.

Przedrukowywanie materiałów lub ich fragmentów wymaga pisemnej zgody redakcji. Redakcja ma prawo do korzystania z utworu, rozporządzania nim i udostępniania dowolną techniką, w tym też elektroniczną oraz ma prawo do rozpowszechniania go dowolnymi kanałami dystrybucyjnymi.