

# Zastosowanie metod odpornościowych w analizie dokładności pomiarów międzylaboratoryjnych (1)

## Zasady statystyki odpornościowej, metoda Hubera czyli Algorytm-A

Zygmunt Lech Warszawa

Przemysłowy Instytut Automatyki i Pomiarów, Al. Jerozolimskie 202, 02-486 Warszawa, PL

Evgeniy T. Volodarsky

Narodowy Techniczny Uniwersytet Ukrainy „Politechnika Kijowska”, 03056 Kijów, Al. Pobedy 37, UA

**Streszczenie:** W dwuczęściowej pracy omówiono zastosowanie statystyki odpornościowej do oceny wartości i niepewności mierzonych uzyskiwanych na podstawie próbki danych doświadczalnych, gdy niektóre z tych danych różnią się istotnie od pozostałych, czyli są outlierami. Metodami odpornościowymi wyznaczono parametry statystyczne wyniku pomiaru ze wszystkich danych, ale wpływ outlierów potraktowano odmiennie. Dla próbek o niewielkiej liczności uzyskano wyniki bardziej wiarygodne niż w sposób klasyczny z odrzuceniem outlierów. Ilustrują to przykłady z porównań międzylaboratoryjnych. W części 1 omówiono podstawowe zasady statystyki odpornościowej oraz iteracyjną metodę odporną podaną przez Hubera, którą w normie ISO 5725-5 nazwano Algorytm-A. Jako ilustrację, w symulowanym przykładzie liczbowym, wyznaczono niepewność procedury pomiarowej testowanej przez porównanie wyników badania jednorodnych obiektów w kilku laboratoriach akredytowanych. Oszacowano średnią niepewność metodą klasyczną dla wszystkich danych. Po usunięciu outlierów zastosowano dwie metody odpornościowe – przeskalowanego odchylenia medianowego  $MAD_S$  i metodę Hubera, czyli iteracyjny Algorytm-A, którego wyniki były najbardziej wiarygodne.

**Słowa kluczowe:** outlier, niepewność pomiarów, odchylenie standardowe, mediana, odporna wartość średnia, rozstęp międzykwartylowy

## 1. Wprowadzenie

W praktyce pomiarowej, szczególnie w badaniach technicznych i przemysłowych, zdarza się, iż wartości jednej lub kilku obserwacji w próbce istotnie odstają od pozostałych danych, czyli są outlierami (ang. *outliers*). Nawet niewielka ich liczba może znacząco zmienić klasycznie obliczane parametry statystyczne próbki, szczególnie o małej liczbie elementów. Wartość wyniku pomiaru i jego niepewność typu A, obliczone zgodnie

z zasadami międzynarodowego Przewodnika GUM (ang. *Guide Uncertainty Measurement*) [1], mogą okazać się niewiarygodne statystycznie oraz niepoprawne, tj. obarczone błędami systematycznymi wskutek asymetrii próbki. W wielu przypadkach nie ma możliwości powtórzeń lub uzupełnienia pomiarów. Aby uniknąć takiej sytuacji, tradycyjnie stosuje się różne testy statystyczne, np. test Grubbsa. Pozwalają one zidentyfikować, a następnie usunąć nietypowe, tj. odstające wartości danych. Takie postępowanie jest skuteczne tylko dla dużych próbek, gdyż dla małych służące do tego celu testy 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 pomiary są niemożliwe do powtórzenia, wówczas należy wykorzystać wszystkie otrzymane dane. Usunięcie dowolnej obserwacji z próbki zmniejsza wiarygodność jej oceny statystycznej. Dla próbek o  $n$  niezależnych obserwacjach z rozkładu normalnego oszacowanie stosunku odchylenia standardowego  $s(u_A)$  niepewności standardowej  $u_A$  do jej wartości

### Autor korespondujący:

Zygmunt Lech Warszawa, zlw1936@gmail.com

### Artykuł recenzowany

nadesłany 30.03.2017 r., przyjęty do druku 01.06.2017 r.



Zezwala się na korzystanie z artykułu na warunkach licencji Creative Commons Uznanie autorstwa 3.0

oczekiwanej (czyli względnej niepewności standardowej niepewności  $u_A$ ) według tabeli E1 w dodatku E Przewodnika GUM [1] wynosi  $[2(n-1)]^{-1/2}$ . Na przykład dla  $n = 10$  stosunek  $s(u_A)/u_A$  wynosi 24%, dla  $n = 4$  – już 42%, a dla  $n = 3$  – rośnie do 52%. Po usunięciu jednej obserwacji z próbki o  $n = 4$  odchylenie standardowe  $s(u_A)$  wzrasta o około 24%. Bardziej odporne niż  $u_A$  jest medianowe odchylenie  $MAD$  ale jest niemal 1,5 razy większe i jego zastosowanie też nie jest korzystne.

Klasyczne, tzw. parametryczne metody statystyczne przetwarzania danych, opierają się na założeniu modelowania ich rozrzutu znanym rozkładem prawdopodobieństwa. Przyjęty kształt tego rozkładu jest tylko proponowanym umownym modelem opisu statystycznego rozrzutu danych doświadczalnych. Jednak w wielu przypadkach liczba pozyskanych w praktyce danych bywa niewystarczająca do wyboru modelu parametrycznego jednoznacznie odpowiadającego określonym rozrzutom. Sam wybór może być nieidealny i wartości obserwacji mogą lepiej podlegać innemu rozkładowi. Poprzednio, głównie ze względu na mniej żmudne obliczenia, zwykle przyjmowano, że jest to rozkład normalny. Opracowywano też tylko dane o wysokiej jakości, tj. po usunięciu wartości odstających od tego rozkładu. Zakorzeniło się to tak głęboko w statystycznym przetwarzaniu danych, że rezygnacja z niego nie wydawała się zasadna. Jednakże już Henri Poincaré wskazywał na nieuzasadnioną, ale panującą powszechnie głęboką wiarę w uniwersalność rozkładu normalnego, gdyż matematycy sądzą, że obserwuje się go w eksperymencie, zaś eksperymenciatorzy – że matematycy mogą udowodnić teoretycznie na podstawie centralnego twierdzenia granicznego, iż musi zachodzić rozkład normalny. W rzeczywistości, w nielicznych tylko przypadkach, np. gdy wynik pomiaru mierzand o stałej wartości wyznacza się z bardzo dużej liczby powtarzanych pomiarów i w systemie pomiarowym nie ma dominującego wśród wielu niezależnych źródeł losowości, to rozkład wyników pojedynczych pomiarów można traktować jako w pełni normalny.

Obecnie przy stosowaniu komputerów, można przetwarzać prawie każde dane. W założeniach teoretycznych przyjmowano też prowadzenie eksperymentu (uzyskiwanie obserwacji) w tych samych stałych warunkach losowości. Jednak w makrorealności często występuje przestrzenno-czasowa zmienność parametrów zarówno samego obiektu badań, jak i systemu pomiarowego oraz warunków otoczenia.

W latach 60. ubiegłego wieku, na podstawie wyników szeregów badań ustalono już, że dane eksperymentalne wymagające przetwarzania w praktyce, zawierają zwykle od 1%, nawet aż do 20% danych odstających od rozkładu normalnego. Dążenie do maksymalnego wykorzystania informacji zawartej we wszystkich danych, łącznie z odstającymi, spowodowało utworzenie nowej gałęzi statystyki (ang. *robust statistics*), której wyniki są odporne na takie dane. Jeden z jej twórców, amerykański statystyk John Tukey [4] prezentował nawet pogląd, że „normalność rozkładu danych uzyskanych w eksperymencie jest mitem i takiego rozkładu nie ma i nigdy się go nie uzyska”.

Jest to zagadnienie nadające się do rozstrzygnięcia przez filozofów. Natomiast w praktyce, szczególnie przy małej liczbie danych pomiarowych, powodem powstania odstających wartości mogą być błędy wynikłe z nieprzebrzegania zasad prowadzenia doświadczenia, niewykryte lub nietrwale uszkodzenia przyrządów pomiarowych, błędy i omyłki w przetwarzaniu danych, chwilowe wpływy zewnętrzne o dużej intensywności itp.

Celem pracy jest przedstawienie oceny dokładności próbek metodami statystyki odpornościowej. Wskutek małej wrażliwości na dane odstające umożliwiają one poprawę wiarygodności wyników, szczególnie przy próbkach o małej liczności. Podstawy tej statystyki opracowano kilkadziesiąt lat temu [4–8,

13–15]. Rozwijana jest dalej do różnych zadań przetwarzania danych, w tym do kalibracji pomiarów wieloparametrycznych [5–9] i do stosowania w wyspecjalizowanych technikach pomiarowych, np. w chemometrii [10–12]. Niektóre z metod odpornościowych są już oprogramowane, m.in. w środowisku obliczeniowym MATLAB. Nie trafiły one jeszcze do prac nad zweryfikowanymi zaleceniami wyznaczania niepewności pomiarów [30] zawartymi w międzynarodowym Przewodniku GUM [1], ale dwie z tych metod są podane jako Algorytm-A i Algorytm-S w załączniku C do norm PN-ISO 5725-5 [2] i PN-ISO 13528:2005 [3] dotyczących badań w laboratoriach akredytowanych. Podstawy statystyki odpornościowej przedstawili polskiemu środowisku metrologicznemu w swoich książkach Janusz Piotrowski i Krystyna Kostyrko [13] w 2000 r. oraz Andrzej Zięba [15] w 2013 r. Ujęcie z perspektywy ekonometrii zawarto w książce Czesława Domańskiego i Krystyny Pruski [14].

## 2. Zasady oceny rozproszenia danych metodami odpornościowymi

Wspomniano już, że w praktyce najbardziej rozpowszechnione są procedury analizy statystycznej, które są optymalne przy założeniu normalności rozkładu danych. Są one jednak dość wrażliwe na niewielkie odchylenia parametrów rozkładu od tego założenia, ze względu na stosowane kryterium minimum kwadratu odchylenia. Występuje to szczególnie przy estymacji wariancji. W rzeczywistości mamy do czynienia z rozkładami różniącymi się od idealnego rozkładu normalnego. Dlatego w przetwarzaniu danych doświadczalnych coraz szerzej stosuje się metody odpornościowe [3–16]. Pod pojęciem *odporność* rozumie się w statystyce niewrażliwość wyznaczanych wartości parametrów próbek pomiarowych na różne niewielkie odchylenia i niejednorodności rozrzutu elementów, czyli ich stabilność. Zwykle powstają one z przyczyn nieznanych.

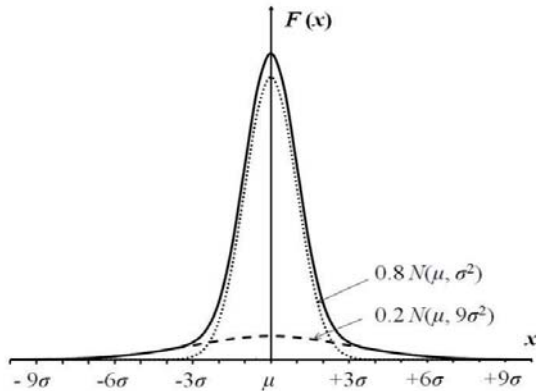
Podstawowym modelem w metodach odpornościowych nie jest zmienna losowa o pojedynczym rozkładzie normalnym, ale model „mieszany”. Dla różnych próbek danych tylko środkową część wartości danych można dość stabilnie modelować rozkładem normalnym. Zaś dolne fragmenty zboczy modelu rzeczywistego rozkładu, czyli jego „ogony” są bardziej rozciągnięte niż rozkład normalny dla części środkowej i mało stabilne. Obserwacje pomiarowe w obszarach krańcowych występują rzadziej. Względem rozkładu normalnego dla części środkowej pojedyncze z nich, szczególnie dla małych próbek, mogą być rozpoznawane jako outliers, choć w rzeczywistości są pseudo-outlierami. Przyjęcie takiego podejścia pozwala zachować tradycyjny pogląd o hipotetycznej jednorodności populacji generalnej, stanowiący dotychczas podstawę wszystkich ocen statystycznych. Jedynie w „ogonach” rzeczywistego rozkładu dopuszcza się możliwość występowania odchylenia odstających od rozkładu normalnego. Równocześnie na „ogony” nakłada się pewne ograniczenia. Modeluje się je rozkładem normalnym o innych parametrach bądź inną statystyką. Do opisu takiej sytuacji J. Tukey [4] zastosował następujące oryginalne podejście. Założył on, że istnieje duża liczba  $n$  danych pomiarowych, wśród których są zmieszane przypadkowo „dobre” i „złe” obserwacje  $x_i$  z populacji o średniej wartości  $\mu$ . „Dobre” i „złe” obserwacje pobrane z tej populacji występują odpowiednio z prawdopodobieństwem  $(1 - \epsilon)$  lub  $\epsilon$ , gdzie  $\epsilon$  to mała liczba. Taka sytuacja występuje w praktyce, gdy część obserwacji pomiarowych ma mniejszą dokładność. Na przykład odchylenie standardowe „złych” jest 3 razy większe niż „dobrych” i oba rodzaje obserwacji  $x_i$  mają różne rozkłady normalne, tj. pierwszy  $N(\mu, \sigma^2)$ , a drugi  $N(\mu, 9\sigma^2)$ , ale o tej samej wartości średniej  $\mu$ . Przy założeniu, że wszystkie war-

tości  $x_i$  są niezależne, otrzymuje się rozkład wypadkowy  $F(x)$  opisany następująco:

$$F(x) = (1 - \varepsilon) \cdot \Phi\left(\frac{x - \mu}{\sigma}\right) + \varepsilon \cdot \Phi\left(\frac{x - \mu}{3\sigma}\right) \quad (1)$$

gdzie  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$ .

Dołne części zboczy wypadkowego rozkładu  $F(x)$  leżą nieco powyżej zboczy rozkładu normalnego sporządzonego dla środkowych obserwacji próbki (rys. 1).



Rys. 1. Rozkład mieszany wg Tukey'a uwzględniający dane odstające  $F(x) = (1 - \varepsilon) N(\mu, \sigma^2) + \varepsilon N(\mu, 9\sigma^2)$  dla  $\varepsilon = 0,2$   
Fig. 1. Tukey joint distribution including outliers  $F(x) = (1 - \varepsilon) N(\mu, \sigma^2) + \varepsilon N(\mu, 9\sigma^2)$  for  $\varepsilon = 0.2$

Wśród opracowanych dotąd i wdrożonych do stosowania metod odpornościowych i ich algorytmów szeroko rozprze-strzeniło się inne podejście, zaproponowane przez Hubera [8, 13–15], również uznane już za klasyczne. Wprowadził on pewną stałą wartość  $c$  zależną od stopnia „zanieczyszczenia” populacji generalnej. Wartość ta służy do określenia granic części środkowej histogramu danych pomiarowych, którą modeluje się podstawowym rozkładem normalnym [5–8, 13–15, 17–24, 28, 29]. Obserwacje o wartościach w obszarach bocznych – „ogonach” histogramu danych, wg jednego ze stosowanych kryteriów (np. testu Grubbsa) mogą zostać uznane za odstające od danych o rozkładzie podstawowym. Ściąga się je na granice obszaru środkowego. Otrzymuje się nowy zbiór wartości obserwacji o mniejszym odchyleniu standardowym i węższym obszarze środkowym. Obliczanie nowych danych skrajnych powtarzane jest iteracyjnie [2, 3, 13–15, 29] tak długo, aż różnice kolejnych średnich i kolejnych standardowych odchyłeń przyjmą wartości poniżej wymaganego poziomu dokładności obliczeń, np. trzeciej cyfry znaczącej. Ta metoda odpornościowa przewidziana jest do stosowania w atestacji dokładności badawczych i kontrolnych procedur pomiarowych przez porównanie wyników w badaniach międzylaboratoryjnych jako „Algorytm A” podany w załączniku C normy PN-ISO 5725-5 [2]. Zasady jej stosowania do oceny wartości średniej wyników w badaniu procedur pomiarowych i biegłości laboratorium przedstawiono z udziałem autora w czasopiśmie [17–24], w materiałach pokonferencyjnych [25–27] i jako rozdział 11 w monografii [29].

Poniżej zostały omówione szczegółowo dwie metody odpornościowe – iteracyjna metoda Hubera (Algorytm-A) oraz metoda przeskalowanego odchylenia medianowego  $MAD_g$ . W dwu przykładach liczbowych porównano wyniki otrzymane klasycznie i przytoczonymi metodami.

### 3. Odpornościowa metoda iteracyjna Hubera, czyli Algorytm-A

Metody statystyki odpornościowej zapewniają mniejszy niż modele klasyczne wpływ danych odstających i innych anomalii danych na wynik pomiaru. Znajdują one szerokie zastosowanie przy szacowaniu odchyłeń standardowych  $\sigma$  populacji na podstawie próbki jej danych. Dalej omówiono szczegółowo iteracyjną metodę przetwarzania danych opartą na podejściu Hubera [9]. W normach [2, 3] metoda ma nazwę Algorytm-A. Zalicza się ona do metod odpornościowych o akronimie IRLS (ang. *iteratively reweighted least squares*). Dzięki zmniejszonej wrażliwości na dane odstające, metoda umożliwia poprawę wiarygodności wyników dla próbek o małej liczbie elementów. Przyjmuje się, że tylko część środkowa rozkładu gęstości prawdopodobieństwa PDF (ang. *probability density function*), tj. dla małych odchyłeń od estymatora wartości mierzanduro, nie odbiega zbyt od przyjętego modelu gaussowskiego.

W obecnie stosowanej statystycznej metodzie obliczania niepewności pomiarów opartej na zaleceniach międzynarodowego przewodnika GUM [1] zakłada się, że dane obserwacji pomiarowych powtarzanych w jednakowych warunkach, po wyeliminowaniu znanych błędów systematycznych, można przetwarzać tak samo jak próbkę pobraną z rozkładu normalnego o wartości średniej  $\mu$  i odchyleniu standardowym  $\sigma$ . Z danych próbki oblicza się estymatory, czyli oszacowania obu tych parametrów, odpowiednio jako wartość średnią mierzanduro  $\bar{x}$  i jej odchylenie standardowe  $s$ . Następnie ocenia się szerokość przedziału, w którym średnia  $\bar{x}$  może znajdować się z określonym prawdopodobieństwem  $P$ , czyli na przykład określa niepewność standardową  $u_A = s$  lub niepewność rozszerzoną  $U_p$ . Ocena ta osiąga minimum dla wartości  $\bar{x}$ , ale jest wrażliwa na dane odstające od modelowanych rozkładem normalnym, czyli outlierów. Zwiększają one wartość niepewności  $u_A$  wraz z kwadratem ich odległości od średniej. W klasycznej procedurze statystycznej wykrywa się takie dane za pomocą testów Grubbsa i usuwa się je z próbki [16, 29]. Dla próbek o bardzo małej liczbie elementów  $n$  wiarygodność wyznaczanych statystycznie jej parametrów jest bardzo mała i zmaleje jeszcze po usunięciu outlierów. Parametry statystyczne obliczone dla pojedynczej małej próbki mogą znacznie różnić się od parametrów populacji, czyli być niepoprawne. Otrzymywane w praktyce wartości obserwacji, i to nawet dla małych próbek pobranych rzeczywiście z populacji o rozkładzie normalnym, zwykle rozkładają się niesymetrycznie względem ich wartości średniej. Badania rozkładów skośności i kurtozy takich próbek, wykonane metodą Monte Carlo opisano w rozdziale 5 monografii [29]. Im próbka była mniejsza, tym średnia wartość jej współczynnika skośności (asymetria) wzrastała i osiągnęła maximum dla  $n = 6$ .

W statystyce odpornościowej rozpatrywane są modele rozkładów wraz z danymi odstającymi. Cechą charakterystyczną takich rozkładów PDF są tylko bardziej rozciągnięte „ogony” niż dla gęstości prawdopodobieństwa rozkładu normalnego, czyli funkcji Gaussa. Skorzystanie z tej właściwości pozwala zachować wygodne do analizy założenie o hipotetycznej jednorodności populacji, na którym opierają się wszelkie statystyczne oszacowania i uwzględnić też częściowo informację zawartą w odchyleniach odstających. W omawianej tu metodzie odpornościowej zakłada się, że jedynie dla centralnej części rozkładu danych doświadczalnych można przyjąć założenie o normalnym rozkładzie populacji. Tylko dla danych z tej części można przeprowadzać uśrednianie, tj. stosować metodę najmniejszych kwadratów. Bardziej odporne na duże odchylenia jest podane przez Laplace'a kryterium minimum modułu. Przy tworzeniu metody odpornościowej Algorytm-A dokonano „symbiozy” obu tych kryteriów. Dla centralnej części wartości danych stosuje

się metodę najmniejszych kwadratów, a poza granicami tego przedziału – kryterium modułowe. W ten sposób zmniejsza się wpływ danych odstających, ale ich się nie usuwa. W celu zmniejszenia czułości na wpływ odstających danych zaproponowano [2, 3, 8], aby w obliczeniach stosować korekcyjne wartości danych o modułach odchylen  $\varepsilon$  spełniających nierówność

$$|\varepsilon_i| = |x_i - \hat{\mu}| > \varphi \quad (2)$$

gdzie:  $\varphi = c\sigma$  – odległość od środka grupowania danych  $\hat{\mu}$ ,  $c$  – stała określająca przyjęty stopień odporności,  $\sigma$  – odchylenie standardowe populacji.

Wartość stałej  $c$  zależy od stopnia „zanieczyszczenia” rozkładu. Dla zanieczyszczenia 1% oraz 5% przyjmuje się odpowiednio  $c = 2$  oraz  $c = 1,4$ . Zwykle stosuje się  $c = 1,5$ .

Przy obliczaniu odchylenia standardowego dla danych w obszarze środkowym  $\mu \pm \varphi$  stosuje się kryterium jak dla rozkładu normalnego, a na zewnątrz minimum modułu błędu. Taki sposób pozwala na mniej „ostre” podejście do danych odstających na wartość  $|\varepsilon| > c\sigma$  od środka grupowania rozkładu danych. Zgodnie z tak wybranym kryterium modyfikuje się skrajne z danych pozyskanych w eksperymencie w następujący sposób:

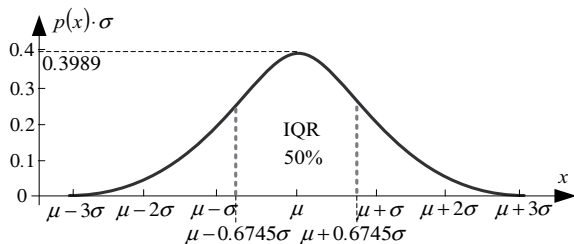
$$x_i^* = \begin{cases} x_i & \text{dla } |x_i - \hat{\mu}| \leq c\sigma, \\ \hat{\mu} - c\sigma & \text{dla } x_i < \hat{\mu} - c\sigma, \\ \hat{\mu} + c\sigma & \text{dla } x_i > \hat{\mu} + c\sigma \end{cases} \quad (3)$$

gdzie:  $\hat{\mu} = \text{med}\{x_i\}$  (przyjęto medianę jako odpornościową ocenę środka grupowania danych  $x_i$  uszeregowanych rosnąco).

Opisana wzorami (3) „przeróbka” wartości danych nosi nazwę *winsoryzacja* (ang. *winsorizing*) od nazwiska amerykańskiego matematyka Charlesa Winsora.

Ponieważ parametry statystyczne populacji generalnej  $\mu$  i  $\sigma$  zwykle nie są znane, to we wzorach (3) należy stosować ich estymaty wyznaczone z danych próbki. Jako początkową ocenę środka grupowania rozkładu  $\hat{\mu}$  próbki preferuje się medianę  $\text{med}\{x_i\}$ , która jest najbardziej odporna na dane odstające.

Badania [9] wykazały też, że najlepsze wyniki uzyskuje się dla przedziału środkowego między trzecim ( $p = 3/4$ ) i pierwszym ( $p = 1/4$ ) kwartyłami (rys. 2). Odległość kwartyli od jego środka nazywa się odchyleniem ćwiartkowym [16]. Długość rozstępu między kwartyłami wynika jednoznacznie z normalnego rozkładu gęstości prawdopodobieństwa przyjętego dla rozrzutu danych próbki. Powierzchnia pod krzywą rozkładu dla tego przedziału wynosi 50%.



Rys. 2. Definicja rozstępu międzykwartyłowego (a, b) (o odchyleniu ćwiartkowym). Linie przerywane – rzędne: pierwszego kwartyła  $a = \mu - 0,6745\sigma$  i trzeciego kwartyła  $b = \mu + 0,6745\sigma$   
Fig. 2. Definition of inter-quartile mid-range (a, b). Dotted lines – ordinates of first quartile  $a = \mu - 0.6745\sigma$  and third quartile  $b = \mu + 0.6745\sigma$

Występujące we wzorze (3) odchylenie standardowe populacji  $\sigma$  zwykle nie jest znane. Jako początkowe oszacowanie zakresu

przejścia z rozkładu pełnego do obciętego wprowadza się do obliczeń bezwzględne odchylenie medianowe  $MAD$  (ang. *median absolute deviation*), które można obliczyć z danych próbki:

$$MAD_n = \text{med}\{|x_i - M_n|\} \quad (4)$$

gdzie:  $M_n = \text{med}\{x_i\}$ ,  $n$  – liczba elementów w próbce.

Dla ustalenia zależności między parametrami obciętego rozkładu PDF oraz przewidywanego dla całej populacji, trzeba przeliczyć odchylenie standardowe  $\sigma$  za pomocą współczynnika korygującego, tj.

$$f(x) = \frac{d}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

Przy rozkładzie nieobcętym  $d = 1$ , zaś dla obciętego w kwartyłach:

$$d = \frac{0,5}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \quad (6)$$

gdzie:  $\Phi(z)$  – znormalizowana dystrybuanta (CPDF) rozkładu normalnego.

Z tabeli rozkładu normalnego wynika, że:

$$d = \frac{1}{0,6745} = 1,4826 \approx 1,483 \quad (7)$$

Zmiana skali przy przejściu między przedziałami o rozkładzie nieobcętym i obcętym wynosi 1,483. Ocena odchylenia standardowego  $s^*$  próbki o dużej liczbie obserwacji  $n$ , wyznaczona na podstawie znormalizowanego rozstępu między kwartyłami, wynosi:

$$s^* = 1,483 \times MAD_n \quad (8)$$

Wartość  $s^*$  różni się od odchylenia standardowego próbki  $s$  wg przewodnika GUM [1], tj.:

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} \quad (9)$$

Huber [9] wykazał, że odchylenie  $s^*$  stanowi nieobciążoną ocenę odchylenia standardowego populacji  $\sigma$ . Jest ono bardzo stabilne nawet, jeśli próbka zawiera do 50% odstających wartości obserwacji, czyli outlierów. Aby uzyskać stabilność oszacowania parametrów próbki przy występowaniu takich danych, należy wyznaczyć medianę zestawu danych surowych. Wartość tę przyjmuje się za ocenę początkową  $x_0^*$  środka grupowania danych jako bardziej odporną na wartości odstające niż ich średnia. Następnie oblicza się wartość  $MAD$  z (4) i  $s_1^*$  z (8) oraz po wyborze postaci funkcji  $\varphi$ , np. dla  $c = 1,5$  znajduje się odległości od  $x_0^*$  do granic przejścia  $\pm \varphi_1 = c \cdot s_1^*$  między obszarami stosowania metody najmniejszych kwadratów MNK i najmniejszego modułu MNM.

Na podstawie warunków (3) otrzymuje się zmodyfikowane dane  $x_{i(1)}^*$  o wartości średniej:

$$\bar{x}_1^* = \frac{1}{n} \sum_{i=1}^n x_{i(1)}^* \quad (10)$$

Po wstawieniu wartości  $x_i^*$  i  $\bar{x}_1^*$  wyznacza się z (9) odchylenie standardowe  $s_2(x^*)$ . Obliczoną wartość  $s_2(x^*)$  wykorzystuje się z kolei do obliczenia nowego stabilnego odchylenia standardowego:

$$s_2^* = 1,134 \cdot s_2(x^*) \quad (11)$$

Współczynnik 1,134 odpowiada wartości  $c = 1,5$  dla rozkładu normalnego. Wartość  $s_2^*$  wykorzystuje się do obliczenia nowej granicy  $\varphi_2$  i kontynuuje się procedurę w taki sam sposób, jak omówiony powyżej. Na podstawie wartości  $\bar{x}_j^*$  i  $\bar{x}_{j-1}^*$  obliczonych w bieżącym  $j$  i poprzednim kroku iteracji  $j-1$  określa się zbieżność algorytmu. Procedurę powtarza się, aż zmiany  $\bar{x}_j^*$  i  $s_j^*$  między kolejnymi krokami staną się dopuszczalnie małe.

Zastosowanie metod odpornościowych będzie dalej w tekście zilustrowane przykładami liczbowymi. Dotyczyć będą one wykorzystania tych metod do walidacji lub weryfikacji pewnej procedury pomiarowej przez porównanie wyników pomiarów (wartości średnich i niepewności) uzyskanych tą procedurą w badaniach parametrów jednorodnego obiektu w wielu akredytowanych laboratoriach. Należy więc omówić istotę takiego porównania.

## 4. Istota porównań międzylaboratoryjnych

W celu sprawdzenia lub walidacji [13] określonej procedury pomiarów daną metodą pomiarową prowadzi się wspólny eksperyment, w którym uczestniczą wybrane laboratoria o odpowiednim przygotowaniu merytorycznym. Rozkład wyników badań laboratoryjnych wykonywanych poprawnie powinien z założenia zbliżać się asymptotycznie do rozkładu normalnego. Zwykle tak nie jest i właśnie takie porównanie ujawnia występowanie wartości odstających. Wstępnie przyjmuje się, że pomiary wykonywane przez wszystkie uczestniczące w eksperymencie laboratoria mają zbliżoną powtarzalność wyników. Jednak zdarza się w praktyce, że z powodów obiektywnych w niektórych z nich powtarzalność jest gorsza niż w pozostałych. Przy klasycznym podejściu stosuje się testy statystyczne wydzielające z danych wyniki odstające [16, 29], a z pozostałych danych wyznacza się wariancję jako ocenę odtwarzalności. Bardzo często niektóre z tych wyników są bliskie wartości granicznej oddzielającej odchylenia quasi-odstające i odstające. Przy małej liczbie powtórzeń pomiaru nieuwzględnienie kilku wyników istotnie obniża wiarygodność oszacowania wariancji. Aby ten wpływ złagodzić, stosuje się metody odpornościowe.

Porównania międzylaboratoryjne (ang. *interlaboratory comparisons*) polegają na określaniu poprawności badanej metody [2–4] i jej powtarzalności, szacowanej wg średniej dyspersji wyników pomiarów otrzymanych w laboratoriach. Dyspersja ta jest obiektywną oceną rzeczywistych możliwości technicznych i organizacyjnych danego laboratorium. Dotychczas przy ocenie wskaźników precyzji procedury badań (jej powtarzalności i odtwarzalności) usuwa się odstające wartości wyników obserwacji, np. za pomocą testu Grubbsa. Usunięcie danych „najgorszych” uzyskanych przez niektóre z laboratoriów idealizuje jednak rzeczywiste warunki badań i zwiększa niepewność wspólnego wyniku. Natomiast nieuwzględnienie odstających danych 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 uzyskanych w każdej z próbek jest ograniczona. W przykładzie poniżej, dla serii wyników z wartościami odstającymi uzyskanych przez różne laboratoria, wyznaczy się i porównuje wartości i niepewności menzurandu obliczone w sposób klasyczny z wszystkich danych

i po odrzuceniu danych odstających, czyli outlierów oraz dwoma metodami odpornościowymi.

## 5. Przykład obliczeniowy 1

W dziewięciu laboratoriach przeprowadzono wspólny eksperyment polegający 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  $n = 9$  wartości średnich  $x_i$ , podanych w postaci następującego uporządkowanego szeregu

$$\underline{7,81} \quad 7,93 \quad 8,13 \quad 8,14 \quad 8,38 \quad 8,40 \quad 8,44 \quad 8,52 \quad \underline{9,31}$$

Dwie wyróżnione wartości  $x_1$  i  $x_9$  istotnie odstają od pozostałych. W modelu tradycyjnym (kontaminacji) zakłada się, że poprawne obserwacje pochodzą tylko z rozkładu normalnego. Konsekwencją jest zastosowanie testu, np. Grubbsa [16, 28], do znalezienia wartości odstających i ich odrzucenie. Z pozostałych wyznacza się wynik wspólny dla całego eksperymentu jako wartość średnią  $\bar{x} = 8,28$  i standardowe odchylenie  $s = 0,213$ . Tak obliczone  $s$ , tj. tylko na podstawie siedmiu wartości, będzie mniej wiarygodną oceną statystyczną niż uzyskana dla dziewięciu danych (z [1] – E.4 wynika  $\sigma[s(x)]/s(x) = 29\%$  i  $24\%$ ).

Jeśli natomiast założy się, że wszystkie dziewięć 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. Stosuje się to w dwu rozpatrywanych metodach odpornościowych: o przeskalowanym odchyleniu medianowym  $MAD_s$  i iteracyjnej dwukryterialnej czyli Algorytm-A [2, 3].

### Obliczenie metodą przeskalowanego odchylenia medianowego $MAD_s$

Sposób postępowania w tej metodzie odpornościowej jest bardzo prosty:

- ze wszystkich  $n = 9$  danych  $x_i$  wyznacza się medianę  $med$  jako estymator wartości wyniku pomiaru,
- następnie oblicza się odchylenie medianowe  $MAD$  wg wzoru (4),
- za standardową niepewność pomiaru  $s(x)$  uznaje się przeskalowane odchylenie medianowe  $MAD_s$ :

$$s(x) \equiv MAD_s = \kappa \cdot MAD \quad (12)$$

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

Tabela 1. Wartości mnożnika  $\kappa(n)$  dla prób losowych o liczebności  $n$  wg [15]

Table 1. Values of coefficient  $\kappa(n)$  for random samples of  $n$  elements [15]

$n$	$\kappa(n)$	$n$	$\kappa(n)$	$n$	$\kappa(n)$
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	$\infty$	.
9	1,633	25	1,530		1,483

Według Zięby [15] taką odpornościową procedurę proponowali Randa (2000), Burke (2001), Müller (2001) i nieco inną Cox (2002). W rozdziale 5 swojej książki [15] Zięba, w ślad za opublikowanym w Internecie preprintem Randy z 2005 r. [16], podał modyfikację tej metody. Zamiast wartości asymptotycznej  $\kappa(\infty) = 1,483$ , która dla próbek o skończonej liczbie elementów  $n$  zaniża oceny wyniku pomiarów, należy stosować  $\kappa(n)$  z tabeli 1.

Dla próbki o  $n$  danych spełniony jest warunek  $s(x_n) > s(x_\infty)$ . Stąd dla analizowanych dziewięciu danych otrzymuje się:

$$med = 8,38; MAD = 0,24$$

$$s(x_9) \equiv \kappa(9) \times MAD = 1,633 \cdot 0,24 = 0,39$$

Według obliczeń tą metodą otrzymuje się estymatę wartości mierzand:  $med = 8,38$  i estymatę jej odchylenia standardowego  $s(x_9) = 0,39$ .

### Obliczenia ocen mierzand metodą iteracyjną Hubera

W tabeli 2 podano obliczenia odpornościowego wyniku pomiarów dla tych samych danych wg metody Hubera, nazwanej Algorytmem-A w aneksie C normy PN-ISO 5725-5 [2].

W kolumnie 0 tej tabeli zamieszczono w porządku rosnącym wyniki pomiarów otrzymane w  $i = 1, \dots, 9$  laboratoriach jako dane  $x_{i(0)}$  stanu początkowego iteracji. Natomiast kolumny 1-4 zawierają dane dla kolejnych kroków iteracji Algorytmu-A.

Dla wszystkich danych początkowych obliczana jest: wartość średnia:

$$\bar{x}_0 = \frac{1}{n} \sum_{i=1}^9 x_{i(0)} = 8,340$$

i odchylenie standardowe próbki:

$$s_0 = \sqrt{\frac{1}{n-1} \sum_{i=1}^9 (x_{i(0)}^* - \bar{x}_0)^2} = 0,436$$

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

Jako początkowe (dla surowych danych  $x_{i(0)}^*$ ) oszacowanie położenia centrum grupowania populacji danych przyjmuje się medianę:

$$x_0^* \equiv med(x_{i(0)}^*) = x_{5(0)}^* = 8,380$$

tworzy się szereg różnic  $|x_{i(0)}^* - x_0^*|$  przyjmujących wartości:

$$0,57 \quad 0,45 \quad 0,25 \quad \mathbf{0,24} \quad 0 \quad 0,02 \quad 0,06 \quad 0,14 \quad 0,93$$

i z (8) wyznacza się jej odchylenie standardowe

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

Otrzymuje się  $s_1^* = 1,483 \cdot 0,24 = 0,356$  i następnie:

$$\varphi_1 = 1,5 \cdot s_1^* = 1,5 \cdot 0,356 = 0,53$$

Wartość  $\varphi_1$  określa granice przejścia dla odchyżeń z metody o najmniejszej sumie kwadratów do metody najmniejszego modułu:

**Tabela 2. Przykład obliczeń odpornościowego wyniku pomiarów metodą Algorytm-A**  
Table 2. Example of calculation of the robust measurement result by Algorithm-A

Nr iteracji $j$	0	1	2	3	4
$\varphi_j$		0,53	0,56	0,55	0,54
$x_j^* - \varphi_j$		7,85	7,74	7,74	7,75
$x_j^* + \varphi_j$		8,91	8,86	8,84	8,83
$x_{1(j)}^*$	7,81	7,85	7,81	7,81	7,81
$x_{2(j)}^*$	7,93				
$x_{3(j)}^*$	8,13				
$x_{4(j)}^*$	8,14				
$x_{5(j)}^*$	8,38				
$x_{6(j)}^*$	8,40				
$x_{7(j)}^*$	8,44				
$x_{8(j)}^*$	8,52				
$x_{9(j)}^*$	9,31	8,91	8,86	8,84	8,83
$\bar{x}_j^*$	8,340	8,300	8,290	8,288	8,287
$s_j^*$	0,436	0,326	0,322	0,317	0,315
$\bar{x}_{j+1}^*$	8,380	8,300	8,290	8,288	8,287
$s_{j+1}^*$	0,356	0,370	0,365	0,359	0,357

$$x_0^* - \varphi_1 = 8,380 - 0,53 = 7,85$$

$$x_0^* + \varphi_1 = 8,380 + 0,53 = 8,91$$

Ponieważ  $x_{1(0)}^* < (x_0^* - \varphi_1)$  oraz  $x_{9(0)}^* > (x_0^* + \varphi_1)$ , to wartości  $x_{1(0)}^*$ ,  $x_{9(0)}^*$  wykraczają poza te progi i w kolumnie 1 zmienionym danym  $x_{1(1)}^*$ ,  $x_{9(1)}^*$  przypisuje się wartości progów:

$$x_{1(1)}^* = 7,85; \quad x_{9(1)}^* = 8,91$$

zaś pozostałe dane  $x_{2(1)}^* = x_{2(0)}^*, \dots, x_{8(1)}^* = x_{8(0)}^*$  nie zmieniają się.

Po podobnych jak poprzednio obliczeniach otrzymuje się wartość średnią:

$$\bar{x}_1^* = \frac{1}{n} \sum_{i=1}^9 x_{i(1)} = 8,300$$

i odchylenie standardowe:

$$s_1^* = 0,326.$$

Do ich obliczenia użyto wartości  $x_{i(1)}^*$  pobrane z kolumny 1 tabeli 2. Centrum grupowania rozkładu będzie teraz:

$$x_1^* = \bar{x}_1^* = 8,300$$

Następnie wyznacza się nową wartość:

$$s_2^* = 1,134 \sqrt{\frac{1}{n-1} \sum_{i=1}^9 (x_{i(1)}^* - x_1^*)^2} = 0,370$$

gdzie:  $med\{x_{i(1)}^* - x_1^*\} = 0,326$ .

Górna i dolna granice wyniosą teraz odpowiednio:

$$x_1^* - \varphi_2 = 7,74 \text{ oraz } x_1^* + \varphi_2 = 8,86,$$

gdzie:  $\varphi_2 = 1,5 \cdot s_2^* = 1,5 \cdot 0,370 = 0,555$ .

W drugim kroku iteracji ( $j = 2$ ) otrzymuje się wartości zmodyfikowane:

$$x_{1(2)}^* = x_1^* - \varphi_2 = 7,74$$

$$x_{9(2)}^* = x_1^* + \varphi_2 = 8,86$$

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 tym etapie iteracji. Wyniki obliczeń umieszcza się w kolejnej kolumnie 3 dla kroku (3) i dalej realizuje się ten sam algorytm jak poprzednio. Wartości średnie już dla kroków (4) i (5), tj.  $\bar{x}_4^* = 8,288$  i  $\bar{x}_5^* = 8,287$ , praktycznie nie różnią się. Odporne oceny odchylenia standardowego  $s_i^*$  wartości średniej pomiarów w laboratoriach są też dla tych kroków zbliżone, tj.  $s_4^* = 0,317$  i  $s_5^* = 0,315$ .

## Analiza wyników obliczeń

Wyniki obliczeń uzyskanych różnymi metodami zestawiono do porównywania w tabeli 3.

Tabela 3. Porównanie wyników dla 2 metod klasycznych i 2 metod odpornościowych

Table 3. Comparison of results obtained by 2 classic and 2 robust methods

Metoda	wszystkie 9 danych	usunięte $x_1$ i $x_9$	odpornościowa $MAD_s$	odpornościowa Algorytm-A
Estymator wartości	$\bar{x}_0 = 8,340$	$m = 8,277$	$med = 8,38$	$\bar{x}_5^* = 8,287$
Odchylenie standardowe	$s_0 = 0,436$	$s = 0,213$	$s(x_9) = 0,39$	$s_5^* = 0,315$

Tradycyjne podejście polega na zidentyfikowaniu wartości odstających jako outlierów, np. za pomocą kryterium Grubbsa i wyeliminowaniu ich z dalszego przetwarzania. W tym przykładzie usuwa się wartości krańcowe  $x_{1(0)}^* = 7,81$ ,  $x_{9(0)}^* = 9,31$  i bierze się pod uwagę tylko wyniki z pozostałych  $n = 7$  laboratoriów.

Wówczas dla wspólnych badań otrzymuje się oszacowanie średniej wartości  $m = 8,277$  i jej odchylenia standardowego  $s = u_A = 0,213$ . Niepewność  $u_A$  będzie mniej wiarygodna o 16%, gdyż wg dodatku E.4 do Przewodnika GUM [1] ocenę własnego względnego odchylenia standardowego niepewności  $s(u_A)$  dla próbek z rozkładu normalnego wyznacza się z wyrażenia  $1/\sqrt{2(n-1)}$ . Więc dla  $n = 7$  i  $n = 9$  otrzymuje się odpowiednio 0,29% i 0,24%.

Wartości centrum grupowania danych  $med$  i  $\bar{x}_5^*$  oraz ich niepewności  $s(x_9)$  i  $s_5^*$ , które wyznaczono obiema metodami odpornościowymi, są bardzo zbliżone, gdyż oceniano je z danych tego samego badania. Dla próbki o 9 danych dokładność oceny niepewności  $u_A$  wynosi 24%. Zbieżność wyników uzyskanych obiema metodami odpornościowymi można w praktyce traktować jako zupełną, ale tylko dla danych tego przykładu. Estymator  $s_5^* = 0,315 > s$  dla metody Algorytm-A jest najbliższy dyspersji  $\sigma$  populacji możliwych wyników badań.

W podejściu klasycznym, po usunięciu dwu krańcowych wyników badań  $x_1$  i  $x_9$  otrzymano najmniejszą wartość odchylenia standardowego  $s$ . Odbiega ona jednak od rzeczywistości, gdyż wyniki pomiaru wyidealizowano przyjmując założenie, że wykorzystuje się tylko te dane eksperymentalne, które są blisko siebie i mogą należeć do populacji generalnej o rozkładzie normalnym. Zmniejszyła się też wiarygodność takiego wyniku. Dla wykazania, że teza ta jest poprawna rozpatrzone zostały jeszcze prostszy przykład 2.

## 6. Przykład obliczeniowy 2

Pobrano próbkę danych z populacji o rozkładzie normalnym o wartości średniej  $\mu = 75,238$  i odchyleniu standardowym  $\sigma = 13,475$ . Próbkę zawierała tylko cztery wartości:

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

Według kryterium Grubbsa, największa wartość: 102,1 została uznana jako odstająca, chociaż  $\mu + 2\sigma = 102,2$ . Metodą odchylenia medianowego  $MAD_s$  otrzymuje się estymator  $s^*(x_4) = 1,0095$ . Odbiega on znacznie od  $\sigma = 13,475$  dla populacji.

Natomiast stosując Algorytm-A otrzymuje się stabilną wartość  $s_{28}^* = 14,882$  estymatora odchylenia standardowego dopiero po 28 iteracjach. Nie stanowi to problemu przy stosowaniu współczesnej techniki obliczeniowej. Wartość  $s_{28}^* = 14,882$  jest dużo bliższa  $\sigma$  i bardziej wiarygodnie opisuje możliwy rozrzut wartości populacji obserwacji pomiarowych niż  $s^*$ . Mimo że dla próbki o  $n = 4$  względna niepewność odchylenia  $s_{28}^*$  wynosi tylko 42%, to metodą odpornościową Algorytm-A uzyskano istotnie lepszy wynik niż metodą  $MAD_s$ .

## 7. Podsumowanie

Zastosowana w obu przykładach metoda przeskalowanego odchylenia medianowego  $MAD_s$  jest bardzo prosta, ale nie daje wystarczająco poprawnych rezultatów, gdy wartość odstająca znacznie odbiega od grupy pozostałych danych. Uwidacznia się to szczególnie dla mało licznych próbek.

Metoda odpornościowa Hubera, czyli Algorytm-A [2, 3] jest iteracyjna i bardziej skomplikowana, ale łatwa do zautomatyzowania. Dzięki wprowadzeniu dla dużych odchyleń od estymatora menzurandu granic obszarów zewnętrznych o mniejszej czułości, wyznacza się oceny statystyczne odporne na dane odstające.

Analiza obu przykładów liczbowych wykazała większą przydatność metody odpornościowej Algorytm-A do wyznaczania parametrów statystycznych mało licznych próbek danych.

Przyjęto, że próbki te są pobrane z populacji generalnej o rozkładzie normalnym, ale zawierają nieliczne wyniki odstające od pozostałych, czyli outliery. Metoda ta pozwoliła lepiej niż w sposób klasyczny oraz metodą przeskalowanego odchylenia medianowego  $MAD_s$  ocenić wartość i dokładność wyniku procedury pomiarowej, którą kalibruje się za pomocą porównań międzylaboratoryjnych.

Występujące w przykładzie obliczeniowym 1. różnice między ocenami niepewności wyników pomiarów przeprowadzonych metodą badaną eksperymentalnie w poszczególnych laboratoriach i niepewnością wyniku całego międzylaboratoryjnego eksperymentu, można wykorzystać jako podstawę analizy rzeczywistego stanu organizacji procesu badań i sposobu prowadzenia działalności w tych laboratoriach, w których uzyskano wyniki odstające.

Metoda nadaje się do stosowania we wszystkich przypadkach, gdy w małych próbkach danych pomiarowych występują odstające odchylenia ich wartości, niezależnie czy powstały one wskutek rozrzutu wartości mierzonych w obiekcie badanym, czy przy przetwarzaniu sygnału w systemie pomiarowym.

W części 2 omówiona zostanie metoda odpornościowa Algorytm-A i przykład jej zastosowania w wyznaczaniu niepewności wyników eksperymentu międzylaboratoryjnego.

## Bibliografia

- Guide to the Expression of Uncertainty in Measurement (GUM), revised and corrected version of GUM 1995, BIPM JCGM 100:2008.
- PN-ISO 5725-5 Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów. Załącznik C.
- PN-ISO 13528:2005 Statistical methods for use in proficiency testing by inter-laboratory comparisons (IDT), Attachment C.
- Tukey J.W., *Exploratory Data Analysis*. Addison-Wesley, 1978.
- Frosch M.S., von Frese J., Bro R., *Robust methods for multivariate data analysis*. "Journal of Chemometrics", Vol. 19, Iss. 10, 2005, 549–563, DOI: 10.1002/cem.962.
- Doksum K.A., *Mathematical Statistics: Basic and Selected Topics*. Second updated edition, Pearson Prentice-Hall, 2007.
- Olive D.J., *Applied Robust Statistics*. Southern Illinois University Department of Mathematics. June 23, 2008.
- Huber P.J., Ronchetti E.M., *Robust Statistics*. Second edition. Wiley 2011.
- Buchczik D., *Procedura kalibracji przy wykorzystaniu metody najmniejszej mediany w przypadku modeli wielowymiarowych*. „Pomiary Automatyka Kontrola”, R. 54, Nr 5, 2008, 294–297.
- Daszykowski M., Kaczmarek K., Vander H.Y., Walczak B., *Robust statistics in data analysis – A review: Basic concepts*. "Chemometrics and Intelligent Laboratory Systems", Vol. 85, No. 2, 2007, 203–219, DOI: 10.1016/j.chemolab.2006.06.016.
- Stanimirova I., Serneels S., van Espen P.J., Walczak B., *How to construct a multiple regression model for data with missing elements and outlying objects*. "Analytica Chimica Acta", Vol. 581, No. 2, 2007, 324–332, DOI: 10.1016/j.aca.2006.08.014.
- Stanimirova I., Walczak B., *Classification of data with missing elements and outliers*. "Talanta", Vol. 76, No. 3, 2008, 602–609, DOI: 10.1016/j.talanta.2008.03.049.
- Piotrowski J., Kostyry K., *Wzorcowanie Aparatury Pomiarowej*. Wydawnictwo Naukowe PWN, Warszawa 2000, 2012.
- Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*. Polskie Wydawnictwo Ekonomiczne, 2000.
- Zięba A., *Analiza danych w naukach ścisłych i technice*. PWN Warszawa 2013.
- Randa J., *Proposal for KCRV and Degree of Equivalence for GTRF Key Comparisons*. NIST (2000) and update (2005).
- Volodarsky E., Warsza Z.L., Kosheva L., *Odporna ocena dokładności metod pomiarowych*. „Pomiary Automatyka Kontrola”, R. 58, Nr 4, 2012, 396–401.
- Volodarsky E., Warsza Z., Kosheva L., Palanychenko D., *Zastosowanie estymacji odpornej w badaniach bieglności laboratorium przy niewielkiej liczbie pomiarów*. „Pomiary Automatyka Kontrola”, R. 59, Nr 6, 2013, 554–557.
- Volodarsky E., Warsza Z.L., *Zastosowanie statystyki odpornościowej na przykładzie badań międzylaboratoryjnych*. „Przegląd Elektrotechniczny”, R. 89, Nr 11, 2013, 260–267.
- Volodarsky E., Warsza Z.L., Kosheva L., *System oceny i zapewnienia jakości badań bieglności laboratoriów przy akredytacji*. „Przemysł Chemiczny”, T. 93, Nr 8, 2014, 1252–1254, DOI: dx.medra.org/10.12916/przemchem.2014.1252.
- Volodarsky E., Warsza Z.L., Kosheva L., *Ocena precyzji procedury pomiarowej w badaniach międzylaboratoryjnych metodą odporną z wykorzystaniem „Algorytmu S”*. „Przemysł Chemiczny”, T. 94, Nr 6, 2015, 1008–1011, DOI: 10.15199/62.2015.6.30.
- Volodarsky E.T., Warsza Z.L., *Ocena precyzji badań międzylaboratoryjnych metodą odporną „S-algorytm”*. „Przegląd Elektrotechniczny”, R. 91, Nr 10, 2015, 192–196.
- Volodarsky E.T., Warsza Z.L., *Robust estimation in inter-laboratory measurements with small number of measurements*. "Measurement Automation Monitoring", Vol. 61, No. 4, 2015, 104–110.
- Volodarsky E., Warsza Z.L., Kosheva L., Idźkowski A., *Robust Algorithm S to assess precision of interlaboratory measurements*, "Measurement Automation Monitoring", Vol. 61, No. 4, 2015, 111–114.
- Volodarsky E.T., Warsza Z.L., *Application of two robust methods on the example of interlaboratory comparison*. [in:] "Advanced Mathematical and Computational Tools in Metrology and Testing X" (AMCTM X) (F. Pavese at all, Eds.), Vol. 10, Series on Advances in Mathematics for Applied Sciences vol. 86, 2015, 385–391, DOI: 10.1142/9789814678629\_0046.
- Volodarsky E., Warsza Z.L., *Examples of robust estimation with small number of measurements*. Progress in Automation, Robotics and Measuring Techniques. Series of Advances in Intelligent Systems and Computing, Vol. 352, 2015, 285–291, DOI: 10.1007/978-3-319-15835-8\_31.
- Volodarsky E., Warsza Z.L., Kosheva L., Idźkowski A., *Assessment of precision of the interlaboratory test data by using robust "Algorithm S"*. Advanced Mechatronics Solutions, Vol. 393. Series of Advances in Intelligent Systems and Computing. 2016, 87–96, DOI: 10.1007/978-3-319-23923-1\_13.
- Volodarsky E., Kosheva L., *Technicheskiye Aspekty Akredytacii Ispytatelnykh Laboratorii*. Winnicki Narodowy Uniwersytet Techniczny Ukrainy, Vinnica 2013.
- Warsza Z.L., *Metody rozszerzenia analizy niepewności pomiarów*. Monografie • Studia • Rozprawy. Oficyna Wydawnicza PIAP, Warszawa 2016.
- Bich W., Cox M., Michotte C., *Towards a new GUM - an update*. "Metrologia", Vol. 53, No. 4, 2016, S149–S159, IOP publishing | BIPM.



# Application of Robust Methods in Evaluation the Accuracy of Interlaboratory Measurements

## Part 1. Bases of Robust Statistics. Huber Method, i.e. Algorithm-A

**Abstract:** This two-part paper discusses the use of robust statistics to assess the value and uncertainty of measurand obtained from a sample of experimental data when some of these data differ significantly from the others, i.e. are outliers. The statistical parameters of the measurement result are determined by robust methods from all data, but influence of outliers is treated differently. For small sample sizes results are more reliable than obtained by classical methods with exclusions of outliers. This is illustrated by examples from the interlaboratory key comparisons. Part 1 discusses the basic principles of the robust statistics and the iterative robust method given by Huber, which is called Algorithm-A in ISO 5725-5. As illustration in the simulated numerical example, the uncertainty of some measurement method was estimated based on measurements of homogeneous object in several accredited laboratories. The mean uncertainty of this experiment is estimated by classic method for all data and with exclusion of outliers and by two robust methods: rescaled median deviation and by Algorithm-A. The result of last method is the most reliable.

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

### doc. dr inż. Zygmunt Lech Warszawa

zlw1936@gmail.com

Ukończył Wydział Elektryczny Politechniki Warszawskiej 1959, doktorat 1967, docent od 1970. Praca: Instytut Elektrotechniki 1958–1963 i 1994–1995, Politechnika Warszawska 1960–70, Politechnika Świętokrzyska 1970–1978 (organizator i dziekan Wydziału Transportu w Radomiu), Organizator i kierownik: Ośrodka Aparatury Pomiarowej w Instytucie Meteorologii i Gospodarki Wodnej 1978–1982 oraz Zakładu Automatykacji i Pomiarów w Instytucie Chemii Przemysłowej 1983–1992. Doradca Ministra Edukacji Narodowej 1992–1995, Politechnika Radomska 1983–2002. Obecnie główny specjalista w Przemysłowym Instytucie Automatyki i Pomiarów PIAP. Autor ponad 250 publikacji, 6 monografii, kilkadziesiąt prac badawczych i konstrukcyjnych, 11 patentów oraz promotor 2 doktorów. Prezes Polskiego Towarzystwa Metrologicznego. Członek stowarzyszenia PolSPAR oraz Akademii Metrologii Ukrainy.



### prof. dr nauk techn. Evgeniy T. Volodarsky

et-1@ukr.net

Praca: Narodowy Techniczny Uniwersytet Ukrainy – Politechnika Kijowska „KPI”, Katedra Automatykacji Badań Eksperymentalnych; Przewodniczący Podkomisji „Metrologia” w Państwowej Agencji Akredytacyjnej Ukrainy. Senior member of IEEE. Prezydent Ukraińskiej Akademii Metrologii. Zainteresowania naukowe: podstawy teoretyczne systemów pomiarowych do monitorowania i testowania. Autor kilku książek i ponad 300 publikacji.

