Eugenij T.VOLODARSKY[1], Zygmunt L.WARSZA[2]

Narodowy Uniwersytet Techniczny Ukrainy -Politechnika Kijowska (1), Przemysłowy Instytut Automatyki i Pomiarów (PIAP) Warszawa (2)

doi:10.15199/48.2015.10.40

Ocena precyzji badań międzylaboratoryjnych metodą odporną "S-algorytm"

Streszczenie. *Na przykładzie międzylaboratoryjnych badań porównawczych precyzji (niepewności) pewnej metody pomiarowej, omówiono jak na dokładność oceny statystycznej ich wyników wpływają outliery (dane odstające), jeśli pojawią się w tych wynikach. Rozpatrzono możliwość zastosowania odpornych metod oszacowania jako alternatywę do tradycyjnie stosowanego odrzucania danych odstających. Uwzględniają one wyniki wszystkich pomiarów wraz z outlierami. Pozwalają też na bardziej wiarygodne statystycznie oszacowanie rozkładu normalnego modelującego dane eksperymentalne, szczególnie dla małych próbek. Jako ilustrację, oszacowano wspólne odchylenie standardowe precyzji pewnej metody pomiarowej dla wyników badań tą metodą otrzymanych w 9-ciu laboratoriach. Odchylenie to, obliczone tradycyjnie bez odrzucenia outliera, było 1,5 razy większe niż z odrzuceniem, zaś dla metody odpornej "S- algorytm" jest bliskie mniejszej z obu wartości, lecz ma większą od niego wiarygodność.*

Abstract. *The influence of outliers in measurement results on the accuracy of resulting estimates is shown. Implementation of robust estimation methods is considered. These methods take into account all measurement results including outliers and the corresponding to them normal distribution could be choose better. Then it allows to provide a more reliable statistical estimates than classic methods with eliminating outliers, especially for samples of small volume. As the example the estimates of the common standard deviation of all numerical data from comparing tests of the measurement precision of some method in 9 labs are calculated by traditional methods and robust method "S-Algorithm". Results confirm the better efficiency of this robust method. (**Estimation of the precision in inter-laboratory control experiment by robust method S-algorithm**).*

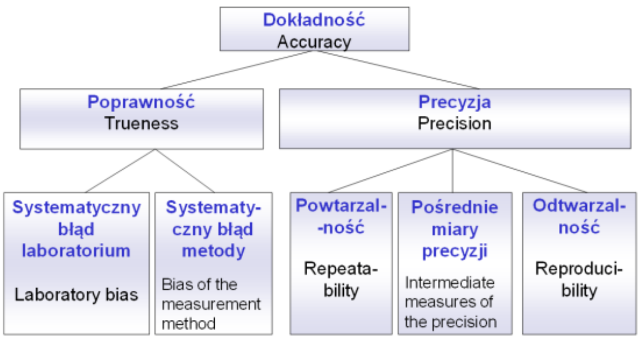
Słowa kluczowe: precyzja pomiarów, dane odstające, metody odporne, statystyczna wiarygodność ocen.
Keywords: precision of measurements, outliers, robust methods, the statistical reliability of estimates

1. Wprowadzenie

Ocenę jakości produktów, w tym elektrycznych, dokonuje się głównie na podstawie wyników wiarygodnych badań laboratoryjnych. Badania tych samych produktów mogą być dokonywane w różnych upoważnionych laboratoriach, nawet ulokowanych w różnych krajach. Porównywalność wyników uzyskuje się dzięki, prowadzeniu badań według tej samej metody i jej jednolitej procedury o znanych cechach dokładności. Badania prowadzi się w określony sposób i w określonych warunkach. Obejmują też one wielostopniowe przygotowanie homogenicznych próbek badanego obiektu, które trzeba uwzględniać w procesach certyfikacji i walidacji metody badań i jej procedury. Ocena i normalizacja parametrów dokładności badań powinna obejmować zarówno wpływy możliwych zmian warunków badania, jak i specyfikę organizacji eksperymentu w danym laboratorium. Zwykle w praktyce nie można rozwiązać tego zadania na drodze analitycznej, gdyż brakuje pełnego modelu matematycznego, który uwzględniałby wszystkie zależności mierzonych wielkości od właściwości badanego obiektu oraz warunków i sposobu jego działania. Zastępuje się go międzylaboratoryjnym eksperymentem, który polega na wykonywaniu identycznych badań próbek homogenicznych obiektów w wielu laboratoriach uczestniczących w tym eksperymencie. Z badań tych szacuje się wspólny wynik wraz z estymacją jego parametrów dokładności. Należy też uwzględnić występowania specyfiki organizacji badań w poszczególnych laboratoriach i kombinacje przewidywanych zmian wielkości wpływających w zakresach nie przekraczających wartości dopuszczalnych ujętych w odpowiednich normach. Międzylaboratoryjne porównania stanowią w rzeczywistości eksperymentalną realizację modelu fizycznego procedury badawczej dla kontrolowanej metody pomiarowej. Model ten powstaje na podstawie wyników pomiarów daną metodą uzyskanych wg tej samej procedury w laboratoriach specjalizujących się w danego rodzaju badaniach i o zbliżonym poziomie merytorycznym.

2. Model opisu wyników badań laboratoryjnych

W przepisach międzynarodowych dotyczących badań laboratoryjnych [1] - [3], do opisu dokładności metod i wyników pomiarów stosuje się terminologię wywodzącą się z analizy chemicznej. Ujmuje ona źródła powstawania niepewności wyniku. Przedstawiono to na rys 1.



Rys 1. Miary dokładności metod pomiarowych.

Jest to terminologia inna niż w Przewodniku Wyrażania Niepewności Pomiarów GUM [4]. I tak na przykład poprawność (ang. *trueness*), to heurystyczne oszacowanie szerokości przedziału, w którym może znajdować się wartość średnia B wyniku pomiarów wskutek oddziaływań systematycznych nieznanych co do wartości, a więc i nieusuwalnych przez poprawki. Według GUM jest to niepewność typu B. Natomiast precyzja (ang. *precision*) i niepewność typu A wg GUM jako wyznaczane metodami statystycznymi są swoimi odpowiednikami.

Wyniki eksperymentu między-laboratoryjnego opisuje się następującym modelem statystycznym [1]

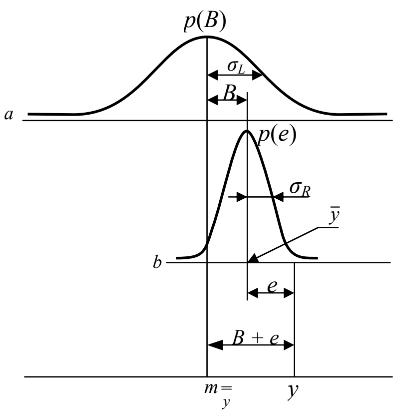
$$y = m_{\bar{y}} + B + e \tag{1}$$

gdzie: $m_{\bar{y}} = \mu + \delta$ - wartość średniej wspólnej dla wyników pomiarów we wszystkich laboratoriach; δ - składowa poprawności wyniku, tj. przesunięcie wartości średniej powstałe wskutek niedoskonałości procedury badań; B - składnik poprawności wyników laboratorium w warunkach odtwarzalności; e - składnik losowy błędu pomiarów w warunkach powtarzalności. Związek parametrów tego modelu statystycznego przedstawiono na rys 2.

Model ten pozwala sformalizować badaną procedurę. Dla badań międzylaboratoryjnych przeprowadzanych według atestowanej procedury kontrolowaną metodą pomiarową, wariancja wyniku σ_R^2 jest sumą geometryczną dwu składowych określających jej powtarzalność

(2) $$\sigma_R^2 = \sigma_L^2 + \sigma_r^2$$

Pierwszy składnik σ_L^2 - to wariancja rozrzutów wyników badań jednorodnych obiektów przy tej samej metodzie i procedurze zastosowanej w poszczególnych laboratoriach. Rozrzuty te wynikają z dopuszczalnych różnic w organizacji procesu badań przy tych samych warunkach pomiaru. Zaś składnik σ_r^2 jest wariancją rozrzutu wyników (wyraża ich powtarzalność) wskutek wpływu wielkości oddziałujących losowo na proces eksperymentu. Wielkości te mogą zmieniać się w granicach dopuszczanych przez odpowiednie normy.



Rys 2. Podstawowy model statystyczny wyniku pomiaru.

W między-laboratoryjnym eksperymencie przyjmuje się założenie, że wszystkie laboratoria zapewniają taką samą powtarzalność pomiarów. W praktyce jednak zdarza się, że niektóre z tych laboratoriów mają powtarzalność gorszą z powodów obiektywnych. Występuje to w szczególności, gdy przeprowadza się badania nową metodę i doświadczenie w jej realizacji nie jest jeszcze wystarczające we wszystkich laboratoriach uczestniczących w eksperymencie.

Całość statystycznego przetwarzania danych opiera się na założeniu, że rozrzut wartości pojedynczych pomiarów ma rozkładu normalny. Stanowi to podstawę decyzji podejmowanych za pomocą teorii wnioskowania statystycznego. W praktyce objętość danych surowych może być niewystarczająca do budowy wiarygodnych modeli parametrycznych, w pełni adekwatnych dla danego eksperymentu. W latach 60-tych ubiegłego wieku wiodący statystycy, na podstawie analizy rzeczywistych danych statystycznych wykorzystywanych w badaniach wykazali, że dane te często zawierają od 1 do 20%, tj. średnio około 10% wyników odstających (ang. *outliers*) zarówno o jawnej, jak i ukrytej postaci [2], [4]. Ponadto we wszystkich założeniach teoretycznych przyjmuje się, że wartości obserwacji pomiarowych pozyskuje się z badań eksperymentalnych wykonywanych w tych samych stałych warunkach. W praktyce pomiarowej może istnieć przestrzenno-czasowa zmienność zarówno warunków pomiaru, jak i parametrów badanego obiektu. Należy więc zaakceptować konkluzję Tukeya [5], że w rzeczywistości "normalność - to mit i normalności nie było i nigdy nie będzie". Przyczyną występowania w wynikach badań danych odstających są niesprawności użytych przyrządów pomiarowych, omyłki i nieprzestrzeganie zasad prowadzenia eksperymentu, błędy i omyłki w szacowaniu wyników, wpływy czynników zewnętrznych i inne przyczyny. Ponadto w próbkach o małej liczbie danych pomiarowych, nawet pobieranych z populacji o rozkładzie Gaussa pojawiają się przypadkowe wartości skośności oraz kurtozy różnej od 3, a nawet quasi-outliery [14].

W analizie wyników pomiarów w praktyce, przy ocenie wskaźników precyzji badań dość często występują dane leżące na granicy oddzielającej outliery. Nieuwzględnianie tych danych w obliczeniach, np. oparte na teście Q Cochrena [12] przy estymacji wariancji σ_r^2 oceniającej precyzję, lub według kryterium Grubbsa [12], [14] przy ocenie wariancji σ_L^2 badań międzylaboratoryjnych, może znacząco wpłynąć na dokładność i niezawodność statystycznej oceny precyzji badań. Jednakże klasyczne parametryczne oceny wyników badań eksperymentalnych są oparte na rozkładzie normalnym. Usadowiły się one tak mocno w praktyce, że rezygnowanie z nich nie byłoby właściwe. Metody parametryczne są dość proste i oczywiste intuicyjnie. Istnieje też dla nich gruntownie opracowana i kompletna teoria wnioskowania statystycznego. Wyłoniła się więc potrzeba, by "stary" model dopasować do nowych wezwań, czyli opracować takie metody estymacji, które pod pewnymi warunkami brały by pod uwagę "dane odstające", czyli pozwalały by wystarczająco dokładnie oceniać parametry pomiarów na podstawie wszystkich uzyskanych danych. Zaproponowano szereg takich metod i nazwano je odpornymi (ang. *robust*) [2] -[4], [11].

3. Ocena rozrzutu danych metodami odpornymi

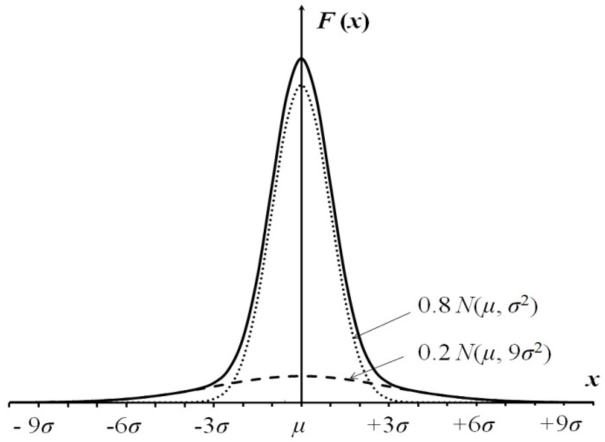
W praktyce najbardziej rozpowszechnione są takie procedury analizy statystycznej, które są optymalne przy założeniu normalności rozkładu danych. Są one jednak dość wrażliwe na niewielkie odchylenia parametrów rozkładu od tego założenia. Występuje to szczególnie przy estymacji wariancji. W rzeczywistości często mamy do czynienia z rozkładami, różniącymi się od idealnego rozkładu normalnego. Dlatego też ostatnio w przetwarzaniu danych doświadczalnych coraz szerzej stosuje się metody odporne [2]-[4], [13], [14]. Pod pojęciem odporność rozumie się w statystyce niewrażliwość wyznaczanych parametrów próbek pomiarowych, czyli stabilność ich wartości, na różne odchylenia i niejednorodności rozrzutu elementów. W ogólnym przypadku powstają one z przyczyn nieznanych.

Podstawowym modelem w metodach odpornych nie jest zmienna losowa o pojedynczym rozkładzie normalnym, ale model "mieszany". Dla różnych próbek danych tylko tę część można dość stabilnie modelować rozkładem normalnym. Zaś dolne fragmenty zboczy modelu rzeczywistego rozkładu, czyli jego "ogony", są bardziej rozciągnięte, niż dla rozkładu normalnego części środkowej i mało stabilne. Obserwacje pomiarowe w tych krańcowych obszarach występują rzadziej. Względem rozkładu normalnego dla części środkowej pojedyncze z nich, szczególnie dla małych próbek, mogą być rozpoznawane jako outliery, pomimo, że 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 odchyleń odstających od rozkładu normalnego części środkowej. Równocześnie jednak na "ogony" nakłada się pewne ograniczenia. Modeluje się je rozkładem normalnym o innych parametrach, bądź inną statystyką. Do opisu takiej sytuacji często stosuje się podejście zaproponowane przez Tukeya [5]. Założył on, że istnieje dużą liczba n danych pomiarowych, wśród których są przypadkowo zmieszane "dobre" i "złe" obserwacje x_i z populacji o średniej wartości μ. "Dobre" i "złe" obserwacje pobrane z tej populacji występują odpowiednio z prawdopodobieństwem $(1 - \varepsilon)$ lub ε, gdzie ε - mała liczba. 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 μ - rys 3. Odchylenie standardowe "złych" jest 3 razy większe niż "dobrych". Przy założeniu, że wszystkie wartości x_i są niezależne, rozkład ten jest opisany następującym wzorem:

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

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

Otrzymuje się rozkład wypadkowy o dolnych częściach zboczy nieco powyżej zboczy rozkładu normalnego dla środka próbki - rys 3.



Rys 3. Rozkład wypadkowy wg Tukey'a uwzględniający dane odstające.

Wśród opracowanych dotąd metod odpornych i algorytmów wdrożonych do ich stosowania szeroko rozprzestrzeniło się też inne podejście zaproponowane przez Hubera [3], również uznane już za klasyczne. Wprowadził on pewną wartość k zależną od stopnia "zanieczyszczenia" populacji generalnej. Wartość ta służy do określenia granic części środkowej histogramu danych pomiarowych. Część tę modeluje się podstawowym rozkładem normalnym [6], [8]. Obserwacje o wartościach w obszarach bocznych występują rzadziej i wg jednego z kryteriów zastosowanych dla podstawowego rozkładu, mogą być odstające. Skrajne obserwacje ściąga się na granice obszaru środkowego, ale dla nowego ich zbioru następuje zmiana odchylenia standardowego i zwężenie tego obszaru. Dostosowywanie należy powtórzyć i przebiega ono dalej iteracyjnie [6]. Zastosowanie metody Hubera do oceny wartości średniej wyników w badaniu procedur laboratoryjnych i biegłości laboratorium autorzy opisali w kilku czasopismach polskich [5] -[8]. Różnicę pomiędzy średnimi wyznaczonymi w badaniach między-laboratoryjnych wykorzystuje się dla oceny odtwarzalności wyniku. Podstawę użytych w tych pracach algorytmów stanowiła duża stabilność przedziału międzykwartylnego (rys 2) przy "zanieczyszczeniach" dochodzących nawet do 50%. Autorzy przedstawili też jednolite systemowe ujęcie procedur badania biegłości laboratorium [9], [10].

4. Wyznaczenie parametrów rozrzutu powtarzalności

Celem przeprowadzania omawianych tu porównawczych badań międzylaboratoryjnych jest oszacowanie i standaryzacja wariancji opisującej powtarzalność wyników uzyskiwanych badaną metodą pomiarową. Wyznacz się więc parametry łącznego rozkładu wariancji uzyskanych przez poszczególne laboratoria uczestniczące w tym wspólnym eksperymencie. Pozwala to, uwzględnić wpływ możliwych kombinacji zmian warunków w granicach dopuszczalnych dla badań tą metodą. W większości przypadków w praktyce, wskutek różnych ograniczeń (np. koszt lub czas trwania eksperymentu, badania niszczące) oddzielne szacunki trzeba określać na podstawie małej liczby obserwacji. Otrzymane wartości poszczególnych pomiarów zwykle są wówczas rozłożone asymetrycznie i znacznie mogą odbiegać od rozkładu Gaussa. Bliższy dla nich może być opis wg zależności (3). Ponadto, jeśli zastosować tradycyjny test Q Cochrana [12], to niektóre z wartości tych obserwacji zostały by uznane za dane odstające, czyli jako "outliery". Należało by więc je usunąć z dalszego przetwarzania statystycznego. Podejście takie można często zaakceptować, gdy poszukuje się wartości średniej. Zaś celem omawianego tu międzylaboratoryjnego eksperymentu jest ocena dopuszczalnego rozproszenia wyników laboratorium na podstawie otrzymanych danych doświadczalnych. Ocenę te wykorzystuje się następnie do standaryzacji powtarzalności procedury badanej w tym eksperymencie. Uzasadnione staje się tu zastosowanie metod odpornych. Są one oparte na wszystkich dostępnych danych doświadczalnych, a zatem dają bardziej wiarygodne oszacowanie statystyczne odpowiadające rzeczywistemu rozkładowi rozproszenia wyników pomiarów.

Do uzyskania stabilnego oszacowania wariancji powtarzalności wyników (czyli ich precyzji) najbardziej odpowiednia jest metoda odporna oparta na "S-algorytmie" [11]. Niezbędnym warunkiem realizacji tej metody jest równe zeru przesunięcie oceny średniej S_j^* dla wariancji wyników otrzymanych w laboratoriach, czyli jej poprawności otrzymanej metodą odporną. Ocena ta, w każdym j -tym kroku iteracji zbliża się do standardowego odchylenia rozkładu normalnego σ odpowiadającego rzeczywistym danym doświadczalnym. Do oszacowania przesunięcia wariancji danej próbki wprowadza się współczynnik dopasowania ξ. Powinien zachodzić warunek

(4) $$E\left\{\left(\xi S^*\right)^2\right\} = \sigma^2$$ [1)]

Jednocześnie odporna ocena odchylenia standardowego S^* powinna być stabilna z pewnym prawdopodobieństwem, tj. zawierać się w określonych granicach. W tym więc celu wprowadza się w postaci nierówności ograniczenie na maksymalne odchylenie $\eta\sigma$ od rozkładu preferowanego

(5) $$P\left\{S^* > \eta\sigma\right\} = \alpha$$

gdzie: σ - odchylenie standardowe populacji o rozkładzie normalnym, która odpowiadała by danym doświadczalnym przy założeniu ich "czystego" rozkładu normalnego; η - czynnik ograniczający, zależny od liczby danych w próbce; $P=(1\text{-}\alpha)$ - prawdopodobieństwo spełnienia warunku ograniczenia dopuszczalnego oszacowania odchylenia standardowego S^* dla oczekiwanego rozkładu normalnego.

Wartości współczynników, dopasowującego ξ i ograniczającego η, wyznacza się zwykle dla $\alpha = 0{,}1$. Wykorzystuje się tu fakt, że skumulowane krzywe rozkładów jednomodalnych przecinają się w pobliżu punktu o prawdopodobieństwie $0{,}9$. Należy zbadać analitycznie to podejście i ocenić jego skuteczność. Współczynnikowi ograniczającemu η odpowiada górna wartość $(1\text{-}\alpha)100\%$ rozkładu dla rozrzutu odchylenia standardowego S^*. Oceną

[1] w tym tekście gwiazdką * oznacza się oszacowanie wiarygodne.

tego rozrzutu może być odchylenie standardowe tego rozkładu. Dla liczby elementów n w próbce jest ono zależne od liczby stopni swobody $\nu=n$ -1.Uwzględnia się to mnożąc obie strony równania (4) przez ν i uzyskuje

(6) $$E\left\{\nu\left(\frac{S^*}{\sigma}\right)^2\right\}=\frac{\nu}{\xi^2} \quad \text{lub} \quad E\left\{\chi^2_{\nu\ робаст}\right\}=\frac{\nu}{\xi^2}$$

Według (5) ekwiwalentne (zastępcze) prawdopodobieństwo ograniczenia od góry zmiennej losowej χ^2 wynosi

(7) $$P\left\{\chi^2_{\nu\ robast} > \nu\cdot\eta^2\right\}=\alpha$$

Zakłada się, że "ogon" rozkładu χ^2, zawierający $\alpha\cdot 100\%$ możliwych wartości danej zmiennej losowej, daje się w przybliżeniu aproksymować rozkładem równomiernym o gęstości $\nu\cdot\eta^2$ (jego początkowa rzędna). Stąd

(8) $$\int_{\nu\cdot\eta^2}^{+\infty} x\cdot p_{\chi_\nu}(x)dx=\alpha\cdot\nu\cdot\eta^2$$

Wartość $\chi^2_{\nu,P=1-\alpha}$ otrzymuje się z tablic rozkładu Pearsona [15], a następnie znajduje się współczynnik ograniczający η, dla którego zachodzi warunek (4). Stąd

(9) $$\eta^2=\frac{\chi^2_{\nu,\ P=0,1}}{\nu}$$

Wychodząc z zależności $P\left(\chi^2_\omega\le\nu\cdot\eta^2\right)=1-\alpha$, dla głównej części rozkładu znajduje się z tablic wartość z dla danego prawdopodobieństwa. Stąd

(10) $$\xi=\frac{1}{\sqrt{z+0{,}1\eta^2}}$$

Jest to współczynnik dopasowania, przy którym dla wybranego współczynnika ograniczającego η, odporna ocena odchylenia standardowego nie będzie przesunięta.

W realizacji algorytmu iteracyjnego, zapewniającego oszacowanie odporne odchylenia standardowego, wartości *k* ustalającej lokalne odchylenie odpowiada

(11) $$\psi_j=\eta\cdot S_j^*$$

gdzie - S_j^* odporne oszacowanie odchylenia standardowego, obliczone dla *j*-tego kroku iteracji.

Z uporządkowanego szeregu ocen wariancji wyników pomiarów w wszystkich laboratoriach uczestniczących w eksperymencie, jako ocenę początkową odchylenia standardowego przewidywanej normalnej populacji wybiera się medianę, czyli

(12) $$S_0^{*2}=Me\{S_i^{*2}\},$$

gdzie $i = 1, .., n$ - kolejny numer wyrazu w uporządkowanym szeregu badań laboratoryjnych.

Następnie dokonuje się zmiany odchyleń standardowych części laboratoriów według zależności

(13) $$S_{ij}^*=\begin{cases}\psi_j,\ gdy\ S_i>\psi_j\\ S_i-w\,innych\ przypadkach\end{cases} \qquad j=0,1,...$$

Na podstawie wartości ψ_j w bieżącym kroku iteracji znajduje się zmodyfikowane wartości zbioru odchyleń S_{ij}^*, a następnie uściślone odporne oszacowanie odchylenia standardowego opisującego powtarzalność

(14) $$S_{j+1}^*=\xi\sqrt{\sum_{i=1}^{n}\frac{\left(S_{ij}^*\right)^2}{n}}$$

gdzie - S_{ij}^* odporne oszacowanie odchylenia standardowego w *j*-tym kroku dla *i*-tego laboratorium uczestniczącego we wspólnym eksperymencie (*n* - ogólna liczba laboratoriów).

Obliczoną wartość S_{j+1}^* wykorzystuje się do wyznaczenia nowej wartości granicznej ψ_{j+1}. Procedura iteracyjna trwa dotąd, aż wszystkie odchylenia standardowe dla pomiarów z laboratoriów uczestniczących we wspólnym eksperymencie znajdą się w granicach przedziałów bieżącego ograniczenia.

5. Przykład liczbowy

Do przeprowadzenia wspólnego eksperymentu wybrano dziewięć laboratoriów o nagromadzonym długoletnim pozytywnym doświadczeniu w tego rodzaju badaniach. W każdym z laboratoriów wykonano po 2 badania jednorodnych obiektów. Bezwzględne różnice wyników tych badań w *i*-tym laboratorium wynoszą

$$w_i=\left|x_{i1}-x_{i2}\right|,\ i=\overline{1,n},$$

gdzie: x_{i1}, x_{i2} - wyniki 2-u pomiarów w *i*-tym laboratorium.

Obliczone wartości rozbieżności dla wszystkich laboratoriów podano poniżej:

$w_1 = 0{,}28$; $w_2 = 0{,}49$; $w_3 = 0{,}40$; $w_4 = 0{,}00$; $w_5 = 0{,}35$; $w_6 = 1{,}98$; $w_7 = 0{,}80$; $w_8 = 0{,}32$; $w_9 = 0{,}95$.

Odchylenia standardowe dla różnicy dwu wyników z *i*-tego laboratorium wynosi $S_i^2=\frac{1}{2}\left|x_{i1}-x_{i2}\right|^2$.

Ocenę powtarzalności rozpatrzy się dla $\sum_{i=1}^{n}w_i^2$. Rozbieżność średniokwadratowa wyznaczona dla wszystkich danych

$$w_0=\sqrt{\frac{1}{9}\sum_{i=1}^{9}w_i^2}=0{,}827.$$

Analizując wartości bezwzględnych różnic w_i zauważa się, że wartość w_6 =1,98 znacznie różni się od pozostałych. Hipotezę o statystycznie odstającym wyniku w laboratorium nr 6 (różnica w_6=1,98) testuje się stosując się test Q Cochrana [15], wg którego $G_p=\frac{1{,}98^2}{6{,}1663}=0{,}636$. Z Tabeli rozkładu typu G [15] znajduje się wartości krytyczne: $G_{kr}(5\%)=0{,}638$ i $G_{kr}(10\%)=0{,}754$. Tak więc, G_P dla w_6 leży tuż poniżej granicy tego przedziału i w_6 należy traktować jako quasi-outlier. Według reguł podejścia tradycyjnego rozbieżność w_6= 1,98 należało by pominąć w dalszym statystycznym przetwarzaniu danych. Wówczas dla *n*=8 otrzyma się z pozoru "bardziej precyzyjne" odchylenie standardowe $w_0^{'}=0{,}530$. Jest ono znacznie mniejsze niż wartość w_0=0,827 obliczona ze wszystkich (*n*=9) danych źródłowych ($w_0^{'}=0{,}64w_0$). Z tych dwu oszacowań wynika więc, że wykluczenie ze źródłowych danych jednej tylko różnicy, leżącej na granicy oddzielającej outliery, ma istotny wpływ na wynik analizy, tj. na ocenę odchylenia standardowego dla rozrzutu powtarzalności testowanej procedury pomiarowej.

Jak już wspomniano, metody odporne wykorzystują wszystkie dane eksperymentalne, a więc również i dane odstające, ale w sposób zmodyfikowany. Ocenę powtarzalności odchylenia standardowego badań daną metodą, czyli jej precyzji w oparciu o wyniki uzyskane we wszystkich (n = 9) laboratoriach umożliwia metoda odporna o nazwie "S-algorytm" [1-część 5], [2]. Jej podstawowe zależności i sposób realizacji podano już powyżej.

Liczba stopni swobody $\nu = 1$. Wartości współczynników, dopasowującego ξ i ograniczającego η, według (10) i (9), wynoszą odpowiednio $\xi = 1097$ i $\eta = 1645$. Poniżej omówi się przebieg procedury iteracyjnej.

W pierwszym kroku iteracji wyznacza się $\psi_1 = w_5^* \cdot \eta = 0{,}40 \cdot 1{,}645 = 0{,}658 \approx 0{,}66$. Jest to dla tego kroku wartość ograniczająca. Z danych surowych należy zmodyfikować $w_{70}^*, w_{80}^*, w_{90}^*$, gdyż są one większe niż ψ_1. Nowy zbiór różnic w_{i1}^*, zapisano w kolumnie 1. Z ich wartości oblicza się nieprzesuniętą odporną ocenę odchylenia średniokwadratowego po kroku 1 iteracji $w_1^* = \xi \cdot \sqrt{\frac{1}{9}\sum_{i=1}^{9}\left(w_{i1}^*\right)^2} = 1{,}097 \cdot 0{,}47 = 0{,}52$. Otrzymuje się "nową" wartość ograniczającą $\psi_2 = 1{,}645 \cdot 0{,}52 \approx 0{,}86$. Podobnie postępuje się dalej. Już w czwartym kroku iteracji otrzymuje se odporną wartość $w_4^* = 0{,}68$ różniącą się od $w_3^* = 0{,}66$ o $\frac{0{,}02}{0{,}66} \cdot 100 \approx 3\%$. Jako końcowy wynik można przyjąć $w^* = 0{,}68$. Wartość ta znajduje się pomiędzy dwoma ocenami obliczonymi tradycyjnie, tj. wartością średnią $w_0 = 0{,}8270$ wyników precyzji ze wszystkich 9-ciu laboratoriów i $w_0^{'} = 0{,}530$ - dla 8-u tylko wyników po odrzuceniu wartości odstającej rozpoznanej jako quasi-outlier. Ostatecznie więc należy jako wspólne przyjąć odporne odchylenie standardowe $S_r = \frac{1}{\sqrt{2}} w^* = \frac{1}{\sqrt{2}} 0{,}68 = 0{,}48$. To oszacowanie odporne precyzji metody pomiarowej, zbadanej w tym eksperymencie międzylaboratoryjnym, jest bardziej wiarygodne statystycznie niż wyznaczone tradycyjne z wyników w 8-u laboratoriach po odrzuceniu odstającego wyniku w 9-tym laboratorium jako outliera, gdyż uzyskano je z wyników badań we wszystkich 9-ciu laboratoriach. Stosunek względnych niepewności obu obliczonych precyzji, wyznaczony po zastosowaniu wzoru z GUM [4 - dodatek E1], wynosi ok. $\sqrt{7/8} = 0{,}935$.

Oszacowanie precyzji dla wszystkich danych tego przykładu metodą odporną Algorytm-S jest więc o 6,5% dokładniejsze, niż obliczone klasycznie, tj. po odrzuceniu jednego outliera.

6. Podsumowanie

Omówiono pokrótce sposób standaryzacji oceny precyzji pewnej metody pomiarowej w przypadku, gdy pełny opisujący ją model nie jest znany. Przeprowadza się wówczas badania obiektów jednorodnych wg tej samej procedury w grupie laboratoriów o podobnych kompetencjach. Można wówczas założyć, że rozrzut wyników badań pomiędzy laboratoriami modeluje zmienna losowa o rozkładzie normalnym i na tej podstawie tworzy model statystyczny. Precyzję takiego międzylaboratoryjnego eksperymentu wyznacza sie ze zbioru precyzji poszczególnych laboratoriów. Jednakże w praktyce mogą wśród nich wystąpić odstające wyniki badań, czyli outliery. Odrzucanie ich z dalszego przetwarzania dla małej liczby danych pozyskanych eksperymentalnie istotnie zmniejsza wiarygodność oceny. W takich przypadkach powinno się stosować statystyczne metody odporne. Ilustruje to podany przykład numeryczny. Wartość standardowego odchylenia wyniku badań w jednym z 9-ciu laboratoriów zauważalnie odstawała od pozostałych, czyli była outlierem, tj. stanowiła "zanieczyszczenie". Dokonano ocen precyzji w sposób tradycyjny, bez i z odrzuceniem outliera oraz metodą odporną o nazwie "S-algorytm" [1], [2].

W metodzie tej wykorzystano wszystkie uzyskane dane doświadczalne, czyli bez odrzucenia outliera. Otrzymana ocena wspólnego dla wszystkich wyników odchylenia standardowego jest zbliżona do oceny tradycyjnej z odrzuceniem outliera, lecz nieco bardziej wiarygodną statystycznie.

Wniosek końcowy: Ocenę powtarzalności wyników badań kontrolowaną metodą pomiarową prowadzonych wg określonej procedury, czyli ocenę nominalnej precyzji (niepewności) tych badań, uzyskuje się na podstawie wyników uzyskanych przez grupę laboratoriów, czyli w wyniku eksperymentu między-laboratoryjnego. Jeśli zbiór wartości precyzji poszczególnych laboratoriów otrzymanych w tym eksperymencie jest niejednorodny, tj. dane te są "zanieczyszczone" outlierem, to ocenę wspólnej precyzji należy szacować metodą odporną "S-algorytm". Daje ona bardziej wiarygodną ocenę niż metody tradycyjne.

LITERATURA

1. PN-ISO 5725-1...-6: 2002: Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów. Części: 1. Zasady i definicje; 2. Basic method for the determination of repeatability and reproducibility of a standard measurement method.; ... 5. Alternative methods for the determination of the precision ...; 6. Stosowanie w praktyce wartości określających dokładność.
2. ISO 13528: (2005) Statistical methods for use in proficiency testing by interlaboratory comparisons (IDT), attachment C2
3. ISO/TS 21748: (2010). Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.
4. Guide to the Expression of Uncertainty in Measurement.GUM. First ed. 1993 ISO Switzerland, last corrected ed. JCGM BIPM (2008); + polskie tłum. wydania 1995: Wyrażanie niepewności pomiarów. Przewodnik. GUM Alfavero (1999, 2002)
5. Tukey J.W.: Exploratory Data Analysis. Addison-Wesley (1978)
6 Huber P.J. Robust Statistics, J. Wiley & Sons, New York... 1981
7. Willinik R.: *What is robustness in data analysis?* Metrologia 45 (2008), 442-447.
8. Volodarsky E.T., Koshevaya L.A., Warsza Z.: *Niepewność jako miara poziomu zaufania do wyników niektórych procedur doświadczalnych.* PAK (Pomiary Automatyka Kontrola) vol. 57, nr 5 (2011) s. 483-486
9. Volodarsky E.T., Warsza Z.L., Koshevaya L. A.: *Odporna ocena dokładności metod pomiarowych.* PAK vol. 58, n.4/2012 s. 396 -401
10. Volodarsky E.T., Warsza Z. L., Koshevaya L.A., Palanyczenko D.: *Zastosowanie estymacji odpornej w badaniach biegłości laboratorium przy niewielkiej liczbie pomiarów.* Pomiary Automatyka Kontrola vol.59 nr 6 2013 s. 554 -557
11. Volodarski E. T, Warsza Z. L., *Zastosowanie odpornościowej statystyki na przykładzie badań międzylaboratoryjnych.* Przegląd Elektrotechniczny - El. Review 11' 2013 s. 260 - 267.
12. Volodarsky E. T., Warsza Z. L., Koshevaya L.A., *System oceny i zapewnienia jakości badań biegłości laboratoriów przy ich akredytacji.* Przemysł Chemiczny vol. 93 n.8 (2014) s. 1252 -54
13. Volodarsky E. T., Warsza Z. L., Koshevaya L.A., *System oceny statystycznej w badaniu biegłości laboratoriów badawczych.* PAK (Pomiary Automatyka Kontrola) 2014 nr 10, s. 816 -821
14. Warsza Z. L., Korczyński J., *Statystyki skośności i kurtozy małych próbek pomiarowych z populacji o rozkładzie normalnym i kilku innych.* PAK vol. 60 n.12 (2014) s.1119 -23
15. Zieliński R.. Tablice statystyczne", PWN, Warszawa 1972
16. Piotrowski J.(†), Kostyrko K.: Wzorcowanie aparatury pomiarowej. Nowe wydanie. PWN Warszawa 2012
17. Zięba A. Analiza danych w naukach ścisłych i technice. Wydawnictwo Naukowe PWN Warszawa 2013

***Autorzy:** prof. dr hab. inż. Evgeniy T. VOLODARSKY, Uniwersytet Narodowy Ukrainy " Politechnika Kijowska", email: vet-1@ukr.net Doc. (em.) dr inż. Zygmunt L. WARSZA,* Przemysłowy Instytut Automatyki i Pomiarów PIAP, 02 486 Warszawa Al.Jerozolimskie 202, email: zlw@op.pl