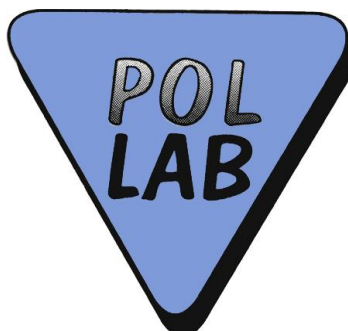


Klub Polskich Laboratoriów Badawczych POLLAB



**Wytyczne dotyczące oceny wyników
badań biegłości i porównań międzylaboratoryjnych
z udziałem poniżej trzydziestu uczestników**

W niniejszych Wytycznych zastosowano odniesienia do normy PN-EN ISO/IEC 17043:2011 obowiązującej organizatorów i uczestników badań biegłości i porównań międzylaboratoryjnych (PT/ILC).

Klub POLLAB wyraża podziękowanie dr inż. Sabinie Żebrowskiej-Łucyk za wkład pracy włożony w opracowanie niniejszych Wytycznych.

Prezes Klubu POLLAB – Andrzej Brzyski

1. Wprowadzenie

Za podstawę do przygotowania niniejszego opracowania, zwanego dalej w skrócie *Wytycznymi*, posłużyły przede wszystkim normy: PN-EN ISO/IEC 17043:2011 [1] i ISO 13528:2015 [2]. Uwzględnione zostały także niektóre zalecenia, wskazówki i sugestie zawarte w PN-ISO 5725:2002 [3] oraz w innych dokumentach i publikacjach wymienionych w Bibliografii [4 -13].

Dokumenty te opisują wiele możliwości planowania programów badania biegłości i statystycznej analizy wyników, ale nie przedstawiają sztywnych i jednoznacznych reguł postępowania. Szerokie pole decyzyjne pozostawione organizatorom programów porównań międzylaboratoryjnych oznacza, że powinni oni nie tylko dysponować bardzo dużą wiedzą merytoryczną w danej dyscyplinie, ale także dobrze rozumieć założenia i specyfikę wybranych metod statystycznych.

Warunkiem miarodajnej oceny kompetencji uczestników i sformułowania użytecznych konkluzji jest zastosowanie odpowiednio dobranych metod statystycznych na wszystkich etapach realizacji programu - począwszy od scharakteryzowania właściwości obiektów badań rozsyłanych do uczestników, poprzez analizę wyników pomiarów, obliczanie wskaźników aż do interpretacji i komentarzy. Jedną z kwestii kluczowych, która ma wpływ na kształt planu badań i sposób analizy wyników, jest liczba uczestników programu. Niektóre narzędzia statystyczne przewidziane do programów z udziałem wielu laboratoriów nie nadają się do stosowania w rundach, w których bierze udział niewielu uczestników. Niniejsze *Wytyczne* adresowane są głównie do organizatorów programów z udziałem do 30 uczestników.

Wytyczne dotyczą przede wszystkim programów badania biegłości (PT), ale mogą być przydatne również w innych porównaniach międzylaboratoryjnych (ILC), jeżeli służą one do oceny zgodności wyników uzyskanych przez poszczególnych uczestników programu podczas niezależnych badań tego samego (lub takiego samego) obiektu.

Symbole stosowane w tekście *Wytycznych* są zgodne z normą ISO 13528:2015 i mają następujące znaczenie:

x_{pt} - wartość przypisana;

$u(x_{pt})$ - niepewność standardowa wartości przypisanej;

$U(x_{pt})$ - niepewność rozszerzona wartości przypisanej;

δ_E - błąd maksymalny dopuszczalny;

σ_{pt} - odchylenie standardowe do badania biegłości.

Wytyczne są adresowane do osób odpowiedzialnych za opracowanie planu programu porównań, jego koordynację, weryfikację oraz zatwierdzenie. Osoby te, odpowiednio do sprawowanej funkcji, a także kompetencji zawodowych i wiedzy statystycznej, wpływają na plan programu PT/ILC, a także na ocenę wyników uzyskanych przez poszczególnych uczestników, podsumowanie efektów danej rundy programu i sformułowanie wniosków dotyczących organizacji przyszłych rund.

Zarówno osoby zaangażowane w organizację programów PT/ILC jak i ich uczestnicy powinni pamiętać, że choć najbardziej oczywistym celem takich programów jest weryfikacja kompetencji uczestniczących laboratoriów, to równie ważne jest aby służyły one doskonaleniu metod badawczych, wiedzy i umiejętności personelu laboratoryjnego oraz identyfikowaniu problemów w laboratoriach. Oczekuje się [1, 2], że programy PT/ILC spełniają też funkcje edukacyjne, zatem opracowanie wyników powinno być wnikliwe, rzetelne i wieloaspektowe, a komentarze zamieszczone w sprawozdaniu końcowym powinny ułatwiać laboratorium uzyskiwanie wyników o odpowiedniej dokładności.

Występujący w dalszej części tekstu termin *organizator programu badania biegłości* nie ma tutaj znaczenia prawnego, lecz odnosi się również do osób, którym Klub POLLAB przekazał określone uprawnienia i obowiązki (np. koordynator programu, weryfikator).

2. Liczba uczestników rundy programu badania biegłości

W bardzo wielu porównaniach międzylaboratoryjnych obserwuje się znaczący udział wyników odstających (nierzadko powyżej 20%). Z tego powodu, jeżeli wartości przypisane x_{pt} albo odchylenia standardowe do badania biegłości σ_{pt} , albo obie te wielkości wyznacza się na podstawie danej rundy, to plan programu PT musi zawierać opis procedury postępowania z danymi istotnie odbiegającymi od pozostałych.

Możliwe są dwa podejścia:

Podejście 1. Wykorzystanie wszystkich wyników dostarczonych przez uczestników i zastosowanie takich metod przetwarzania danych, które są niepodatne na obecność nawet dużego odsetka wyników odstających i dlatego zwane są odpornymi (lub odpornościowymi, ang. *robust methods*). **To podejście jest w programach PT/ILC preferowane.**

Podejście 2. Użycie testów statystycznych, które umożliwiają wskazanie w zbiorze wyników dostarczonych przez uczestników danych odstających i pominięcie tych danych przy

wyznaczaniu wartości przypisanej x_{pt} oraz odchylenia standardowego do badania biegłości σ_{pt} . Podejście to można stosować, gdy spodziewana frakcja wyników odstających jest znikoma.

W celu uniknięcia niejasności dotyczących liczby uczestników programu jest celowe wprowadzenie dwóch symboli:

p_0 – liczba uczestników rundy programu, którzy przedstawili wyniki pomiaru badanej cechy obiektu i podlegają ocenie,

p – liczba uczestników rundy programu, których wyniki zostały uwzględnione przy wyznaczaniu wartości x_{pt} i σ_{pt} .

Oczywiście, zawsze zachodzi relacja $p \leq p_0$.

Rozróżnienie liczb p_0 i p jest potrzebne nie tylko w związku z usuwaniem wyników odstających, ale również w przypadkach, gdy organizator badań biegłości zapowiedział w planie programu, że do obliczenia wartości x_{pt} i σ_{pt} wykorzysta jedynie wyniki laboratoriów o określonym statusie akredytacyjnym lub tylko od tych uczestników, którzy uzyskali zadowalające wyniki w poprzednich rundach. Wtedy p może być mniejsze od p_0 , nawet jeżeli zastosowano metody obliczeniowe odporne na wartości odstające. W normie [2] symbol p_0 nie występuje. Czytając ten dokument, można na ogół wywnioskować z kontekstu, kiedy symbol p oznacza liczbę wszystkich uczestników rundy, a kiedy tylko liczebność określonego podzbioru uczestników.

Programy organizowane przez Klub POLLAB skupiają zwykle poniżej 30 uczestników, a nierzadko zaledwie kilku. Na potrzeby niniejszego opracowania przyjęto następujące umowne określenia liczebności danych wykorzystywanych do wyznaczenia dla badanej cechy wartości uzgodnionych:

- $20 \leq p < 30$ – liczebność niewielka,
- $8 \leq p < 20$ – liczebność mała,
- $p < 8$ – liczebność bardzo mała.

Wytyczne odnoszą się m.in. do trzech ważnych etapów tworzenia planu programu badania biegłości. Są to:

- ustalenie dla obiektu badań wartości przypisanej x_{pt} i jej niepewności $u(x_{pt})$ z uwzględnieniem niejednorodności i niestabilności obiektów badań przekazanych uczestnikom (pkt. 3),
- dobór odpowiednich wskaźników do oceny wyników laboratoriów (pkt. 4),
- ustalenie sposobu pozyskania wartości potrzebnych do wyznaczenia wskaźników, np. δ_E bądź σ_{pt} (pkt. 5).

Jeżeli wartość przypisana i jej niepewność, a także inne wartości potrzebne do obliczenia wskaźników ocen, są wyznaczone takimi metodami, że nie zależą od wyników dostarczonych przez uczestników, to liczba p może być bardzo mała, dopuszcza się nawet $p = 2$.

Jeżeli wartość przypisana x_{pt} ma być wyznaczona na podstawie wyników bieżącej rundy, to ustalając minimalną liczbę p należy wziąć pod uwagę możliwość spełnienia warunków podanych w punkcie 3.

Jeżeli odchylenie standardowe do badania biegłości σ_{pt} ma być wyznaczone na podstawie wyników bieżącej rundy, to powinien być spełniony warunek $p \geq 20$ (wykluczona jest liczebność bardzo mała i mała).

3. Wyznaczenie wartości przypisanej x_{pt} i jej niepewności $u(x_{pt})$

3.1. Metody wyznaczenia wartości przypisanej

Organizator programu badania biegłości jest zobowiązany przedstawić w planie programu metodę wyznaczenia wartości przypisanej x_{pt} . Jeżeli organizator wyznaczył wartość przypisaną przy użyciu metod niezależnych od wyników, jakie podadzą uczestnicy planowanej rundy programu, to powinien on podać w planie programu również wartość niepewności $u(x_{pt})$. W przeciwnym razie powinien przedstawić sposób, w jaki obliczy niepewność $u(x_{pt})$ po otrzymaniu wyników od uczestników.

Norma ISO 13528:2015 wyszczególnia 5 metod wyznaczenia wartości przypisanej:

- I. Wytworzenie próbek modelowych, przez zmieszanie składników w znanych i ściśle określonych proporcjach (ich udział jest wtedy dokładnie określony)
- II. Zastosowanie certyfikowanego materiału odniesienia
- III. Wynik uzyskany metodą o wysokiej dokładności w laboratorium wybranym przez organizatora
- IV. Wartość uzgodniona przez laboratoria eksperckie
- V. Wartość wyznaczona na podstawie wyników uczestników

Cztery pierwsze metody (I, II, III, IV) można stosować niezależnie od liczby uczestników rundy.

Metoda V, oparta na wynikach pomiarów podanych przez uczestników, jest najmniej zalecana. Organizator wybierający tę metodę, powinien obliczyć minimalną liczbę uczestników p , jaka pozwoli spełnić zalecenia dotyczące niepewności $u(x_{pt})$ określone za pomocą nierówności (1) i (2) podanych w punkcie 3.2. Wpływ liczby uczestników p na wartość niepewności $u(x_{pt})$ wyrażają wzory (6) i (7) zamieszczone w punkcie 3.4.

3.2. Dopuszczalna niepewność wartości przypisanej

Stosownie do zaleceń normy ISO 13528:2015, niepewność wartości przypisanej można uznać za pomijalną i nie trzeba jej włączać do interpretacji wyników danej rundy, jeżeli jest spełniony warunek:

$$u(x_{pt}) < 0,1\delta_E \quad (1)$$

gdzie:

$u(x_{pt})$ - niepewność standardowa wartości przypisanej,

δ_E - błąd maksymalny dopuszczalny; kryterium oceny wskaźnika D (pkt. 4).

Błędowi maksymalnemu dopuszczalnemu δ_E można przypisać wartość, jaką organ regulujący, jednostka akredytacyjna lub eksperci techniczni organizatora PT uważają za uzasadnioną dla celu danej rundy.

Warunek (1) można przedstawić w innej postaci:

$$u(x_{pt}) < 0,3\sigma_{pt} \quad (2)$$

gdzie:

$u(x_{pt})$ - niepewność standardowa wartości przypisanej,

σ_{pt} - odchylenie standardowe do oceny biegłości,

Warunki (1) i (2) są równoważne, jeżeli przyjmuje się założenie, że

$$\sigma_{pt} = \frac{1}{3}\delta_E \quad (3)$$

Jeżeli zastosowana przez organizatora metoda wyznaczenia wartości przypisanej nie spełnia wymagania (1) lub (2), to należy wybrać jeden z poniższych wariantów postępowania.

- Poszukać innej metody wyznaczenia x_{pt} , która wymagania to spełni.
- Uwzględnić niepewność wartości przypisanej $u(x_{pt})$ podczas interpretacji wyników badania biegłości. Wyznaczając wskaźniki D , $D\%$ lub P_A , należy zastąpić błąd δ_E błędem δ_E' (definicja symbolu opisana wzorem 18), zamiast *wskaźnika z* należy stosować *wskaźnik z'* bądź posługiwać się tylko wskaźnikami ζ lub E_n .
- Jeżeli x_{pt} jest wartością obliczoną z wyników uczestników i duża wartość $u(x_{pt})$ jest spowodowana różnicami pomiędzy wykrytymi podpopulacjami uczestników, to może być zasadne rozdzielenie wyników na grupy i wyznaczenie niepewności osobno dla każdej grupy (np. uczestników stosujących różne metody pomiarowe).
- Poinformować uczestników rundy, że nie można zaniedbać $u(x_{pt})$, a to może wpłynąć na ocenę kompetencji laboratoriów.

Jeżeli działania a) - d) nie są możliwe, należy powiadomić uczestników, że nie można wyznaczyć miarodajnej wartości x_{pt} , a więc nie można wyznaczyć żadnych wskaźników.

3.3. Niepewność $u(x_{pt})$ wartości przypisanej x_{pt} wyznaczonej niezależnie od wyników rundy

Podstawą do wyznaczenia składowych niepewności jest równanie pomiaru:

$$x_{pt} = x_{char} + \delta_{hom} + \delta_{trans} + \delta_{stab} \quad (4)$$

Przyjmuje się, że wszystkie składniki prawej strony równania (4) są od siebie niezależne, wobec czego niepewność standardową wartości przypisanej opisuje wzór:

$$u(x_{pt}) = \sqrt{u_{char}^2 + u_{hom}^2 + u_{trans}^2 + u_{stab}^2} \quad (5)$$

x_{char} - wartość badanej właściwości przyjęta za wartość przypisaną;

δ_{hom} , δ_{trans} , δ_{stab} - składowe błędy związane z niejednorodnością obiektów PT, zmianami obiektów podczas transportu, niestabilnością w czasie trwania danej rundy programu;

u_{char} - niepewność standardowa wnoszona przez wyznaczenie x_{char} ;

u_{hom} , u_{trans} , u_{stab} - składowe niepewności wnoszone przez δ_{hom} , δ_{trans} , δ_{stab} .

Jeżeli organizator programu badania biegłości jest w stanie zapewnić, że zmiany spowodowane niestabilnością δ_{stab} lub powstałe w transporcie δ_{trans} są pomijalne w porównaniu do σ_{pt} , to można przyjąć, że $u_{stab} = 0$ i $u_{trans} = 0$.

Niepewność u_{char} należy wyznaczyć zgodnie z zaleceniami *Przewodnika* [4].

3.4. Wyznaczenie wartości przypisanej x_{pt} oraz niepewności $u(x_{pt})$ na podstawie wyników otrzymanych od uczestników

Jeżeli wartość przypisana x_{pt} , wobec niemożności zastosowania innych metod, musi być wyznaczona na podstawie wyników danej rundy, to zaleca się użycie jednego spośród estymatorów położenia rozkładu odpornych na wartości odstające. Może to być:

- mediana,
- średnia x^* obliczona według iteracyjnego algorytmu A,
- estymator Hampela.

Dopuszcza się również stosowanie miary nieodpornej, jaką jest

- średnia arytmetyczna,

po uprzednim odrzuceniu wartości odstających wykrytych za pomocą uznanego testu statystycznego (np. Grubbsa).

Niezależnie od zastosowanego sposobu obliczeń, metoda ustalenia wartości przypisanej oparta na danych od uczestników rundy ma istotne wady:

- a) Wyniki uczestników mogą być niewystarczająco zgodne.
- b) Wartość uzgodniona x_{pt} może zawierać nieznanne obciążenie, które nie jest uwzględnione w $u(x_{pt})$ z powodu stosowania przez wielu uczestników niewłaściwej procedury badawczej.
- c) Wartość uzgodniona x_{pt} może być obciążona z powodu obciążenia metody obliczenia wartości przypisanej.
- d) Wykazanie spójności metrologicznej wartości uzgodnionej jest trudne; organizator programu badania biegłości musiałby posiadać pełną informację o wzorcach używanych przez uczestników i innych istotnych warunkach pomiarów.

W programach z udziałem niewielu uczestników, dochodzi do tego bardzo poważny problem: nadmierna niepewność wartości przypisanej, wynikająca z ograniczonego zbioru danych, które posłużyły do obliczenia x_{pt} .

Jeżeli wyniki pomiaru mają rozkład zbliżony do normalnego i wartość przypisana jest wyznaczona jedną z metod odpornych, to jej niepewność $u(x_{pt})$ szacuje się według wzoru:

$$u(x_{pt}) = \frac{1,25}{\sqrt{p}} s^* \quad (6)$$

Oznacza to, że w przypadku przyjęcia $\sigma_{pt} = s^*$, warunek (2) może być spełniony tylko wtedy, jeżeli liczba danych p uwzględnionych w obliczeniach wynosi co najmniej 18.

Jeżeli wartość przypisaną wyznaczono jako średnią arytmetyczną, po odrzuceniu wartości odstających, to jej niepewność oblicza się ze wzoru:

$$u(x_{pt}) = \frac{1}{\sqrt{p}} s \quad (7)$$

Metoda obliczania wartości przypisanej jako średniej arytmetycznej jest uznawana za gorszą niż metody odporne z uwagi na wpływ wartości odstających, które w programach międzylaboratoryjnych należy brać uwagę, a które mogą pozostać niewykryte z powodu małej mocy testów statystycznych dla małych liczebności p . Jeżeli jednak doświadczenia organizatora badań pokazują, że ryzyko wystąpienia wartości odstających jest niewielkie, to w programach z udziałem małej i bardzo małej liczby uczestników można preferować średnią arytmetyczną, z racji mniejszej niepewności $u(x_{pt})$.

Ponieważ niepewność wartości przypisanej obliczonej na podstawie wyników rundy z udziałem małej liczby uczestników nie spełnia na ogół warunków (1) lub (2), należy dołożyć wszelkich starań, aby wartość przypisaną ustalić innymi metodami, a w razie konieczności postąpić zgodnie z wariantami b), c) lub d) w punkcie 3.2, a nawet odstąpić od oceny wyników, ograniczając sprawozdanie do prezentacji graficznej wyników otrzymanych od uczestników i ich opisowego skomentowania.

3.5. Porównanie wartości przypisanej x_{pt} z niezależną wartością odniesienia

Jeżeli x_{pt} wyznaczono metodą V (na podstawie wyników bieżącej rundy) i jest dostępne niezależne miarodajne oszacowanie x_{ref} (np. metodą I), to po każdej rundzie programu PT należy porównać x_{pt} z x_{ref} , wyznaczając ich różnicę x_{diff} .

$$x_{diff} = x_{ref} - x_{pt} \quad (8)$$

Niepewność standardowa różnicy x_{diff} wynosi:

$$u_{diff} = \sqrt{u^2(x_{ref}) + u^2(x_{pt})} \quad (9)$$

Jeżeli

$$x_{diff} > 2u_{diff} \quad (10)$$

to należy poszukać przyczyn niezgodności.

Podobnie, jeżeli x_{pt} wyznaczono stosując metody I - IV, to należy porównać średnią odporną x^* z danej rundy z wartością x_{pt} :

$$x_{diff} = x^* - x_{pt} \quad (11)$$

i sprawdzić czy nie zachodzi nierówność (10).

4. Wskaźniki zalecane do oceny wyników podanych przez laboratoria

4.1. Aspekty oceny

Zależnie od celu programu badania biegłości, różnicę pomiędzy wynikiem uczestnika a wartością przypisaną ($x_i - x_{pi}$) można ocenić, uwzględniając:

- a) kryteria zewnętrzne (wymagania),
- b) wyniki innych uczestników,
- c) deklarowaną przez uczestnika niepewność.

Podane przez uczestników wyniki badań należy przekształcić w unormowane wskaźniki rezultatów działania, aby przeprowadzić jednolitą ocenę i interpretację w całym zakresie wielkości mierzonych.

Zalecane obecnie wskaźniki oceny oraz wartości kryterialne [2] i nazwy ocen zestawiono w tabeli 1. Są one nieco inne niż podane we wcześniejszych normach. Istotną nowością jest wprowadzenie wskaźnika P_A . Ponadto zmieniono niektóre nazwy ocen, wprowadzono kosmetyczne zmiany w znakach nierówności liczb kryterialnych dla wskaźnika z , zamieniono nazwę *liczba* E_n na *wskaźnik* E_n , a w jego wzorze definicyjnym występuje teraz wartość przypisana zamiast wartości referencyjnej.

Tablica 1 Wskaźniki zalecane do oceny rezultatów uzyskanych przez uczestników

Symbol wskaźnika	Wzór	Interpretacja
D	$D_i = x_i - x_{pt}$	Wynik jest zadowalający ^{1/} jeżeli $ D_i < \delta_E$ lub jeżeli $ D_i < \delta'_E$.
D%	$D_i\% = 100 \frac{D_i}{x_{pt}} \%$	Kryteria analogiczne jak dla D_i
P_A	$P_{A_i} = 100 \frac{D_i}{\delta_E} \%$	Wynik jest odpowiedni jeżeli $ P_A < 100 \%$
z	$z_i = \frac{x_i - x_{pt}}{\sigma_{pt}}$	$ z_i \leq 2$ - wynik zadowalający $2 < z_i < 3$ – wynik wątpliwy (sygnał ostrzegawczy) ^{2/} $ z_i \geq 3$ - wynik niezadowalający ^{3/}
z'	$z'_i = \frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}}$	Kryteria liczbowe jak dla wskaźnika z
ζ	$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}}$	Kryteria jak dla wskaźnika z lub $ \zeta_i < k$ - wynik zadowalający k – stosowany współcz. rozszerzenia
E_n	$(E_n)_i = \frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}}$	$ (E_n)_i < 1,0$ - wynik zadowalający, o ile niepewności są poprawne oraz wartość $ x_i - x_{pt} $ jest mniejsza od wymaganej przez klientów

^{1/} in. dopuszczalny/dozwolony/akceptowalny (ang. *acceptable*); wg ISO 17043 taki wynik jest „zadowalający” (ang. *satisfactory*)

^{2/} ang. *warning signal*; wg ISO 17043 wynik jest „wątpliwy” (ang. *questionable*)

^{3/} in. niedopuszczalny/niedozwolony/nieakceptowalny / (ang. *unacceptable*);
wg ISO 17043 jest to wynik „niezadowolający” (ang. *unsatisfactory*)

4.2. Wskaźniki D , $D\%$ i P_A . Błąd maksymalny dopuszczalny δ_E (δ_E').

Wskaźniki D , $D\%$ oraz P_A służą do oceny zgodności wyniku uczestnika z kryterium zewnętrznym, jakim jest wartość błędu maksymalnego dopuszczalnego. Błąd ten oznaczony jest symbolem δ_E (3.2), gdy jest wyrażony w jednostkach wielkości mierzonej, a symbolem $\delta_E\%$ wtedy, gdy określa procent wartości wielkości mierzonej.

Wskaźnik D jest podstawową miarą stopnia dokładności wyniku [ISO 13528, 9.3.1]. Stanowi oszacowanie błędu pomiaru. Dla i -tego wyniku pomiaru x_i wskaźnik ten wynosi:

$$D_i = x_i - x_{pt} \quad (12)$$

Wartość przypisaną x_{pt} należy wyznaczyć jedną z metod wymienionych w p. 3.1 i przedstawionych szczegółowo w normie ISO 13528:2015.

Wynik uznaje się za zadowolający (dozwolony, ang. *acceptable*), tzn. ‘nie ma sygnału’, jeżeli

$$-\delta_E < D_i < \delta_E. \quad (13)$$

sposób ustalenia błędu maksymalnego dopuszczalnego δ_E podano w punkcie 3.2.

Zaletami wskaźnika D są intuicyjność oraz ściśle powiązane oceny wyniku z jego przydatnością do przewidywanego wykorzystania. Posługując się wskaźnikiem D , trudno jest jednak analizować sprawozdania obejmujące wiele wielkości mierzonych oraz zawierające zróżnicowane wartości kryterialne (np. zależne od poziomu badanej cechy).

Wskaźnik $D\%$ odnosi różnicę do wartości przypisanej, co ułatwia porównywanie wyników pomiaru różnych cech.

$$D_i\% = 100 \frac{D_i}{x_{pt}} \% \quad (14)$$

Wynik uznaje się za zadowolający, jeżeli

$$-\delta_E \% < D_i < \delta_E \%. \quad (15)$$

Wskaźnik P_A (*Percentage of Allowed Deviation* [ISO 13528, 9.3.6]) wyraża błąd D_i jako procent błędu dopuszczalnego.

$$P_{Ai} = 100 \frac{D_i}{\delta_E} \% \quad (16)$$

Wynik uznaje się za zadowolający (dozwolony), tzn. ‘nie ma sygnału’, jeżeli

$$-100\% < P_{Ai} < 100\%. \quad (17)$$

Wskaźnik P_A jest tak unormowany, że ma jednakowe wartości kryterialne dla wszystkich cech obiektu badań, niezależnie od ich wartości. Jest więc bardzo przydatny do porównywania

wyników badania różnych cech na wielu poziomach wielkości mierzonej. Otrzymane w kolejnych rundach wartości wskaźnika P_{Ai} można wykreślać na kartach (podobnie jak wartości wskaźnika z_i), co jest szczególnie użyteczne, gdy częstość rund PT jest duża i bada się wiele cech obiektu. Co ważne, wskaźnik ten nie ma interpretacji probabilistycznej, a więc ocena uczestnika nie zależy od wartości wyników innych uczestników rundy ani od rozkładu tych wyników.

Wartość bezwzględna wskaźnika $|P_{Ai}|$ pozwala jednolicie oceniać wyniki położone w danej odległości od wartości przypisanej, niezależnie od znaku D_i .

Kryteria oceny wskaźników D , $D\%$ oraz P_A można złagodzić przez zastąpienie błędu δ_E większą wartością - δ'_E , obliczoną według wzoru:

$$\delta'_E = \sqrt{\delta_E^2 + U^2(x_{pt})} \quad (18)$$

gdzie $U(x_{pt})$ niepewność rozszerzona wartości przypisanej; $k = 2$.

4.3. Wskaźniki z i z'

Wskaźnik z [ISO 13528, 9.4] jest w programach badania biegłości najbardziej rozpowszechnioną miarą oceny kompetencji laboratoriów. Do jego obliczenia konieczna jest znajomość dwóch parametrów charakteryzujących rozkład (teoretyczny lub obserwowany), którymi są: wartość przypisana x_{pt} oraz odchylenie standardowe do badania biegłości σ_{pt} .

$$z_i = \frac{x_i - x_{pt}}{\sigma_{pt}} \quad (19)$$

Koncepcja wykorzystania tego wskaźnika bazuje na założeniu, że wyniki poprawnie wykonanych pomiarów opisuje (niekiedy dopiero po przekształceniu) rozkład normalny $N(x_{pt}; \sigma_{pt})$ i że jego parametry x_{pt} oraz σ_{pt} zostały przez organizatora dokładnie określone (na podstawie rozważań teoretycznych lub dostatecznie rozległych badań empirycznych). Przy takim założeniu wskaźnik z ma rozkład normalny $N(0; 1)$, a co za tym idzie, można oczekiwać, że około 5% wartości z_i wykroczy poza zakres $[-2; 2]$ i tylko 0,3% znajdzie się poza zakresem $[-3; 3]$. Ponieważ prawdopodobieństwo przekroczenia przedziału $[-3; 3]$ jest bardzo małe, to również jest mało prawdopodobne, by sygnał do działania pojawił się przypadkowo. Przyjmuje się więc, że nierówność $|z| \geq 3$ jest przejawem jakiejś nieprawidłowości w procedurze badawczej uczestnika i należy ją zidentyfikować.

Rozumowanie powyższe jest podstawą sformułowania kryteriów oceny wyników od uczestników [ISO 13825:2015, 9.4.2]:

$$\begin{aligned} |z| \leq 2 & \quad - \text{wynik zadowolający;} \\ 2 < |z| < 3 & \quad - \text{wynik wątpliwy (sygnał ostrzegawczy);} \\ |z| \geq 3 & \quad - \text{wynik niezadowolający (sygnał do działania).} \end{aligned} \quad (20)$$

Organizator programu może przyjąć, że już przekroczenie wartości 2 generuje sygnał działania.

Norma [2] przedstawia, oprócz 5 metod wyznaczenia x_{pt} , również 5 metod ustalenia odchylenia standardowego do badania biegłości σ_{pt} (patrz punkt 5). Bardzo często (zbyt często) osoby odpowiedzialne za opracowanie planu badań biegłości wybierają podejście najłatwiejsze – obie te wartości obliczają na podstawie wyników dostarczonych przez uczestników w bieżącej rundzie programu. Podejście to jest najmniej zalecane, a jeżeli liczba uczestników p jest mniejsza od 20, to nie powinno ono być stosowane.

Występują wtedy trzy naruszenia założeń dotyczących wskaźnika z :

- niedokładne oszacowanie x_{pb} ,
- niedokładne oszacowanie σ_{pt} ,
- brak wymaganej niezależności pomiędzy wynikiem uczestnika x_{pt} oraz pomiędzy x_{pt} i σ_{pt} .

Niedostateczna liczba uczestników wpływa najsilniej (najbardziej niekorzystnie) na niedokładne oszacowanie σ_{pt} . „W przypadku mało licznych zbiorów danych nie zaleca się przyjmowania kryteriów oceny opartych na rozproszeniu wyników uczestników, z powodu bardzo dużej zmienności wszelkich estymat rozproszenia. Np. dla $p = 30$, można się spodziewać, że estymaty odchylenia standardowego σ_{pt} danych o rozkładzie normalnym zmieniają się po około 25% po obu stronach jego prawdziwej wartości (przy poziomie ufności 95%) [ISO 13825:2015, p. D.1.4.1]. Jeżeli p zmaleje do 20, to spodziewany przedział zmienności s wzrośnie do $0,68\sigma < s < 1,31\sigma$, czyli fluktuacje s mogą wtedy przekraczać 30% prawdziwej wartości σ . Dla $p = 7$ przedział ten wynosi $0,45\sigma < s < 1,55\sigma$ (a więc jest około dwukrotnie szerszy niż dla $p = 30$), a gdy $p = 3$, to $0,16\sigma < s < 1,92\sigma$, czyli z rundy na rundę s może się zmieniać o rząd wielkości! Zastosowanie zamiast klasycznego odchylenia standardowego z próby s jednej z miar odpornych (punkt 5) nie może zmniejszyć wpływu losowości wyników w rundzie.

Reasumując, nie wolno wyznaczać σ_{pt} na podstawie wyników rundy programu z udziałem bardzo małej bądź małej liczby p uczestników. Jeżeli $p \geq 20$, to dopuszcza się stosowanie tej metody, analogicznie jak przy projektowaniu kart kontrolnych.

Organizator badań biegłości powinien ustalić i podać uczestnikom zasady zaokrąglania wskaźnika z , biorąc pod uwagę liczbę cyfr znaczących w wyniku oraz w wartościach x_{pt} i σ_{pt} . Rzadko jest użyteczne podawanie więcej niż dwóch cyfr po przecinku.

Wskaźnik z' [ISO 13528, 9.5] jest efektem modyfikacji wzoru (19), polegającej na uwzględnieniu niepewności standardowej wartości przypisanej $u(x_{pt})$:

$$z'_i = \frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}} \quad (21)$$

Wskaźnik z' interpretuje się w taki sam sposób jak wskaźnik z , stosując te same wartości krytyczne 2,0 i 3,0.

Można go preferować w tych sytuacjach, gdy wartość przypisaną x_{pt} oraz/lub wartość σ_{pt} wyznacza się niezależnie od wyników uczestników. Jeżeli zalecenie (2) dotyczące $u(x_{pt})$ jest spełnione, to z i z' różnią się od siebie znikomo:

$$0,96 < z'/z < 1. \quad (22)$$

4.4. Wskaźnik ζ , zeta [ISO 13528:2015, 9.6]

Wskaźnik ζ zaleca się stosować wtedy, gdy celem programu badania biegłości jest ocena zdolności uczestnika do uzyskania wyników bliskich wartości przypisanej, zgodnie z deklarowaną niepewnością

$$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}} \quad (23)$$

$u(x_i)$ – oszacowanie przez i -tego uczestnika niepewności standardowej jego wyniku x_i ,

$u(x_{pt})$ – niepewność standardowa wartości przypisanej.

Wskaźnik ζ pozwala ocenić podany przez uczestnika wynik pomiaru wraz z niepewnością. Wartości wskaźnika ζ wykraczające poza przedział $[-2, 2]$ mogą oznaczać albo duże odchylenie x_i od x_{pt} , albo niedoszacowanie niepewności przez uczestnika, albo obie przyczyny równocześnie.

Można też przyjąć te same wartości krytyczne co dla z (2 i 3) i stosować wskaźniki z i ζ łącznie, aby łatwiej zidentyfikować problemy uczestników. Jeżeli uczestnik otrzymuje wskaźnik z , który stale przekracza wartość 3, to jest zasadne przeanalizowanie jego procedury badań krok po kroku, aby wprowadzić konieczne korekty. Jeżeli wskaźnik ζ tego uczestnika również stale przekracza wartość 3, to znaczy, że uczestnik nie uwzględnił istotnych składników niepewności. Jeśli zaś uczestnik otrzymuje $z \geq 3$ ale $\zeta < 2$, to może oznaczać, że poprawnie oszacował niepewność swoich wyników, jednak nie spełniają one oczekiwań programu badania biegłości.

4.5. Wskaźnik E_n [ISO 13528, 9.7]

Wskaźnik E_n jest użyteczny w programach badania biegłości, których celem jest ocena zdolności uczestnika do uzyskiwania wyników bliskich wartości przypisanej wewnątrz deklarowanej niepewności.

$$(E_n)_i = \frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}} \quad (24)$$

$U(x_i)$ – niepewność rozszerzona wyniku uczestnika,

$U(x_{pt})$ – niepewność rozszerzona wartości x_{pt}

We wcześniejszych normach wskaźnik E_n nazywany był liczbą E_n . Ponieważ wskaźnik ten utworzono z myślą o ocenie laboratoriów wzorcujących, to w jego wzorze definicyjnym miejsce x_{pt} zajmowała wartość referencyjna (wartość odniesienia) a miejsce $U(x_{pt})$ - niepewność wartości referencyjnej U_{ref} . Obecna postać wzoru jest zgodna z wzorami na inne wskaźniki, a przeznaczenie wskaźnika E_n jest bardziej uniwersalne - można go stosować także w programach przeznaczonych dla laboratoriów badawczych.

Uwaga: Dodawanie kwadratów niepewności rozszerzonych nie jest zgodne z ISO/IEC Guide 98-3 i jest niepoprawne, jeżeli współczynniki rozszerzenia k dla $U(x_i)$ i $U(x_{pt})$ są różne.

Wartości wskaźnika $E_n \geq 1,0$ lub $E_n \leq -1,0$ świadczą o potrzebie przeanalizowania przez uczestnika oszacowanej przez niego niepewności lub o konieczności poprawienia procedury pomiarowej.

Wartości $-1,0 < E_n < 1,0$ świadczą o uzyskaniu wyników zadowalających tylko wtedy, gdy niepewności są oszacowane poprawnie i gdy odchylenie ($x_i - x_{pt}$) jest mniejsze od wymaganego przez klientów.

Włączenie do interpretacji wyników porównań międzylaboratoryjnych informacji o niepewności może odegrać ważną rolę w doskonaleniu jej zrozumienia przez uczestników i rozwijaniu umiejętności szacowania. Więcej informacji na temat oceny niepewności deklarowanej przez uczestników podano w p. 6.

5. Odchylenie standardowe do badania biegłości σ_{pt} (lub/i błąd δ_E)

5.1. Metody ustalenia σ_{pt} (lub/i błędu maksymalnego dopuszczalnego δ_E)

Organizator, w porozumieniu z uczestnikami programu i jednostkami akredytacyjnymi, biorąc pod uwagę istniejące przepisy, wybiera metodę wyznaczenia σ_{pt} i przygotowuje na ten temat szczegółową informację.

Można stosować następujące metody [ISO 13528:2015, p. 8]:

- 1) jako wartość wymagana (przepisy, opinia ekspertów),
- 2) na podstawie wyników z poprzednich rund,
- 3) na podstawie modelu ogólnego,
- 4) na podstawie eksperymentalnie wyznaczonych wartości odchylenia standardowego odtwarzalności σ_R i powtarzalności σ_r ,
- 5) na podstawie danych otrzymanych w tej samej rundzie programu badania biegłości.

Ad 1) Odchyleniu standardowemu σ_{pt} , podobnie jak błędowi maksymalnemu dopuszczalnemu δ_E , można przypisać wartość, jaką organ regulujący, jednostka akredytacyjna lub eksperci techniczni organizatora PT uważają za uzasadnioną dla uczestników (racjonalną). Jeżeli w odnośnych wymaganiach podana jest wartość błędu maksymalnego dopuszczalnego δ_E , to

można go przekształcić w σ_{pt} , dzieląc δ_E przez liczbę sygnalizującą, że wynik jest niedopuszczalny. Jeżeli stosuje się wskaźnik z , to zgodnie ze wzorem (3) właściwą liczbą jest 3. Wówczas wymaganie $z_i < 3$ jest tożsame z wymaganiami $D_i < \delta_E$ oraz $P_A < 100\%$.

Ad 2) Wartość σ_{pt} (a więc i δ_E) można wyznaczyć na podstawie poprzednich rund PT, w których mierzono tę samą cechę o zbliżonej wartości przy zastosowaniu tej samej procedury pomiarowej. Wskazane jest kooperacja z innymi organizatorami, którzy prowadzą podobne badania. Metoda ta jest bardzo użyteczna, jeżeli opinie ekspertów nie są zgodne. Ma ona następujące zalety:

- oceny bazują na uzasadnionych oczekiwaniach;
- kryteria ocen nie zmieniają się z rundy na rundę;
- organizatorzy programów badania biegłości, jeżeli jest kilku w danym obszarze badań lub wzorcowania, stosują te same kryteria ocen.

Przegląd poprzednich rund programu badania biegłości powinien obejmować wyniki tylko od kompetentnych uczestników, niezakłócone z powodu mniejszych liczebnie grup czy przez udział nowych uczestników. Obliczenia należy przeprowadzić zgodnie z normą ISO 13528:2015, pkt. 8.3.

Ad 3) Z modelu ogólnego wynika, że odchylenie standardowe odtwarzalności metody pomiarowej zależy tylko od poziomu mierzonej cechy (a nie zależy od rodzaju mierzonej wielkości, procedury pomiarowej ani od liczebności próbki). Przykładem modelu ogólnego jest zależność opracowana przez Horwitza i Thompsona [ISO 13528:2015, pkt. 8.4], szeroko stosowana w chemii analitycznej.

Wartość σ_{pt} wyznaczona za pomocą modelu ogólnego musi być użyteczna w zastosowaniu praktycznym (tzn. uzasadniona, biorąc pod uwagę istniejące uwarunkowania). Należy więc sprawdzić czy model ogólny jest adekwatny do specyficznego zastosowania. Jeśli okaże się, że w efekcie przyjęcia obliczonej wartości σ_{pt} bardzo duża lub bardzo mała część uczestników ma przypisane sygnały do działania lub sygnały ostrzegawcze, to organizator programu powinien upewnić się czy jest to zgodne z celem tego programu.

Ad 4) Jeżeli metoda pomiarowa, która ma być stosowana w programie badania biegłości, jest metodą standardową, a ponadto jest dostępna informacja o powtarzalności i odtwarzalności, to stosuje się wzór [ISO 13528, p. 8.5]:

$$\sigma_{pt} = \sqrt{\sigma_R^2 - \sigma_r^2 \left(1 - \frac{1}{m}\right)} \quad (25)$$

gdzie:

σ_R - odchylenie standardowe odtwarzalności,

σ_r - odchylenie standardowe powtarzalności,

m - liczba powtórzeń wykonanych przez każde laboratorium przy wyznaczaniu wyniku pomiaru.

W programach badania biegłości najczęściej stosuje się $m = 1$. Wtedy:

$$\sigma_{pt} = \sigma_R \quad (26)$$

Jeżeli wartości σ_R i σ_r zależą od wartości mierzonej wielkości, to należy ustalić użyteczną w praktyce zależność funkcyjną. Badania wspólne, które posłużyły do wyznaczenia σ_R i σ_r , muszą być wykonane zgodnie z ISO 5725-2 lub według równoważnej procedury.

Zaletami tej metody są: obiektywność, jednakowe podejście do wielu wielkości mierzonych, bazowanie na rozległym i starannie zanalizowanym eksperymencie.

Uwaga: Na ogół organizator programu badań biegłości wymaga od uczestników podania jednego wyniku pomiaru dla badanej cechy. Nie jest wtedy dozwolone podawanie przez niektóre laboratoria więcej niż jednego wyniku. Niekiedy jednak dodatkowym celem badań biegłości może być ocena precyzji pomiarów. Wtedy organizator informuje uczestników jaka jest wymagana liczba wyników pomiarów powtórzonych i wszystkie laboratoria powinny ściśle to wymaganie spełnić.

Ad 5) Wyznaczanie σ_{pt} na podstawie wyników w rundzie nie jest zalecane nawet w programach z dużą liczbą uczestników, ponieważ parametry rozproszenia wyników zmieniają się z rundy na rundę. Zwłaszcza należy tej metody unikać, jeżeli $p < 20$.

Jeżeli organizator decyduje, że - wobec braku lepszych metod - odchylenie standardowe σ_{pt} zostanie wyznaczone na podstawie wyników rundy, to zaleca się użycie jednego z odpornych na dane odstające estymatorów szerokości (skali) rozkładu. Mogą to być:

- przeskalowane odchylenie medianowe $MADe$,
- unormowany rozstęp międzykwartyłowy $nIQR$,
- odchylenie standardowe s^* obliczone według iteracyjnego algorytmu A,
- estymator Q lub Q_n [ISO 13528:2015, C.5].

Dla liczebności $p \geq 15$ algorytm A tworzy lepszy estymator σ_{pt} niż parametry $MADe$ i $nIQR$.

Dopuszcza się również stosowanie nieodpornej miary rozrzutu, jaką jest odchylenie standardowe s , po uprzednim odrzuceniu wartości odstających wykrytych za pomocą uznanego testu statystycznego (np. Grubbsa).

5.2. Wprowadzenie ograniczeń dla σ_{pt} w przypadku wyznaczania tej wielkości na podstawie wyników podanych przez uczestników

Organizator badania biegłości powinien ustawić dolną granicę σ_{pt} , czyli najmniejszą wartość, jaka będzie używana w przypadku, gdyby odporne odchylenie standardowe s^* okazało się bardzo małe. Granica ta powinna być taka, że kiedy błąd pomiaru spełnia najbardziej ostre wymagania spodziewanego zastosowania, to $|z| < 3,0$.

Przykład [wg ISO 13528]

W programie badania biegłości dotyczącym wytwarzania tkaniny, wyznacza się m.in. liczbę wątków przypadających na centymetr. Odporne odchylenie standardowe może być w niektórych rundach bardzo małe ($s^* < 1$ wątek/cm), podczas gdy błędy do 4 wątki/cm uznaje się za nieistotne. Wobec tego organizator badania biegłości ustala, że:

- jeżeli $s^* \geq 1,3$ wątki/cm, to $\sigma_{pt} = s^*$,
- w przeciwnym razie stosuje się $\sigma_{pt} = 1,3$ wątki/cm.

Organizator programu badań biegłości powinien również ustawić maksymalną wartość σ_{pt} , jaka będzie używana, lub maksymalną wartość wyników pomiaru, które będą oceniane jako zadowalające (brak sygnału), na wypadek gdy s^* jest bardzo duże. Granica ta powinna być tak wybrana, aby wyniki niezadowalające do przeznaczenia otrzymały sygnał do działania.

W pewnych przypadkach, w których symetryczne przedziały włączałyby wyniki niespełniające oczekiwań programu (nieprzydatne w aspekcie przewidywanego zastosowania), organizator PT może ustawić dwie granice, górną i dolną, pomiędzy którymi wyniki będą ocenione jako „zadowalające”, to znaczy że nie będą generować sygnałów ostrzegawczych czy sygnałów działania.

Przykład [wg ISO 13528]

Przepisy odnośnie systematycznie wykonywanych programów badania biegłości wody niezdatnej do picia podają, że wyniki muszą się zawierać w przedziale $x^* \pm 3\sigma_{pt}$. Ponieważ jednak w pewnych przypadkach symetryczny zakres akceptowanych wartości mógłby zawierać 0 $\mu\text{g/L}$ (a nawet wartości ujemne), ustalono dodatkowo, że wynik mniejszy niż 10 % wartości zadanej podczas sporządzania próbki (czyli wartości przypisanej wyznaczonej metodą I) będzie generował sygnał działania.

Przyjmijmy, że sporządzony obiekt badania biegłości zawiera substancję o zawartości 4,0 $\mu\text{g/L}$. Średnia odporna od uczestników rundy wynosi $x^* = 3,2$ $\mu\text{g/L}$, zaś $s^* = \sigma_{pt} = 1,1$ $\mu\text{g/L}$. Zatem jeżeli uczestnik dostarczy wynik 0,0 $\mu\text{g/L}$, to będzie on wewnątrz przedziału $x^* \pm 3\sigma_{pt}$, jednak zostanie oceniony jako niezadowalający, gdyż jest mniejszy od 0,4 $\mu\text{g/L}$.

6. Ocena niepewności podanych przez laboratoria uczestniczące w programie badania biegłości

6.1. Argumenty za wymaganiami od uczestników podawania niepewności pomiarów

Organizator może zobowiązać uczestników rundy do podania oprócz wyniku pomiaru x_i również niepewności standardowej tego wyniku $u(x_i)$ lub jego niepewności rozszerzonej $U(x_i)$ – pod warunkiem podania wartości k .

Argumenty za gromadzeniem informacji o niepewności, nawet jeśli nie jest ona używana do oceny wyników uczestników za pomocą wskaźników ζ czy E_n , są następujące:

- 1) jednostki akredytacyjne mogą zapewnić, że uczestnicy podają niepewności zgodne z ich zakresem akredytacji;
- 2) uczestnicy mogą porównać niepewność podaną przez siebie z podaną przez innych; jest szansa wykrycia, że w budżecie nie uwzględniono znaczących składników niepewności lub że niektóre z nich przeszacowano;
- 3) programy badania biegłości mogą służyć do potwierdzenia deklarowanych niepewności, a jest to najprostsze wtedy, gdy niepewność wyników jest podana jawnie.

Ponadto organizator może sprawdzić poprawność podawania wyników pomiaru pod względem formalnym, np. liczby cyfr znaczących w niepewności i zgodności zapisu niepewności z zapisem estymaty wartości wielkości mierzonej.

6.2. Ustalenie dla niepewności $u(x_i)$ granicy dolnej u_{min}

Jeżeli wartość x_{pt} jest wyznaczona przy użyciu metod niezależnych od wyników w rundzie i jest spełnione kryterium $u(x_{pt}) < 0,3\sigma_{pt}$, to jest mało prawdopodobne, aby odchylenie standardowe uczestnika $u(x_i)$ było mniejsze niż $u(x_{pt})$. Zatem wartość $u(x_{pt})$ powinna być stosowana jako u_{min} , czyli dolna granica przesiewania (selekcji) niepewności.

$$u_{min} = u(x_{pt}) \quad (27)$$

Jeżeli wartość przypisana jest wyznaczona z wyników uczestników rundy, to organizator programu powinien ustalić praktyczną granicę przesiewania niepewności u_{min} .

Zastrzeżenie: Jeżeli $u(x_{pt})$ zawiera zmienność spowodowaną niejednorodnością lub niestabilnością badanego obiektu, to niepewność wyniku uczestnika $u(x_i)$ może być mniejsza niż u_{min} .

6.3. Ustalenie dla niepewności $u(x_i)$ granicy górnej u_{max}

Jest również mało prawdopodobne, aby niepewność wyniku któregośkolwiek uczestnika była większa niż 1,5-krotność odpornego odchylenia standardowego, które mierzy rozproszenie wyników od wielu uczestników. Zatem za górną granicę odsiewu podanych niepewności można przyjąć:

$$u_{max} = 1,5 s^* \quad (28)$$

Współczynnik 1,5 jest górną granicą odchylenia standardowego, jakiego można oczekiwać dla s^* wyznaczonego z 10 lub więcej wyników. Bazuje on na pierwiastku kwadratowym kwantyli rozkładu F. Organizator programu badania biegłości może przyjąć inny mnożnik.

6.4. Interpretacja niepewności wyniku uczestnika $u(x_i)$ w odniesieniu do u_{min} i u_{max}

Jeżeli organizator badań biegłości, w celu wykrycia odbiegających niepewności, stosuje wartości graniczne u_{min} lub u_{max} lub inne kryteria, to powinien wytłumaczyć uczestnikom, że nawet wtedy, gdy niepewność $u(x_i)$ wykracza poza ustalone granice, może być poprawna i wyjaśnić, że jeśli to się zdarzyło, to uczestnicy i strony zainteresowane powinni sprawdzić szacowanie niepewności.

Analogicznie, podana niepewność może zawierać się w granicach od u_{min} do u_{max} , a mimo to może nie być poprawna. Są to informacje, które wymagają wnikliwego zanalizowania.

6.5. Ocena niepewności $u(x_i)$ na podstawie kwantyli rozkładu

Organizator badań biegłości może również zwracać uwagę na wyjątkowo małe i bardzo duże wartości niepewności, opierając ocenę np. na:

- centylach podanych niepewności, np. poniżej 5 centyla i powyżej 95 centyla niepewności standardowych lub rozszerzonych;
- granicach, bazujących na oczekiwanym rozkładzie o szerokości wynikającej z rozproszenia podanych niepewności.

Jeśli niepewności od uczestników nie mają rozkładu normalnego, a używa się granic, które bazują na założeniu o normalności, to może być konieczne przekształcenie danych. Np. granice wykresu ramkowego z wąsami obliczone przy użyciu odchylenia ćwiartkowego mają interpretację probabilistyczną tylko wtedy, gdy rozkład jest normalny.

7. Jednorodność i stabilność obiektów badań

7.1. Możliwe podejścia do oceny jednorodności i stabilności

Organizator badania biegłości powinien zapewnić dostateczną jednorodność i stabilność obiektów badań. Do oceny tych właściwości może przyjąć następujące podejścia:

7.2. Procedura oceny jednorodności

- 1) Wybór cechy lub cech do badania pod kątem jednorodności.
- 2) Wybór laboratorium do przeprowadzenia kontroli jednorodności i metody pomiaru, która będzie stosowana.
- 3) Przeprowadzenie badań w celu wyznaczenia odchylenia standardowego międzypróbkowego s_s (punkt 7.3).
- 4) Porównanie s_s z przyjętą wartością kryterialną (punkt 7.4).

Ad 1. W zasadzie wszystkie właściwości obiektu badane podczas rundy powinny być sprawdzone pod kątem jednorodności i stabilności. Jeśli jednak istnieje podzbiór cech, których zachowanie jest dobrym sprawdzianem jednorodności i/lub stabilności wszystkich właściwości badanych w rundzie, to kontrolę jednorodności można ograniczyć do tego podzbioru. Wytypowane cechy powinny być szczególnie czułe na zmiany stopnia niejednorodności lub niestabilności obiektów badania.

Ad 2. Odchylenie standardowe powtarzalności s_r wybranej metody powinno być na tyle małe, aby można było wykryć niejednorodność. Zaleca się [2, 7], aby:

$$s_r < \sigma_{pt}/2 \quad (29)$$

czyli
$$s_r < \delta E/6 \quad (30)$$

Jeżeli podany warunek nie może być spełniony, należy wykonać więcej pomiarów powtórzonych.

7.3. Wyznaczenie miary niejednorodności obiektów badania biegłości – odchylenia standardowego międzypróbkowego s_s

- Należy przygotować i zapakować tyle obiektów badania biegłości, aby ich liczba wystarczyła dla uczestników rundy programu i do badania jednorodności.
- Stosując zasadę randomizacji, należy pobrać g zapakowanych obiektów, gdzie $g \geq 10$. Liczba obiektów badania biegłości sprawdzanych pod względem jednorodności może być zmniejszona, jeśli dostępne są dane z wcześniejszych sprawdzeń jednorodności na podobnych obiektach badań przygotowanych przy użyciu tych samych procedur.
- Przygotować po $m \geq 2$ porcji testowych z każdego obiektu badania biegłości, w taki sposób aby różnice między porcjami były jak najmniejsze.
- Zachowując warunki powtarzalności, wykonać w kolejności losowej pomiary wszystkich $g \cdot m$ próbek.
- Obliczyć średnią ogólną $\bar{\bar{x}}$, odchylenie standardowe wewnątrzpróbkowe s_w i odchylenie standardowe międzypróbkowe s_s .

Obliczenie s_s dla zbioru g próbek mierzonych dwukrotnie ($m = 2$) przebiega następująco:

Odchylenie standardowe średnich z próbek:

$$s_{\bar{x}} = \sqrt{\frac{1}{g-1} \sum_{t=1}^g (\bar{x}_t - \bar{\bar{x}})^2} \quad (31)$$

Odchylenie standardowe wewnątrz próbek:

$$s_w = \sqrt{\frac{1}{2g} \sum_{t=1}^g w_t^2} \quad (32)$$

Odchylenie standardowe między próbkami:

$$s_s = \sqrt{\max\left\{0, s_{\bar{x}}^2 - \frac{s_w^2}{2}\right\}} \quad (33)$$

gdzie:

\bar{x}_t - średnia dla t -tej próbki ($t = 1, \dots, g$)

$\bar{\bar{x}}$ - średnia ogólna

w_t - rozstęp wyników w próbce (między porcjami)

Jeżeli $m > 2$, należy stosować wzory bardziej złożone, podane w Aneksie B3 do normy ISO 13528.

Jeżeli nie można wykonać pomiarów powtórzonych ($m = 1$), np. gdy badania są niszczące, można przyjąć jako s_s odchylenie standardowe g wyników pomiaru. W takiej sytuacji jest szczególnie ważne, aby stosować metodę o wystarczająco małym odchyleniu standardowym powtarzalności s_r .

7.4. Podstawowe kryterium oceny jednorodności

Próbki można uznać za wystarczająco jednorodne, jeżeli jest spełniony warunek:

$$s_s \leq 0,3\sigma_{pt} \quad (34)$$

Współczynnik 0,3 zapewnia, że odchylenie standardowe międzypróbkowe s_s wnosi mniej niż 10% zmienności wprowadzonej przez σ_{pt} , a więc niejednorodność próbek nie wpływa znacząco na oceny wyników uczestników.

Jeżeli odchylenie σ_{pt} nie jest znane przed rozpoczęciem rundy, np. będzie wyznaczane z wyników od uczestników, to organizator badania biegłości powinien zastosować inny sposób oceny jednorodności, jak np.:

- sprawdzenie istotności statystycznej różnic pomiędzy obiektami badania biegłości, np. przy użyciu analizy wariancji (test F; $\alpha = 0,05$);
- oszacowanie σ_{pt} na podstawie informacji z poprzednich rund;
- wykorzystanie wartości σ_R z eksperymentu precyzji (wg ISO 5725-2);
- podjęcie ryzyka związanego z dostarczeniem niedostatecznie jednorodnych próbek i sprawdzenie kryterium dopiero po wyznaczeniu σ_{pt} jako wartości uzgodnionej.

Jeżeli warunek (34) nie jest spełniony, organizator może zastosować jeden z poniższych zabiegów:

- a) Do oszacowania σ_{pt} włączyć odchylenie standardowe międzypróbkowe s_s i obliczyć σ'_{pt} ; należy o tym poinformować uczestników;

$$\sigma'_{pt} = \sqrt{\sigma_{pt}^2 + s_s^2} \quad (35)$$

- b) Do oszacowania $u(x_{pt})$ i do oceny rezultatów działania włączyć s_s , stosując wskaźnik z' lub δ_E' ;
- c) Jeśli σ_{pt} jest odpornym odchyleniem standardowym z wyników uczestników, to już zawiera się nim niejednorodność obiektów badania biegłości, a więc można złagodzić kryterium akceptowania niejednorodności (zachowując ostrożność).

Jeżeli nie można zastosować żadnego z podanych sposobów, należy odrzucić obiekty badania biegłości i przygotować nowe, po usunięciu przyczyn niejednorodności.

Warunek wystarczającej jednorodności równorzędny do (34) ma postać:

$$s_s \leq 0,1\delta_E \quad (36)$$

7.5. Procedura i kryteria oceny stabilności

Badania stabilności należy przeprowadzić dwukrotnie, stosując tę samą metodę i ten sam materiał. Czas pomiędzy sprawdzaniem jednorodności i stabilności powinien być zbliżony do czasu, jaki upłynie do badań przez uczestników.

Etapy postępowania:

- Pobrać losowo pewną liczbę g próbek; $g \geq 2$.
- Utworzyć dwie porcje z każdej próbki, stosując tę samą metodę co w badaniach jednorodności.
- Zmierzyć w losowym porządku wszystkie $2g$ porcji w warunkach powtarzalności.
- Obliczyć średnie ogólne z wyników otrzymanych przed rozpoczęciem rundy \bar{y}_1 i po jej zakończeniu \bar{y}_2 .
- Porównać średnie ogólne pomiaru otrzymane przed rozpoczęciem rundy (np. podczas sprawdzania jednorodności) i po jej zakończeniu.

Próbki można uznać za wystarczająco stabilne, jeżeli:

$$|\bar{y}_1 - \bar{y}_2| \leq 0,3\sigma_{pt} \quad (37)$$

lub
$$|\bar{y}_1 - \bar{y}_2| \leq 0,1\delta_E \quad (38)$$

Jeżeli warunek (37) lub (38) nie został spełniony, należy zanalizować możliwości udoskonalenia procedury przygotowywania próbek i ich przechowywania.

Jeżeli jest prawdopodobne, że przyczyną niespełnienia warunku (37) jest niedokładność metody pomiarowej, to można zastosować kryterium złagodzone:

$$|\bar{y}_1 - \bar{y}_2| \leq 0,3\sigma_{pt} + 2\sqrt{u^2(\bar{y}_1) + u^2(\bar{y}_2)} \quad (39)$$

W programach dotyczących wzorcowania, kiedy ten sam obiekt badania jest używany przez wielu uczestników, organizator badania biegłości powinien zapewnić stabilność podczas rundy lub mieć procedury wykrywania niestabilności w trakcie postępu rundy. Należy uwzględnić poszukiwanie tendencji w zmianach zachodzących w poszczególnych obiektach badań, jak dryft.

W niektórych sytuacjach monitorowanie stabilności powinno obejmować również badanie efektów wielokrotnego transportu obiektu badań.

8. Analiza graficzna wyników pomiaru oraz wskaźników

Pierwszym krokiem każdej analizy danych w programach badania biegłości powinien być ogląd wykresów obrazujących rozłożenie wyników dostarczonych przez uczestników. Przegląd może potwierdzić spodziewany rozkład wyników lub ujawnić anomalia czy niespodziewane źródła zmienności.

Narzędzia graficzne przydatne w pierwszym etapie analizy wyników pomiaru:

- wykres punktowy danych posortowanych w kolejności kodów laboratoriów,
- wykres punktowy danych posortowanych rosnąco,
- histogram (dla $p \geq 20$),
- wykres rozkładu skumulowanego (dla $p \geq 20$),
- jądrowy estymator gęstości,
- wykres ramka-wąsy,
- diagram łodyga-liść.

Sprawozdanie z każdej rundy programu powinno zawierać wykresy ilustrujące wartości wyznaczonych wskaźników (P_A , z , z' , ζ , E_n).

Wykresy powinny zawierać kody literowe lub liczbowe, które pozwalają uczestnikowi zidentyfikować własne wyniki, a uniemożliwiają ustalenie, który z uczestników otrzymał jakiś inny wynik.

Do równoczesnego przedstawienia pewnej liczby ocen (dla wyników pomiaru podobnej właściwości obiektu) zaleca się stosowanie zbiorczych wykresów słupkowych. Wykres słupkowy ujawnia prawidłowości we wskaźnikach poszczególnych uczestników. Np. jeżeli laboratorium otrzymało wiele dużych wartości wskaźnika, może ono mieć obciążenie dodatnie, nawet wtedy gdy żadna wartość nie wykroczyła poza przyjęte kryterium. Najczęściej przedstawiany jest w tej formie wskaźnik z .

Podobnie można przedstawiać inne wskaźniki standaryzowane: $D\%$, P_A , z' , ζ , E_n .

- Jeżeli w rundzie programu wykonano badania powtórzone, to może być użyteczne wykonanie wykresu ilustrującego precyzję pomiarów; np. przy użyciu statystyki k zgodnie z ISO 5725-2:2002.
- Jeżeli uczestnicy podali niepewność wyników pomiaru, zaleca się przedstawienie wyników jako punktów połączonych ze słupkami o wysokości równej niepewności rozszerzonej $U(x_i)$ lub niepewności standardowej $u(x_i)$. Wartość przypisaną x_{pt} obrazuje na takim wykresie linia pozioma. Dwie równoległe do niej linie umieszcza się w odległości $U(x_{pt})$ lub $u(x_{pt})$.
- Jeżeli w tej samej rundzie programu badane są dwa podobne materiały (lub dwie cechy o zbliżonej wartości), to użyteczny jest wykres Youdena, na którym przedstawia się dwie wartości *wskaźnika* z lub dwa nieprzetworzone wyniki pomiarów jako współrzędne punktu w układzie kartezjańskim.
- Do śledzenia zmian w czasie zaleca się stosowanie kart kontrolnych obrazujących wartości wskaźnika (np. z , P_A), które sporządza się osobno dla każdego uczestnika.

Bibliografia

1. PN-EN ISO/IEC 17043:2011 Ocena zgodności - Ogólne wymagania dotyczące badania biegłości
2. ISO 13528:2015 Statistical methods for use in proficiency testing by interlaboratory comparison – wyd. 2. Załącznik 1/2016 do BI 1/53/2014 przedstawia spolszczone fragmenty tego dokumentu).
3. PN-ISO 5725:2002 Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów
4. ISO/IEC Guide 98-3:2008 (en) Uncertainty of measurement. Guide to the expression of uncertainty in measurement (JCGM 100:2008 <http://www.iso.org/sites/JCGM/GUM-JCGM100.htm>)
5. PN-EN ISO/IEC 17025:2005 Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcujących.
6. IUPAC Technical report „The International Harmonized Protocol for the PT of analytical chemistry laboratories”, 2006.
7. IUPAC/CITAC Technical Report: Selection and use of proficiency testing schemes for a limited number of participants, Pure Appl. Chem., Vol. 82, No. 5, pp. 1099–1135, 2010
8. ILAC-G13:08/2007 Wytuczne ILAC – Wymagania dotyczące kompetencji organizatorów programów badań biegłości. (ILAC Guidelines for the Requirements for the Competence of Providers of Proficiency Testing Schemes)
9. Selection, use and interpretation of proficiency testing (PT) schemes, Eurachem 2011
10. Guidelines for the assessment of the appropriateness of small interlaboratory comparisons within the process of laboratory accreditation including measurement audits organised by accreditation bodies, EEE/PT(11)03v16. Oprac. EEE-PT Working Group “Proficiency Testing in Accreditation”, wyd. EA
11. DA-05 Polityka dotycząca uczestnictwa w badaniach biegłości, PCA
12. DAPT-01 Akredytacja organizatorów badań biegłości. Wymagania szczegółowe, PCA
13. Della W. M. Sin, Yiu-chung Wong: Use of assigned reference values: revisiting a small-scale inter-laboratory comparison for residual pesticides in tea, Accred Qual Assur, 2015