



Klub Polskich
Laboratoriów
Badawczych
POLLAB

XX Sympozjum

**Metody statystyczne
w praktyce
laboratoryjnej**

Jachranka / Kołobrzeg, 2014 r.



Klub Polskich
Laboratoriów
Badawczych
POLLAB

XX Sympozjum

**Metody statystyczne
w praktyce laboratoryjnej**

Jachranka / Kołobrzeg, 2014 r.

Informacje o wydaniu

Symposium Klubu Polskich Laboratoriów Badawczych POLLAB
o charakterze szkoleniowym

Materiały szkoleniowe – wydanie 1

Praca zbiorowa pod redakcją: Krystyny Krzyśko – Prezesa Klubu POLLAB

Wydrukowano z materiałów dostarczonych przez autorów

Nakład 600 egz.

Wydawca:

Klub Polskich Laboratoriów Badawczych POLLAB

ul. Kłobucka 23A

02-699 Warszawa

www.pollab.pl

Projekt okładki, skład i druk:

Wydawnictwo MALAMUT

ul. Królowej Marysieńki 9/6a, 02-954 Warszawa

www.malamut.pl

ISBN 978-83-923010-7-3

Warszawa, 2014

Szanowni Państwo! Uczestnicy sympozjum

Zarząd Klubu Polskich Laboratoriów Badawczych POLLAB, odnosząc się do potrzeb zgłaszanych przez Członków Klubu, zdecydował się na przybliżenie tematyki metod statystycznych wykorzystywanych w laboratoriach i dołożył wszelkich starań aby tematyka sympozjum przyczyniła się do rozwoju i doskonalenia działalności laboratoriów oraz potwierdzenia ich kompetencji, których owocem będą jedynie miarodajne wyniki pomiarów i badań (wiarygodne, rzetelne i użyteczne).

Słowo statystyka ma obecnie wiele znaczeń. Pierwotnie (do połowy XIX w.) oznaczało ono podany w tabelarycznej formie, zbiór danych liczbowych dotyczących stanu państwa. Z czasem pojęciem tym zaczęto określać wszelkie gromadzone dane.

Z analizą zbiorów danych (wyników badań) związane było powstanie specyficznych metod, które odrębniły się w postaci dyscypliny naukowej określonej mianem statystyki. Tak więc statystyka, to zbiór metod służących pozyskiwaniu, prezentacji i analizie danych liczbowych.

Celem stosowania tych metod jest otrzymywanie użytecznych informacji na temat wyników pomiarów i badań, których dotyczą. Podstawowe zadania statystyki to jednak analiza i interpretacja.

Według Kaoru Ishikawy **„95% problemów w każdej organizacji może być rozwiązane za pomocą tylko 7 narzędzi sterowania jakością”**.

Osiągnięcie postawionego celu sympozjum będzie możliwe za sprawą znakomitych referatów naszych prelegentów, wśród których znajdują się przedstawiciele Polskiego Centrum Akredytacji, Głównego Urzędu Miar, profesory wyższych uczelni oraz praktycy akredytowanych organizacji - członkowie Klubu POLLAB.

W imieniu Zarządu Klubu Polskich Laboratoriów Badawczych POLLAB wyrażam **podziękowanie wszystkim autorom referatów** za ich przygotowanie na wysokim poziomie merytorycznym.

Wszystkim uczestnikom życzę korzyści i satysfakcji z udziału w naszym sympozjum. Wyrażam przekonanie, że zdobyta wiedza oraz wymiana doświadczeń, przyczynią się do pokonywania trudności, z którymi spotykają się Państwo w swojej działalności, podczas pracy w laboratoriach, czy innych instytucjach. Ponadto mam nadzieję, że doskonała atmosfera integracyjna XX Sympozjum będzie kolejną okazją do odnowienia kontaktów, nawiązania nowych oraz wymiany posiadanych doświadczeń.

*„I tak jak gwiazdy są ważniejsze od astronomii a róże od botaniki,
tak ważniejsze są zastosowania metod statystycznych od niej samej...”*

Hogo Dionizy Steinhaus

Krystyna Krzyśko
Prezes Klubu POLLAB

SPIS TREŚCI

1. Analiza statystyczna danych laboratoryjnych z wykorzystaniem arkuszy kalkulacyjnych – <i>Sabina Żebrowska-Łucyk</i>	1
2. Wskaźniki oceny kompetencji laboratoriów w badaniach biegłości / porównaniach międzylaboratoryjnych – <i>Tadeusz Bulski</i>	13
3. Zastosowanie statystyk Mandela w badaniach międzylaboratoryjnych – <i>Sabina Żebrowska-Łucyk</i>	21
4. Praktyczne wykorzystanie normy ISO 13528:2005 – doświadczenia Ośrodka Badań Biegłości Instytutu Chemicznej Przeróbki Węgla – <i>Mariusz Mastalerz</i>	33
5. Praktyczne wykorzystanie normy ISO 13528 w laboratorium badawczym przy organizacji ILC niewielkiej liczby uczestników – <i>Aneta Paduchowicz</i>	39
6. Podstawowe testy statystyczne wykorzystywane w trakcie walidacji – <i>Piotr Konieczka</i>	47
7. Proste statystyczne podejście do szacowania niepewności pobierania i przygotowania próbki do badań – <i>Waldemar Korol, Sławomir Walczyński, Grażyna Bielecka, Jolanta Rubaj</i>	55
8. Metodyka opracowania danych pomiarowych w świetle dokumentów Międzynarodowego Biura Miar – <i>Paweł Fotowicz</i>	61
9. Nowe techniki kalibracyjne oparte na regresji liniowej w analizie chemicznej – <i>Wojciech Hyk</i>	65

Analiza statystyczna danych laboratoryjnych z wykorzystaniem arkuszy kalkulacyjnych

Przedstawiono możliwości zastosowań starych i nowych wersji programu MS Excel w pracach laboratoriów pomiarowych i wzorcujących. Szczególną uwagę zwrócono na: a) często używane funkcje statystyczne, które mogą sprawiać użytkownikom kłopoty z powodu różnic pomiędzy programami MS OFFICE 2010 i MS OFFICE 2013 a ich poprzednimi wersjami, b) sposób postępowania się funkcjami opisującymi modele matematyczno-statystyczne (na przykładzie rozkładu t-Studenta), c) testowanie hipotez statystycznych z użyciem funkcji Excela.

1. WPROWADZENIE

1.1 Funkcje statystyczne w arkuszach kalkulacyjnych

Trudno wyobrazić sobie pracę nowoczesnego laboratorium badawczego czy wzorcującego bez narzędzi komputerowych wspomagających dokumentowanie i przetwarzanie rozmaitych danych laboratoryjnych, zwłaszcza wyników pomiaru. Bez wątplenia do takich nieodzownych dzisiaj narzędzi należą wszechobecne arkusze kalkulacyjne. Ich intuicyjność, łatwa obsługa i wszechstronność sprawia, że mogą być używane zarówno do wykonywania prostych obliczeń, sporządzania zestawień i raportów, jak i do zaawansowanych analiz, zgłębiania danych, symulacji, testowania hipotez badawczych i formułowania prognoz. Najpowszechniej stosowanym obecnie arkuszem kalkulacyjnym jest Excel, jeden z dwóch kluczowych pakietów środowiska MS Office [1, 2, 3]. Pakiet ten jest wyposażony w liczne i bardzo użyteczne w laboratorium funkcje statystyczne. Praktyka pokazuje jednak, że większość użytkowników, w tym pracownicy laboratoriów, wykorzystują zaledwie znikomą część możliwości, jakie daje to narzędzie. Jednym z powodów takiej sytuacji jest niedopracowany opis funkcji, szczególnie w polskojęzycznej wersji pakietu, gdzie występuje wiele błędów i niepoprawnie stosowanych terminów, komentarze do przykładów są zdawkowe a nawet mylące, często brakuje jakichkolwiek objaśnień.

Mimo występowania oczywistych niedostatków, uciążliwych przede wszystkim dla bardziej zaawansowanych odbiorców, kolejne edycje Excela, łącznie z MS Office 2007, nie były doskonalone w zakresie odnoszącym się do funkcji statystycznych [2, 3]. Dopiero w wersji MS Office 2010 wprowadzono liczne zmiany, mające na celu: poprawę dokładności obliczeń, osiągnięcie zgodności z funkcjami stosowanymi w innych pakietach statystycznych używanych w środowisku naukowym, poszerzenie możliwości analizy danych oraz uzyskanie spójności nazw różnych funkcji programu Excel. Trzeba jednak stwierdzić, że ten ostatni cel nie został osiągnięty. Dla użytkowników polskich, którzy postępowali się wcześniejszymi wersjami programu, nazwy niektórych funkcji nowego produktu mogą być dezorientujące i stać się źródłem groźnych w skutkach pomyłek. Nierzadko bowiem brakuje konsekwencji i porządku, co obrazują przykłady zamieszczone w kolejnych rozdziałach.

W trosce o zachowanie kompatybilności MS Excel 2010 oraz MS Excel 2013 z poprzednimi wersjami tego programu, funkcje statystyczne, które zostały zastąpione innymi, są nadal dostępne pod stary-

mi nazwami. Zostały one umieszczone w specjalnie do tego celu utworzonej grupie funkcji o nazwie *Zgodność*. Producent pakietu zachęca jednak do korzystania z nowego, poszerzonego zestawu funkcji umiejscowionych w grupie o nazwie *Statystyczne*, która istniała też we wcześniejszych wersjach programu.

1.2 Zastosowania arkuszy kalkulacyjnych w laboratoriach badawczych i wzorcujących

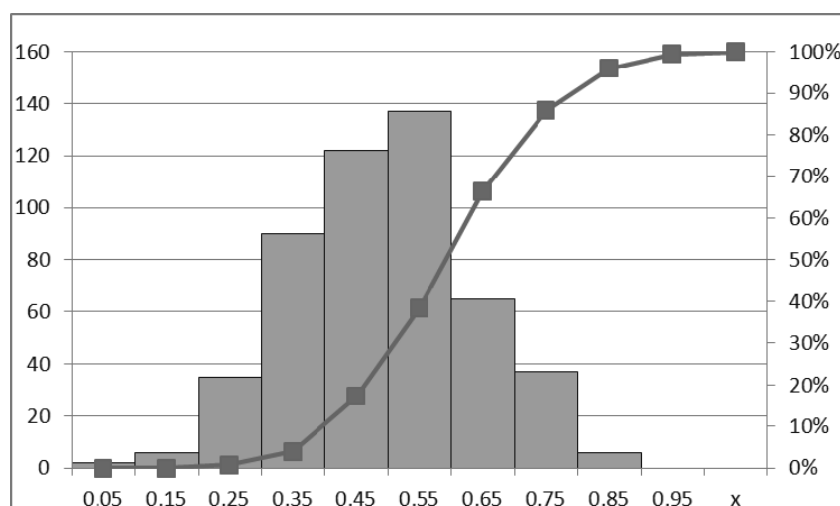
Jednym z podstawowych zastosowań arkuszy kalkulacyjnych jest gromadzenie uporządkowanych w logiczny sposób danych liczbowych oraz tekstowych. Pojemność pojedynczego arkusza nowych edycji Excela wynosi obecnie ponad milion wierszy (2^{20}) i ponad 16 tysięcy kolumn (2^{14}). Sprawnemu wprowadzaniu danych sprzyjają liczne mechanizmy automatyzujące ich wpisywanie do komórek arkusza [1, 6] oraz łatwy import danych zapisanych w plikach o różnych formatach.

Zastosowanie narzędzi automatycznie wyróżniających dane, które spełniają sformułowane przez użytkownika warunki, pozwala wykryć symptomy niepokojące a także zauważyć pojawienie się stanów szczególnie korzystnych. Podobne korzyści daje sortowanie danych, wielokryterialna filtracja zbiorów, dołączenie komentarzy do komórek. Właściwe użycie funkcji z grupy *bazodanowych* oraz funkcji tablicowych może zaoszczędzić użytkownikowi wiele godzin pracy.

Odpowiednio dobrane wykresy pozwalają nie tylko zidentyfikować problemy i ułatwić ich diagnozowanie (np. wykresy Pareto-Lorentza, wykresy we współrzędnych logarytmicznych, zależności korelacyjne, wykresy trójwymiarowe itp.), ale też poglądowo zilustrować osiągnięcia laboratorium.

Zastosowanie funkcji dotyczących rozkładów prawdopodobieństwa eliminuje potrzebę używania tablic statystycznych, gdyż pozwala szybko uzyskać dokładne wartości kwantyli oraz poszukiwanych prawdopodobieństw. Szczególnie użyteczny w laboratoriach jest rozkład t-Studenta (rozdz. 3), który służy m.in. do ustalenia wartości współczynnika k , niezbędnego do obliczenia niepewności rozszerzonej.

Dzięki zastosowaniu funkcji tablicowych REGLINW oraz REGLINP sprawą trywialną staje się wyznaczenie w arkuszu krzywych kalibracyjnych oraz przedziałów ufności dla tych krzywych. Inna funkcja tablicowa CZĘSTOŚĆ, podobnie jak nieco trudniejsza w użyciu funkcja statystyczna LICZ.WARUNKI automatyzują sporządzanie histogramów i krzywych kumulacyjnych (rys. 1).



Rys. 1. Histogram i kumulacyjna krzywa prawdopodobieństwa (dystrybuanta empiryczna) wyznaczone w pakiecie Excel

Do tworzenia budżetów niepewności arkusz jest wręcz nieodzowny. Niewiele jednak osób wie, że także w odniesieniu do pomiarów wyznaczanych metodami pośrednimi można łatwo zautomatyzować

szacowanie niepewności. Zamiast wyznaczać pochodne cząstkowe, wystarczy przeprowadzić nieskomplikowane działania w arkuszu przy zastosowaniu metody przyrostowej [4] (w obu przypadkach: gdy wynik badania jest funkcją zmierzonych wielkości od siebie niezależnych lub wielkości o błędach z sobą skorelowanych).

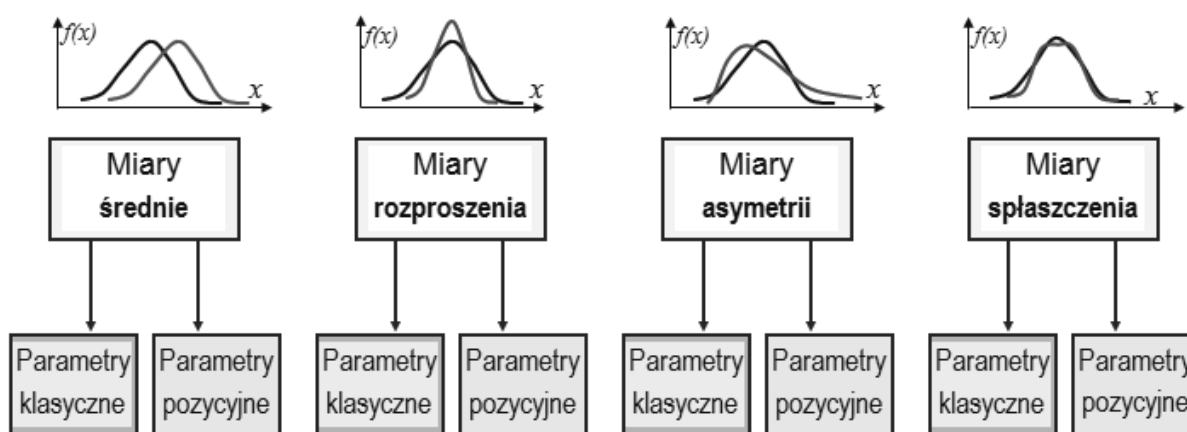
Arkusze nadają się również do prowadzenia na użytek laboratoriów kart kontrolnych, a jeśli są odpowiednio zaprojektowane, automatycznie wykrywają sygnały rozregulowania procesu.

Trzeba też choćby wspomnieć o szerokiej gamie dostępnych testów statystycznych, potrzebnych m.in. do wykrywania wartości odstających, porównania precyzji pomiarów, wykrywania obciążenia laboratoriów, porównania wartości średnich dwóch lub wielu zbiorów danych, oceny statystycznej istotności trendów. Niektóre z testów są zaimplementowane w kompletnej formie w postaci jednej funkcji, jak testy t (rozd. 4) czy test Fishera-Snedecora, inne można włączyć do arkusza przy stosunkowo niewielkim nakładzie pracy (Grubbsa, Mandela h , Mandela k , Dixona, Cochрана). Wreszcie, niektóre z testów można bardzo łatwo przeprowadzić przy użyciu algorytmów umieszczonych w zakładce *Analiza danych* (np. ANOVA jednoczynnikowa i dwuczynnikowa).

Znaczna część wymienionych wyżej funkcji arkuszy była już szczegółowo omówiona we wcześniejszych pracach autorki, prezentowanych na sympozjach klubów POLLAB [4] oraz Polskie Forum ISO [5]. W niniejszym artykule nacisk położono na kwestie związane z: a) często używanymi funkcjami, które mogą sprawiać użytkownikom kłopoty z racji niekonsekwentnych nazw spowodowanych różnicami pomiędzy programami MS OFFICE 2010 i MS OFFICE 2013 a ich poprzednimi wersjami, b) korzystaniem z funkcji opisujących modele matematyczno-statystyczne, c) stosowaniem funkcji służących do testowania hipotez.

2. WYZNACZANIE PARAMETRÓW STATYSTYCZNYCH NA PODSTAWIE PRÓBY

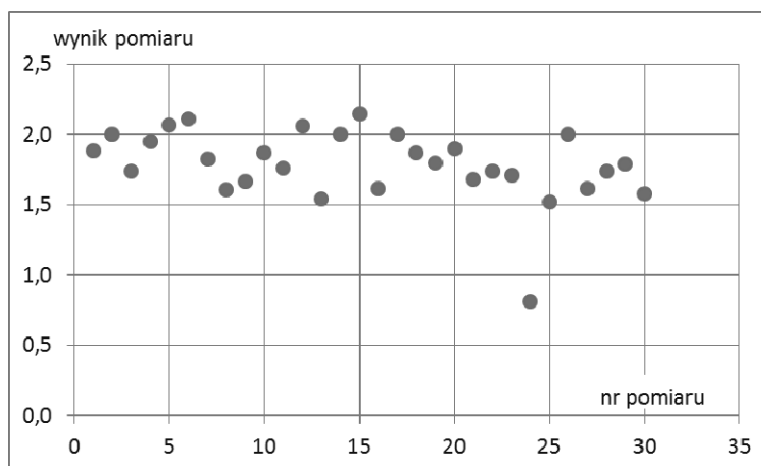
Jednym z podstawowych sposobów przetwarzania zgromadzonych danych pomiarowych jest wyznaczenie ich parametrów statystycznych. Na ogół wyznacza się co najmniej jedną miarę tendencji centralnej i jedną miarę rozproszenia, czasem dodatkowo jeszcze parametry charakteryzujące skośność i spłaszczenie empirycznego rozkładu danych. W każdej z wymienionych czterech grup parametrów wyróżnić można dwie odmiany: parametry klasyczne i parametry pozycyjne (rys. 2).



Rys. 2. Klasyfikacja statystycznych miar charakteryzujących rozkład wyników pomiaru

Parametry klasyczne bazują na wszystkich danych w zbiorze, są więc wrażliwe na wartości skrajne. Właściwość tę mają najczęściej używane w laboratoriach parametry: *średnia arytmetyczna* i *odchylenie standardowe* z próby. Ich reakcję na wartości odstające ilustruje przykład przedstawiony na rysunku

3. Dla zbioru 30 punktów wartość średnia i odchylenie standardowe wynoszą odpowiednio $\bar{x} = 1,79$ i $s = 0,26$. Natomiast po usunięciu punktu o numerze kolejnym 19 wartości te zmieniają się na $\bar{x} = 1,82$ oraz $s = 0,18$.



Rys. 3. Przykład zbioru danych z wartością odstającą

W niektórych analizach statystycznych może być więc celowe obliczanie średnich po wcześniejszym usunięciu ze zbioru ustalonego procenta wartości skrajnych (służy do tego funkcja ŚREDNIA.WEWN), bądź z uwzględnieniem jedynie tych danych, które spełniają określone kryteria, dotyczące albo ich wartości albo też charakteryzujące źródło czy warunki pozyskania danych (funkcja ŚREDNIA.WARUNKÓW), co ilustruje rys. 4.

	A	B	C	D	E
1	Wynik	Analitik		średnie wyniki	
2	4,7	Nowak		5,33	=ŚREDNIA(A2:A12)
3	5,5	Bzowski		4,72	=ŚREDNIA.WEWN(A2:A12;0,2)
4	6,6	Bzowski		6,10	=ŚREDNIA.WARUNKÓW(A2:A12;A2:A12;">5";A2:A12;"<10")
5	7,0	Bzowski		3,28	=ŚREDNIA.WARUNKÓW(A2:A12;B2:B12;"Nowak")
6	2,7	Nowak		7,78	=ŚREDNIA.WARUNKÓW(A2:A12;B2:B12;"Bzowski")
7	3,3	Nowak			
8	14,5	Bzowski			
9	5,3	Bzowski			
10	1,6	Nowak			
11	4,3	Nowak			
12	3,1	Nowak			

Rys. 4. Przykład obliczania średnich na podstawie pełnego i niepełnych zbiorów danych

Obliczając odchylenie standardowe, należy zachować staranność, aby spośród kilku funkcji o bardzo zbliżonych nazwach (tabela 1) wybrać odpowiednią do danego celu.

Tabela 1. Funkcje służące do obliczania klasycznych miar rozrzutu zaimplementowane w programach MS Excel 2010 i MS Excel 2013

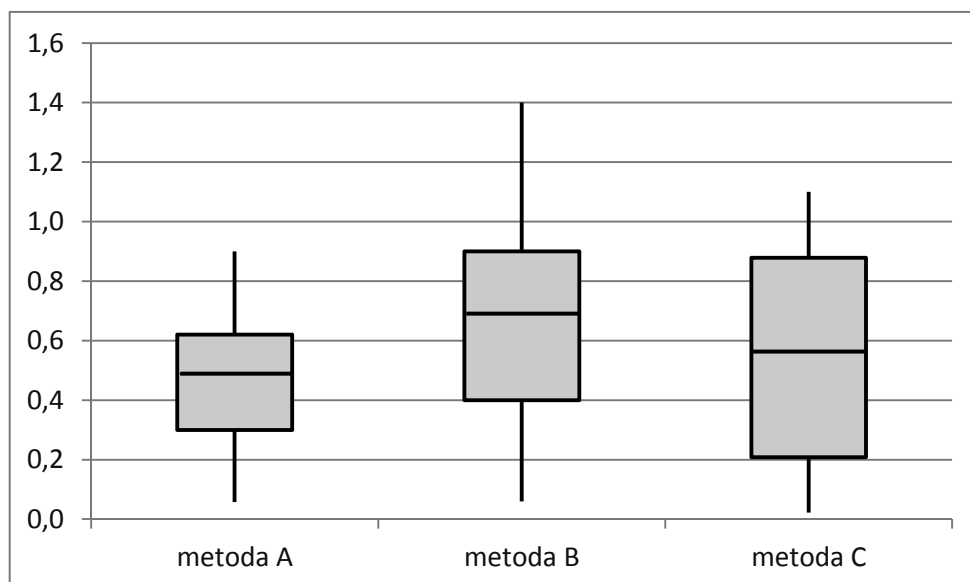
Funkcje w grupie <i>Statystyczne</i> f_x	Funkcje w grupie <i>Zgodność</i> f_{y1}	Wzór
ODCH.STAND.POPUL ODCH.STANDARD.POPUL.A	ODCH.STANDARD.POPUL	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$
ODCH.STANDARD.PRÓBK ODCH.STANDARDOWE.A	ODCH.STANDARDOWE	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$
ODCH. ŚREDNIE		$s = \frac{\sum_{i=1}^n x_i - \bar{x} }{n-1}$

Dodatkowym źródłem pomyłek może być funkcja ODCH.KWADRATOWE, która służy do wyznaczania sumy kwadratów odchyleń poszczególnych wartości w zbiorze od wartości średniej.

Większość parametrów pozycyjnych jest znacznie bardziej niż klasyczne odporna na wartości odstające (wyjątkiem jest rozstęp, który bardzo silnie reaguje na wartości nietypowe). Najczęściej stosowanymi w laboratoriach parametrami odpornymi są: mediana med – miara tendencji centralnej oraz rozstęp ćwiartkowy IQR (*Interquartile range*) – miara rozproszenia punktów pomiarowych.

Mediana dzieli uporządkowany monotonicznie zbiór danych na dwa równoliczne podzbiory. Można ją wyznaczyć w arkuszu Excel przy użyciu dowolnej z następujących funkcji: mediana, percentyl lub kwartył. Mediana i kwartyły są bowiem szczególnymi przypadkami percentyli, a mediana szczególnym przypadkiem kwartyła. P-ty percentyl zbioru danych ($0 \leq P \leq 100$) to, nieco upraszczając, taka wartość $x(P)$, przed którą znajduje się P procent uporządkowanych niemalejąco elementów tego zbioru. Mediana jest więc 50. percentylem, kwartył dolny to 25. percentyl, a kwartył górny to 75. percentyl. Na przykład, jeśli dane znajdują się w komórkach A1:A30, to ten sam skutek przyniesie użycie formuł: MEDIANA(A1:A30), PERCENTYL(A1:A30;0,5) oraz KWARTYL(A1:A30;3). Jak łatwo zauważyć, jako argument funkcji PERCENTYL wpisuje się wartość P/100. Argumentem funkcji KWARTYL może być liczba tylko całkowita spośród wartości 0, 1, 2, 3, 4, które odpowiadają kolejno percentyloom rzędu 0 (minimum zbioru), 25 (kwartył dolny), 50 (mediana), 75 (kwartył górny) i 100 (maksimum zbioru).

Wyznaczone pięć wartości można wykorzystać do sporządzenia wykresów ramka-wąsy (ang. Box & Whisker Plot) ilustrujących rozkłady danych w porównywanych zbiorach (rys. 3, [4]). Funkcje PERCENTYL i KWARTYL należą do grupy *Zgodność*. Ich odpowiednikiem w grupie funkcji o nazwie *Statystyczne* są funkcje KWARTYL.PRZEDZ.ZAMK i PERCENTYL.PRZEDZ.ZAMK. Oprócz nich istnieją jeszcze funkcje liczone dla przedziałów, które nie obejmują wartości granicznych 0 i 1. Są to: KWARTYL.PRZEDZ.OTW i PERCENTYL.PRZEDZ.OTW.



Rys. 5. Przykład wykorzystania funkcji KWARTYL.PRZEDZ.ZAMK do sporządzenia wykresów typu ramka-wąsy

Parametr IQR jest równy różnicy pomiędzy górnym a dolnym kwartylem. W podanym przykładzie można go więc wyznaczyć jako: $KWARTYL(A1:A30;3) - KWARTYL(A1:A30;1)$ lub $PERCENTYL(A1:A5;0,75) - PERCENTYL(A1:A30;0,25)$.

Do szacowania odchylenia standardowego zbiorów zawierających wartości odstające wykorzystywany jest czasem związek

$$\sigma \cong 0,741 \cdot IQR \quad (1)$$

Podstawą jego stosowania jest założenie o oczekiwanej normalności rozkładu danych.

W programach badania biegłości laboratoriów decyzję o tym, czy dany wynik jest akceptowalny, podejmuje się na ogół na podstawie jego odległości od wartości przypisanej, a przedział akceptacji jest zwykle względem niej symetryczny ($\pm \Delta$). W pewnych jednak sytuacjach (zwłaszcza, przy dużym rozproszeniu wyników bądź występowaniu skośnych rozkładów danych lub kiedy liczba różnych odpowiedzi jest ograniczona) właściwym postępowaniem może być ocena wyników oparta na percentylach, np. 5 % wyników najbardziej odległych od wartości modalnej lub od wartości przypisanej może być uznawane za nieakceptowalne. Zaleca się zachowanie ostrożności przy stosowaniu tej metody [6].

W międzylaboratoryjnych badaniach porównawczych jako miarę rozrzutu danych (precyzji) stosuje się również odchylenie medianowe MAD (*median absolute deviation*), zdefiniowane następująco:

$$MAD\{x_i\} = med\{|x_i - med(x_i)|\} \quad (2)$$

Dla rozkładu Gaussa relacja między s i MAD ma postać $s = 1,483 MAD$, co pozwala szacować niepewność standardową pojedynczego pomiaru jako

$$u = 1,483 \cdot MAD \quad (3)$$

Jeśli wynik końcowy jest ustalony na podstawie n -elementowej próby, to

$$u = 1,859 \cdot MAD/n. \quad (4)$$

Mnożnik występujący we wzorze (3) wynika z zastosowania współczynnika uwzględniającego stratę efektywności estymatora z powodu zastąpienia średniej medianą, a mianowicie:

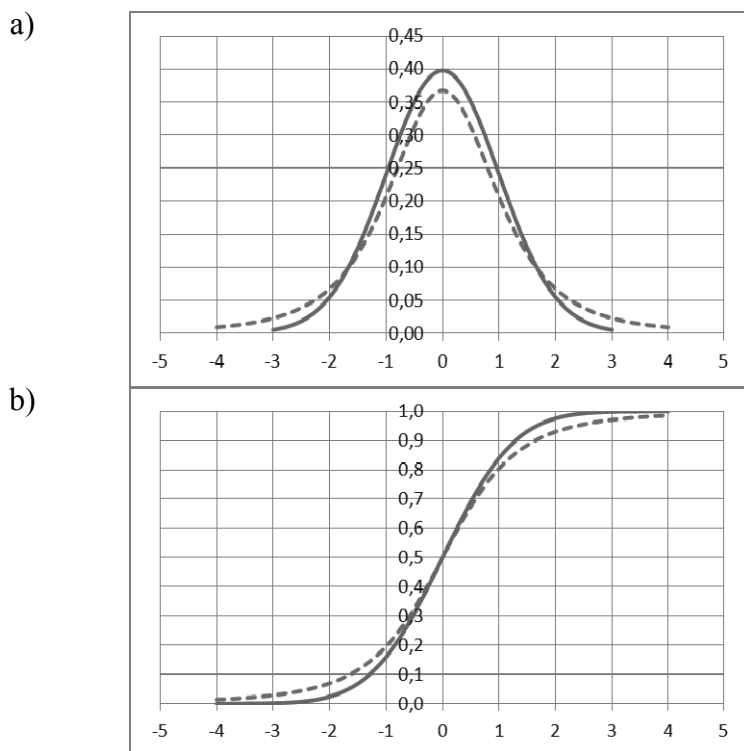
$$1,859 = 1,483 \cdot (0,5\pi)^{1/2} \quad (5)$$

W badaniach biegłości dane dostarczone przez poszczególne laboratoria przetwarza się, stosując iteracyjne procedury wyznaczania miar średnich (algorytm A) i miar rozrzutu (algorytm S) [7]), bazujące na medianie. Jednym z ważnych parametrów rozkładów empirycznych jest *wartość modalna* (in. moda, dominanta, ang. *mode*), czyli ta, która w zbiorze danych pojawia się najczęściej. Do jej znalezienia służy funkcja WYST.NAJCZĘŚCIEJ.WART. Zdarza się nierzadko, że w zbiorze danych występuje nie jedna lecz dwie lub więcej wartości o tak samo dużej liczności (rozkład jest dwumodalny lub wielomodalny). Do wykrycia wszystkich wartości modalnych służy funkcja tablicowa WYST.NAJCZĘŚCIEJ.TABL.

3. ROZKŁADY PRAWDOPODOBIEŃSTWA – MODELE MATEMATYCZNO-STATYSTYCZNE. ROZKŁAD T-STUDENTA

Program Excel zawiera pokazny zbiór funkcji opisujących rozkłady prawdopodobieństwa. Nazwy tych funkcji w wersjach, które ukazały się w roku 2010 i 2013 są inne niż w wersjach wcześniejszych, jest ich też więcej. Ponieważ dokładność obliczeń w poprzednich programach spotykała się z krytyką, algorytmy zostały poprawione, m.in. w rozkładzie Chi-kwadrat. Do niektórych funkcji włączono dodatkowy argument rozszerzający zakres ich stosowania. Można teraz, operując parametrem logicznym, uzyskać bezpośrednio wartość dystrybuanty rozkładu (lewostronne prawdopodobieństwo rozkładu skumulowanego) albo wartość funkcji gęstości prawdopodobieństwa, co wcześniej wymagało wykonania dodatkowych działań na dostępnych funkcjach.

Przykłady wykresów rozkładu t-Studenta, odpowiadające 3 i 100 stopni swobody, pokazano na rys. 6. Uzyskano je za pomocą funkcji ROZKŁ.T(t;3;0) i ROZKŁ.T(t;3;1) (linia kreskowa przerywana) oraz ROZKŁ.T(t;100;0) i ROZKŁ.T(t;100;1) (linia ciągła), gdzie t przyjmuje wartości z przedziału $[-4, 4]$. Natomiast w tabeli 2 zestawiono wszystkie funkcje związane z rozkładem Studenta, które są dostępne w starszych i nowych wersjach Excela, posiłkując się przykładami ich zastosowania w rozkładzie o 3 stopniach swobody.



Rys. 6. Przykłady wykresów ilustrujących rozkład t-Studenta dla liczby stopni swobody $\nu = 100$ (linia ciągła) i $\nu = 3$ (linia kreskowa); a) funkcja gęstości, b) dystrybuanta

Tabela 2. Przykłady użycia funkcji zaimplementowanych w nowych programach MS Excel, dotyczących rozkładu t-Studenta; liczba stopni swobody wynosi 3.

Lp	Przeznaczenie/ zwracana wartość	Funkcje w grupie Statystyczne f_x	Funkcje w grupie Zgodność f_{λ}	Wynik
1	funkcja gęstości prawdopodobieństwa $f(t)$ dla $t = 2$	ROZKŁ.T(2;3;0)	–	0,068
2	wartość dystrybuanty $F(t)$ dla $t = 2$	ROZKŁ.T(2;3;1)	–	0,930
3	$1 - F(t)$ dla $t = 2$	ROZKŁ.T.PS(2;3)	ROZKŁAD.T(2;3;1)	0,070
4	$2[1 - F(t)]$ dla $t = 2$	ROZKŁ.T.DS(2;3)	ROZKŁAD.T(2;3;2)	0,139
5	kwantyl t dla $F(t) = 0,05$	ROZKŁ.T.ODWR(0,05;3)	–	-2,353
6	wartość bezwzględna t dla $F(t) = 0,025$	ROZKŁ.T.ODWR.DS(0,05;3)	ROZKŁAD.T.ODW(0,05;3)	3,182

W analizach laboratoryjnych największe zastosowanie ma funkcja ROZKŁ.T.ODWR.DS(*alfa*; *liczba stopni swobody*), która służy do wyznaczenia współczynnika rozszerzenia k , gdzie liczba stopni swobody wynika z liczby powtórzonych pomiarów i rozkładów prawdopodobieństwa wielkości uwzględnianych w budżecie niepewności (wzór Welcha-Satterthwaite'a).

4. TESTOWANIE HIPOTEZ STATYSTYCZNYCH W LABORATORIUM Z UŻYCIEM FUNKCJI DOSTĘPNYCH W ARKUSZACH EXCELA

W laboratoriach pomiarowych często przeprowadza się badania oparte na porównaniach. Do tyczą one działania różnych przyrządów, różnych metod, poszczególnych analityków, różnych laboratoriów, tej samej procedury ale w różnych okresach itp. Najczęściej porównuje się ze sobą dwa zbiory danych pod względem wartości średnich oraz stopnia rozproszenia wyników.

Warunkiem wyciągnięcia miarodajnych konkluzji jest zastosowanie odpowiednich metod statystycznych. Wnioskowanie o populacji, oparte na użyciu testów statystycznych, rozpoczyna się od sformułowania dwóch wykluczających się przypuszczeń dotyczących tej populacji, tj. od postawienia hipotez: zerowej, oznaczanej zwykle symbolem H_0 , oraz będącej jej zaprzeczeniem hipotezy alternatywnej, oznaczanej symbolem H_1 . Hipoteza zerowa jest hipotezą prostą, zaś hipoteza alternatywna jest zwykle złożona. Hipotezę alternatywną jesteśmy skłonni przyjąć wtedy, gdy na podstawie wyników próby należy odrzucić H_0 . Kolejnym krokiem jest wybór tzw. statystyki testowej, zwanej również sprawdzianem hipotezy, która jest funkcją wyników próby. Uzyskana w próbie wartość statystyki jest podstawą do odrzucenia lub do przyjęcia H_0 . Rozkład statystyki testowej wyznacza się przy założeniu, że hipoteza zerowa jest prawdziwa. Cały obszar zmienności statystyki T_n dzieli się na dwa dopełniające się zbiory: zbiór przyjęć i zbiór krytyczny. Jeśli hipoteza H_0 jest prawdziwa, to prawdopodobieństwo zaobserwowania wartości T_n ze zbioru krytycznego W wynosi α . Prawdopodobieństwo α to poziom istotności testu. Najczęściej przyjmuje się $\alpha = 0,05$. Jeśli zaś prawdziwa jest hipoteza H_1 , to prawdopodobieństwo zaobserwowania wartości T_n ze zbioru krytycznego W jest maksymalne.

Przykład testowania hipotezy dotyczącej dwóch wartości średnich

Dwóch analityków wykonało pomiary tych samych pięciu próbek (pył kontrolny dla krzemionki; wyniki w tabeli w procentach). Z badać, czy różnice pomiędzy wartościami średnimi są statystycznie istotne. Założyć $\alpha = 0,05$.

Pracownik \ Nr próbki	1	2	3	4	5	6	średnia aryt. \bar{d}	s_d
pracownik A	2,8	2,4	2,7	3,2	2,9	3,8		
pracownik B	2,9	2,7	3,0	3,1	3,0	3,8		
$d = A - B$	-0,1	-0,3	-0,3	0,1	-0,1	0,0	-0,12	0,16

Hipotezy zerowa i alternatywna mają postać:

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$$

gdzie μ_1 i μ_2 – wartości oczekiwane badanych populacji.

W rozpatrywanym przykładzie poszczególne wyniki tworzą pary, odpowiadające pomiarom kolejnych próbek. Dlatego podstawą testu będą różnice w parach oznaczone symbolem \bar{d} . Przyjmując, że rozkłady w obu populacjach są normalne, można do rozstrzygnięcia, która z postawionych hipotez jest słuszna, zastosować statystykę testową:

$$t = \frac{\bar{d}}{s_d} \sqrt{n} \quad (5)$$

Ma ona rozkład t-Studenta o liczbie stopni swobody $\nu = n - 1$. Przyjmując poziom istotności testu $\alpha = 0,05$ i posługując się funkcją =ROZKŁ.T.ODWR.DS(0,05;5), a w starszych wersjach MS Excel funkcją =ROZKŁAD.T.ODW(0,05;5), otrzymuje się

$$t_{kr} = t_{0,05;5} = 2,571$$

Wartość obliczona ze wzoru (4) wynosi $t = \frac{-0,12}{0,16} \sqrt{6} = -1,784$.

Ponieważ $|t| < t_{kr}$, nie ma podstaw do stwierdzenia, że różnice pomiędzy średnimi są statystycznie istotne. Można dodatkowo, używając funkcji =ROZKŁ.T.DS(1,784;5), wyznaczyć obustronne prawdopodobieństwo, że jeśli hipoteza zerowa jest prawdziwa, to pojawi się wynik $t \geq |-1,784|$. Prawdopodobieństwo to (ang. nazwa i oznaczenie p-value) wynosi $\pi = 0,1345$. Jest więc ono wyraźnie większe od przyjętego poziomu istotności $\alpha = 0,05$.

Wszystkie powyższe obliczenia można zastąpić użyciem zaledwie jednej funkcji T.TEST (a w starszych wersjach programu TEST.T), co ilustruje rys. 5. W przykładowym użyciu funkcja ta ma postać T.TEST(A20:A25;B20:B25;2;1). Dwa pierwsze argumenty to zakresy porównywanych komórek, trzeci argument oznacza, że test jest dwustronny, ostatni wskazuje rodzaj testu – wartości sparowane. Podana funkcja od razu zwraca obliczoną wcześniej krok po kroku wartość $p = 0,1345$, co – jak już wiadomo – oznacza, że nie ma podstaw by twierdzić, że wyniki uzyskane przez dwóch analityków różnią się od siebie pod względem wartości oczekiwanych (brak przesunięcia, wzajemnego obciążenia).

	A	B	C	D	E	F
1	Analityk A	Analityk B				
2	2,8	2,9				
3	2,4	2,7		=T.TEST(A2:A7;B2:B7;2;1)		
4	2,7	3,0				
5	3,2	3,1		0,1345		
6	2,9	3,0				
7	3,8	3,8				

Rys. 5. Użycie testu par wartości do porównania średnich uzyskanych przez dwóch analityków

Tę samą funkcję można stosować do porównania wartości oczekiwanych zbiorów niezależnych, np. do oznaczania zawartości substancji w dwóch partiach materiału. Zbiory takie nie muszą być wtedy jednakowo liczne. Zmieniając wartość ostatniego argumentu z 1 na 2 otrzymuje się test zmiennych niezależnych dla populacji o jednakowych wariancjach, a gdy argument ten wynosi 3, wykonywany jest test dla populacji o różnych wariancjach, zwany testem Cochran-Coxa.

Równie łatwa w użyciu jest funkcja F.TEST, którą można stosować w laboratoriach do porównania precyzji oraz funkcja Z.TEST do porównania wartości średniej zbioru wyników z zakładaną (wymaganą) wartością oczekiwaną.

5. PODSUMOWANIE

Arkusze kalkulacyjne mogą w większym niż obecnie stopniu wspomagać pracę laboratoriów, usprawniając wiele działań, w tym wszelkie analizy i obliczenia, a tym samym ułatwiać wdrażanie a potem monitorowanie procesu walidacji metod pomiarowych.

Zaletami arkuszy są m.in.: ogromna łatwość ich dostosowania do potrzeb laboratorium, możliwość tworzenia makr (VBA) automatyzujących operacje, a także import i eksport danych.

Nowe wersje programu MS Excel (2010 i 2013) oferują szerszą od poprzednich, udoskonaloną paletę funkcji statystycznych. Poprawiono również dokładność obliczeń. Do wad zaliczyć trzeba brak konsekwencji w nazwach niektórych funkcji. Nadal też opisy w plikach pomocy są wysoce niezadowolające, a w bibliotece wykresów ciągle brakuje bardzo przydatnego wykresu ramka-wąsy (ang. *Box & Whisker plot*).

Wpisane przez użytkownika formuły, służące do statystycznego przetwarzania danych, powinny być starannie sprawdzone. Należy też zadbać o wprowadzeniu zabezpieczeń, jakie daje ochrona komórek zawierających formuły, aby zapobiec ich mimowolnym zmianom, a w razie potrzeby ochrona całych plików. Wysitek włożony w zaplanowanie arkusza szybko procentuje uzyskaniem efektywnych narzędzi, niezastąpionych pod względem uniwersalności, a pod wielu innymi względami dorównujących specjalistycznym programom statystycznym.

LITERATURA

- [1] J. C. Nash: Teaching statistics with Excel 2007 and other spreadsheets. Computational Statistics & Data Analysis, vol.52, 2008, 4602–4606.
- [2] A. Talha Yalta: The accuracy of statistical distributions in Microsoft® Excel 2007. Computational Statistics & Data Analysis, vol. 52, 2008, 4579–4586.

- [3] Yu-Sung Su: It's easy to produce chartjunk using Microsoft® Excel 2007 but hard to make good graphs. *Computational Statistics & Data Analysis*, vol. 52, 2008, 4594–4601.
- [4] Żebrowska-Łucyk S.: Analiza wyników badań laboratoryjnych z wykorzystaniem arkuszy kalkulacyjnych XII Sympozjum Klubu POLLAB „Doskonalenie systemu zarządzania w laboratorium”. Ustroń, 10-12.2006, Mat. konf. 39-51.
- [5] Żebrowska-Łucyk S.: Przedziały ufności dla krzywych kalibracyjnych a niepewność pomiarów. *Mat. Symp. Metrologia w systemach zarządzania – 6*. Dymaczewo k/Poznania, 15-17.09.2010. Wyd. Polskie Forum ISO 9000, Warszawa 2010, 121-129.
- [6] PN-ISO 5725-2:2002 Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów. Część 2. Podstawowe metody określania powtarzalności i odtwarzalności standardowej metody pomiarowej.
- [7] ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparisons.

Wskaźniki oceny kompetencji laboratoriów w badaniach biegłości/porównaniach międzylaboratoryjnych

W wystąpieniu omówiono zasady oceny kompetencji technicznych laboratoriów na podstawie wyników udziału w badaniach biegłości/porównaniach międzylaboratoryjnych (PT/ILC). Przedstawiono wskaźniki stosowane przez organizatorów PT/ILC, sposób ich wyznaczania oraz interpretację ich zastosowania w kontroli jakości. Wskazano przy tym na konieczność wyboru rodzaju wskaźnika do celu prowadzonych badań oraz zasadność podawania niepewności wyników przesyłanych do organizatora PT/ILC.

1. WPROWADZENIE

Raport przygotowany przez organizatora porównań międzylaboratoryjnych powinien być ważnym źródłem informacji na temat jakości wyników, pozwalającym bądź na potwierdzenie kompetencji laboratorium, bądź na analizę ewentualnych problemów. Udział laboratorium w PT/ILC może być dobrowolny lub konieczny ze względu na wymagania, np. prawne. We wszystkich przypadkach udział w PT/ILC powinien być traktowany jako narzędzie pozwalające na identyfikację sytuacji potencjalnie wpływających na wynik prowadzonych badań.

Wymagania odnośnie organizacji i prowadzenia badań biegłości są opisane w normie PN-EN/IEC 17043:2011 „Ocena zgodności – Ogólne wymagania dotyczące badań biegłości”. Warto przy tym podkreślić, że w normie tej przywołana jest norma ISO 13528:2005, w której opisane są metody statystyczne stosowane do oceny wyników przesłanych przez uczestników PT/ILC.

Podstawowym aspektem udziału w PT/ILC jest przeprowadzenie badania dla dostarczonego przez organizatora porównań, obiektu (próbki), w warunkach identycznych z codzienną praktyką postępowania w danym laboratorium. Uzyskane wyniki są następnie porównywane, z przyjętą przez organizatora wartością odniesienia, a do oceny kompetencji laboratorium stosowany jest wskaźnik liczbowy pozwalający na obiektywne porównanie danego wyniku w stosunku do wartości odniesienia. W przypadku uzyskania wyników spełniających określone kryteria, laboratorium potwierdza swoje kompetencje w zakresie prowadzonych badań. W przypadku uzyskania wyników niezadawalających, laboratorium powinno przeprowadzić szczegółową analizę stosowanej procedury pomiarowej.

Porównania międzylaboratoryjne (IRC) polegają na zorganizowaniu, wykonaniu i ocenie pomiarów tego samego typu lub podobnych obiektów przez co najmniej dwa laboratoria, zgodnie z uprzednio ustalonymi warunkami – DA-05 PCA Wyd.5.

2. PODSTAWOWE ZASADY PT/ILC

Podstawową zasadą organizacji badań biegłości / porównań międzylaboratoryjnych jest wybór i przygotowanie obiektu badań, czyli próbek spełniających wysokie wymagania jakościowe. Przy doborze materiału konieczne jest spełnienie dwóch podstawowych kryteriów:

– obiekty wykorzystywane w PT/ILC powinny być w jak największym stopniu zbliżone pod względem fizycznym i chemicznym do badanych w laboratoriach próbek;

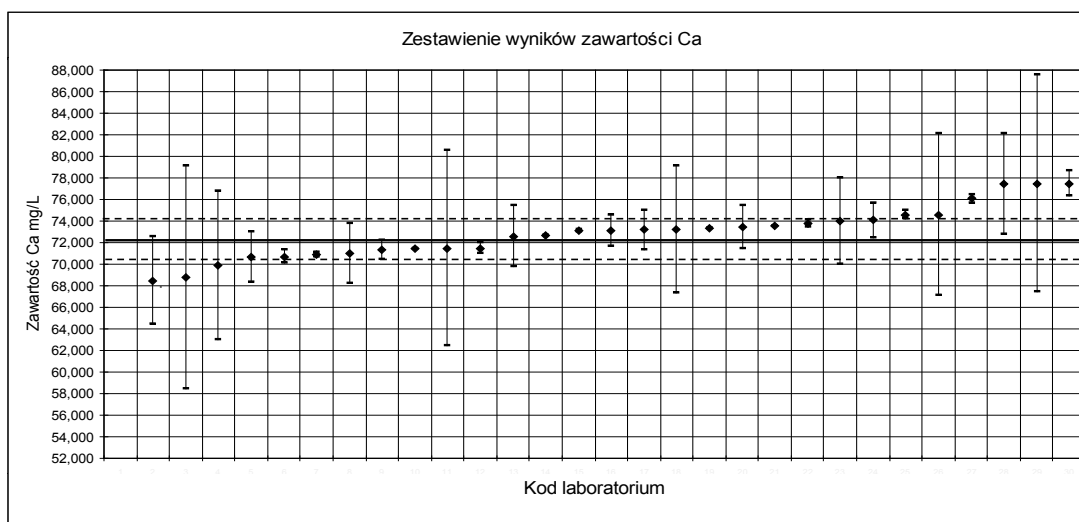
Idealnym materiałem są oczywiście naturalne obiekty, zawierające naturalne zawartości oznaczanych składników, zgodne z rodzajem próbek badanych w laboratoriach uczestniczących w porównaniu.

– materiał wykorzystywany do przygotowania obiektów powinien spełniać wysokie kryteria jednorodności.

Wymóg odnośnie jednorodności wynika z tego, że właściwa ocena wyników porównań może być przeprowadzona jedynie wtedy, gdy każdy z uczestników otrzyma próbkę o identycznym składzie.

Organizator PT/ILC zobowiązany jest do przygotowania materiału do badań (jednej bądź kilku obiektów), który jest jednorodny oraz stabilnym w okresie prowadzenia danej rundy badań. Jednorodność materiału wykorzystywanego do przygotowania obiektów do badań, jak również jego stabilność powinny być wykazane za pomocą odpowiednich pomiarów, zgodnie z wcześniej opisanymi procedurami. Próbkę (obiekty badań) są rozsyłane do laboratoriów – uczestników danej rundy badań, wraz z odpowiednimi informacjami, zgodnie z wymaganiami normy PN-EN/IEC 17043:2011. Organizator określa termin, w którym uczestnicy powinni dostarczyć wyniki badań, a następnie przeprowadza statystyczną ocenę otrzymanych wyników. Pierwszym etapem jest porównanie wyników uzyskanych od laboratoriów z przypisaną wartością odniesienia.

Na rysunku 1 pokazano typowy wykres, na którym zebrane zostały wyniki przesłane przez laboratoria. Linia pozioma (pogrubiona) oznacza wartość odniesienia, a linie poziome przerywane oznaczają przedział niepewności wartości odniesienia. Symbole (♦) wraz z pionowymi liniami to wyniki otrzymane od uczestników porównania, z podanymi przez laboratoria wartościami odchylenia standardowego lub niepewności. Zamieszczony rysunek dotyczy wyników oznaczania stężenia wapnia w próbkach wody powierzchniowej. Na podstawie tak przedstawionych wyników, każdy uczestnik może ocenić uzyskany wynik poprzez porównanie z wartością odniesienia, jak również poprzez porównanie z wynikami uzyskanymi przez pozostałych uczestników danej rundy. Warto zwrócić uwagę, że w kilku przypadkach wartość średnia podana przez laboratoria jest zbliżona, natomiast przypisane wartości niepewności różnią się znacznie. Jak zostanie to pokazane w dalszej części, może to mieć istotny wpływ na wartość wskaźników oceny jakości wyników.



Rysunek 1. Typowy wykres pokazujący zestawienie wyników przesłanych przez uczestników danej rundy PT/ILC względem przypisanej przez organizatora wartości odniesienia.

2.1 Wyznaczanie wartości odniesienia i jej niepewności

Do oceny kompetencji laboratorium na podstawie udziału w PT/ILC stosuje się **wskaźniki klasyfikujące**, pozwalające przede wszystkim na porównanie podanych przez uczestników wyników względem przyjętej wartości odniesienia, w odniesieniu do ustalonych przez organizatora kryteriów. Jak wynika z przeglądu różnych raportów oraz literatury, w praktyce stosowane są różne kryteria, a w konsekwencji różne wskaźniki klasyfikujące. W każdym jednak przypadku wskaźnik klasyfikujący jest wyrażeniem ułamkowym, a podstawowym i wspólnym elementem jest zapisywana w mianowniku różnica pomiędzy wynikiem uzyskanym w danym laboratorium (x_{lab}) a wartością odniesienia (X_{odn}) podaną przez organizatorów.

$$\Delta = x_{lab} - X_{odn}$$

Gdzie:

x_{lab} – wynik podany przez laboratorium;

X_{odn} – wartość odniesienia podana przez organizatorów ILC.

W normie ISO 13528:2005 podanych jest kilka sposobów wyznaczenia wartości odniesienia i jej niepewności.

1. Próbka do badań jest przygotowana ze składników o znanej zawartości danej substancji, a wartość odniesienia jest wyznaczana poprzez obliczenie zawartości danej substancji na podstawie znanych wartości (między innymi: masa, objętość, czystość substancji, gęstość, itp.). Wartość niepewności jest wyznaczana na podstawie równania modelowego zgodnie z przewodnikiem GUM (p. 5.2 normy 13528:2005);
2. Próbka do badań jest to materiał odniesienia z certyfikowaną wartością danego składnika (CRM); a wartość certyfikowana jest przyjęta jako wartość odniesienia w danej rundzie PT/ILC, jednocześnie jako niepewność wartości odniesienia jest przyjmowana wartość podana w świadectwie materiału odniesienia (p. 5.3 normy 13528:2005);
3. Próbka do badań jest to materiał odniesienia (RM), odpowiednio jednorodny i stabilny. Wartość odniesienia jest wyznaczana przez laboratorium eksperckie, które stosuje zwalidowaną procedurę pomiarową. W celu zapewnienia spójności pomiarowej laboratorium stosuje właściwy CRM, a wartość odniesienia jest wyznaczana na podstawie wielokrotnych pomiarów dla losowo wybranych próbek

w warunkach powtarzalności. Wartość niepewności jest wyznaczana jako niepewność złożona dla wartości niepewności CRM oraz niepewności wyznaczania wartości odniesienia na podstawie wielokrotnych pomiarów (p. 5.4 normy 13528:2005);

4. Wartość odniesienia jest wyznaczana na podstawie wyników uzyskanych przez laboratoria eksperckie, do których przesyłane są losowo wybrane próbki materiału przygotowanego do danej rundy PT/ILC. Na podstawie uzyskanych wyników wyznaczana jest wartość średnia oraz niepewność, przy czym organizator stosuje do obliczeń statystykę odporną (p. 5.5 normy 13528:2005);
5. Wartość odniesienia jest wyznaczana na podstawie wyników uzyskanych przez wszystkich uczestników danej rundy PT/ILC. W tym przypadku również stosowana jest statystyka odporna, jako wartość odniesienia przyjmuje się wartość średnią odporną z wyników uczestników, zaś niepewność standardowa wartości przypisanej wyraża się wzorem: $u_x = 1,25 \times s^* / \sqrt{p}$, gdzie s^* jest odpornym odchyleniem standardowym (p. 5.5 normy 13528:2005).

2.2 Wyznaczanie wartości odchylenia standardowego dla kryteriów oceny udziału w badaniach biegłości

W normie ISO 13528:2005 opisano również pięć sposobów wyznaczania odchylenia standardowego, stosowanego do oceny udziału laboratorium w badaniach biegłości (σ).

1. **Wartość zalecana.** Wartość odchylenia standardowego jest ustalona zgodnie z wymaganiami dla prowadzonych w laboratoriach badań lub na podstawie odnośnych wymagań prawnych.
2. **Wartość wymagana.** Wartość odchylenia standardowego wynika z wymagań stawianych laboratorium uczestniczącym w danej rundzie badań biegłości i odpowiada oczekiwany wymaganiom dla potrzeb danych badań.
3. **Na podstawie modelowania.** Wartość odchylenia standardowego dla badań biegłości może być wyznaczona na podstawie modelu opisującego zależność wartości odchylenia standardowego od stężenia oznaczanej substancji. Często stosuje się do tego celu ogólne równanie Horowitza:

$$\sigma_R = 0,02 c^{0,8495}$$

gdzie c oznacza stężenie danej substancji w procentach (ułamek masowy).

4. **Na podstawie precyzji pomiarów.** Wartość odchylenia standardowego może być wyznaczona eksperymentalnie w danym laboratorium w warunkach powtarzalności (σ_p) lub w warunkach odtwarzalności (σ_R) na podstawie wielokrotnych pomiarów.
5. **Na podstawie wyników uczestników w danej rundzie badań biegłości.** Wartość odchylenia standardowego odpowiada wartości odchylenia standardowego odpornego dla wyników dostarczonych przez uczestników $\sigma = s^*$.

2.3 Wskaźniki oceny wyników laboratorium

Jak już wspomniano wcześniej do oceny kompetencji laboratorium stosowane są różne **wskaźniki klasyfikujące**. W każdym przypadku wskaźnik klasyfikujący jest wyrażeniem ułamkowym, w którym w podana jest w liczniku wartość różnicy pomiędzy wynikiem uzyskanym w danym laboratorium (x_{lab}) a wartością odniesienia (X_{odn}) podaną przez organizatorów. Tą różnicę odnosi się względem klasyfikatora, czyli względem wartości zapisywanej w mianowniku.

Różnica procentowa

Najprostszym sposobem porównania dwóch wartości liczbowych jest wyznaczenie różnicy procentowej, co w przypadku ILC można opisać następującym równaniem:

$$D\% = \frac{(x_{lab} - X_{odn})}{X_{odn}} \times 100\%$$

gdzie:

$D\%$ – różnica wyrażona w procentach;

x_{lab} – wynik podany przez laboratorium;

X_{odn} – wartość odniesienia

Kryterium oceny jest ustalane przez organizatorów, na przykład można przyjąć, że wartości zadowalające spełniają kryterium $D\% \leq 20\%$.

Wskaźnik z

W przypadku stosowania wskaźnika z stosuje się następujące równanie:

$$z = \frac{x_{lab} - X_{odn}}{\sigma_p}$$

gdzie: σ_p oznacza najczęściej wartość odchylenia standardowego dla wyników uzyskanych od uczestniczących laboratoriów lub przyjętą przez organizatorów wartość dla danego porównania (wynikająca na przykład z wymagań prawnych dla danych badań).

W tym przypadku kryterium oceny jest zgodne z powszechnie przyjętymi i podanymi poniżej zakresami:

$|z| \leq 2$ wynik zadowalający
 $2 < |z| \leq 3$ wynik wątpliwy
 $|z| > 3$ wynik niezadowalający

W niektórych przypadkach organizatorzy zamiast wartości odchylenia standardowego wszystkich wyników podanych przez uczestniczące laboratoria, stosują arbitralnie przyjętą wartość wynikającą z wymagań prawnych dla danych oznaczeń, na przykład σ_p może odpowiadać 0,2 wartości odniesienia X_{odn} .

W takim przypadku stosowane równanie przyjmuje postać:

$$z = \frac{x_{lab} - X_{odn}}{0,2 X_{odn}}$$

Warto podkreślić, że zastosowany zabieg pozwala na rzeczywistą ocenę kompetencji laboratorium względem wymagań prawnych przyjętych.

Wskaźnik z'

W równaniu opisującym wskaźnik z nie jest uwzględniona niepewność wartości odniesienia. W tych przypadkach, kiedy jest to możliwe zaproponowano stosowanie zmodyfikowanego wskaźnika, oznaczanego symbolem z' , w którym uwzględniono zarówno odchylenie standardowe przyjęte dla danego porównania (tak jak dla wskaźnika z) jak i niepewność standardową wartości odniesienia.

$$z' = \frac{x_{lab} - X_{odn}}{\sqrt{\sigma_p^2 + u_x^2}}$$

Gdzie u_x oznacza wartość niepewności standardowej dla wartości odniesienia.

Kryterium oceny wyniku przy stosowaniu wskaźnika z' jest takie samo jak w przypadku wskaźnika z :

$|z'| \leq 2$ wynik zadowolający
 $2 < |z'| \leq 3$ wynik wątpliwy
 $|z'| > 3$ wynik niezadowolający

Wskaźnik zeta

Kolejną modyfikacją wskaźnika z jest wskaźnik **zeta**, w którym uwzględniona została niepewność standardowa złożona wyniku podanego przez laboratorium, co nie było uwzględniane we wcześniej omawianych wskaźnikach. Stosowanie tego wskaźnika jest możliwe wtedy, gdy laboratorium podaje wynik z przypisaną niepewnością.

$$zeta = \frac{x_{lab} + X_{odn}}{\sqrt{u_x^2 + u_X^2}}$$

Kryterium oceny jest identyczne z podanym, wcześniej kryterium dla wskaźników z i z' .

Wskaźnik **zeta** może być stosowany zamiast wskaźnika z w tych przypadkach, gdy laboratorium podaje wyniki z oszacowaną niepewnością.

Liczby E_n

Kolejnym kryterium oceny, coraz częściej stosowanym przez organizatorów porównań, są liczby E_n , nazywane też błędem znormalizowanym. Kryterium to jest stosowany przede wszystkim do oceny porównań dwustronnych, gdzie dane laboratorium porównuje swoje wyniki względem laboratorium o wyższych kompetencjach metrologicznych i jest powszechnie stosowany przez laboratoria wzorcujące.

$$E_n = \frac{x_{lab} - X_{odn}}{\sqrt{U_x^2 + U_X^2}}$$

gdzie: U_x oraz U_X to wartości niepewności rozszerzonej (dla $k = 2$) odpowiednio wyniku podanego przez laboratorium i wartości odniesienia.

Warto zwrócić uwagę, że podstawową różnicą między liczbami E_n w stosunku do wskaźnika **zeta** jest to, że w przypadku liczb E_n w mianowniku zamieszczone są wartości niepewności rozszerzonej, stąd wynika zmiana wartości liczbowych kryterium oceny.

$|E_n| \leq 1$ wynik zadowalający
 $|E_n| > 1$ wynik niezadowalający

Liczby E'_n

W niektórych przypadkach organizatorzy zamiast wartości niepewności rozszerzonej ($k = 2$) wartości odniesienia stosują arbitralnie przyjętą wartość wynikającą z wymagań prawnych dla danych oznaczeń, na przykład może odpowiadać 0,2 wartości odniesienia X_{odn} .

Liczby E'_n , w którym stosowana jest ustalona przez organizatora wartość niepewności (zgodnie np. z wymaganiami odpowiedniej dyrektywy) jest oznaczany symbolem E'_n .

$$E'_n = \frac{x_{lab} + X_{odn}}{\sqrt{u_{lab}^2 + (0.1X_{odn})^2}}$$

gdzie:

$$|E'_n| \leq 2 = (+)$$

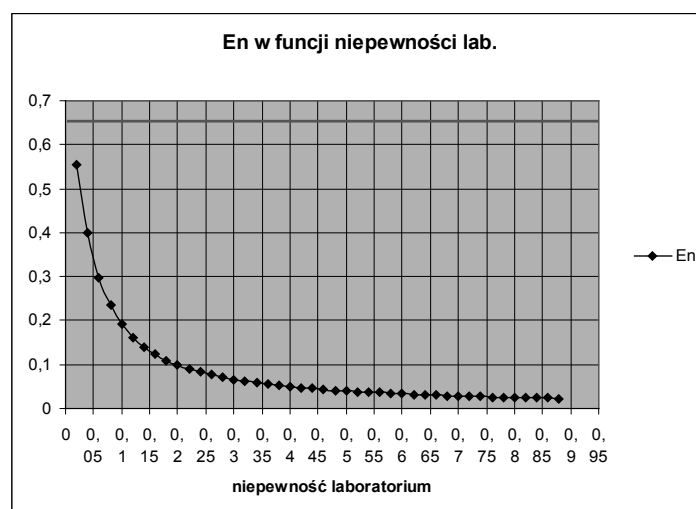
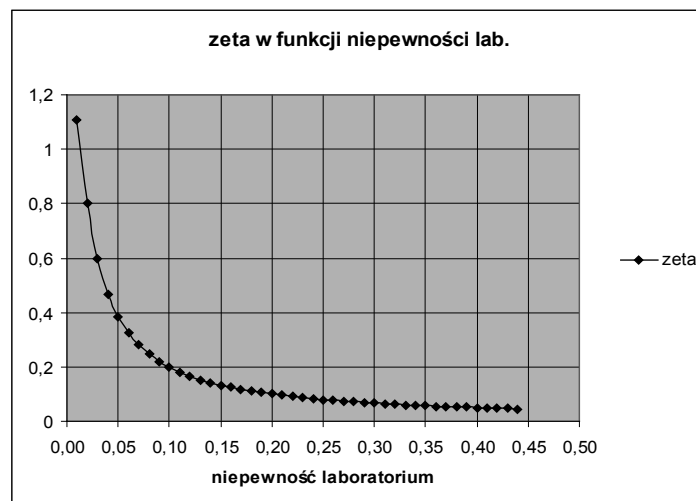
$$2 < |E'_n| < 3 = (?)$$

$$|E'_n| \geq 3 = (-)$$

Przedstawione powyżej omówienie najważniejszych wskaźników stosowanych przez organizatorów porównań (ILC/PT) do oceny wyników wskazuje, że wynik oceny może zależeć od zastosowanego kryterium. Stąd niezmiernie ważnym elementem raportów przygotowywanych przez organizatorów powinno być jednoznaczne opisanie stosowanych wskaźników, łącznie z podaniem wartości. Z drugiej strony laboratorium oceniając swój udział w danym porównaniu powinno we własnym zakresie krytycznie ocenić, na ile zastosowane kryterium odpowiada przyjętym wymaganiom odnośnie jakości wyników oraz w razie potrzeby wykorzystać zamieszczone w raporcie dane do zastosowania własnych kryteriów.

2.4 Wpływ wartości niepewności wyniku na ocenę uczestnika badań biegłości

Jak wynika z przedstawionych w rozdziale 2.3 opisów wskaźników, ocena danego laboratorium może istotnie zależeć od zastosowanego kryterium. Warto w tym miejscu podkreślić, że rolą organizatora jest wybór zastosowanego algorytmu statystycznej oceny wyników przesłanych przez uczestników danej rundy oraz wybór wskaźnika, biorąc pod uwagę przede wszystkim cel danego badania. Konsekwencją tego jest to, że przekazany przez laboratorium wynik (wartość średnia wraz z przypisaną niepewnością) może być oceniony jako zadowalający lub nie, w zależności od przyjętego kryterium. Na rysunku 2 pokazano dwa przykłady, jak zmienia się wartość wskaźników zeta (rys. 2A) oraz E_n (rys. 2B) w zależności od podanej przez laboratorium wartości niepewności.



Rysunek 2: Zależność wartości wskaźników zeta oraz En od wartości niepewności wyniku przesłanego przez laboratorium. Pozioma linia pogrubiona odpowiadająca wartości 0,67 oznacza wartość wskaźnika z (bez uwzględnienia wartości niepewności wyniku laboratorium).

LITERATURA

- [1] PN-EN ISO/IEC 17025:2005 „Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcowych”
- [2] ISO/IEC 13528:2005 „Statistical methods for use in proficiency testing by interlaboratory comparisons”
- [3] PN-EN ISO/IEC 17043:11 „Ocena zgodności; Ogólne wymagania dotyczące badań biegłości”
- [4] DA-05, PCA „Polityka uczestnictwa w badaniach biegłości”, Wyd.5,Warszawa, 17.11.2011
- [5] Ewa Bulska „Metrologia chemiczna”, Wydawnictwo Malamut 2012
- [6] Elizabeth Prichard, Vicki Barwick „Quality Assurance in Analytical Chemistry” Wiley 2007
- [7] Andrzej Zięba „Analiza danych w naukach ścisłych i technice”, PWN 2013

Zastosowanie statystyk Mandela w badaniach międzylaboratoryjnych

Przedstawiono krótki przegląd technik i narzędzi statystycznych stosowanych w laboratoriach oraz w badaniach międzylaboratoryjnych do wykrywania wartości niepewnych i odstających i na tym tle omówiono statystyki Mandela, służące do oceny wariancji wyników w laboratoriach (precyzji pomiarów) oraz wariancji międzylaboratoryjnej. Podano sposób wyznaczenia wartości krytycznych testów Mandela k i h oraz przykłady obliczeń statystyk Mandela wraz z interpretacją wyników.

1. WPROWADZENIE

Badania międzylaboratoryjne są stosowane coraz powszechniej i obejmują coraz to szersze obszary tematyczne, gdyż pozwalają obiektywnie i w dobrze udokumentowany sposób ocenić sprawność wykonywania w laboratoriach pomiarów i analiz oraz wiarygodność uzyskiwanych wyników [1-6]. Udział laboratoriów w badaniach wspólnych stał się nieodzownym elementem doskonalenia ich pracy.

Wśród badań prowadzonych łącznie przez grupy laboratorium czołowe miejsce zajmują bez wątpienia badania biegłości [2-5] oraz eksperymenty służące oszacowaniu dokładności metod pomiarowych [6]. Badania biegłości mają na celu ocenę zdolności laboratoriów do prowadzenia określonych badań lub pomiarów. Uczestnictwo w takich pomiarach oznacza systematyczne monitorowanie uzyskiwanych efektów, ułatwia laboratoriom identyfikowanie problemów oraz zapewnienie należytej dokładności pomiarów. Natomiast w przypadku drugiego ze wspomnianych rodzajów badań zakłada się, że kompetencje laboratorium są uznane, a efektem wspólnego eksperymentu ma być określenie cechy stosowanej w metody pomiarowej.

Typowy plan badań międzylaboratoryjnych zakłada na ogół eksperyment zrównoważony, w którym każde z p laboratoriów uczestniczących bada n -krotnie próbki na q poziomach. Poziom opisywany jest przez średnią wyników wziętą ze wszystkich laboratoriów, odnoszącą się do danego materiału lub do próbki. Ponieważ precyzja pomiaru zależy na ogół od wartości badanej cechy lub materiału (np. od stężenia, rodzaju matrycy), dlatego walidację metody czy też badania biegłości trzeba wykonać w szerokim zakresie zmienności wpływających czynników. Na ogół zakłada się, że aby pokryć interesujący zakres wystarcza $q = 5$. W efekcie pomiarów przewiduje się otrzymanie łącznie pqn wyników. Jednak z powodu różnych czynników, takich jak dane niekompletne, wyniki odstające czy laboratoria odstające, sytuacja rzeczywista może być inna.

W eksperymencie wyznaczania precyzji należy na podstawie danych z poszczególnych poziomów oszacować odchylenie standardowe odtwarzalności i odchylenie standardowe powtarzalności. Badania powtórzone wykonuje się niezależnie, tak jakby to było n badań na różnych materiałach, stosując ten sam przyrząd. Grupy odpowiadające różnym poziomom można badać w różnych dniach, ale przez tego samego operatora. Jeśli materiał na q poziomach rozestano do p laboratoriów, z których każde wykonuje pomiary ze zmianą czynnika (czynników) precyzji pośredniej między każdym z n pomiarów,

to analizę pomiarów przeprowadza się tak samo, ale zamiast odchylenia standardowego powtarzalności szacuje się odchylenie standardowe precyzji pośredniej.

Chociaż cele badań międzylaboratoryjnych mogą być rozmaite i różna jest też organizacja badań, to w każdym przypadku konieczna jest krytyczna analiza wyników dostarczonych przez laboratoria pod kątem ich wiarygodności.

2. IDENTYFIKACJA WARTOŚCI ODSTAJĄCYCH

2.1 Etapy analizy wyników pomiaru w eksperymentach międzylaboratoryjnych

Zgodnie z normą międzynarodową [6] analiza danych dostarczonych przez laboratoria powinna obejmować następujące etapy:

- a) weryfikacja danych w celu zidentyfikowania wartości odstających lub innych nieprawidłowości,
- b) obliczenia wstępnych wartości określających precyzję i średnie dla każdego poziomu osobno,
- c) ustalenie końcowych wartości określających precyzję i średnie oraz zależności między precyzją a poziomem (jeśli taka zależność istnieje).

Każdy organizator badań musi się liczyć z uzyskaniem od laboratoriów wyników znacznie odbiegających od przeciętnych i mieć opracowaną procedurę postępowania z takimi wartościami. Możliwe są różne podejścia do tej kwestii. Jedno z nich polega na sformułowaniu kryteriów odrzucania wartości nietypowych i wyeliminowaniu ich ze zbioru, zanim zostaną ustalone ostateczne miary położenia rozkładu eksperymentalnego (średnia globalna) oraz rozrzutu danych (np. granica powtarzalności i granica odzwierciedlenia). Innym podejściem jest zastosowanie parametrów statystycznych odpornych na wartości skrajne, jak mediana czy rozstęp ćwiartkowy IQR. Wreszcie kolejnym sposobem postępowania jest zabieg przycinania danych skrajnych, zwany winsoryzacją (ang. *winsorizing*), np. zgodnie z algorytmami A i S, opisanymi w załączniku do normy [3]. Są też takie metody odporne, które nie odrzucają ani nie przycinają danych podejrzanych, ale przypisują im mniejszą wagę (np. dwuwagowa funkcja Tukey'a). Te dwa ostatnie podejścia biorą pod uwagę fakt, że nie zawsze jest słuszne twierdzenie, że to większość laboratoriów dostarcza „prawdziwe” wyniki.

Analiza wyników powinna uwzględniać wnioskowanie statystyczne, ale ponadto wiele aspektów całego kontekstu badań. Służy temu połączenie graficznych (wizualnych) i liczbowych metod sprawdzania spójności wyników dostarczonych przez laboratoria.

2.2 Testy stosowane w badaniach międzylaboratoryjnych do wykrywania wartości nietypowych i odstających

Niemal wszystkie procedury analizy danych stosowane w badaniach międzylaboratoryjnych zalecają wykrywanie najpierw odstających wariacji. Do sprawdzania homogeniczności wariacji służą trzy testy: Bartletta, Hartley'a oraz Cochra. Dwa pierwsze z wymienionych są jednak wrażliwe na bardzo małą wariację, więc nie powinny być stosowane w badaniach międzylaboratoryjnych, gdzie nierzadko zdarza się, że jedna lub więcej wariacji wewnątrzlaboratoryjnych jest równa zeru.

Za pomocą testu Cochra ocenie statystycznej poddawana jest maksymalna wartość odchylenia standardowego powtarzalności. Po wyeliminowaniu tej wartości można test stosować wielokrotnie, sukcesywnie odrzucając wartości skrajne. Test Cochra został opracowany dla eksperymentów zrównoważonych (ang. *uniformlevel designs*) i jest ścisły tylko wtedy, gdy wszystkie odchylenia standardowe policzono dla takiej samej liczby wyników [7]. Jeśli jednak odstępstwa od tego warunku, spowodowane odrzuceniem danych lub ich niedostarczeniem, są niewielkie, to test może być nadal używany. Dlatego w porównaniach międzylaboratoryjnych test Cochra jest stosowany bardzo często.

Inne podejście do wykrywania nadmiernych wariancji laboratoryjnych podaje niemiecka norma DIN 38 402, w której zaleca się stosowanie testu F do porównania poszczególnych wewnątrzlaboratoryjnych wariancji z wariancją odtwarzalności. Ponieważ udział składnika międzylaboratoryjnego może być silnie dominujący w wariancji odtwarzalności, konieczne jest poprzedzenie tego testu odrzuceniem wyników, które dały odstające wartości średnie.

W badaniach międzylaboratoryjnych testy wariancji przeprowadza się zawsze jako jednostronne, a więc tylko duża wariancja jest traktowana jako odstająca.

Natomiast do wykrywania odstających średnich arytmetycznych, uzyskanych przez poszczególnych uczestników, stosuje się testy wykrywające zarówno małe jak i duże wartości skrajne (oba krańce uporządkowanych ciągów danych). W dawniejszych publikacjach i normach polecano stosowanie do tego celu testu Dixona, który jest bardzo łatwy w użyciu. Jego wadą jest słaba moc statystyczna a także maskowanie efektów, gdy występuje wiele wartości odstających. Nowsze dokumenty [6] zalecają głównie test Grubbsa, choć stosuje się też test Grafa–Henninga. Główna różnica między nimi polega na tym, że test Grubbsa w pierwszym szacowaniu s i średniej arytmetycznej uwzględnia wszystkie dane, zaś drugi z testów pomija w obliczeniach wstępnych wartość podejrzaną. W badaniach biegiłości z udziałem dużej liczby laboratoriów i powtórzeń wykazano, że wielokrotne stosowanie testu Grubbsa pojedynczych wartości wykrywa więcej wartości odstających niż test Grafa–Henninga [8]. Istnieją też inne testy o podobnym przeznaczeniu, nie są one jednak przywoływane w oficjalnie publikowanych procedurach zalecanych do stosowania w badaniach międzylaboratoryjnych.

Często łączy się testy wykrywania pojedynczych wartości odstających z wykrywaniem dwóch takich wartości. F.E. Grubbs opracował, oprócz testu dla pojedynczych wartości odstających (wartość najwyższa lub najniższa), test dla dwóch wartości (dwie wartości najwyższe lub dwie najniższe). Modyfikacją tego ostatniego jest test dla dwóch wartości odstających, które mogą znajdować się albo po jednej, albo po obu stronach uporządkowanego ciągu danych.

Przy dużym rozproszeniu wyników lub występowaniu skośnych rozkładów stosuje się niekiedy percentyle. Zaleca się jednak zachowanie ostrożności przy stosowaniu tej metody.

Wartości zidentyfikowane jako odstające na poziomie istotności między 1% and 5% są traktowane jako niepewne (ang. *stragglers*), a tylko te, które są istotne na poziomie mniejszym od 1% są traktowane jako odstające (ang. *outliers*) i eliminowane z dalszych analiz. Dane niepewne są usuwane tylko wtedy, gdy ustalono przyczynę ich pojawienia się. W każdym przypadku powinny one być jednak zaznaczone, aby ułatwić późniejsze interpretacje i wyciąganie wniosków z badań.

2.3 Graficzne metody przedstawiania wyników

Interpretację wyników badań międzylaboratoryjnych bardzo ułatwia posiłkowanie się różnorodnymi metodami graficznym. Powszechnie do tego celu stosowane są wykresy punktowe oraz słupkowe, a także karty kontrolne.

Ważną formą prezentacji danych eksperymentalnych są też histogramy częstości – szczególna postać wykresów słupkowych – które obrazowo przedstawiają rozkład wyników i pozwalają wykryć wielomodalność. Coraz częściej stosowane są wykresy ramkowe (ang. Box & Whisker Plot), które pokazują rozrzuty w poszczególnych laboratoriach oraz rozkład średnich laboratoryjnych; na ogół są one tworzone jako uzupełnienie analizy wariancyjnej (ANOVA).

Z kolei wykres Youdena, wraz z elipsami obrazującymi przedziały ufności, ułatwia interpretację eksperymentów z zastosowaniem próbek dzielonych (rozszczepionych) – dwóch grup wyników o znikomą różniących się poziomach [3]. Ponadto stosuje się wykresy przedstawiające odchylenia standardowe jako funkcję średniej wartości cechy (np. stężenia w klasie).

Typowymi dla badań międzylaboratoryjnych są wykresy słupkowe uporządkowanych rosnąco wartości wskaźnika z (ang. z -score) oraz wykresy obrazujące statystyki zaproponowane przez Johna Mandela [9-11, 6], omówione bardziej szczegółowo w dalszej części artykułu.

3. STATYSTYKI MANDELA

3.1 Charakterystyka statystyk Mandela

Jedną z ważnych zalet wykresów obrazujących statystyki Mandela jest ich zwięzłość. Przy użyciu zaledwie dwóch wykresów można zestawić osiągnięcia wszystkich laboratoriów uczestniczących w projekcie wspólnym i porównać je tak pod względem wartości średnich jak i rozrzutu wyników na wszystkich poziomach badania. Dlatego statystyki Mandela są stosowane nie tylko w eksperymentach wyznaczenia precyzji, ale też w innych rodzajach porównań międzylaboratoryjnych.

Przyjmuje się oznaczenia:

p – liczba laboratoriów ($1, \dots, i, \dots, p$)

q – liczba poziomów ($1, \dots, j, \dots, q$)

n – liczba badań w każdej klasie ($1, \dots, k, \dots, n$).

Wstępne analizy wyników prowadzi się osobno dla każdego j -tego poziomu badania. Na początku analizuje się klasy wyników, czyli wyniki na danym poziomie otrzymane w jednym laboratorium. W każdej ij -tej klasie oblicza się średnie i odchylenia standardowe, na podstawie których wylicza wskaźniki k_{ij} i h_{ij} .

Następnie należy sporządzić wykresy wartości k_{ij} i h_{ij} z wartościami uporządkowanymi w kolejności numerów przypisanych laboratoriom. Obliczone wartości porównuje się z wartościami krytycznymi odpowiadającymi poziomowi istotności $\alpha = 0,05$ oraz $\alpha = 0,01$. Wynik uznaje się za niepewny, jeśli znajduje się w zakresie pomiędzy wartościami granicznymi ustalonymi dla powyższych poziomów istotności. Jeśli wynik przekracza wartość graniczną dla poziomu istotności $\alpha = 0,01$, to uznaje się go za odstający.

3.2 Przeznaczenie i budowa statystyki Mandela k

Statystyka k służy do badania zgodności wewnątrzlaboratoryjnej. Odnosi odchylenia standardowe poszczególnych laboratoriów s_{ij} do odchylenia standardowego powtarzalności s_j .

Wynikiem eksperymentu wspólnego jest zbiór wartości k_{ij} obliczonych ze wzoru:

$$k_{ij} = \frac{s_{ij}}{s_j} \quad (1)$$

s_i – odchylenie standardowe i -tego laboratorium,

s_r – odchylenie standardowe powtarzalności (łącznie odchylenie standardowe wewnątrzklasowe)

$$s_j = \sqrt{\frac{\sum_{i=1}^p (n_j - 1) s_j^2}{\sum_{i=1}^p (n_j - 1)}} \quad (2)$$

Jeśli wszystkie laboratoria dostarczyły na każdym poziomie taką samą liczbę wyników, tzn. $n_j = n$, to wzór (2) przyjmuje prostszą postać:

$$s_j = \sqrt{\frac{\sum_{i=1}^p s_{ij}^2}{p}} \quad (3)$$

Ponieważ wartości odchylenia standardowego są dodatnie, lub w skrajnym przypadku równe zero, wskaźnik k_{ij} również nie może być ujemny.

Aby wskazać te odchylenia standardowe, które są istotnie większe od s_{ij} , stosuje się test jednostronny. Wartość krytyczną testu k_c określa wzór:

$$k_c = \sqrt{\frac{p \cdot F_{\alpha, f_1, f_2}}{F_{\alpha, f_1, f_2} + (p-1)}} \quad (4)$$

w którym:

F_{α, f_1, f_2} – kwantyl rzędu $1-\alpha$ rozkładu Fishera-Snedecora o f_1 i f_2 stopniach swobody,
 α – poziom istotności; zwykle $\alpha = 0,05$ lub $\alpha = 0,01$.

Liczba stopni swobody f_1 i f_2 wynosi odpowiednio:

$$f_1 = n - 1 \quad f_2 = (n - 1)(p - 1) \quad (5)$$

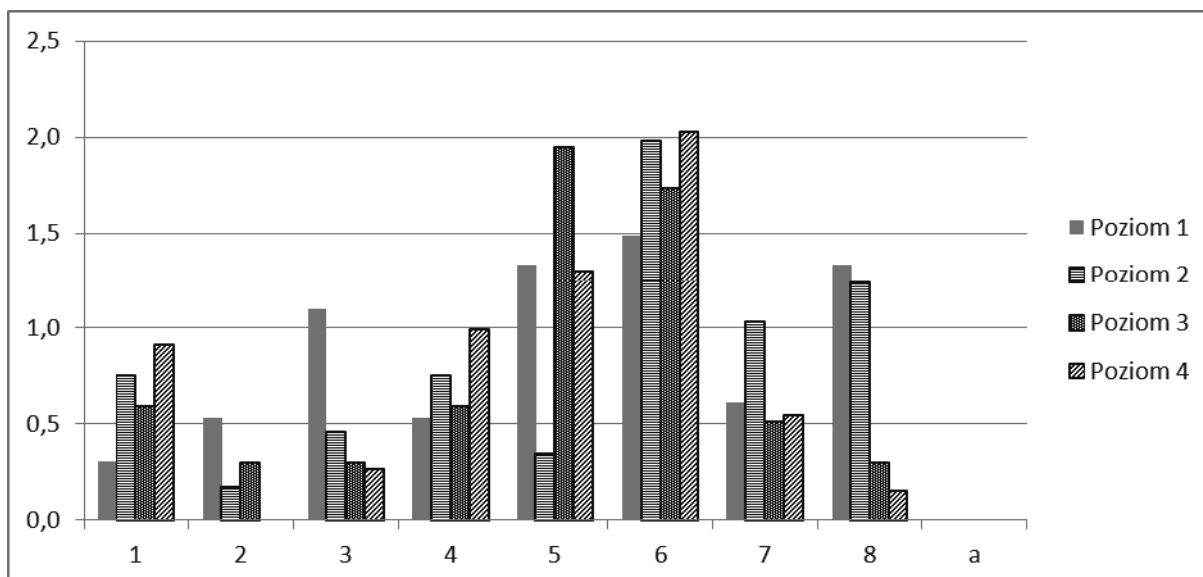
Jeśli $k_{ij} > k_c$, to odchylenie s_{ij} w i -tym laboratorium jest raczej nadmierne i przyjmuje się, że:

- dla $\alpha = 0,05$ wartość jest niepewna, in. wątpliwa,
- dla $\alpha = 0,01$ wartość jest odstająca.

W przypadkach, gdy z wykresu obliczonych wartości statystyki k wynika, że dane laboratorium uzyskało wiele dużych wyników i odróżnia się tym od pozostałych, zaleca się wyjaśnienie tego faktu.

Na rys. 1 przedstawiono wyniki badań międzylaboratoryjnych dotyczących oznaczenia siarki w węglu. W eksperymencie wzięło udział 8 laboratoriów, pomiary przeprowadzono na 4 poziomach, powtarzając pomiary trzykrotnie. Wartości krytyczne k_c , równe dla poziomów istotności 0,05 i 0,01 odpowiednio 1,67 i 1,96 oznaczono liniami poziomymi. Sposób obliczeń wartości k_c pokazano w p.3.3, można też je znaleźć w formie tabelarycznej [6, 12])

Trzeba zaznaczyć, że chociaż testem Mandela k sprawdza się tylko duże wartości statystyki k_{ij} , porównując je z wartościami krytycznymi, to jednak na wykresach łatwo jest zauważyć również bardzo małe wartości tego wskaźnika, w tym zerowe. Warto na nie zwrócić uwagę, bowiem ich przyczyną może być niedostateczna rozdzielczość urządzenia pomiarowego lub nadmierne zaokrąglenie wyników.



Rys. 1. Przykład wyników badań międzylaboratoryjnych zinterpretowanych za pomocą statystyki Mandela k (zawartość siarki w węglu; $p = 8$, $n = 3$)

3.3 Przykład zastosowania statystyki k z interpretacją

Należy wyznaczyć wartości krytyczne statystyki Mandela k dla programu badań porównawczych z udziałem 11 laboratoriów. Przyjąć dwa poziomy istotności testu: $\alpha = 0,05$ oraz $\alpha = 0,01$. Przewiduje się, że każde laboratorium dostarczy dwa wyniki pomiaru na danym poziomie.

Obliczenia: $p = 11$ $n = 2$ $f_1 = n - 1 = 1$ $f_2 = (n - 1)(p - 1)$

Dla $\alpha = 0,05$, stosując funkcję Excela =ROZKŁAD.F.ODW(0,05;1;10), otrzymujemy $F_{0,05;1;10} = 4,965$.
Zatem:

$$k_c = \sqrt{\frac{11 \cdot 4,965}{4,965 + 11 - 1}} = 1,91$$

Podobnie dla $\alpha = 0,01$, stosując funkcję =ROZKŁAD.F.ODW(0,01;1;10), otrzymujemy $F_{0,01;1;10} = 10,044$.
Zatem:

$$k_c = \sqrt{\frac{11 \cdot 10,044}{10,044 + 11 - 1}} = 2,35$$

Aby rozróżnić wartości k_c obliczone dla różnych poziomów istotności α , wprowadzono dodatkowy indeks dolny umieszczony w nawiasie:

$$k_{c(0,05)} = 1,91 \quad k_{c(0,01)} = 2,35$$

Te same wartości można odnaleźć w jednej z tabel normy ISO 5725-2.

Zestawienie wyników dostarczonych przez laboratoria oraz obliczenia wykonane na klasach wyników zawiera tab. 1. Ponieważ dane dotyczą tylko jednego poziomu, w oznaczeniach pominięto indeks dolny j .

Tabela 1. Przykład wyznaczenia statystyk Mandela k

Nr lab i	pomiar 1	pomiar 2	s_i	k_i
1	3,52	3,57	0,035	0,408
2	3,66	3,54	0,085	0,979
3	3,75	3,66	0,064	0,735
4	2,99	3,12	0,092	1,061
5	3,29	3,11	0,127	1,469
6	3,56	3,44	0,085	0,979
7	4,04	3,95	0,064	0,735
8	3,45	3,46	0,007	0,082
9	3,63	3,39	0,170	1,959
10	3,56	3,51	0,035	0,408
11	3,31	3,22	0,064	0,735
			$s_r = 0,087$	

W podanym przykładzie w każdej klasie znajdują się tylko po dwa wyniki. W takich przypadkach do obliczenia odchylenia standardowego można stosować wygodniejszą postać ogólnie znanego wzoru, a mianowicie:

$$s_i = \frac{|y_{i1} - y_{i2}|}{\sqrt{2}}$$

$$s_r = \sqrt{\frac{\sum_{i=1}^{11} s_i^2}{11}} = 0,087$$

Jak wynika z ostatniej kolumny tabeli 1, wynik laboratorium 9 jest niepewny.

3.4 Przeznaczenie i budowa statystyki Mandela h

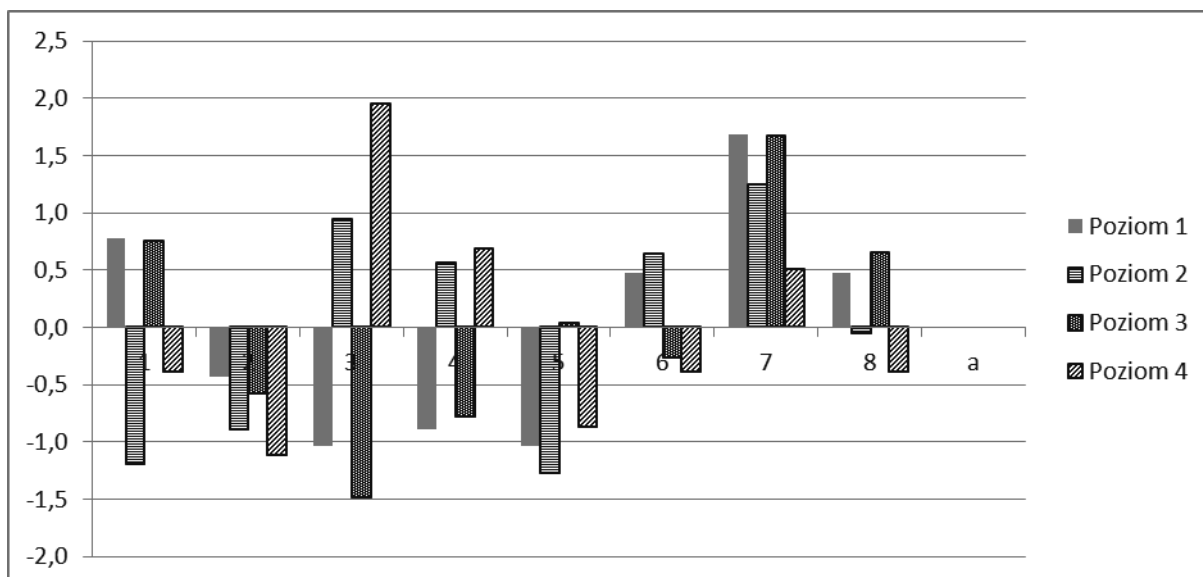
Statystyka h służy do badania zgodności międzylaboratoryjnej. Miarą zgodności wyników jest wartość standaryzowanej różnicy między średnim wynikiem danego laboratorium \bar{y}_i a średnią ogólną $\bar{\bar{y}}$.

$$h_{ij} = \frac{\bar{y}_{ij} - \bar{\bar{y}}}{s_{mj}} \quad (6)$$

gdzie:

s_{mj} – odchylenie standardowe średnich wyników z poszczególnych laboratoriów

$$s_{mj} = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (\bar{y}_{ij} - \bar{\bar{y}})^2} \quad (7)$$



Rys. 2. Przykład wyników badań międzylaboratoryjnych zinterpretowanych za pomocą testu Mandela h (zawartość siarki w węglu; $p = 8$, $n = 3$)

Ponieważ wartość odchylenia standardowego s_m jest dodatnia, wskaźnik h_i może przyjmować zarówno wartości dodatnie jak ujemne, zależnie od tego czy średnia z danego laboratorium \bar{y}_i jest większa czy też mniejsza od średniej ogólnej \bar{y} .

Aby wskazać odstające wartości średnie z laboratoriów \bar{y}_i (istotnie różne od \bar{y}), stosuje się test dwustronny. Wartość krytyczną testu h_c określa wzór:

$$h_c = \frac{(p-1)t_{\alpha,f}}{\sqrt{p(t_{\alpha,f}^2 + p-2)}} \quad (8)$$

gdzie:

α – poziom istotności; zwykle $\alpha = 0,05$ lub $\alpha = 0,01$

$t_{\alpha,f}$ – kwantyl rzędu $1 - \alpha/2$ rozkładu t -Studenta o f stopniach swobody, gdzie $f = p - 2$

Jeśli $|h_i| > h_c$, to wynik w i -tym laboratorium można uznać za niepewny (gdy $\alpha = 0,05$) lub odstający (gdy $\alpha = 0,01$).

Na rys. 2 przedstawiono wyniki obliczeń wskaźnika h_{ij} dla oznaczenia siarki w węglu, posługując się tym samym zbiorem danych wejściowych co w punkcie 2.2. Wartości krytyczne h_c , dla poziomów istotności 0,05 i 0,01 równe odpowiednio 1,75 i 2,06 ($p = 8$, $n = 3$), oznaczono dwiema parami poziomych linii. Na wykresach statystyki h mogą pojawić się różne konfiguracje słupków. Na przykład wszystkie laboratoria mogą na różnych poziomach eksperymentu uzyskać zarówno dodatnie jak i ujemne wartości h . Może też dane laboratorium mieć na wszystkich poziomach wartości tylko dodatnie lub tylko ujemne. Taka sytuacja może sugerować obciążenie laboratorium, zwłaszcza jeśli wartości h dla innych laboratoriów są przeciwnego znaku. Wyjaśnienia wymaga również sytuacja, gdy wartości h dla któregoś z laboratoriów są ekstremalne i wydaje się że w sposób systematyczny zależą od wartości poziomu.

3.5 Przykład zastosowania statystyki Mandela h z interpretacją

Należy wyznaczyć wartości krytyczne testu Mandela h dla programu badań porównawczych z udziałem 11 laboratoriów. Każde laboratorium dostarcza dwa wyniki pomiaru. Przyjść dwa poziomy istotności testu: $\alpha = 0,05$ oraz $\alpha = 0,01$.

Tablica 2. Przykład wyznaczenia statystyk Mandela h

Nr lab <i>i</i>	pomiar 1	pomiar 2	\bar{y}_i	h_i
1	3,52	3,57	3,545	0,226
2	3,66	3,54	3,600	0,442
3	3,75	3,66	3,705	0,856
4	2,99	3,12	3,055	-1,705
5	3,29	3,11	3,200	-1,134
6	3,56	3,44	3,500	0,048
7	4,04	3,95	3,995	<u>1,999</u>
8	3,45	3,46	3,455	-0,129
9	3,63	3,39	3,510	0,088
10	3,56	3,51	3,535	0,186
11	3,31	3,22	3,265	-0,878
			$\bar{\bar{y}} = 3,488$	
			$s_m = 0,254$	

$$\bar{\bar{y}} = \frac{1}{11} \sum_{i=1}^{11} \bar{y}_i = 3,488$$

$$s_m = \sqrt{\frac{1}{10} \sum_{i=1}^{11} (\bar{y}_i - \bar{\bar{y}})^2} = 0,254$$

Obliczenia wartości krytycznych: $p = 11$ $n = 2$ $f = p - 2 = 9$

Dla $\alpha = 0,05$, stosując funkcję Excela =ROZKŁAD.T.ODW(0,05;9), otrzymujemy $t_{0,05;9} = 2,262$.

Zatem:

$$h_c = \frac{(11-1) \cdot 2,262}{\sqrt{11(2,262^2 + 11-2)}} = 1,82$$

Podobnie, dla $\alpha = 0,01$, stosując funkcję =ROZKŁAD.T.ODW(0,01;9), otrzymujemy $t_{0,01;9} = 3,250$.

Zatem:

$$h_c = \frac{(11-1) \cdot 3,250}{\sqrt{11(3,250^2 + 11-2)}} = 2,2$$

Analogicznie jak w przypadku wartości krytycznych k_c , aby rozróżnić wartości h_c obliczone dla różnych poziomów istotności α , wprowadzono dodatkowy indeks dolny umieszczony w nawiasie:

$$h_{c(0,05)} = 1,82 \quad h_{c(0,01)} = 2,22$$

Jak pokazują wartości w ostatniej kolumnie tabeli 2, wynik laboratorium 7 zawiera się pomiędzy $h_{c(0,05)} = 1,82$ a $h_{c(0,01)} = 2,22$, a więc jest niepewny, zaś pozostałe nie budzą zastrzeżeń.

4. POSTĘPOWANIE Z WYNIKAMI NIEPEWNYMI I ODSTAJĄCYMI

Z analizy wykresów statystyk Mandela k i h może wynikać, że niektóre laboratoria wykazują układ wyników wyraźnie różny od konfiguracji pozostałych wyników podlegających analizie. W przypadkach wątpliwych zaleca się zgrupowanie wyników nie tylko według laboratoriów lecz niezależnie od tego w innym układzie – według poziomów badań.

Norma ISO 5725 zaleca następujące połączenie graficznych (wizualnych) i liczbowych metod sprawdzania zgodności wyników (tab. 3). Dokument ten, rekomendując zastosowanie statystyk Mandela do przedstawiania wyników uzyskanych na wszystkich poziomach, nie postuluje automatycznego odrzucania wartości odstających na podstawie testów k i h . Zaleca aby decyzję końcową podejmować dopiero po zastosowaniu testów Cochraha i Grubbsa.

Tablica 3. Techniki statystyczne zalecane do stosowania w celu oceny zgodności w badaniach międzylaboratoryjnych [6]

	Ocena zgodności wewnątrzlaboratoryjnej	Ocena zgodności międzylaboratoryjnej
Techniki graficzne	statystyka Mandela k	statystyka Mandela h
Techniki numeryczne (obliczeniowe)	test Cochraha C	test Grubbsa G

Jeśli w niektórych laboratoriach obserwuje się na wielu poziomach duże lub bardzo małe wartości wariancji wewnątrzklasowej lub ekstremalne wartości średnie dla wielu poziomów, to zaleca się aby organizator nawiązał z takimi laboratoriami kontakt w celu wyjaśnienia zaobserwowanych odstępstw. Specjalista statystyk może w zależności od ustaleń:

- zaaprobować chwilowo dane uzyskane od laboratorium,
- zaproponować powtórzenie pomiarów przez dane laboratorium (jeśli jest to wykonalne),
- odrzuć dane z tego laboratorium z dalszego opracowywania.

Jeśli w teście zgodności zostaje wykryty jakiś punkt podejrzany, to laboratorium, w którym powstała ta wartość, powinno być proszone o wyjaśnienie czy nie powstał błąd techniczny (przy przepisywaniu, zła próbka, zła kalibracja itp.). Jeśli nie można zastąpić wyniku podejrzanego, to odstający należy usunąć (chyba, że jest uzasadniony powód, by go zatrzymać), a wynik niepewny można pozostawić.

Jeśli w laboratorium wykryto testem Mandela k (lub testem Cochraha) wiele wartości odstających dla różnych poziomów badania, może to świadczyć o wyjątkowo dużej wariancji w tym laboratorium (nie-dostatecznie opanowana metoda, różni operatorzy, niewłaściwe przyrządy, warunki otoczenia, niedbałe wykonanie) i wszystkie wyniki z tego laboratorium powinny być odrzucone.

Jeśli w laboratorium wykryto testem Mandela h (lub testem Grubbsa) wiele wartości odstających dla różnych poziomów badania, może to wskazywać na wyjątkowo duże obciążenie tego laboratorium. W przypadku znalezienia przyczyny (duży błąd systematyczny przy kalibracji, błędy obliczeniowe, zła formuła) dane mogą być poprawione po powtórnej kalibracji lub przeliczeniach. W przeciwnym razie dane należy odrzucić.

Założeniem opisanych wyżej metod jest jednomodalny rozkład populacji. Aby sprawdzić słuszność tego założenia warto jest oprócz wykresów k i h sporządzić histogramy dla średnich w klasach oraz dla rozstępów lub odchyłeń standardowych. Ujawnienie, że rozkład danych jest dwumodalny, może świadczyć o zmieszanej populacji wyników, z powodu różnych metod, zanieczyszczonych próbek lub źle opracowanych instrukcji.

5. PODSUMOWANIE

Występowanie wartości, które wydają się niezgodne z wynikami z innych laboratoriów, może znacząco zmienić oszacowanie wyznaczanych parametrów. W związku z tym organizator powinien, mieć opracowane jednoznaczne procedury dotyczące wykrywania takich wartości i postępowania z nimi.

Statystyki Mandela są bardzo użytecznym i nieskomplikowanym narzędziem analizy danych pozyskanych od laboratoriów. Wskazane jest jednak aby wyniki obliczeń były zinterpretowane przez specjalistę z doświadczeniem technicznym i statystycznym, który potrafi dostrzec i zbadać anomalie, jakie mogą wystąpić w pojedynczych eksperymentach międzylaboratoryjnych czy też w kolejnych rundach badań biegłości.

LITERATURA

- [1] PN-EN ISO/IEC 17025:2005 (2007) Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcujących.
- [2] PN-EN ISO/IEC 17043:2011 Ocena zgodności -- Ogólne wymagania dotyczące badania biegłości.
- [3] ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparisons.
- [4] Przewodnik ILAC-G13:2000. Wytyczne dotyczące wymagań odnośnie kompetencji organizatorów badania biegłości.
- [5] Thompson M., Ellison S.L.R., Wood R.: The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report), 2006.
- [6] PN-ISO 5725-2:2002 Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów. Część 2. Podstawowe metody określania powtarzalności i odtwarzalności standardowej metody pomiarowej.
- [7] Pestiaux P., Pic Ch.: The limitations of the Cochran and Grubbs outlier tests in round robin testing. SAE Paper, No. 2004-01-1894, June 2004.
- [8] Linsinger T.P.J., Kandler W., Krska R., Grasserbauer M., Accred. Qual. Assur. 3 (1998) 322–327.
- [9] Mandel J.: The validation of measurement through interlaboratory studies. Chemometrics and Intelligent Laboratory Systems; vol. 11, 1991, 109-119.
- [10] Mandel J.: Structure and outliers in interlaboratory studies. Journal of Testing and Evaluation, Vol. 25, 1995, 364-369.
- [11] Mandel J.: The Statistical Analysis of Experimental Data.
- [12] Wilrich P.-T.: Critical values of Mandel's h and k, the Grubbs and the Cochran test statistic. Advances in Statistical Analysis, 97(1), 2013, 1-10.

Mariusz Mastalerz
Instytut Chemicznej
Przeróbki Węgla

Praktyczne wykorzystanie normy ISO 13528:2005 – doświadczenia Ośrodka Badań Biegłości Instytutu Chemicznej Przeróbki Węgla

Ośrodek Badań Biegłości (OBB) Instytutu Chemicznej Przeróbki Węgla (IChPW) posiada wdrożony system zarządzania zgodny z wymaganiami zawartymi w normie PN-EN ISO/IEC 17043:2011 potwierdzony certyfikatem akredytacji organizatora badań biegłości PT 004, wydanym przez Polskie Centrum Akredytacji. OBB jest pierwszym i jedynym w Polsce akredytowanym organizatorem programów badań biegłości w zakresie analiz: biomasy, węgla brunatnego, koksu i odpadów paleniskowych. Ośrodek Badań Biegłości IChPW również jako jedyny w kraju organizuje międzynarodowe badania biegłości w zakresie analiz koksu z udziałem laboratoriów z Polski, Słowacji, Czech, Ukrainy, Niemiec, Francji, Włoch, Hiszpanii, Finlandii, Austrii, Wielkiej Brytanii. W publikacji przedstawiono doświadczenia OBB z wykorzystania metod analiz statystycznych stosowane podczas realizacji rund badań biegłości. Podsumowano słabe i mocne strony normy ISO 13528:2005.

1. WPROWADZENIE

1.1 Ośrodek Badań Biegłości – akredytowany organizator badań biegłości

Instytut od 2005 roku jest organizatorem badań biegłości w zakresie analiz węgla kamiennego, biomasy, węgla brunatnego, koksu i odpadów paleniskowych. W 2008 roku PCA, w dokumencie DAPT-01 [1] określiło wymagania dotyczące uzyskania certyfikatu akredytacji przez organizatorów badań biegłości. W tym samym roku w IChPW wyodrębniono strukturę Ośrodka Badań Biegłości (OBB) oraz rozpoczęto prace nad opracowaniem dokumentacji i wdrożeniem systemu zarządzania zgodnego z ILAC-G13:08/2007 [2]. W grudniu 2010 roku Ośrodek Badań Biegłości IChPW otrzymał certyfikat akredytacji ORGANIZATORA BADAŃ BIEGŁOŚCI nr PT 004. Od stycznia 2012 roku OBB posiada akredytację na organizację badań biegłości PT zgodnie z normą PN-EN ISO/IEC 17043:2011 [3], która powołuje normę ISO 13528:2005 „Statistical methods for use in proficiency testing by interlaboratory comparisons” [4] jako odpowiednią do prowadzenia analizy statystycznej.

Zakres akredytacji obejmuje programy badań biegłości w zakresie analiz: węgla kamiennego, węgla brunatnego, biomasy, koksu, odpadów paleniskowych. Należy zaznaczyć, że OBB jest pierwszym i jedynym w Polsce akredytowanym organizatorem programów badań biegłości w zakresie analiz biomasy, węgla brunatnego, koksu i odpadów paleniskowych. Ośrodek Badań Biegłości IChPW również jako jedyny w kraju organizuje międzynarodowe badania biegłości w zakresie analiz koksu z udziałem laboratoriów z Polski, Słowacji, Czech, Ukrainy, Niemiec, Francji, Włoch, Hiszpanii, Finlandii, Austrii, Wielkiej Brytanii.

2. ANALIZA STATYSTYCZNA RUND BADANIA BIEGŁOŚCI ORGANIZOWANYCH PRZEZ OŚRODEK BADAŃ BIEGŁOŚCI ICHPW

2.1 Ocena jednorodności obiektów badań

Zgodnie z normą ISO 13528:2005 [4] Załącznik B próbki uważane są za wystarczająco jednorodne, jeżeli spełnione jest kryterium:

$$s_s \leq 0,3 \hat{\sigma}$$

gdzie:

s_s – jest odchyleniem standardowym pomiędzy próbkami jednostkowymi,

$\hat{\sigma}$ – jest odchyleniem standardowym do oceny biegłości.

Odchylenie standardowe do oceny biegłości jest pojęciem nowym, którego wprowadzenie pozwoliło zunifikować przedstawione w normie modele prowadzenia rund badań biegłości oraz analiz statystycznych niezależnie od dziedziny, w której to badanie jest przeprowadzane.

Na tym etapie realizacji rundy badania biegłości nie dysponujemy wynikami badań próbek, dlatego przyjęto, że odchylenie standardowe biegłości $\hat{\sigma}$, równe jest odpornemu odchyleniu standardowemu s^* wyznaczonemu zgodnie z normą ISO 13528:2005 [4], załącznik C, z wyników uzyskanych przez Uczestników w poprzedniej rundzie. Spełnienie tego warunku potwierdza jednorodność próbek.

Stosując w praktyce kryterium opisane powyższym wzorem stwierdzono, że zastosowany współczynnik 0,3 powoduje zaostrenie stosowanego kryterium do poziomu niemożliwego w realizacji pracy badawczej, laboratorium prowadzącego analizę próbek. Na przykład dla próbek węgla kamiennego badanie jednorodności przeprowadza się na podstawie analizy statystycznej wyników oznaczania zawartości popiołu. Zgodnie z normą przedmiotową warunek powtarzalności tj. dopuszczalna różnica pomiędzy dwiema próbkami analizowanymi równolegle w warunkach powtarzalności wynosi: dla zawartości popiołu do 10% – 0,2%, dla zawartości popiołu powyżej 10% – 2% względnie, ale nie więcej niż 0,5%. Obliczone zgodnie z normą ISO 13528:2005 [4] z wyników uzyskanych w rundzie WK/17/2014 (45 uczestników) kryterium jednorodności dla rundy kolejnej wynosi 0,013%. Z tego powodu w OBB odstąpiono od analizy statystycznej proponowanej w normie na rzecz wzorów zmodyfikowanych opisanych w „The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories” (IUPAC Technical Report 2006) [5]. Zgodnie z zapisami protokołu IUPAC próbki uważane są za wystarczająco jednorodne, jeżeli spełnione jest kryterium (na poziomie istotności 95%):

$$s_{sam}^2 < c$$

Wartość c uwzględnia współczynniki rozkładu normalnego wyników w populacji poprzez powiązanie ze wskaźnikami statystyki χ^2 i F -Snedecora, co w praktyce sprowadza kryterium do poziomu realnego.

2.2 Ocena stabilności obiektów badań

Metody obliczeniowe zastosowane do oceny stabilności obiektów badań są zgodne z normą ISO 13528:2005, Załącznik B.

2.3 Wyznaczanie wartości przypisanej jej niepewności i odchylenia standardowego do oceny biegłości

Prowadzone przez OBB programy dotyczą badań próbek materiałów naturalnych, dlatego jako podstawowy sposób wyznaczania wartości przypisanej przyjmuje się, że jest to wartość uzgodniona z wynikami otrzymanymi przez uczestników. Ze względu na warunek dotyczący niepewności wartości przypisanej:

$$u_x \leq 0,3 \hat{\sigma}$$

gdzie:

u_x – niepewność wartości przypisanej,

$\hat{\sigma}$ – odchylenie standardowe do oceny biegłości.

i jej uwzględniania w wskaźnikach osiągnięć oraz w konsekwencji skorelowania wzorów wskaźników oceny osiągnięć z metodami numerycznymi opisywanymi w normie, wypracowano dwa modele wyznaczania wartości przypisanej, jej niepewności i odchylenia standardowego do oceny biegłości.

2.3.1 Rundy badań biegłości o liczbie uczestników powyżej dziesięciu

Zgodnie z załącznikiem C normy ISO 13528:2005 [4] wartości przypisane oraz wartość odchylenia standardowego do oceny biegłości dla wszystkich badanych parametrów wyznacza się statystyczną metodą odporną. Metoda odporna jest jedną z metod numerycznych. Wynik końcowy uzyskuje się poprzez kolejne coraz to bliższe oszacowania – iteracje. Pierwszym przybliżeniem wartości przypisanej jest wartość mediany. Iteracje prowadzi się tak długo, dopóki wartość przypisana i wartość odchylenia standardowego w danym kroku są równe bądź mniejsze od uzyskanych wartości w kroku poprzednim. Norma ISO 13528:2005 [4] podaje parę wariantów obliczania wartości niepewności wyznaczania wartości przypisanej. W OBB przyjęto, że najodpowiedniejszym z punktu widzenia realizacji celów organizowanych rund badań biegłości jest wariant opisany w punkcie 5.6.2. normy.

2.3.2 Rundy badań biegłości o liczbie uczestników poniżej dziesięciu

Ośrodek Badań Biegłości stosuje metodę zgodną z normą PN-ISO 5725-2:2002 „*Dokładność (poprawność i precyzja) metod pomiarowych i wyników pomiarów. Część 2: Podstawowa metoda określania powtarzalności i odtwarzalności standardowej metody pomiarowej*” [6], dotyczącą technik minimalizacji wpływu wyników skrajnych, którą przywołuje norma ISO 13528:2005 [4]. Wartość przypisaną, jej niepewność i odchylenie standardowe do oceny biegłości wyznacza się po wcześniejszym przeprowadzeniu testu Grubbsa identyfikującego wartości odstające i niepewne. Zidentyfikowanych wartości odstających i niepewnych nie uwzględnia się przy wyznaczaniu wyżej wymienionych parametrów.

2.4 Obliczenia statystyki osiągnięć

OBB stosuje dwa spośród zaproponowanych w normie wskaźników osiągnięć w zależności od spełnienia kryterium dotyczącego niepewności wartości przypisanej i jej uwzględniania w wskaźnikach osiągnięć: Dla rund, w których kryterium niepewności wartości przypisanej jest spełnione, oblicza się wskaźnik z-score. Natomiast w przypadku rund badań biegłości, w których kryterium niepewności wartości przypisanej nie jest spełnione OBB stosuje do oceny osiągnięć uczestników wskaźnik z'-score.

2.5 Ocena wartości niepewności z jaką uczestnik przeprowadził badanie

Znajomość niepewności pomiaru podawana wraz z wynikami badań jest bardzo ważna dla laboratoriów, ich klientów i wszystkich instytucji, wykorzystujących te wyniki do podejmowania kluczowych decyzji. Zgodnie z normą ISO 13528:2005 [4], w przypadku wyznaczania wartości przypisanej na podstawie wyników uzyskanych od uczestników, wskaźnik E_z -score jest jedynym możliwym do zastosowania,

wskaźnikiem wykorzystującym niepewność wartości przypisanej i niepewność, z jaką uczestnik przeprowadził analizę. Wskaźnik E_z -score (E_{z-} -score i E_{z+} -score) wyrażony jest wzorami:

$$E_{z-} = \frac{x - (X - U_x)}{U_x} \quad E_{z+} = \frac{x - (X + U_x)}{U_x}$$

gdzie:

x – wartość średnia dla danego parametru uzyskana przez uczestnika,

X – wartość przypisana,

U_x – niepewność rozszerzona wartości przypisanej X ,

U_x – niepewność rozszerzona wyniku x uzyskanego przez uczestnika.

Interpretacja oceny wyników badań uzyskanych przez uczestników jest następująca:

$$\begin{array}{lll} -1 \leq E_z \leq 1 & \text{i} & -1 \leq E_{z+} \leq 1 & = \text{poziom zadowolający} \\ -1 > E_z > 1 & \text{lub} & -1 > E_{z+} > 1 & = \text{poziom wątpliwy} \\ -1 > E_z > 1 & \text{i} & -1 > E_{z+} > 1 & = \text{poziom niezadowolający} \end{array}$$

3. PODSUMOWANIE I WNIOSKI

W krótkim zarysie przedstawiono kluczowe elementy obliczeń statystycznych wykonywanych na potrzeby realizacji badań biegłości stosowane przez Ośrodek Badań Biegłości Instytutu Chemicznej Przeróbki Węgla. Prowadzone przez OBB programy dotyczą badań próbek materiałów naturalnych, dlatego jako podstawowy sposób wyznaczania wartości przypisanej przyjmuje się, że jest to wartość uzgodniona z wyników otrzymanych przez uczestników. Analiza statystyczna opisana w normie ISO 13528:2005 [4] dla takiego modelu organizacji badań biegłości jest niewystarczająca. Proponowane kryterium jednorodności nie ma przełożenia praktycznego i w żaden inny sposób nie uwzględnia powtarzalności wykonywania analiz z materiału rozsyłanego w ramach badania biegłości. Stosowanie analizy odpornej do wyznaczania wartości przypisanej i odchylenia standardowego biegłości dla rund o większej ilości uczestników, zgodnie z zapisami normy poprzez skorelowanie wzorów, ogranicza stosowanie wskaźników osiągnięć praktycznie tylko do z-score. Dla rund o małej liczbie uczestników ($p < 10$) norma nie podaje żadnej propozycji analizy statystycznej, co w efekcie prowadzi do dowolności stosowanych przez organizatorów badań PT modeli statystycznych. Jedynym możliwym do zastosowania, wskaźnikiem wykorzystującym niepewność wartości przypisanej i niepewność, z jaką Uczestnik przeprowadził analizę, w przypadku wyznaczania wartości przypisanej na podstawie wyników uzyskanych od uczestników zgodnie z normą ISO 13528:2005 [4], jest wskaźnik E_z -score. Ustalone w normie kryteria poprawności dla tego wskaźnika często powodują wystąpienie sygnału do działania chociaż w rzeczywistości uczestnik poprawnie oszacował wartość niepewności z jaką wykonał badanie.

Z uwagi na fakt, że norma ISO 13528:2005 [4] opisuje modele analizy statystycznej w wyniku których oceniana jest biegłość laboratoriów do prowadzenia określonych badań a otrzymane wyniki są zasadniczym potwierdzeniem kompetencji technicznych laboratorium, bezwzględnie wymaganym przez jednostkę akredytującą, istnieje realna potrzeba jej znowelizowania.

LITERATURA

- [1] DAPT – 01 wydanie 2 Warszawa, 04.12.2008r.
- [2] ILAC-G13:08/2007 Wymagania dotyczące kompetencji organizatorów badań biegłości
- [3] PN-EN ISO/IEC 17043:2011 Ocena zgodności – Ogólne wymagania dotyczące badania biegłości

- [4] ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparisons
- [5] The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories", IUPAC Technical Report 2006
- [6] PN-ISO 5725-2 Dokładność (poprawność i precyzja) metod pomiarowych wyników pomiarów Część 2: Podstawowa metoda określania powtarzalności i odtwarzalności standardowej metody pomiarowej

Praktyczne wykorzystanie normy ISO 13528 w laboratorium badawczym przy organizacji ILC niewielkiej liczby uczestników

W niniejszym referacie przedstawiono korzyści wynikające z uczestnictwa w badaniach biegłości/porównaniach międzylaboratoryjnych (PT/ILC) oraz wytyczne normy ISO 13528:2005 „Statistical methods for use in proficiency testing by interlaboratory comparison”, a także opisano problemy pojawiające się przy organizacji przez laboratorium porównań międzylaboratoryjnych dla niewielkiej liczby uczestników.

1. WPROWADZENIE

Laboratoria chcąc otrzymywać wiarygodne wyniki muszą mieć wdrożony system sterowania jakością (QC; ang. *Quality Assurance*). Badania międzylaboratoryjne, a wśród nich porównania międzylaboratoryjne (ILC; ang. *Interlaboratory Comparison*) są jednym ze sposobów monitorowania miarodajności wyników oznaczeń poprzez [1]:

- **identyfikację problemów dotyczących pomiarów**
wynik niezadowolający jest sygnałem do podjęcia działań korygujących mających na celu zidentyfikowanie potencjalnych źródeł błędów, które mogłyby pozostać niezauważone,
- **porównanie metod lub procedur**
udział w ILC daje możliwość porównania wyników uzyskiwanych w laboratorium przy zastosowaniu różnych metod analitycznych,
- **potwierdzenie kompetencji analityków**
- **zwiększenie osiągnięć**
ILC stwarza możliwość zidentyfikowania obszarów, które wymagają np. dodatkowych szkoleń analityków, modyfikacji metod analitycznych lub wdrożenia nowych, a także daje sygnał do zwiększenia wewnętrznej kontroli jakości w danym obszarze,
- **konsultacje z organizatorem ILC**
po otrzymaniu sprawozdania, uczestnik porównania ma możliwość skontaktowania się z organizatorem ILC w celu otrzymania informacji na temat wyniku lub uzyskania porad dotyczących potencjalnych przyczyn otrzymania wyniku niezadowolającego.

Udział w porównaniach międzylaboratoryjnych jest narzędziem „samopomocy”, które umożliwia uczestnikom zidentyfikowanie nieoczekiwanych źródeł błędów [2], a także pozwala ocenić różne aspekty kompetencji w tym: „charakterystyki jakościowej, porównania uczestników, oznaczeń ilościowych oraz interpretacji zbiorów wyników” [3].

2. ZASADY ORGANIZACJI ILC WEDŁUG ISO 13528:2005

Za całość działań związanych z organizacją badań biegłości/ porównań międzylaboratoryjnych (PT/ ILC) odpowiada KOORDYNATOR, jest on odpowiedzialny za ustalenie planu porównania oraz wybór odpowiedniego obiektu badania. Ponadto ocenia jednorodność, stabilność, określa wartość przypisaną obiektu badania oraz jej niepewność, a także identyfikuje wartości odbiegające i przeprowadza statystyczną ocenę wyników każdego z laboratoriów uczestniczących w (PT/ILC) stosując odpowiednie parametry statystyki indywidualnej oraz sporządza sprawozdanie z badań.

Norma ISO 13528 określa schemat postępowania przy organizacji badań biegłości/ porównań międzylaboratoryjnych, wg którego laboratorium powinny starać się przeprowadzić także wewnętrzne porównania międzylaboratoryjne:

2.1 Jednorodność i stabilność

Ocena wyżej wymienionych parametrów odbywa się zgodnie z procedurą opisaną w Aneksie B normy ISO 13528. Ważnym elementem oceny, zarówno jednorodności, jak i stabilności jest przeprowadzenie jej w warunkach powtarzalności (wybrany parametr oznaczamy podwójnie). Norma ISO 13528 zaleca oceniać jednorodność obiektu badania dla nie mniej niż dziesięciu próbek, jednakże dopuszcza zmniejszenie ilości próbek objętych kontrolą, jeżeli zachowane są wyniki z ostatniej oceny jednorodności i próbki zostały przygotowane wg tych samych procedur. Natomiast ocenę stabilności należy wykonać dla nie mniej niż trzech próbek.

Ocena jednorodności odbywa się przed rozestaniem obiektu badania do laboratoriów uczestniczących w porównaniu, a ocena stabilności w momencie, w którym uczestnicy rozpoczynają analizę obiektu badania.

W celu oceny odpowiedniego stopnia jednorodności należy obliczyć międzypróbkowe odchylenie standardowe:

$$S_s = \sqrt{S_x^2 - \left(\frac{S_w^2}{2}\right)}$$

gdzie: S_x – odchylenie standardowe dla analiz powtarzanych; S_w – odchylenie standardowe wewnątrz próbek (odchylenie powtarzalności próbek podwójnie wykonanych).

Obiekt badań jest wystarczająco jednorodny, jeżeli spełnione jest poniższe kryterium:

$$S_s \leq 0,3\sigma$$

gdzie: σ oznacza odchylenie standardowe dla porównań międzylaboratoryjnych.

Zastosowanie faktora 0,3 oznacza, że w przypadku spełnienia kryterium jednorodności wartość międzypróbkowego odchylenia standardowego przyczynia się nie więcej niż o ok 10% do wzrostu odchylenia standardowego dla porównań międzylaboratoryjnych.

Wartość S_s wynikająca z niejednorodności obiektu badania uwzględniana jest w niepewności wartości przypisanej, zarówno w przypadku spełnienia jak i niespełnienia powyższego kryterium [5].

Natomiast obiekt badań jest wystarczająco stabilny, gdy spełnione jest poniższe kryterium:

$$|\bar{x} - \bar{y}| \leq 0,3\sigma$$

gdzie: \bar{x} – średnia wartość oznaczanego parametru dla próbek jednostkowych otrzymana podczas badania jednorodności; \bar{y} – średnia wartość oznaczanego parametru dla próbek jednostkowych otrzymana podczas badania stabilności.

2.2 Wartość przypisana i jej niepewność

Przedmiotowa norma opisuje pięć sposobów określania wartości przypisanej obiektowi badania:

- certyfikowany materiał odniesienia CRM

wartość przypisana oraz jej niepewność odczytane są z certyfikatu użytego certyfikowanego materiału odniesienia,

- materiał odniesienia RM

w oparciu o wyniki analiz, pomiaru próbek w celu wyznaczenia wartości RM (materiału odniesienia nie posiadającego certyfikatu) poprzez porównanie obiektu badania z certyfikowanym materiałem odniesienia lub wzorcem, mającym odniesienie do wzorca krajowego lub międzynarodowego,

- wartość wynikająca z określonego sposobu przygotowania obiektu badania (*ang. Formulation*)

wartość przypisana określana jest na podstawie znajomości szczegółów dotyczących przygotowania obiektu badań – znanych i dokładnych ilości analitów wykorzystanych w tym celu.

Niepewność wartości przypisanej wyliczana jest na podstawie pomiarów wagowych i objętościowych wynikających z określonego sposobu przygotowania obiektu badania.

- wartość uzgodnioną na podstawie wyników badań otrzymanych przez laboratoria eksperckie w tym przypadku wartość przypisana obliczana jest w oparciu o procedurę opisaną w aneksie C normy ISO 13528 jako średnia odporna, natomiast niepewność standardowa wartości przypisanej obliczana jest na podstawie poniższego wzoru:

$$u_x = \frac{1,25}{p} \cdot \sqrt{\sum_{i=1}^p u_i^2}$$

gdzie: p – liczba laboratoriów eksperckich biorących udział w porównaniu; u_i – niepewność standardowa wyników pomiaru laboratoriów biorących udział w porównaniu.

- wartość uzgodniona na podstawie wyników uczestników ILC

wartość przypisaną wyznacza się przy użyciu metody odpornej opisanej w ISO 13528 – aneks C. Wynik końcowy uzyskuje się przez kolejne iteracje – oszacowania. Iteracje prowadzi się do uzyskania zbieżności (*ang. convergence*). Zgodnie z wytycznymi normy zbieżność zostaje osiągnięta, kiedy po dwóch kolejnych iteracjach spełnione jest założenie o identyczności nowych/ odpornych wartości x^* i s^* .

Dla każdego wyniku uczestnika x_i obliczyć nowe/ odporne x^* i s^* wg wzorów:

$$x^* = \frac{\sum x_i}{p} \quad s^* = 1,134 \sqrt{\frac{\sum (x_i - x^*)^2}{p-1}}$$

Niepewność standardowa wartości przypisanej wyznacza się zgodnie ze wzorem:

$$u_x = \frac{1,25 \cdot s^*}{\sqrt{p}}$$

gdzie: p – liczba laboratoriów biorących udział w porównaniu międzylaboratoryjnym; 1,25 – stała reprezentująca maksymalny stosunek odchylenia standardowego mediany do odchylenia standardowego średniej arytmetycznej dla danych o rozkładzie normalnym.

2.3 Ocena normalności rozkładu

Weryfikacja założenia o normalności wyników, uzyskanych przez laboratoria uczestniczące w porównaniu, przeprowadza się z zastosowaniem odpowiedniego testu np. Shapiro-Wilka. Wynikiem testu jest wartość prawdopodobieństwa testowego p , dla którego wartością graniczną jest 0,05. Otrzymane wyniki mają rozkład normalny, jeżeli otrzymana wartość p jest większa od wartości granicznej.

2.4 Wartości odbiegające

Aby prawidłowo wyznaczyć wartość przypisaną istnieje konieczność wyeliminowania wpływu wyników odbiegających. W tym celu należy zastosować odpowiedni test statystyczny np. Hampel'a, Grubbs'a lub Q-Dixona [5]. Wytyczne dotyczące stosowania testów dla wartości odbiegających podano w normie PN-ISO 5725-2.

Wyniki uznane za odbiegające nie są brane pod uwagę przy obliczaniu wartości przypisanej, natomiast są włączane do obliczeń statystyki osiągnięć.

2.5 Statystyka osiągnięć

Przed dokonaniem wyboru odpowiedniego wskaźnika spośród zaproponowanych w ISO 13528 należy zbadać warunek:

$$u_x \leq 0,3 \sigma$$

gdzie:

u_x – niepewność standardowa wartości przypisanej;

σ – odchylenie standardowe dla badań biegłości/ porównań międzylaboratoryjnym (PT/ILC)

Jeśli niepewność standardowa wartości przypisanej jest zbyt duża w porównaniu z odchyleniem standardowym dla PT/ILC, istnieje ryzyko, że niektóre laboratoria uczestniczące w porównaniu będą otrzymywały wyniki wątpliwe lub niezadowolające z powodu niedokładności w wyznaczeniu wartości przypisanej, a nie z jakiegokolwiek przyczyny wynikającej z błędu danego laboratorium.

Według zasad opisanych w ISO 13528, w przypadku niespełnienia powyższego kryterium koordynator powinien:

- rozważyć wybór innego sposobu wyznaczenia wartości przypisanej, tak żeby jej niepewność standardowa spełniała powyższy warunek,
- użyć niepewności wartości przypisanej w interpretacji wyników porównań międzylaboratoryjnych,
- poinformować laboratoria uczestniczące w porównaniu, o tym że niepewność wartości przypisanej jest istotna.

Wśród przytoczonych w przedmiotowej normie wskaźników klasyfikujących znajdują się:

- **Różnica procentowa**

$$D \% = 100 (x - X) / X$$

gdzie:

x – wynik podany przez laboratorium;

X – wartość przypisana

Kryterium akceptacji opisanego wskaźnika ustalane jest przez koordynatora na podstawie np. cech charakterystycznych wyznaczonych w toku walidacji metody badawczej lub wartości przyjętej na podstawie specyficznych wymagań prawnych [3].

- **Wskaźnik z-scores**

$$z = \frac{x - X}{\sigma}$$

gdzie:

σ – odchylenia standardowe dla badań biegłości/porównań międzylaboratoryjnych.

Norma ISO 13528 podaje, że wskaźnik z-score nie powinien być stosowany gdy liczba laboratoriów uczestniczących w porównaniu międzylaboratoryjnym jest mała.

Ocena wartości obliczonego wskaźnika z-scores odbywa się według poniższych zakresów:

$|z| \leq 2$ wynik uzyskany jest „zadowolający”

$2 < |z| < 3$ wynik uzyskany jest „wątpliwy”

$|z| \geq 3$ wynik uzyskany jest „niezadowolający”

- **Wskaźnik z'**

$$z' = \frac{x - X}{\sqrt{\sigma^2 + u_x^2}}$$

gdzie:

u_x – niepewność standardowa wartości przypisanej

- **Wskaźnik zeta**

$$zeta = \frac{x - X}{\sqrt{u_x^2 + u_X^2}}$$

gdzie:

u_x – niepewność standardowa wyniku uczestnika;

u_X – niepewność standardowa wartości przypisanej.

Kryteria akceptacji dla wskaźników z' i zeta są takie same jak dla z-scores.

W normie ISO 13528 znajduje się informacja stwierdzająca, że wskaźników z' i zeta nie powinno się stosować gdy wartość przypisana obliczana jest z wyników otrzymanych od uczestników porównania międzylaboratoryjnego.

- **Wskaźnik E_n**

$$E_n = \frac{x - X}{\sqrt{U_x^2 + U_{ref}^2}}$$

gdzie:

U_x – niepewność rozszerzona wyniku uczestnika;

U_X – niepewność rozszerzona wartości przypisanej określona w laboratorium referencyjnym.

Kryterium akceptacji rezultatów dla E_n :

$|E_n| \leq 1$ wynik uzyskany jest „zadowolający”

$|E_n| > 1$ wynik uzyskany jest „niezadowolający”

- **Wskaźnik E_z**

Wskaźnik E_z można zdefiniować jako:

$$E_{z-} = \frac{x - (X - U_X)}{U_x} \quad \text{oraz} \quad E_{z+} = \frac{x - (X + U_X)}{U_x}$$

Kryterium akceptacji rezultatów dla E_z :

$-1 \leq |E_{z-}| \leq 1$ i $-1 \leq |E_{z+}| \leq 1$ – wynik uzyskany jest „zadowolający”

$-1 < |E_{z-}| < 1$ lub $-1 < |E_{z+}| < 1$ – wynik uzyskany jest „wątpliwy”

$-1 > |E_{z-}| > 1$ i $-1 > |E_{z+}| > 1$ – wynik uzyskany jest „niezadowolający”

2.6 Odchylenie standardowe dla badań biegłości/porównań międzylaboratoryjnych

Bardzo ważnym i problematycznym tematem jest ustalenie odpowiedniego odchylenia standardowego dla badań międzylaboratoryjnych. Za wybór właściwej metody odpowiada koordynator, który może konsultować się w tej sprawie z uczestnikami badania oraz konsultantami technicznymi.

Zgodnie z normą ISO 13528 odchylenie standardowe dla porównań międzylaboratoryjnych można wyznaczyć jako:

- wartość przyjętą na podstawie specyficznych wymagań np. aktów prawnych,
- wartość przyjętą tak, aby spełnić określone parametry np. zgodnie z oczekiwaną biegłością laboratoriów. Jest to tak zwane odchylenie adekwatne do celu (*ang. fitness for purpose*),
- wartość przyjętą na podstawie modelu ogólnej odtwarzalności metody pomiarowej np. z krzywej Horwitz'a,
- wartość obliczoną na podstawie wyników badań precyzji. Dla metod znormalizowanych z danych dotyczących powtarzalności i odtwarzalności,
- wartość obliczoną na podstawie wyników otrzymanych od laboratoriów uczestniczących w porównaniu. Tego sposobu nie stosuje się gdy liczba laboratoriów uczestniczących w porównaniu jest mała.

3. PROBLEMY PRZY ORGANIZACJI ILC NIEWIELKIEJ LICZBY UCZESTNIKÓW

Organizowanie porównań międzylaboratoryjnych dla niewielkiej liczby uczestników niesie za sobą trudności, z którymi osoby odpowiedzialne za ich zorganizowanie muszą się zmierzyć. Jedną z głównych bolączek organizatorów jest brak ogólnie przyjętych zasad, które opisywałyby w sposób jasny schemat postępowania oraz pozwalałyby na rozwianie mnożących się wątpliwości. Problemy pojawiają się już na etapie planowania, a jednym z pierwszych jest:

3.1 Poufność

Zachowanie wiarygodności, a tym samym zapobieganie znowiu jest jedną z kwestii, z którą zmagają się jednostki organizujące porównania, w których ponadto biorą udział laboratoria będące, częścią tej samej organizacji, grupy organizacji o tym samym profilu czy laboratoria zaprzyjaźnione.

W praktyce trudno jest wyeliminować ryzyko związane z podejrzeniem fałszowania wyników, gdyż standardowe procedury polegające np. na nadawaniu w sposób losowy indywidualnych kodów, znanych tylko uczestnikowi nie mają, w tym przypadku zastosowania. Jednakże podjęcie działań minimalizujących rozwój opisanych powyżej praktyk jest niezbędne. Jednym ze sposobów ograniczających możliwość podważenia przez osoby trzecie otrzymanych rezultatów jest wyodrębnienie niezależnej osoby, która poza wypełnianiem swoich bieżących obowiązków zajmowałaby się przeprowadzaniem porównań począwszy od planowania, przez rozesłanie obiektów badań do opracowania oceny statystycznej i sprawozdania z ILC. Ponadto kluczowa jest świadomość personelu dotycząca celu porównania oraz odpowiedzialności za profesjonalne postępowanie [5].

3.2 Obiekt badań

Zalecane jest, aby materiał do badań był jak najbardziej zbliżony do próbek poddawanych rutynowym analizom, tzn. organizator wybierając obiekty badań powinien kierować się maksymalnie dużym stopniem podobieństwa pod względem stężenia i matrycy do próbek rzeczywistych. Jednak w celu osiągnięcia oczekiwanej jednorodności lub stabilności, a także dostępności próbek rzeczywistych spełnienie po-

wyższych wymagań może być niemożliwe. W takich przypadkach zaleca się stosowanie zatężonych roztworów wzbogaconych, materiałów odniesienia oraz certyfikowanych materiałów odniesienia.

3.3 Analiza danych i ocena rezultatów

Zdecydowana większość dostępnych dokumentów zorientowana jest w kierunku organizacji ILC dla dużej liczby uczestników, stąd problemy podczas organizacji ILC niewielkiej liczby laboratoriów uczestniczących w porównaniu.

Zgodnie z przewodnikiem IUPAC [6] odnoszącym się do badań biegłości organizowanych dla niewielkiej liczby uczestniczących laboratoriów, w przypadku gdy statystyczny rozmiar próbki jest mniejszy niż $N=20$ użyte metody statystyczne stają się niewiarygodne ponieważ:

- niepewność miar położenia (mediana lub średnia) nie jest wystarczająco mała by można ją było bezpiecznie pominąć,
- odchylenia od normalnego rozkładu są trudniejsze do zidentyfikowania,
- stosowanie statystyk odpornych w takim przypadku nie jest zalecane.

W związku z powyższym wartość przypisana nie może być bezpiecznie obliczana z wyników nadesłanych od uczestników jako wartość konsensusu, ponieważ jej niepewność staje się na tyle duża by wpłynąć na ocenę osiągnięć laboratoriów uczestniczących w porównaniu.

Ponadto różnica między parametrami charakteryzującymi populację i próbę znacznie wzrasta wraz z malejącą liczbą N . Średnia dla próby otrzymana od uczestników porównania z N wyników może różnić się od średniej z populacji nawet o $\pm 1,96 \sigma / \sqrt{N}$ dla 0,95 poziomu ufności (gdzie: 1,96 to właściwy percentyl rozkładu normalnego dla 0,95 poziomu ufności, σ jest to odchylenie standardowe populacji). Także odchylenie standardowe z próby różni się od odchylenia standardowego z populacji: przy $N=30$ o ponad 25%, przy poziomie ufności 0,95. Natomiast dla $N < 30$ różnica między estymatorem opisującym próbę i populację wzrasta wraz z zmniejszającym się N , szczególnie dramatyczny wzrost różnic między odchyleniami standardowymi obserwuje się gdy $N < 20$.

3.3.1 Wybór właściwego wskaźnika klasyfikującego

Ocena osiągnięć w sytuacji gdy w porównaniu międzylaboratoryjnym bierze udział niewielka liczba uczestników staje się problemem, ponieważ zdecydowana większość dokumentów poświęconych temu zagadnieniu zaleca stosowanie powszechnie znanych wskaźników, które niestety posiadają pewne obwarowania i wymagania dotyczące swojego zastosowania.

Najczęściej stosowanym parametrem oceny indywidualnej jest wskaźnik z-scores, z przyczyn opisanych we wcześniejszych rozdziałach, stosowanie jego nie jest zalecane przy porównaniach międzylaboratoryjnych organizowanych dla małej ilości uczestników. Jednak, w sytuacji gdy zostanie ustalone odchylenie tzw. adekwatne do celu istnieje możliwość zastosowania wskaźnika z-score nawet w przypadku niewielkiej liczby laboratoriów uczestniczących w porównaniu. W tej sytuacji ważne jest, żeby wartość ta była ustalona przez organizatorów wcześniej, nie powinna być ona zależna od wyników uzyskanych przez uczestników porównania. Zaleca się aby organizator, jako wartość adekwatną do celu, określił właściwy poziom niepewności standardowej, który powinien być zaakceptowany przez uczestników oraz końcowych użytkowników wyniku czyli klientów. „Właściwa” wartość niepewności to taka, która jest na tyle mała, aby decyzje, które będą podejmowane na jej podstawie, nie były błędne, a z drugiej strony nie zbyt niska, aby nie generowała kosztów analizy. Pojęcie adekwatności do celu oznacza, że użyta wartość niepewności będzie minimalizowała koszty związane z analizą i kary związane z błędnymi decyzjami [6]. Jeśli informacje niezbędne do ustalenia odchylenia standardowego adekwatnego do

celu nie są dostępne lub jeśli niepewność standardowa wartości przypisanej jest znacząca, wówczas informacje zawarte w niepewności pomiaru wyników raportowanych przez uczestników porównania powinny być pomocne przy ocenie wyników za pomocą wskaźnika zeta oraz En.

Ma to również zastosowanie w przypadku ILC niewielkiej liczby laboratoriów uczestniczących, które pracują według własnych kryteriów adekwatnych do celu i deklarowana przez nich wartość niepewności może być wykorzystana [6].

4. PODSUMOWANIE

W referacie podjęto próbę opisanie wytycznych przedstawionych w normie ISO 13528:2005, niewątpliwie treść zawarta w niej w sposób uporządkowany przedstawia schemat organizacji badań biegłości/ porównań międzylaboratoryjnych, a także informuje jak rozwiązać kluczowe kwestie związane z wyborem odpowiednich metod zarówno związanych z obiektem badań (ocena jednorodności i stabilności), jak i z oceną statystyczną wyników uzyskanych w trakcie porównań międzylaboratoryjnych. Norma ISO 13528:2005 wskazuje kierunek, ale niestety nie rozwiązuje wszystkich problemów związanych z organizacją porównań międzylaboratoryjnych przeprowadzanych dla małej grupy laboratoriów uczestniczących w nich. Niemniej jednak działając świadomie można, a raczej powinno się ją zaadaptować na potrzeby wewnętrznych porównań międzylaboratoryjnych.

LITERATURA

- [1] EURACHEM, Selection, Use and Interpretation of Proficiency Testing (PT) Schemes, Second Edition 2011.
- [2] Pure Appl. Chem., Vol. 78, No 1, pp. 145-196, 2006.
- [3] Konieczka P., Namieśnik J.: Ocena i kontrola jakości wyników pomiarów analitycznych, Wyd. Naukowo – Techniczne, Warszawa 2007.
- [4] ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparison.
- [5] „Ogólne wymagania dotyczące porównań międzylaboratoryjnych”, instrukcja systemowa Centrum Badań Jakości Sp. z o.o.
- [6] Pure Appl. Chem., Vol. 82, No. 5, pp. 1099-1135; 2010.

Podstawowe testy statystyczne wykorzystywane w trakcie walidacji

Na etapie opracowywania wyników pomiarów analitycznych, w trakcie przeprowadzania procesu walidacji wykorzystywane są różnego rodzaju testy statystyczne. Każde wnioskowanie dotyczące wartości otrzymywanych wyników, dotyczących zwłaszcza ich zgodności z wartościami normowanymi, granicznymi, założonymi, obarczone jest pewnym marginesem niepewności. W wystąpieniu zostaną omówione zasady stosowania oraz opisy najczęściej wykorzystywanych testów statystycznych.

1. WPROWADZENIE

Stosowanie testów statystycznych wymaga znajomości podstaw statystyki matematycznej. Jest ona dziedziną matematyki, która wykorzystuje rachunek prawdopodobieństwa i jest ukierunkowana na badanie prawidłowości pojawiania się określonych cech w przypadku obiektów materialnych lub zjawisk, występujących masowo, tzn. dowolną ilość razy. Statystyka przedstawia te prawidłowości za pomocą liczb. Statystyka stanowi bardzo użyteczne narzędzie, dzięki któremu można znaleźć odpowiedź na wiele pytań. Pozwala bowiem rozstrzygnąć wiele wątpliwości i odpowiedzieć na niemało pytań związanych z naturą procesu analitycznego, np.:

- jak dokładny jest wynik oznaczenia,
- ile oznaczeń powinno się przeprowadzić, żeby zwiększyć precyzję pomiaru,
- czy badany produkt spełnia stawiane mu wymagania i/lub normy?

2. WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Hipoteza to sąd (przypuszczenie) o populacji (badanym obiekcie) oparty na prawdopodobieństwie, przyjęty w celu wyjaśnienia jakiegoś zjawiska, prawa lub faktu, wymagający sprawdzenia. Weryfikacją hipotez nazywa się sprawdzanie sądów o populacji, sformułowanych bez zbadania jej całości. Taka sytuacja ma przecież zawsze miejsce w przypadku przeprowadzenia analiz – wnioskowanie o badanej próbce (zawartość analitu) odbywa się w oparciu o wyniki kilku oznaczeń.

Przebieg procedury weryfikacyjnej wygląda następująco:

1. Sformułowanie hipotezy zerowej i alternatywnej

Hipoteza zerowa H_0 to prosta postać hipotezy poddana testom, z kolei hipoteza alternatywna H_1 jest hipotezą przeciwstawną hipotezie zerowej. Prawdopodobieństwo sumaryczne wystąpienia obu hipotez wynosi 100 %.

2. Wybór odpowiedniego testu

Test służy do sprawdzania hipotezy zerowej. Niezwykle ważny jest wybór testu, który z jednej strony pozwala na sprawdzenie danej hipotezy a z drugiej może być stosowany przy danym zbiorze danych wejściowych.

3. Określenie poziomu istotności α

Poziom istotności α określa maksymalną wartość błędu jaki jest możliwy to popełnienia w trakcie jego stosowania. Jest on funkcją prawdopodobieństwa poprawności wnioskowania $P = 1 - \alpha$.

W trakcie przeprowadzania procesu weryfikacji hipotezy z wykorzystaniem testu statystycznego istnieje jednak możliwość popełnienia dwóch rodzajów błędów:

błędy I rodzaju – odrzucenie hipotezy zerowej H_0 , gdy jest ona prawdziwa;

błędy II rodzaju – przyjęcie hipotezy zerowej H_0 gdy jest ona fałszywa.

4. Wyznaczenie obszaru krytycznego testu

Wielkość obszaru krytycznego wyznacza dowolnie mały poziom istotności α , natomiast jego położenie określa jest przez hipotezę alternatywną. Najczęściej wartość tego obszaru jest określana przez parametr krytyczny testu, którego wartość jest odczytywana, dla danych parametrów badanego zbioru danych, z tablicy rozkładów krytycznych wartości testu.

5. Obliczenie parametru testu na podstawie próby

Wyniki będące podstawą do obliczania parametru testu opracowuje się w odpowiedni sposób, zgodnie z procedurą wybranego testu. Następnie, w oparciu o charakterystyczne dla danego testu równanie obliczany jest parametr, którego wartość będzie porównywana z wartością krytyczną.

6. Podjęcie decyzji

Wyznaczona na podstawie obliczeń w oparciu o wyniki wartość parametru testu jest porównywana jest z wartością krytyczną testu:

- jeżeli wartość ta znajdzie się w obszarze krytycznym, to hipotezę zerową należy odrzucić jako nieprawdziwą – praktycznie oznacza to, że wartość obliczonego parametru testu jest większa od wartości krytycznej testu (odczytanej z odpowiedniej tablicy);
- jeżeli natomiast wartość ta znajdzie się poza obszarem krytycznym, oznacza to brak podstaw do odrzucenia hipotezy zerowej – praktycznie oznacza to, że wartość obliczonego parametru testu nie jest większa od wartości krytycznej testu (odczytanej z odpowiedniej tablicy); stąd wniosek, że hipoteza zerowa może być prawdziwa.

Testy parametryczne służą do weryfikacji hipotez parametrycznych, odnoszących się do parametrów rozkładu badanej cechy w populacji generalnej. Najczęściej za ich pomocą dokonuje się weryfikacji sądów o takich parametrach populacji jak średnia arytmetyczna i wariancja. Testy te konstruowane są przy założeniu znajomości postaci dystrybuanty dla populacji generalnej.

Testy nieparametryczne są stosowane do sprawdzania różnorodnych hipotez, dotyczących m.in. zgodności rozkładu cechy w populacji z określonym rozkładem teoretycznym, zgodności rozkładów w dwóch populacjach, a także losowości doboru próby.

3. PODSTAWOWE TESTY STATYSTYCZNE

Najczęściej stosowanymi w trakcie opracowywania wyników pomiarów analitycznych i/lub walidacji procedur analitycznych wykorzystywanymi testami statystycznymi są testy, które służą do:

- sprawdzania czy w danej serii pomiarowej nie występują wyniki obciążone błędem grubym (odbiegające),
- porównywanie precyzji (wariancji) danej serii pomiarowej z wartością zadaną lub wartościami wariancji dla kilku serii,

– porównywania wartości wyniku pomiaru (średnia arytmetyczna) z wartością zadaną (odniesienia, przypisaną) lub wartości średnich dla kilku serii.

Opis najczęściej stosowanych testów statystycznych zestawiono w poniższych tabelach.

3.1 Test Q-Dixona

cel	sprawdzenie, czy w danym zbiorze wyników nie ma wyniku obciążonego błędem grubym
hipotezy	H_0 – w zbiorze wyników brak obciążonego błędem grubym H_1 – w zbiorze wyników znajduje się wynik obciążony błędem grubym
wymagania	liczność zbioru 3 – 10 z danego zbioru można odrzucić tylko jeden wynik
sposób postępowania	uszeregować wyniki w ciąg niemalejący: $x_1 \dots x_n$ obliczyć wartość rozstępu R zgodnie ze wzorem: $R = x_n - x_1$ obliczyć wartość parametrów Q_1 i Q_n wg wzorów: $Q_1 = \frac{x_2 - x_1}{R} \quad Q_n = \frac{x_n - x_{n-1}}{R}$ porównać otrzymane wartości z wartością krytyczną Q_{kr} , odczytaną dla wybranego poziomu istotności α i liczby stopni swobody $f = n$
wnioskowanie	jeśli, któryś z obliczonych parametrów przekracza wartość krytyczną Q_{kr} , to wynik na podstawie którego został obliczony (x_1 lub x_n) należy odrzucić jako obciążony błędem grubym i dopiero wtedy policzyć wartości x_{sr} i s .

3.2 Test Hampela

cel	wykrycie w danym zbiorze wyników odbiegających
wymagania	liczność wyników w serii (zbiorze) większa od 2
sposób postępowania	uszeregować wartości w porządku rosnącym; obliczyć wartość mediany Me ze wszystkich wyników x_i , gdzie x_i obejmuje zakres od x_1 do x_n ; obliczyć wartości odchyłeń r_i od wartości mediany dla każdego wyniku, korzystając z wzoru: $r_i = (x_i - Me)$ obliczyć wartości bezwzględne $ r_i $; uszeregować wartości $ r_i $ w porządku rosnącym; obliczyć wartość mediany odchyłeń $Me_{ r_i }$; porównać wartości $ r_i $ z wartością $4,5 \cdot Me_{ r_i }$
wnioskowanie	jeżeli spełniony jest warunek $ r_i \geq 4,5 \cdot Me_{ r_i }$ wynik x_i jest uznawany za wynik odbiegający.

3.3 Test χ^2 – Chi kwadrat

cel	sprawdzenie, czy wartość wariancji (odchylenia standardowego) dla danej serii wyników różni się od wartości zadanej
hipotezy	H_0 – obliczona wartość wariancji dla serii wyników nie różni się od wartości zadanej w sposób statystycznie istotny H_1 – obliczona wartość wariancji dla serii wyników różni się od wartości zadanej w sposób statystycznie istotny
wymagania	rozkład normalny wyników w serii
sposób postępowania	obliczyć wartość odchylenia standardowego dla serii wyników obliczyć wartość parametru testu chi-kwadrat χ^2 wg wzoru: $\chi^2 = \frac{n \cdot s^2}{s_o^2}$ gdzie: s – obliczona wartość odchylenia standardowego dla zbioru wyników, s_o – wartość zadana odchylenia standardowego, n – liczba wyników w badanym zbiorze. porównać obliczoną wartość χ^2 z wartością krytyczną parametru χ_{kr}^2 dla przyjętego poziomu istotności α oraz obliczonej liczby stopni swobody $f = n - 1$
wnioskowanie	jeżeli obliczona wartość χ^2 nie przewyższa wartości krytycznej ($\chi^2 \leq \chi_{kr}^2$), wówczas można wnosić, że obliczona wartość odchylenia standardowego nie różni się w sposób statystycznie istotny od wartości zadanej – przyjęcie hipotezy H_0 ; jeżeli natomiast obliczona wartość χ^2 jest większa niż odczytana z tablic wartość krytyczna ($\chi^2 > \chi_{kr}^2$), należy wówczas wyciągnąć wniosek, że porównywane wartości odchylenia standardowego różnią się w sposób statystycznie istotny – odrzucenie hipotezy H_0 .

3.4 Test F-Snedecora

cel	porównanie wartości wariancji (odchylen standardowych) dla dwóch zbiorów wyników
hipotezy	H_0 – obliczone wartości wariancji dla porównywanych serii wyników nie różnią się w sposób statystycznie istotny H_1 – obliczone wartości wariancji dla porównywanych serii wyników różnią się w sposób statystycznie istotny
wymagania	rozkłady normalne wyników w serii
sposób postępowania	obliczyć wartości odchylenia standardowego dla porównywanych serii wyników obliczyć wartość parametru testu F-Snedecora wg wzoru: $F = \frac{\frac{n_1}{n_1 - 1} \cdot s_1^2}{\frac{n_2}{n_2 - 1} \cdot s_2^2}$ gdzie: s_1, s_2 – obliczone wartości odchylenia standardowego dla dwóch zbiorów wyników, n_1, n_2 – liczba wyników dla dwóch zbiorów. UWAGA: wartość wyrażenia należy tak skonstruować, aby licznik był większy od mianownika – wartość F powinna być zawsze większa od 1 porównać obliczoną wartość z wartością krytyczną parametru dla przyjętego poziomu istotności α oraz obliczonej liczby swobody f_1 i f_2 (gdzie $f_1 = n_1 - 1$ i $f_2 = n_2 - 1$)

wnioskowanie	jeżeli obliczona wartość F nie przewyższa wartości krytycznej ($F \leq F_{kr}$), wówczas można wnosić, że obliczone wartości odchylenia standardowego nie różnią się w sposób statystycznie istotny – przyjęcie hipotezy H_0 jeżeli natomiast obliczona wartość F jest większa niż odczytana z tablic wartość krytyczna ($F > F_{kr}$) należy wówczas wyciągnąć wniosek, że porównywane wartości odchylenia standardowego różnią się w sposób statystycznie istotny – odrzucenie hipotezy H_0
--------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

3.5 Test Hartleya

cel	porównanie wartości wariancji (odchyłeń standardowych) dla wielu zbiorów wyników
hipotezy	H_0 – obliczone wartości wariancji dla porównywanych serii wyników nie różnią się w sposób statystycznie istotny H_1 – obliczone wartości wariancji dla porównywanych serii wyników różnią się w sposób statystycznie istotny
wymagania	rozkłady normalne wyników w serii liczności wyników w każdej serii zbiorów większe od 2 liczności zbiorów są jednakowe liczba serii nie większa niż 11
sposób postępowania	obliczyć wartości odchylenia standardowego dla porównywanych serii wyników obliczyć wartość parametru testu F_{max} wg wzoru: $F_{max} = \frac{S_{max}^2}{S_{min}^2}$ gdzie: s_{max}, s_{min} – odpowiednio największa i najmniejsza wartość spośród obliczonych wartości odchylenia standardowego dla zbiorów wyników. porównać obliczoną wartość z wartością krytyczną parametru dla przyjętego poziomu istotności α obliczonej liczby swobody $f = n - 1$ oraz liczby porównywanych serii k
wnioskowanie	jeżeli obliczona wartość F_{max} nie przewyższa wartości krytycznej ($F_{max} \leq F_{max0}$), wówczas można wnosić, że obliczone wartości odchylenia standardowego nie różnią się w sposób statystycznie istotny – przyjęcie hipotezy H_0 jeżeli natomiast obliczona wartość F_{max} jest większa niż odczytana z tablic wartość krytyczna ($F_{max} > F_{max0}$) należy wówczas wyciągnąć wniosek, że porównywane wartości odchylenia standardowego różnią się w sposób statystycznie istotny – odrzucenie hipotezy H_0

3.6 Test t-Studenta

cel	porównanie wartości średnich dla dwóch serii (zbiorów) wyników
hipotezy	H_0 – obliczone wartości średnie dla porównywanych serii wyników nie różnią się w sposób statystycznie istotny H_1 – obliczone wartości średnie dla porównywanych serii wyników różnią się w sposób statystycznie istotny
wymagania	rozkłady normalne wyników w serii liczności wyników w każdej serii zbiorów większe od 2 nieistotność różnic wariancji dla porównywanych zbiorów wyników – test F-Snedecora

sposób postępowania	<p>obliczyć wartości średnie i wartości odchylenia standardowego dla serii wyników; obliczyć wartość parametru testu t-Studenta zgodnie z zależnością:</p> $t = \frac{(x_{1sr} - x_{2sr})}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$ <p>gdzie: x_{1sr}, x_{2sr} – wartości średnie obliczone dla dwóch porównywanych zbiorów wyników s_1, s_2 – wartości odchylenia standardowego dla zbiorów wyników. porównać obliczoną wartość z wartością krytyczną parametru dla przyjętego poziomu istotności α obliczonej liczby swobody $f = n_1 + n_2 - 2$</p>
wnioskowanie	<p>jeżeli wartość t nie przewyższa wartości krytycznej t_{kr}, ($t \leq t_{kr}$), to można wysnuć wniosek, że uzyskane wartości średnie nie różnią się w sposób statystycznie istotny – przyjęcie hipotezy H_0 jeżeli natomiast obliczona wartość t jest większa niż odczytana z tablic wartość krytyczna ($t > t_{kr}$) należy wówczas wnosić, że porównywane wartości średnie różnią się w sposób statystycznie istotny – odrzucenie hipotezy H_0</p>

cel	porównanie wartości średniej z założoną wartością
hipotezy	<p>H_0 – obliczona wartość średnią nie różni się w sposób statystycznie istotny od wartości założonej H_1 – obliczona wartość średnią różni się w sposób statystycznie istotny od wartości założonej</p>
wymagania	rozkład normalny wyników w serii liczność wyników w serii zbiorów większa od 2
sposób postępowania	<p>obliczyć wartość średnią i wartość odchylenia standardowego dla serii wyników; obliczyć wartość parametru testu t-Studenta zgodnie z zależnością:</p> $t = \frac{ x_{sr} - \mu }{s} \sqrt{n}$ <p>gdzie: x_{sr} – wartość średnia obliczona dla zbioru wyników, μ – wartość odniesienia (np. certyfikowana), s – jednostka odchylenia np.: odchylenie standardowe zbioru wyników na podstawie, których obliczano wartość średnią, n – liczba wyników. porównać obliczoną wartość z wartością krytyczną parametru dla przyjętego poziomu istotności α obliczonej liczby swobody $f = n - 1$</p>
wnioskowanie	<p>jeżeli wartość t nie przewyższa wartości krytycznej t_{kr}, ($t \leq t_{kr}$), to można wysnuć wniosek, że uzyskana wartość średnia nie różni się od wartości zadanej w sposób statystycznie istotny – przyjęcie hipotezy H_0 jeżeli natomiast obliczona wartość t jest większa niż odczytana z tablic wartość krytyczna ($t > t_{kr}$) należy wówczas wnosić, że wartość średnia różni się od wartości zadanej w sposób statystycznie istotny – odrzucenie hipotezy H_0</p>

4. PODSUMOWANIE

Wszystkie podejmowane decyzje i wnioskowania są z definicji obarczone pewnym ryzykiem błędu. Każdy wynik pomiaru analitycznego nie jest wartością stałą, lecz zawiera parametr „niepewny”. W konsekwencji wykorzystywanie wartości obciążonych niepewnością do porównań, wnioskowania itp. musi być także zależne od tego parametru – jest tylko w skończony sposób pewne. Pamiętać także należy, że na etapie porównywania wartości (niezależnie od źródeł ich pochodzenia) powinno być brane pod uwagę prawdopodobieństwo. Konsekwencją tego jest wnioskowanie w rodzaju: „wartości nie różnią się między sobą w sposób statystycznie istotny” zamiast wnioskowania: „wartości są takie same”. Dziedziną wiedzy wykorzystywaną podczas tego rodzaju opracowywania wyników pomiarów jest statystyka matematyczna, a narzędziami testy statystyczne. Warto jednak pamiętać, że ze statystyki należy korzystać w sposób rozsądny.

LITERATURA

- [1] Konieczka P., Namieśnik J.: *Quality Assurance and Quality Control in the Analytical Chemical Laboratory: A Practical Approach*, Boca Raton; London; New York: CRC Press, Taylor & Francis Group, 2009
- [2] Dobecki M. (red): „Zapewnienie jakości analiz chemicznych”, Instytut Medycyny Pracy im. Prof. J.Nofera, Łódź 2004.
- [3] Koronacki J, Mielniczuk J.: „Statystyka dla studentów kierunków technicznych i przyrodniczych”, Warszawa, WNT, 2001.
- [4] Bożyk Z., Rudzki W.: „Metody statystyczne w badaniu jakości produktów żywnościowych i chemicznych”, Warszawa, WNT, 1977.
- [5] Kozłowski E.: „Statystyczne kryteria oceny wyników i metod analitycznych w: Bobrański B.: *Analiza ilościowa związków organicznych*”, Warszawa, PWN, 1979.
- [6] Czermiński J.B., Iwasiewicz A., Paszek Z., Sikorski A.: „Metody statystyczne dla chemików”, Warszawa, PWN, 1986.
- [7] Konieczka P., Namieśnik J. (red.): „Kontrola i zapewnienie jakości wyników pomiarów analitycznych”, WNT, Warszawa 2007.

**Waldemar Korol, Sławomir Walczyński,
Grażyna Bielecka, Jolanta Rubaj**
Instytut Zootechniki Państwowy Instytut Badawczy
Krajowe Laboratorium Pasz w Lublinie

Proste statystyczne podejście do szacowania niepewności pobierania i przygotowania próbki do badań

W przypadku podejmowania decyzji o zgodności badanej partii materiału z wymaganiami, pobieranie i przygotowanie próbki może mieć wpływ na całkowitą niepewność pomiaru. Przedstawiono sposób szacowania niepewności pobierania i przygotowania próbki do badań na przykładzie pasz. Istotą proponowanego sposobu oceny niepewności pobierania i przygotowania próbki do badań jest rozdzielenie dwóch składowych zmienności wyniku badania: zmienności analitycznej i zmienności technicznej odpowiadającej niejednorodności badanego składnika. Do szacowania całkowitej niepewności pomiaru wykorzystano zmienną techniczną pobierania próbki odpowiadającą homogeniczności badanej partii materiału i zmienności próbkowania, zmienną techniczną przygotowania próbki do badań odpowiadającą homogeniczności próbki laboratoryjnej i zmienności jej przygotowania oraz niepewność postępowania analitycznego i pomiaru. Do obliczeń wykorzystano prawo propagacji błędów Gauss'a. Podano przykłady szacowania niepewności.

1. WPROWADZENIE

Wg Przewodnika Eurachem/Citac [1] dotyczącego wykorzystania informacji o niepewności pomiaru do oceny zgodności, pobieranie próbki może mieć wpływ na całkowitą niepewność wyniku pomiaru, zwłaszcza gdy decyzja o zgodności dotyczy partii materiału. Przewodnik zaleca uwzględnianie składowej wynikającej z niepewności pobierania próbki zwłaszcza gdy zostały określone wymagania dotyczące pobierania próbki. Opis obliczania niepewności związanej z pobieraniem próbki jest przedmiotem innego przewodnika Eurachem [2]. Zgodnie z aktualnymi wymaganiami w uzasadnionych przypadkach do oceny zgodności zaleca się oszacowanie całkowitej niepewności pomiaru. Całkowita niepewności pomiaru obejmuje niepewność pobierania próbki, przygotowania próbki do badań w laboratorium oraz niepewność postępowania analitycznego łącznie z pomiarem.

Często stawiane jest pytanie, czy w każdym przypadku należy szacować niepewność całkowitą. Norma PN-EN ISO/IEC 17025 [3] nie zawiera pojęcia niepewności całkowitej. Cytowana norma w pkt 5.4.1 stanowi, że laboratorium powinno stosować oszacowanie niepewności pomiaru tam gdzie to właściwe i zaleca w pkt 5.10.3.1 aby niepewność pomiaru była podawana „gdy ma to znaczenie dla miarodajności wyników badania lub ich zastosowania, gdy takie wymaganie jest w wytycznych klienta lub gdy niepewność ma znaczenie dla zgodności z wyspecyfikowanymi wartościami granicznymi”. Norma PN-EN ISO/IEC 17025 jest podstawą do oceny kompetencji laboratorium w zakresie badań wraz z pobieraniem próbek i badań bez pobierania próbek (przy czym PCA wskazuje na ograniczenia związane z przydatnością wyniku do oceny zgodności zwłaszcza w obszarze regulowanym). Z kolei laboratoria badawcze nie pobierające próbek poddawanych badaniom są zobowiązane do przedstawiania w sprawozda-

niach z badań informacji od klienta dotyczących procedur i planu pobierania próbek oraz podmiotów wykonujących pobieranie. Jednocześnie podkreśla się, że laboratorium powinno stosować właściwe procedury postępowania z próbką laboratoryjną, a zwłaszcza związane z przygotowaniem próbki do badań pomimo, że dokumenty te nie są wskazywane w zakresie akredytacji.

Wymagania PCA zostały ostatnio sprecyzowane w dokumencie DAB-07 [4], w którym podkreśla się potrzebę dokładnego określenia celu badań w porozumieniu z klientem i, w przypadku gdy laboratorium nie pobiera próbek, zamieszczenia w sprawozdaniu informacji od klienta o pobieraniu próbek. Żądane informacje mogą być zamieszczone w sprawozdaniu na podstawie danych od klienta, jednak nie rozwiązuje to problemu oszacowania np. niepewności pobierania próbki, która może być określona we współpracy klienta z laboratorium lub gdy podmiot pobierający próbki posiada własne laboratorium. Inny przykład to obowiązek podawania informacji, że niepewność wyniku badania w sprawozdaniu z badań nie uwzględnia niepewności pobierania próbki. Jeżeli taką informację, zgodnie z wytycznymi DAB-07, podaje laboratorium upoważnione do badań w ramach urzędowego nadzoru, powstaje pytanie czy wynik badania może być podstawą do oceny zgodności i podejmowania decyzji przez upoważnioną inspekcję.

Nowe wymagania stawiane laboratoriom, zwłaszcza funkcjonującym w obszarach regulowanych, stwarzają potrzebę lepszej współpracy pomiędzy klientem a laboratorium wykonującym badania. Laboratorium powinno dążyć do poznania właściwości fizykochemicznych nowych materiałów i produktów, zwłaszcza homogeniczności badanych substancji w materiale/produkcie. Z kolei w interesie klienta jest dobra współpraca z laboratorium, która może zaowocować np. określeniem niepewności pobierania próbek.

Celem pracy było zaprezentowanie praktycznego podejścia do określenia niepewności pobierania i przygotowania próbki do badań w laboratorium, polegającego na rozdzieleniu dwóch składowych zmienności wyniku badania zarówno na etapie szacowania niepewności pobierania i niepewności przygotowania próbek, tj. zmienności analitycznej i zmienności technicznej odpowiadającej niejednorodności badanego składnika.

2. MATERIAŁY I METODY

Materiał do szacowania niepewności pobierania próbek stanowiły produkty paszowe sypkie i granulowane (sypka mieszanka paszowa uzupełniająca MPU, granulowana mieszanka paszowa pełnoporcjowa MPP). Próbki pobierano z badanych partii pasz. W przypadku szacowania niepewności przygotowania próbki w laboratorium, materiał do badań stanowiły sypkie MPP i MPU i granulowane MPP oraz próbki premiksów (przedmieszka mineralno-witaminowa): bez rozdrabniania i po rozdrobnieniu.

Niepewność pobierania i przygotowania próbki oszacowano wykorzystując prawo propagacji błędów Gauss'a. Istotą proponowanego sposobu oceny niepewności przygotowania próbki jest rozdzielenie dwóch składowych zmienności wyniku pomiaru wyrażonej jako współczynnik zmienności CV_p (%): zmienności analitycznej CV_a (%) i zmienności technicznej CV_h (%) odpowiadającej niejednorodności badanego składnika podczas pobierania próbki i/lub podczas przygotowania próbki do badań w laboratorium. Zmienność analityczną metody obliczano z rozstępu, na przykładzie Raportu Technicznego Nordtest [5]. Do obliczeń wykorzystano poniższe wzory, (1) i (2).

$$CV_p = \sqrt{CV_h^2 + CV_a^2} \quad (1)$$

$$CV_h = \sqrt{CV_p^2 + CV_a^2} \quad (2)$$

Przyjęto, że współczynnik zmienności technicznej CV_h jest miarą niepewności pobierania próbki i/lub niepewności przygotowania próbki i może być uznany jako standardowa niepewność pobierania próbki ($CV_h = u_s$) i/lub jako standardowa niepewność przygotowania próbki ($CV_h = u_p$) w całkowitym budżecie niepewności pomiaru u_t , obliczonym wg wzoru (3),

$$u_t = \sqrt{u_s^2 + u_p^2} + u_a \quad (3)$$

w którym u_a jest niepewnością postępowania analitycznego i pomiaru. Niepewność rozszerzoną U_T dla $k = 2$ ($P=95\%$) obliczano jak następuje: $U_T = u_t \cdot 2$.

Obliczono niepewności przygotowania próbki takich składników pasz jak białko ogólne, popiół surowy, wapń, sód, chlorki, mangan, cynk i selen. W przypadku niepewności pobierania próbki badano składniki takie jak wapń, chlorki, cynk i miedź oraz pierwiastki śladowe selen i molibden.

Do szacowania niepewności przygotowania próbek, pobierano próbki laboratoryjne pasz o masie około 0,5 kg zgodnie z rozporządzeniem 152/2009 [8]. Z każdej próbki laboratoryjnej przygotowywano wg zaleceń normy EN 6498 [7], 6 próbek do badań o masie około 100 g każda, stosując rozdzielacz szczelinowy do rozdzielania próbki laboratoryjnej. Próbki do badań rozdrabniano w całości z przesiewem przez sito 0,5 mm. W tak przygotowanych próbkach wykonywano oznaczenia zawartości białka, popiołu i składników mineralnych w dwóch powtórzeniach.

W celu określenia niepewności pobierania próbki z partii sypkiej MPU (10%) o masie 2 tony dla kurcząt brojlerów pobrano niezależnie, zgodnie z rozporządzeniem 152/2009 [9] (od 1 stycznia 2014 r. obowiązuje rozporządzenie 691/2013 [6]), 6 próbek zbiorczych, z których wyodrębniono 6 próbek końcowych (laboratoryjnych) o masie ok. 0,5 kg każda. Każdą próbkę laboratoryjną rozdzielano na rozdzielaczu szczelinowym do uzyskania próbki do badań o masie ok. 100 g, które rozdrabniano w młynku laboratoryjnym z sitem 0,5 mm. W tak przygotowanych próbkach wykonywano oznaczenia zawartości chlorków, cynku, miedzi i wapnia, w dwóch powtórzeniach.

Zawartości białka ogólnego oznaczono metodą Kjeldahla a popiołu surowego metodą wagową, zgodnie z rozporządzeniem 152/2009 [8]. Zawartości wapnia, sodu, cynku i manganu oznaczono metodą płomieniowej absorpcyjnej spektrometrii atomowej FAAS [9]. Selen badano metodą absorpcyjnej spektrometrii atomowej z generowaniem wodorków HGAAS, a molibden przy wykorzystaniu techniki ETAAS. Średnią wielkość cząstek premiksu przed rozdrobnieniem i po rozdrobnieniu mierzono optyczno-elektronicznym miernikiem wielkości cząstek. Próbki do badań białka, popiołu i składników mineralnych rozdrabniano na rozdrabniaczu z sitem 1 mm; premiks rozdrabniano w młynie planetarnym do średniej wielkości cząstek 260 μm .

Z wzoru (2) obliczono niepewności pobierania i przygotowania próbki do badań w laboratorium. Z wzoru (3) obliczono niepewności przygotowania próbki do badań w laboratorium i niepewności postępowania analitycznego i pomiaru oraz, w przypadku wapnia, chlorków, cynku i miedzi, niepewności całkowite wyników badań tych składników w sypkiej MPU i granulowanej MPP. Przykład obliczania niepewności przygotowania próbki oraz pozostałe wyniki szacowania niepewności pobierania i przygotowania próbki zamieszczono w tabelach.

3. WYNIKI I DYSKUSJA

W Tabeli 1 podano przykład szacowania niepewności przygotowania próbki. W analogiczny sposób można oszacować niepewność pobierania próbki.

Tabela 1. Obliczanie niepewności przygotowania próbki w laboratorium na przykładzie wapnia w granulowanej MPP

Wyszczególnienie	Sposób obliczenia – wzory
Współczynnik zmienności $CV_p = 2,05\%$	Dane z laboratorium
Powtarzalność obliczona z rozstępu wg [6] $CV_a = 1,86\%$	Dane z laboratorium - szablon Excel
Współczynnik zmienności technicznej $CV_h = 0,86\% = u_p$	Wzór: $CV_h = \sqrt{CV_p^2 - CV_a^2}$
Standardowa niepewność badania wapnia z budżetu niepewności $u = 5,0\%$; $U = 10,0\%$	Dane z laboratorium
Standardowa niepewność badania Ca z niepewnością przygotowania próbki w laboratorium $u_{p+a} = 5,1\%$	Wzór: $u_{p+a} = \sqrt{u_p^2 + u_a^2}$
Niepewność rozszerzona $U_{p+a} = 10,2\%$	$5,1\% \times 2 = 10,2\%$

Porównanie wyników szacowania niepewności rozszerzonej uwzględniającej niepewność postępowania analitycznego i pomiaru U z wynikami szacowania niepewności rozszerzonej U_{p+a} uwzględniającej dodatkowo składową przygotowania próbki dla sypkiej MPU i granulowanej MPP oraz premiksu nie rozdrobionego i rozdrobionego zamieszczono w Tabeli 2.

Tabela 2. Porównanie wyników szacowania niepewności z uwzględnieniem niepewności przygotowania próbki dla wybranych pasz o różnych właściwościach fizykochemicznych

Składnik	MPU sypka		M-ka granul.		Premiks nrozd. $\varnothing = 416 \mu\text{m}$		Premiks rozd. $\varnothing = 260 \mu\text{m}$	
	U	U_{p+a}	U	U_{p+a}	U	U_{p+a}	U	U_{p+a}
Białko	3	4,3	4	4,8	nb	nb	nb	nb
Popiół	4	6,8	4,2	4,6	nb	nb	nb	nb
Wapń	10	12,4	10	10,2	10	14,8	10	10
Sód	10	13,9	10	10,4	nb	nb	nb	nb
Cynk	15	18,1	15	15,8	10	22,8	10	12,6
Żelazo	nb	nb	15	15,8	10	14,2	10	10,4
Mangan	nb	nb	15	16,2	10	17,2	10	11,0
Miedź	nb	nb	20	23,2	10	22,4	10	11,8
Kobalt	nb	nb	24	24,8	nb	nb	nb	nb
Molibden	20	30,9	20	22,0	nb	nb	nb	nb
Selen	15	28,0	15	16,0	nb	nb	nb	nb

U – niepewność rozszerzona ($k = 2$) z budżetu niepewności; U_{s+h} – niepewność rozszerzona ($k = 2$) uwzględniająca niepewność przygotowania próbki; nb – nie badano

Oszacowane niepewności wyników pomiaru białka, popiołu oraz składników mineralnych z uwzględnieniem niepewności przygotowania próbki U_{p+a} w przypadku sypkiej MPU okazały się wyższe od 20,7% do

86,7%, średnio o 48,3%, od niepewności U oszacowanej bez uwzględnienia składowej przygotowania próbki. Przyczyną istotnego wpływu przygotowania próbki na wyniki szacowania niepewności w sypkiej MPU była niehomogeniczność (niejednorodność) próbki laboratoryjnej i skłonność do segregacji składników zarówno organicznych (białko) jak i mineralnych (popiół, makroelementy, mikroelementy).

W przypadku szacowania niepewności pomiaru składników granulowanej MPP, wpływ składowej niepewności przygotowania próbki okazał się niewielki. Uwzględnienie niepewności przygotowania próbki w niepewności pomiaru spowodowało wzrost oszacowanej niepewności od 2% do 20%, średnio o 8,2% (tab. 2). W rezultacie dla mieszanki granulowanej uzyskano podobne wyniki szacowania niepewności, świadczące o poprawie homogeniczności przez ograniczenie segregacji składników.

Rozdrabnianie premiksu wpłynęło zasadniczo na wyniki szacowania niepewności wapnia, cynku, żelaza, manganu i miedzi pomimo, że test homogeniczności oparty o wyniki badania chlorków był pozytywny dla obu premiksów. W premiksie nie rozdrobnionym wyniki szacowania niepewności składników mineralnych z uwzględnieniem składowej przygotowania próbki były średnio o 82,8% wyższe od niepewności bez składowej przygotowania próbki. W przypadku premiksu rozdrobnionego w młynie planetarnym, różnica pomiędzy niepewnościami wyniosła 11,6% (tab. 2).

Do badań niepewności pobierania próbek wybrano sypką MPU (10%) z uwagi na specyficzne właściwości fizykochemiczne tej paszy, duży udział składników mineralnych, skłonność do segregacji składników, segregacja wtórna nawet podczas mieszania, niska homogeniczność składników. Uzyskano standardowe niepewności pobierania próbek od 2,9% w przypadku cynku do 5,1% w przypadku miedzi, tab. 3. Uwzględnienie niepewności pobierania próbek podwyższyło niepewność całkowitą pomiaru badanych pierwiastków od 7,0% do 47,3%, średnio o 19,6%. Uwzględnienie niepewności pobierania próbek może mieć istotne znaczenie dla oceny zgodności badanej partii paszy z wymaganiami, np. z deklaracją zawartości składnika na etykiecie, uwzględniając dopuszczalne granice tolerancji podane w rozporządzeniu 939/2010 [10].

Tabela 3. Standardowe niepewności pobierania próbek u_s (%) dla składników mineralnych sypkiej MPU oraz niepewności rozszerzone U uwzględniające składową przygotowania próbki (U_{p+a}) i niepewność całkowitą (U_T)

Składnik mineralny	MPU sypka (10%)			
	u_s	U	U_{p+a}	$U_T = U_{s+p+a}$
Wapń	4,9	7,1	9,1	13,4
Chlorki	3,2	10,0	12,4	13,9
Cynk	2,9	10,0	14,2	15,2
Miedź	5,1	15,0	20,0	22,4

Sposób szacowania niepewności pobierania próbek został określony w przewodniku EURACHEM/CITAC [2]. Ten użyteczny przewodnik zalecany jest zwłaszcza, gdy proces badania, od pobrania próbki do końcowego pomiaru badanego składnika, jest prowadzony i kontrolowany na każdym etapie przez laboratorium. Aktualne wymagania DAB-07, zwłaszcza w obszarze regulowanym, zobowiązują laboratorium do zamieszczania informacji nt. pobierania próbek, a proces przygotowania próbki do badań traktują jako element postępowania analitycznego. Przedstawione tu proste statystyczne podejście do szacowania niepewności pobierania i przygotowania próbek materiałów sypkich na przykładzie pasz może być użyteczne dla laboratorium współpracującym z klientem w poznaniu charakterystyk fizykochemicznych badanych materiałów, zwłaszcza ich jednorodności i skłonności do segregacji. Zaletą proponowanego

sposobu jest rozdzielenie składowej zmienności analitycznej od składowej zmienności technicznej dzięki obliczeniu zmienności analitycznej (powtarzalności) z rozstępu pomiędzy powtórzeniami wg Nordtest TR 537 [6] dla każdego badanego materiału. Zastosowane podejście pozwala na oszacowanie niepewności pobierania i/lub przygotowania próbki; jest zatem dostosowane do celu badań.

4. PODSUMOWANIE

Uzyskane wyniki potwierdziły, że w przypadku produktu w stanie sypkim, o dużej podatności na segregację składników (np. MPU – mieszanina składników mineralnych i roślinnych), na niepewność pomiaru składników wpływa w znacznym stopniu niepewność pobierania i przygotowania próbki. Uwzględnienie niepewności pobierania i przygotowania próbki w przypadku takich produktów wydaje się celowe w ocenie zgodności z wymaganiami, pomimo braku jednoznacznych zaleceń w tym zakresie [6]. Stwierdzono, że stopień rozdrobnienia próbki wpływa na niepewność pomiaru. Dokładne rozdrobnienie premiksu znacząco obniżyło niepewności pomiaru badanych składników mineralnych, co uzasadnia potrzebę właściwego przygotowania próbki do badań i generalnie, potrzebę poznania właściwości fizykochemicznych badanych matryc. Proponowane podejście do szacowania niepewności pobierania i przygotowania próbki do badań może być wykorzystane na potrzeby urzędowego nadzoru w budowie spójnego systemu oceny i interpretacji wyników. Przykładem mogą być laboratoria niemieckiego nadzoru paszowego VDLUFA, organizujące badania porównawcze w taki sposób aby laboratorium uczestniczące samo rozdzieliło i rozdrobniło próbkę laboratoryjną (uczestnik PT/ILC otrzymuje próbkę o masie co najmniej 0,5 kg, podobnie jak próbka laboratoryjna zlecana przez klienta). Wyniki takich porównań i badań biegłości uwzględniają zmienność przygotowania próbek [11].

LITERATURA

- [1] EURACHEM/CITAC Guide. Use of uncertainty information in compliance assessment. First edition 2007. Wydanie polskie: Biuletyn POLLAB Nr 1/52/2009.
- [2] EURACHEM/CITAC Guide. Measurement uncertainty arising from sampling. A guide to methods and approaches. First edition 2007. Wydanie polskie: Biuletyn POLLAB Nr 1/52/2009.
- [3] Norma PN-EN ISO/IEC 17025:2005. Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcujących.
- [4] Dokument PCA DAB-7. Akredytacja laboratoriów badawczych. Wymagania szczegółowe. Wydanie 10. Warszawa 16.10.2013.
- [5] Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. Nordtest TR 537, Version 3, 2008. Wydanie polskie: Biuletyn POLLAB Nr 2/51/2008.
- [6] Rozporządzenie Komisji (UE) nr 691/2013 z dnia 19 lipca 2013 r. zmieniającym rozporządzenie (WE) nr 152/2009 w odniesieniu do metod pobierania próbek i dokonywania analiz.
- [7] International Standard ISO 6498:2012. Animal feeding stuffs-Guidelines for sample preparation.
- [8] Rozporządzenie Komisji (WE) nr 152/2009 ustanawiające metody pobierania próbek i dokonywania analiz do celów urzędowej kontroli pasz. Dz. Urz. UE L 54/1 z 26.02.2009.
- [9] Norma PN-EN ISO 6869:2002 Pasze – Oznaczanie zawartości wapnia, miedzi, żelaza magnezu, manganu, potasu, sodu i cynku. Metoda absorpcyjnej spektrometrii atomowej.
- [10] Rozporządzenie Komisji (UE) 939/2010 z dnia 20 października 2010 r. zmieniające załącznik IV rozporządzenia (WE) nr 767/2009 (Dz. Urz. UE L 277/4, 21.10.2010).
- [11] VDLUFA Method Book Volume III, The chemical analysis of feedstuffs, 3. Edition 1976, 8. Suppl. 2012. Analytical tolerances. 32.1. VDLUFA-Publisher, Darmstadt, 2012.

Paweł Fotowicz
Główny Urząd Miar

Metodyka opracowania danych pomiarowych w świetle dokumentów Międzynarodowego Biura Miar

Przedstawiono metodykę opracowania danych pomiarowych zalecaną przez dokumenty wydawane pod egidą Międzynarodowego Biura Miar. Opracowuje je Wspólny Komitet ds. Przewodników w Metrologii. Dokumenty mają charakter normatywny i wyznaczają współczesny standard wyrażania i obliczania niepewności pomiaru. Preferują podejście probabilistyczne w postaci prawa propagacji niepewności i propagacji rozkładów, jako metody walidujące. Propagacja rozkładów realizowana jest metodą Monte Carlo. Celem tych działań jest wypracowanie jednolitego podejścia w dziedzinie opracowania danych pomiarowych, które może mieć zastosowanie we wszystkich obszarach metrologii.

1. PAKIET DOKUMENTÓW

Współczesne podejście metrologiczne w dziedzinie opracowania danych pomiarowych kształtuje pakiet dokumentów wydawanych pod egidą Międzynarodowego Biura Miar przez Wspólny Komitet ds. Przewodników w Metrologii (JCGM – Joint Committee for Guides in Metrology). Pełny pakiet dokumentów, wydawanych od 2008 roku, tworzy zbiór dziewięciu opracowań:

- 1) Guide to the expression of uncertainty in measurement. GUM 1995 with minor corrections. JCGM 100:2008
- 2) Supplement 1 to the Guide – Propagation of distribution using a Monte Carlo method. JCGM 101:2008
- 3) Supplement 2 to the Guide – Extension to any number of output quantities. JCGM 102:2011
- 4) Supplement 3 to the Guide – Modelling. JCGM 103
- 5) An introduction to the Guide and related documents. JCGM 104:2009
- 6) Concepts, principles and methods for the evaluation of measurement uncertainty. JCGM 105
- 7) The role of measurement uncertainty in conformity assessment. JCGM 106:2012
- 8) Applications of the least-squares method. JCGM 107
- 9) Supplement 4 to the Guide – Bayesian methods. JCGM 108

W pakiecie tym kluczową rolę odgrywają pierwsze dwa dokumenty. Naczelnym z nich jest Przewodnik dotyczący wyrażania niepewności pomiaru [1]. Zawiera bazowe rozwiązania niezbędne do obliczania niepewności przy zastosowaniu prawa jej propagacji. Definiuje podstawowe pojęcia stosowane w tym obszarze, takie jak niepewność standardowa, złożona niepewność standardowa czy niepewność rozszerzona. Wyodrębnia dwie metody obliczania niepewności standardowej, typu A i B. Metoda typu A polega na analizie statystycznej serii pojedynczych obserwacji, a metoda typu B innymi sposobami niż analiza serii obserwacji. Przedstawia sposoby zapisu niepewności pomiaru i zawiera wskazówki praktyczne co do sposobu obliczania jej składowych.

Drugim z opracowań jest Suplement 1 do Przewodnika [2]. Dokument zawiera wytyczne co do nowego podejścia w dziedzinie wyrażania niepewności pomiaru. Podstawową jego ideą jest zasada propagacji rozkładów realizowana poprzez model matematyczny pomiaru przy użyciu metody Monte Carlo. Miarą wyniku pomiaru jest rozkład prawdopodobieństwa związany z wielkością wyjściową wyznaczany na podstawie rozkładów wielkości wejściowych. Wynik pomiaru przedstawiany jest w postaci parametrów tego rozkładu: wartości oczekiwanej, odchylenia standardowego i kwantyle rozkładu dla określonego prawdopodobieństwa, jako granic przedziału rozszerzenia. Dokument przedstawia zalecaną procedurę Monte Carlo, umożliwiającą obliczanie wszystkich tych parametrów. Wartość oczekiwana traktowana jest jak najlepsza estymata wielkości wyjściowej, a odchylenie standardowe jako niepewność standardowa związana z tą estymatą. Omawia algorytm postępowania z uwzględnieniem liczby próbkowań, umożliwiając osiągnięcie założonej dokładności obliczeniowej przedziału rozszerzenia. Przyjmuje dwie koncepcje przedziału rozszerzenia: symetrycznego probabilistycznie i najkrótszego dla założonego prawdopodobieństwa. Dokument ponadto opisuje procedurę walidacyjną wyznaczania niepewności pomiaru metodami analitycznymi przy użyciu symulacji Monte Carlo.

2. NIEPEWNOŚĆ POMIARU

Niepewność pomiaru to parametr charakteryzujący rozproszenie wartości wielkości związanych z mierzandem. Mierzand to wielkość, która ma być zmierzona i zdefiniowana jest przy użyciu określonego modelu pomiaru. Modelem pomiaru może być dowolna jawna lub niejawna funkcja pomiaru. Na ogół przedstawiana jest w ogólnej postaci jako:

$$Y = f(X_1, K, X_N) \quad (1)$$

gdzie Y oznacza wielkość wyjściową, a X_1, \dots, X_N oznaczają wielkości wejściowe. Wielkości wejściowe traktowane są jak zmienne losowe opisane określonymi rozkładami prawdopodobieństwa. Wyznacza się parametry tych rozkładów w postaci wartości oczekiwanej i odchylenia standardowego. Miarą wartości oczekiwanej jest estymata x_j wielkości, a miarą odchylenia standardowego jej niepewność standardowa $u(x_j)$. Na ogół przyjmuje się do opisu tych wielkości następujące rozkłady prawdopodobieństwa: normalny lub Studenta dla niepewności obliczanej metodą typu A i prostokątny, trójkątny lub trapezowy dla niepewności obliczanej metodą typu B.

W klasycznym podejściu, wynikającym z wykładni pierwszego dokumentu [1], dla wielkości wyjściowej przyjmuje się rozkład normalny lub Studenta z wypadkową liczbą stopni swobody. W tym podejściu kluczowym parametrem jest niepewność rozszerzona obliczana jako:

$$U = k \cdot u_c(y) \quad (2)$$

gdzie k jest współczynnikiem rozszerzenia, a $u_c(y)$ to złożona niepewność standardowa związana z estymatą y wielkości wyjściowej Y reprezentującej mierzand. Przyjmuje się założenie o liniowej lub linearyzowanej funkcji pomiaru, dlatego do obliczania odchylenia standardowego wielkości wyjściowej stosuje się metodę sumowania par kowariancji wielkości wejściowych, jako miary złożonej niepewności standardowej:

$$u_c^2(y) = \sum_{i=1}^N \sum_{j=1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j) \quad (3)$$

co dla wielkości wejściowych nieskorelowanych daje prostszą jej postać:

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) \quad (4)$$

Obie powyższe zależności w Przewodniku nazywane są prawem propagacji niepewności.

Współczynnik k przyjmuje wartość zbliżoną do wartości kwantyla rozkładu przyjętego dla wielkości wyjściowej. Gdy rozkładem wielkości wyjściowej jest rozkład Studenta, to kwantyl tego rozkładu wyznacza się dla wypadkowej liczby stopni swobody, obliczanej ze wzoru Welcha-Satterthwaite'a:

$$v_{\text{eff}} = \frac{u_c^4(y)}{\sum_{i=1}^N \frac{u_i^4(y)}{v_i}} \quad (5)$$

Przy prawdopodobieństwie $p = 95\%$ dla rozkładu normalnego w przybliżeniu $k = 2$, a dla rozkładu Studenta na ogół $k \geq 2$.

3. NOWE PODEJŚCIE

W podejściu reprezentowanym przez suplement do Przewodnika [2] rozkład prawdopodobieństwa związany z wielkością wyjściową Y wyznaczany jest poprzez model lub funkcję pomiaru. Działanie to nazywane jest propagacją rozkładów. Propagacja rozkładów polega na wyznaczeniu dystrybuanty G dla wielkości wyjściowej i może być realizowana:

- metodami analitycznymi poprzez matematyczne przedstawienie funkcji gęstości prawdopodobieństwa dla Y ,
- metodą propagacji niepewności opartą na przybliżeniu funkcji pomiaru pierwszymi wyrazami szeregu Taylora,
- przez włączenie dodatkowych członów wyższych rzędów wyrazów szeregu Taylora,
- metodami numerycznymi, szczególnie z zastosowaniem metody Monte Carlo.

Metoda Monte Carlo prowadzi do uzyskania numerycznej aproksymacji dystrybuanty G dla wielkości wyjściowej. Procedura realizowana jest w kolejnych krokach postępowania:

- wybór liczby próbkowania M ,
- wygenerowanie M prób N elementowego zbioru wielkości wejściowych,
- dla każdej próby obliczenie, na podstawie funkcji pomiaru, odpowiadającej jej wartości y_r wielkości wyjściowej,
- posortowanie wartości wielkości wejściowych w niemalejącym porządku, używając posortowanych wartości y_r do przybliżenia dystrybuanty wielkości wyjściowej,
- wyznaczenie z dystrybuanty G estymaty wielkości wyjściowej i związanej z nią niepewności standardowej,
- wyznaczenie z dystrybuanty G odpowiedniego przedziału rozszerzenia dla określonego prawdopodobieństwa p .

Ogólnie, przy opracowaniu danych pomiarowych postępowanie składa się z trzech etapów: opisu wielkości, obliczeń i zapisu wyniku. Zgodnie z zaleceniami opis wielkości powinien zawierać:

- definicję wielkości wyjściowej jako menzurandu,
- określenie wielkości wejściowych X_i , od których zależy wielkość wyjściowa Y ,
- model matematyczny określający relacje pomiędzy wielkościami wejściowymi a wielkością wyjściową,
- przyjęcie rozkładów prawdopodobieństwa dla wielkości wejściowych.

Etap obliczeń polega na realizacji zasady propagacji rozkładów wielkości wejściowych poprzez model pomiaru w celu otrzymania rozkładu dla wielkości wyjściowej, a zapis wyniku polega na przedstawieniu:

- a) wartości oczekiwanej jako estymaty y i wielkości wyjściowej,
- b) odchylenia standardowego jako niepewności standardowej $u(y)$, związanej z estymatą y ,
- c) przedziału rozszerzenia dla wielkości wyjściowej przy określonym prawdopodobieństwie, w postaci jego dolnej i górnej granicy y_{low} i y_{high} .

Ponieważ wyniki obliczeń metodą propagacji rozkładów mogą się różnić od wyników obliczeń przy zastosowaniu prawa propagacji niepewności, dlatego zaleca się ich walidację. Polega ona na:

- a) zastosowaniu prawa propagacji niepewności w celu wyznaczenia przedziału rozszerzenia o wartościach granicznych: $y \pm U$ dla wielkości wyjściowej Y ,
- b) zastosowanie zalecanej procedury obliczeniowej dla metody Monte Carlo w celu otrzymania wartości niepewności standardowej $u(y)$ oraz granic y_{low} i y_{high} przedziału rozszerzenia dla wielkości wyjściowej Y .

Następnie należy sprawdzić, czy otrzymane przedziały rozszerzenia zgadzają się co do ustalonej tolerancji numerycznej $\delta = 10^a/2$, gdy $u(y) = A \cdot 10^a$. W tym celu oblicza się:

$$d_{\text{low}} = |y - U - y_{\text{low}}| \quad (6)$$

$$d_{\text{high}} = |y + U - y_{\text{high}}| \quad (7)$$

jako bezwzględne wartości odpowiadające granicom przedziału. Jeżeli wartości te są nie większe niż δ , to można uznać obliczenia wykonane na podstawie prawa propagacji niepewności za zwalidowane.

4. PODSUMOWANIE

Powyższe opracowania wyznaczają podstawowy kanon opracowania danych pomiarowych, w myśl kształtującej się współcześnie koncepcji metrologicznej i dlatego noszą wspólny tytuł „Evaluation of measurement data”. Celem ich bowiem jest wypracowanie wspólnej metodyki opracowania wyniku pomiaru, jednolitej dla całego obszaru zastosowań metrologicznych. Podejście to wyrasta z klasycznej teorii błędu pomiaru opierając się o rachunek zmiennych losowych, współcześnie realizowany już przy użyciu numerycznych metod obliczeniowych wspieranych przez powszechnie dostępne techniki komputerowe. Wspiera się na dwóch filarach obliczeniowych w postaci prawa propagacji niepewności i propagacji rozkładów realizowanej metodą Monte Carlo. Wykorzystuje wyobrażenie wielkości mierzonej w postaci menzurandu formowanego poprzez model pomiaru. Jest uniwersalną propozycją w dziedzinie opracowania danych, mogących mieć zastosowanie we wszystkich obszarach nauk technicznych. Stanowi istotny wkład metrologii do rozwoju tych nauk.

LITERATURA

- [1] Guide to the Expression of Uncertainty in Measurement. ISO 1995
- [2] Evaluation of measurement data – Supplement 1 to the “Guide to the expression of uncertainty in measurement” – Propagation of distributions using a Monte Carlo method. JCGM 101:2008

Nowe techniki kalibracyjne oparte na regresji liniowej w analizie chemicznej

Metoda regresji liniowej jest ważnym narzędziem analizy statystycznej wyników pomiarów stosowanym do różnych aspektów pracy badawczej. Realizacja obliczeń oraz numeryczna implementacja odpowiednich wyrażeń regresji liniowej jest uzależniona od przyjęcia lub odrzucenia pewnych upraszczających założeń. Z punktu widzenia kontroli jakości pomiarów weryfikacja tych założeń jest kluczowa dla poprawności wyników i trafności wniosków generowanych na podstawie przeprowadzonych obliczeń regresji liniowej. W referacie zostaną przedstawione modyfikacje istniejących oraz nowe techniki kalibracyjne bez konieczności stosowania upraszczających założeń. Scharakteryzowane zostaną takie techniki kalibracyjne, jak: krzywa kalibracyjna, metoda wielokrotnych dodatków wzorca oraz metoda kolejnych rozcieńczeń.

1. WPROWADZENIE

Regresja liniowa jest metodą przewidywania wartości oczekiwanych jednej zmiennej na podstawie zależności matematycznej wiążącej ją z inną (innymi) zmienną (zmiennymi). Metoda ta jest ważnym narzędziem analizy statystycznej wykorzystywanym do działań związanych z kontrolą jakości pomiarów w laboratorium badawczym. Zasadniczym obszarem rutynowych zastosowań regresji liniowej w pracach laboratorium jest konstrukcja krzywej kalibracyjnej. Krzywa kalibracyjna jest zależnością między zmienną objaśnianą / zależną, Y (np. sygnałem analitycznym, odpowiedzią instrumentu, własnością fizyczną badanej materii) a zmienną objaśniającą / niezależną, X (np. stężeniem analitu w przygotowanych próbkach, cechą chemiczną badanej materii). Zależność kalibracyjna jest wykorzystywana w wielu ważnych aspektach pracy badawczej, wśród których istotne miejsce zajmują: poszukiwanie matematycznego opisu (modelu) zjawiska będącego podstawą funkcjonowania metody pomiarowej, przewidywania wartości korelowanych zmiennych, testowanie pracy aparatury pomiarowej lub porównywanie dwóch metod badawczych (metodyk postępowania w procesie pomiarowym).

W praktyce analitycznej krzywa kalibracyjna stanowi zależność między odpowiedzią instrumentu a stężeniem analitu w próbkach jak najbardziej zbliżonych składem chemicznym do rzeczywistych próbek poddawanych analizie chemicznej. Krzywa kalibracyjna jest fundamentem funkcjonowania większości metod opartych na technikach instrumentalnych, gdyż umożliwia w stosunkowo prosty sposób przekształcić informacje fizyczne o badanej próbce uzyskane w procesie pomiarowym na pożądane informacje chemiczne zawarte w badanej próbce. Jakość tych ostatnich jest w oczywisty sposób związana z poprawnością obliczeń opartych na metodzie regresji liniowej. Poszukiwanie zatem modelu matematycznego opisującego zależność między zmiennymi Y i X powinno doprowadzić do uzyskania funkcji w postaci $Y = f(X)$, która najlepiej jest dopasowana do wprowadzonych punktów eksperymentalnych $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$. Najprostszą postacią funkcji $f(X)$ jest zależność liniowa: $Y = aX + b$. W ogólnym jednak przypadku $f(X)$ może reprezentować dowolny model, który jest liniowy względem

parametrów funkcji (kombinacje liniowe dowolnych funkcji Y względem X). Celem regresji liniowej jest więc jak najlepsze dopasowanie punktów eksperymentalnych do założonej funkcji, czyli uzyskanie najbardziej prawdopodobnych ocen wartości współczynników tej funkcji (w przypadku prostej – parametrów a i b) oraz ich niepewności standardowych (dla prostej – $u(a)$ i $u(b)$). Procedura obliczeń założonej funkcji $Y = f(X)$ w najprostszym wariancie narzuca dwa upraszczające założenia:

1. zmienna niezależna ma charakter nielosowy, jej wartości są ustalonymi liczbami oraz
2. zmienna zależna ma charakter losowy opisywany rozkładem normalnym, a odchylenia standardowe jej wartości są stałe (jednorodny).

Pierwszy warunek zakłada, że wpływ niepewności skali wartości zmiennej niezależnej (stężeń analitu w przygotowanych próbkach) na niepewność poszukiwanej zależności jest pomijalny. W konsekwencji o niepewności krzywej kalibracyjnej decyduje niepewność pomiaru wartości zmiennej zależnej (sygnału analitycznego). Drugie założenie narzuca dodatkowo uproszczony, bo jednorodny, rozkład niepewności standardowych pomiaru sygnału analitycznego w całym zakresie wartości zmiennej Y .

Aby poszukiwany model matematyczny faktycznie odzwierciedlał relacje między eksperymentalnie uzyskanymi wartościami zmiennej zależnej i niezależnej i, w konsekwencji, generował wiarygodne przewidywania, konieczna jest weryfikacja powyższych założeń. Zastosowanie schematu obliczeń regresji liniowej, który będzie wykorzystywał te upraszczające założenia w przypadku ich niespełnienia w oczywisty sposób będzie prowadzić do fałszywych wyników i wniosków.

Matematycznym wyrazem ww. założeń są następujące warunki:

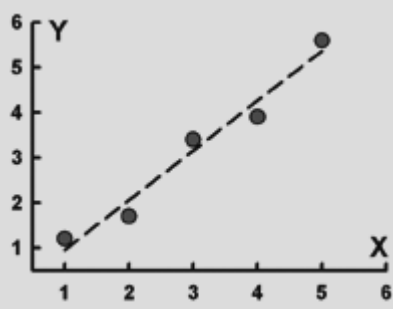
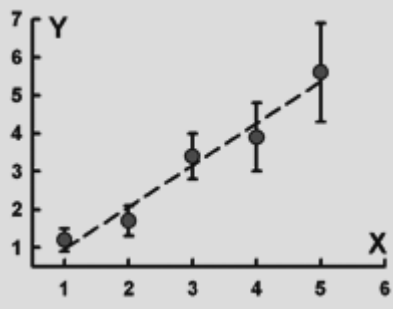
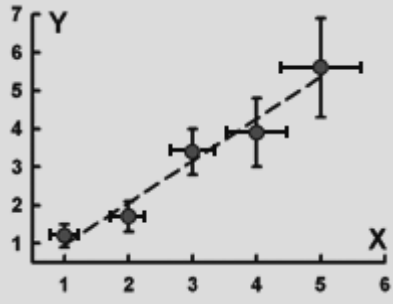
$$1. \quad \frac{u(x_i)}{x_i} < \frac{u(y_i)}{y_i} \text{ w całym przedziale wartości zmiennej } X$$

oraz

$$2. \quad u(y_i) = \text{const} \text{ w całym przedziale wartości zmiennej } Y$$

Proces weryfikacji powyższych warunków można przeprowadzić za pomocą testów statystycznych dedykowanych porównaniom wariancji (np. testem F) [1]. W przypadku spełnienia obydwu tych założeń schemat obliczeniowy może być oparty na wariancie regresji liniowej zwykłej – do obliczeń współczynników poszukiwanej funkcji są wykorzystywane tylko wartości współrzędnych kolejnych punktów eksperymentalnych (x_i, y_i) . W przypadku spełnienia tylko pierwszej relacji i pojawienia się istotnej statystycznie niejednorodności niepewności mierzonych wartości zmiennej Y , aby uzyskać poprawny obraz zależności kalibracyjnej w całym zakresie wartości zmiennej X (w całym zakresie roboczym krzywej kalibracyjnej) wymagane jest zastosowanie schematu obliczeń opartego na wariancie regresji liniowej ważonej (Y). Do obliczeń współczynników poszukiwanej funkcji zostaną wykorzystane wartości współrzędnych kolejnych punktów eksperymentalnych (x_i, \bar{y}_i) (gdzie \bar{y}_i jest uśrednioną wartością zmiennej zależnej dla i -tego punktu) oraz przypisane im wagi, w_i , różnicujące wpływ poszczególnych punktów na przebieg poszukiwanej funkcji w zależności od dokładności pomiaru wartości zmiennej Y . W przypadku niespełnienia obydwu upraszczających założeń w całym zakresie roboczym wartości zmiennej X wymagane jest zastosowanie do obliczeń schematu regresji liniowej ważonej (X, Y) uwzględniającej niepewności obydwu skal oraz istotną niejednorodność niepewności mierzonych wartości zmiennej zależnej. Do obliczeń współczynników poszukiwanej funkcji zostaną wykorzystane uśrednione wartości współrzędnych kolejnych punktów eksperymentalnych (\bar{x}_i, \bar{y}_i) oraz przypisane im wagi, W_i , różnicujące wpływ poszczególnych punktów na przebieg poszukiwanej zależności i będące funkcją niepewności zarówno wartości zmiennej zależnej, jak i niezależnej. Schematyczne zestawienie danych eksperymentalnych niezbędnych do poszukiwania zależności liniowej $Y = aX + b$ najlepiej dopasowanej do punktów eksperymentalnych przy zastosowaniu różnych wariantów regresji liniowej przedstawia Tabela 1 [2].

Tabela 1. Schematyczne zestawienie danych eksperymentalnych niezbędnych do obliczeń zależności kalibracyjnej różnymi wariantami regresji liniowej

Dane eksperymentalne	Wariant metody regresji liniowej	
(x_i, y_i)	Regresja liniowa zwykła 	$w_i = 1$
(x_i, \bar{y}_i, w_i)	Regresja liniowa ważona (Y) 	$w_i = \frac{1}{u^2(y_i)}$ lub $w_i = \frac{n \times u^{-2}(y_i)}{\sum_{i=1}^n u^{-2}(y_i)}$
$(\bar{x}_i, \bar{y}_i, W_i)$	Regresja liniowa ważona (X, Y) 	$W_i = \frac{1}{u^2(y_i) + a^2 u^2(x_i)}$

2. KRZYWA KALIBRACYJNA Z UWZGLĘDNIENIEM NIEPEWNOŚCI DWÓCH KORELOWANYCH ZMIENNYCH

Krzywa kalibracyjna jako istotny komponent metody analitycznej pełni bardzo istotną funkcję w całym procesie pomiarowym. Daje możliwość przewidywania wartości zmiennej niezależnej dla zmierzonych wartości zmiennej zależnej. Miarodajność tych przewidywań mierzona wartością przewidywania i jego niepewnością standardową jest w oczywisty sposób zależna od zastosowanego schematu obliczeń regresyjnych. Ten z kolei jest z jednej strony konsekwencją przyjęcia lub odrzucenia upraszczających założeń, a drugiej – determinuje formułę zależności funkcyjnej między X i Y , czyli wartości współczynników tej zależności. Przegląd wyrażen do obliczenia niepewności standardowej wartości zmiennej niezależnej „odczytanej” z krzywej kalibracyjnej zadanej w postaci funkcji liniowej przedstawia Tabela 2.

Tabela 2. Przewidywania wartości zmiennej niezależnej x_0 na podstawie liniowej zależności kalibracyjnej $Y = aX + b$ według różnych schematów regresji liniowej (problem analizy próbki o nieznannej zawartości analitu na podstawie krzywej kalibracyjnej)

Wariant metody regresji liniowej	Wyrażenie na niepewność standardową x_0
Regresja liniowa zwykła	$u(x_0) = \frac{s_{y/x}}{a} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\bar{y}_0 - \bar{y})^2}{a^2 \sum_{i=1}^n (x_i - \bar{x})^2}}$ $x_0 = \frac{y_0 - b}{a}$
Regresja liniowa ważona (Y)	$u(x_0) = \frac{s_{y/x,w}}{a} \sqrt{\frac{1}{w(y_0)} + \frac{1}{\sum_{i=1}^n w_i} + \frac{(\bar{y}_0 - \bar{y}_w)^2}{a^2 \left(\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2 \right)}}$ $x_0 = \frac{\bar{y}_0 - b}{a}$
Regresja liniowa ważona (X, Y)	$u(x_0) = \frac{1}{a} \sqrt{s_{y/x,w}^2 \left(\frac{1}{w(y_0)} + \frac{1}{\sum_{i=1}^n W_i} \right) + \frac{(\bar{y}_0 - \hat{y}_w)^2}{a^2} u^2(a)}$ $\bar{x}_0 = \frac{\bar{y}_0 - b}{a}$

gdzie a jest współczynnikiem nachylenia liniowej zależności dopasowanej do punktów eksperymentalnych, $s_{y/x}$ oraz $s_{y/x,w}$ oznaczają odpowiednio zwykłe i ważne resztowe odchylenie standardowe, \bar{y} oraz \bar{y}_w to odpowiednio średnia arytmetyczna i średnia arytmetyczna ważona wartości zmiennej zależnej, \hat{y}_w jest wartością oczekiwaną średniej wartości zmiennej Y , \bar{x} oraz \bar{x}_w to odpowiednio średnia arytmetyczna i średnia arytmetyczna ważona wartości zmiennej niezależnej, $u(a)$ jest niepewnością standardową współczynnika a , \bar{y}_0 reprezentuje wartość średnią z m powtórzonych pomiarów zmiennej Y badanej próbki oraz n jest liczbą punktów eksperymentalnych (par obserwacji (x_i, y_i)).

Przedstawione w Tabeli 2 wyrażenia ukazują jak zwiększona ilość informacji o badanym układzie włączona do obliczeń regresyjnych modyfikuje wyrażenia do wyznaczania zarówno przewidywanej wartości zmiennej niezależnej x_0 oraz przypisanej jej niepewności standardowej. W pierwszym przypadku (tj. dla wartości x_0) efekt ten jest uwikłany w wartościach współczynników a i b . W przypadku wyrażen na niepewności standardowe wpływ ten uwidacznia się w postaci dodatkowych członów pod pierwiastkiem. Warto zauważyć, że tylko wariant regresji liniowej ważonej ogólnej (X, Y) daje możliwość włączenia do oszacowanej niepewności przewidywanej wartości x_0 niepewności skali zmiennej X (w praktyce analitycznej wiąże się to z włączeniem do $u(x_0)$ niepewności stężeń przygotowanych wzorców lub próbek kalibracyjnych).

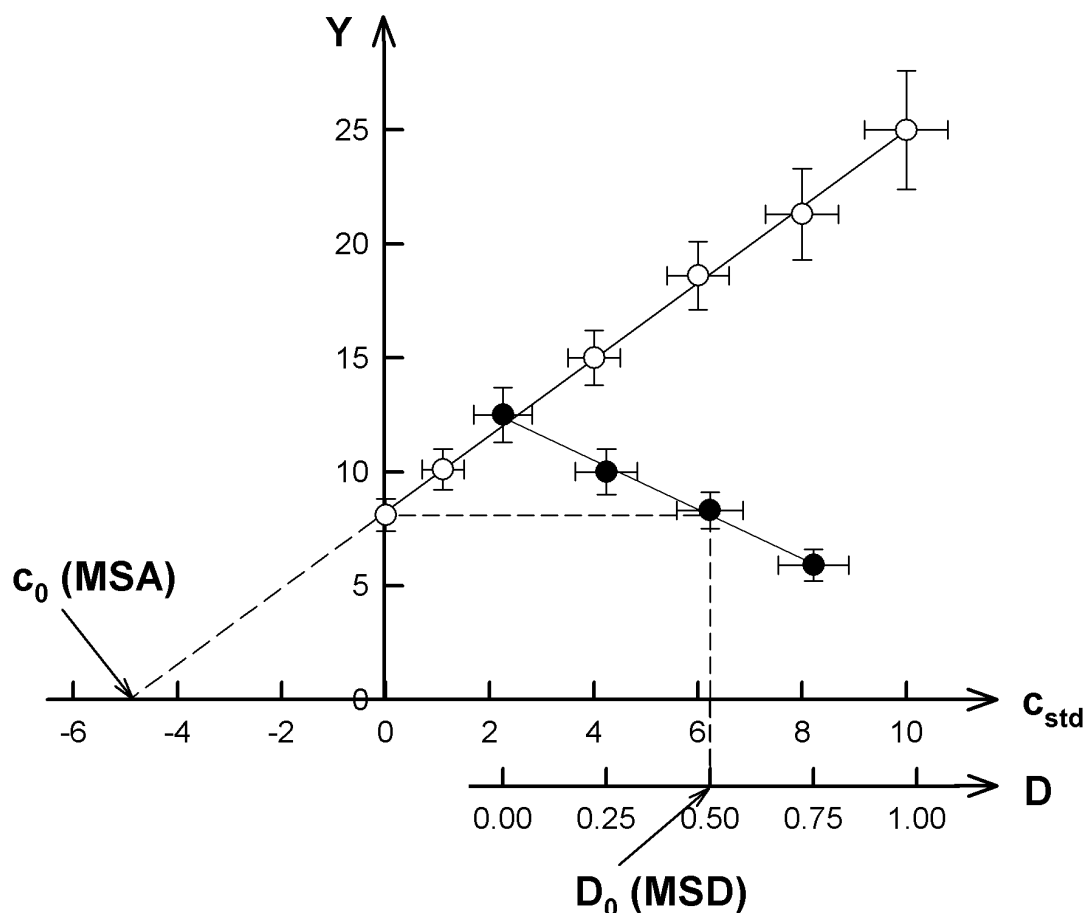
Niewątpliwą zaletą konstrukcji krzywej kalibracyjnej w oparciu o schemat regresji liniowej ważonej (X, Y) jest pełna i właściwa reprezentacja tego elementu w budżecie niepewności metody analitycznej. Ten wariant regresji liniowej pozwala włączyć do budżetu niepewności wszystkie źródła niepewności towa-

rzyszające konstrukcji krzywej kalibracyjnej tj. obok błędu losowego związanego z rozrzutem punktów eksperymentalnych wokół dopasowanej funkcji, niepewności wartości zmiennej zależnej i niezależnej. Krzywa kalibracyjna skonstruowana w oparciu o ten najbardziej ogólny schemat obliczeń regresyjnych może jednak wciąż generować błędne wyniki w pomiarach analitycznych. Ich źródłem jest często niepełne odwzorowanie składu chemicznego próbek kalibracyjnych względem rzeczywistych próbek poddawanych analizie chemicznej. Uwzględnienie wpływów matrycy próbek rzeczywistych umożliwia zastosowanie innych technik kalibracyjnych opartych na regresji liniowej ważonej (X, Y), tj. szeroko rozpowszechnionej metody wielokrotnych dodatków wzorca oraz niedawno zaprezentowanej metody kolejnych rozcieńczeń [3].

3. METODA WIELOKROTNYCH DODATKÓW WZORCA I METODA KOLEJNYCH ROZCIEŃCZEŃ

Typowym podejściem pozwalającym na uniknięcie wpływów niezgodności matrycy próbek kalibracyjnych i rzeczywistych jest zastosowanie metody wielokrotnych dodatków wzorca (MSA) [4,5]. W tym podejściu stężenie analitu w badanej próbce jest wyznaczane na drodze ekstrapolacji krzywej kalibracyjnej dodatków wzorca do zerowej wartości sygnału analitycznego. Metoda wielokrotnych dodatków wzorca wymaga zatem pełnej kalibracji przy każdej analizie chemicznej. Charakteryzuje ją więc duża czasochłonność i znaczące zużycie wzorców. Dodatkowo, technika ta pracuje poprawnie, jeśli oddziaływania między analitem a składnikami matrycy nie zmieniają się przy zwiększaniu stosunku ich stężeń.

Alternatywnym podejściem do metody wielokrotnych dodatków wzorca jest niedawno opracowana metoda kolejnych rozcieńczeń próbki rzeczywistej jedнокrotnie wzbogaconej dodatkiem wzorca (MSD) [3]. W metodzie tej w miejsce kolejnych dodatków wzorca jest zastosowana sekwencja kolejnych dodatków rozpuszczalnika do przygotowanej próbki wzbogaconej. Poszukiwana jest następnie zależność kalibracyjna między sygnałem analitycznym (zmienna zależna Y) a współczynnikiem rozcieńczenia (zmienna niezależna D , zdefiniowana jako stosunek objętości dodanego rozpuszczalnika do całkowitej objętości próbki wzbogaconej). Metoda kolejnych rozcieńczeń zdecydowanie redukuje zużycie wzorców, a przewidywanie stężenia analitu w badanej próbce jest przeprowadzone bardziej dokładnym (w stosunku do ekstrapolacji) działaniem opartym na interpolacji uzyskanej zależności kalibracyjnej. Ideę metody wielokrotnych dodatków wzorca oraz metody kolejnych rozcieńczeń prezentuje graficznie Rysunek 1.



Rysunek 1. Zależności kalibracyjne metody wielokrotnych dodatków wzorca (MSA) oraz metody kolejnych rozcieńczeń (MSD) wraz z graficzną prezentacją sposobu wyznaczenia stężenia analitu w badanej próbce (c_0 oznacza stężenie analitu w badanej próbce, a D_0 – współczynnik rozcieńczenia próbki wzbogaconej odpowiadający stężeniu analitu w niewzbogaconej (badanej) próbce)

Schemat postępowania w metodzie kolejnych rozcieńczeń obejmuje następujące eksperymentalne kroki

1. Pomiar sygnału analitycznego próbki rzeczywistej (badanej)
2. Dodatek wzorca do próbki rzeczywistej
3. Podział próbki wzbogaconej z punktu 2 na kilka równych porcji
4. Dodatek wzrastających objętości rozpuszczalnika do kolejnych porcji próbki wzbogaconej z punktu 3 (przy czym jedna z nich pozostaje nierozcieńczona)
5. Pomiar sygnału analitycznego każdej porcji z punktu 4 (nierozcieńczona i rozcieńczone porcje próbki wzbogaconej)
6. Konstrukcja zależności kalibracyjnej między wartościami sygnału analitycznego Y uzyskanymi w punkcie 5 a wyznaczonymi współczynnikami rozcieńczenia D za pomocą właściwej metody regresji liniowej

Wyznaczanie stężenia analitu w próbce badanej jest przeprowadzone na podstawie wartości D_0 , czyli współczynnika rozcieńczenia próbki wzbogaconej odpowiadającego stężeniu analitu w niewzbogaconej (badanej) próbce. Parametr D_0 przekształca się następnie na stężenie analitu c_0 za pomocą następującej formuły:

$$c_0 = c_{std} \left(\frac{1}{D_0} - 1 \right)$$

gdzie c_{std} jest stężeniem wzorca w przygotowanej próbce wzbogaconej.

Niepewność standardowa przewidywanego stężenia c_0 można oszacować na podstawie reguł propagacji niepewności korzystając z następującego wyrażenia:

$$u(c_0) = c_0 \sqrt{\frac{u^2(D_0)}{D_0^2(1-D_0)^2} + \frac{u^2(c_{std})}{c_{std}^2}}$$

gdzie $u(D_0)$ jest niepewnością standardową D_0 wyznaczoną zgodnie z przyjętym wariantem metody regresji liniowej (Tabela 2), a $u(c_{std})$ jest niepewnością standardową stężenia wzorca dodanego do oryginalnej (badanej) próbki.

4. PODSUMOWANIE

Wyrażenia przedstawione w pracy stanowią istotny suplement matematycznej charakterystyki technik opartych na metodzie regresji liniowej, wykorzystywanych do rozwiązywania wielu różnych problemów w praktyce analitycznej. Problemy te obejmują zagadnienia typowe dla rutynowych działań pomiarowych oraz aspekty związane z obszarem sterowania jakością pomiarów. Wśród nich szczególne miejsce zajmują: wiarygodne oznaczanie analitu w próbkach charakteryzujących się prostymi oraz złożonymi macierzami, kalibracja aparatury pomiarowej, określanie granic wykrywalności i oznaczalności metod analitycznych, porównywanie statystycznych odstępstw między dwoma podejściami w procesie pomiarowym (porównywanie dwóch metod) itp. Wybór techniki kalibracyjnej do rozwiązania postawionego problemu powinien też uwzględniać czynnik ekonomiczny, co świetnie ilustruje zestawienie metod wielokrotnych dodatków wzorca i kolejnych rozcieńczeń.

LITERATURA:

- [1] Hyk W., Stojek Z.: Analiza statystyczna w laboratorium analitycznym, Wyd. 3, Wydział Chemii Uniwersytetu Warszawskiego, Warszawa 2010.
- [2] Hyk W.: Analiza statystyczna online w laboratorium analitycznym, www.e-stat.pl.
- [3] Hyk W., Stojek Z.: Anal. Chem., 85 (2013) 5933.
- [4] Saxberg, Bo E.H., Kowalski B.R.: Anal. Chem., 51 (1979) 1031.
- [5] Kelly W.R., MacDonald B.S., Guthrie, W.F.: Anal. Chem. 80 (2008) 6154.

Sekretariat Klubu POLLAB
ul. Kłobucka 23 A
02-699 Warszawa
www.pollab.pl
e-mail: sekretariat@pollab.pl
tel.: 22 464 55 03
faks: 22 464 55 56