

Automatyzacja walidacji metod pomiarowych

Michał Iwaniec,
Miroslaw Popieluch

StatSoft Polska Sp. z o.o.



Walidacja i monitorowanie wiarygodności metod pomiarowych jest jednym z bardzo ważnych aspektów zapewnienia odpowiedniej jakości produktu końcowego. Podobnie jak to jest w przypadku procesu produkcji, ważne jest, żeby poznać i zrozumieć, jak funkcjonuje system pomiaru, jakie czynniki mogą istotnie wpływać na wynik pomiarów oraz jaka jest niepewność oceny danej wielkości.

W przypadku zastosowań produkcyjnych konieczna jest wiedza na temat zmienności otrzymywanych pomiarów dla wykorzystywanej metody tak, aby była ona odpowiednio mała w stosunku do zmian, które mogą zachodzić w procesie. W przeciwnym przypadku nie będzie możliwości wykrycia niekorzystnych zmian, przez co może ucierpieć jakość produktu. W poniższym artykule zostanie przedstawiona metodyka, która została zaproponowana w polskiej normie i stanowi pełną ocenę przydatności metody.

Trudna droga od danych do wiedzy

Proces walidacji jest niestety czasochłonny i wymaga od przeprowadzających go osób nie tylko wiedzy z zakresu metod pomiarowych, ale także wiedzy z zakresu statystyki. Powoduje to, że walidacja sprawia wiele problemów technicznych i metodycznych. Naturalne jest w związku z tym dążenie do usprawnienia tego procesu tam, gdzie jest to możliwe. W przypadku części procesu walidacji, w której opisywana jest metoda pomiarowa, sposób przygotowania próbki, sprzęt itp. usprawnienie pracy może polegać na przyjęciu odpowiedniego szablonu, ale wszystkie działania muszą zostać opisane ręcznie.

Inaczej jest w przypadku statystycznej oceny danych z eksperymentu walidacyjnego. Po pierwsze osoba, która bardzo dobrze zna wykorzystywane przez siebie metody pomiarowe, niekoniecznie musi być biegła w metodach statystycznych pozwalających na ocenę teje metody. Pojawia się więc pierwsza przeszkoda związana z ustaleniem odpowiednich kroków postępowania podczas statystycznej obróbki danych. Następną przeszkodą są same obliczenia, których przeprowadzenie w sposób ręczny jest bardzo trudne i z dużym prawdopodobieństwem doprowadzi do popełnienia błędów.

Rozwiązaniem przejściowym może być stworzenie odpowiedniego arkusza kalkulacyjnego, który wspomże obliczenia, ale takie podejście ma kilka istotnych wad:

- implementacja metod statystycznych w arkuszu nie jest prosta,
- przed wykorzystaniem takiego rozwiązania należy przeprowadzić procedurę jego walidacji, co jest skomplikowane i czasochłonne.

StatSoft, korzystając z wieloletniego doświadczenia nie tylko na polu metod statystycznych, ale również w dziedzinie wdrażania systemów komputerowych, zaproponował rozwiązanie, które rozwiązuje opisane wyżej trudności. Dostarczane przez nas rozwiązanie z powodzeniem może być wykorzystane do walidacji większości metod pomiarowych.

Statystyka w walidacji – kolejne kroki

W celu potwierdzenia wysokiej jakości systemów pomiarowych niezbędne jest udokumentowanie ich poprawnego działania w oparciu o „twarde” dowody. Takim dowodem jest eksperyment, podczas którego w odpowiedni sposób zbierane są dane pomiarowe. Dane pomiarowe w większości przypadków będą charakteryzować się niepewnością i dlatego właśnie musimy ocenić je, wykorzystując metody statystyczne. Statystyka jest dziedziną subtelną, wymagającą pewnej inwencji i intuicji od badacza, co powoduje, że często nie wystarcza sama „technika obliczeniowa”, ale ważniejsza jest odpowiednia interpretacja wyników. Ponadto nie można w trakcie analizy danych walidacyjnych stosować różnych metod statystycznych w sposób chaotyczny, gdyż wyniki otrzymane w jednych zależą często od innych.

Poniżej przybliżone zostaną metody statystyczne w kolejności, jaka zazwyczaj obowiązuje w trakcie obróbki wyników pochodzących z eksperymentu walidacyjnego. Wykorzystany zostanie przykład, który pozwoli

na przytoczenie większości metod statystycznych wykorzystywanych w walidacji metody.

Zawartość siarki w węglu. Pomiary wzorców na 4 poziomach.					
	Poziom 1	Poziom 2	Poziom 3	Poziom 4	Id grupy
1	0.71	1.2	2.43	4.02	1
2	0.68	1.21	2.58	3.95	1
3	0.75	1.27	2.56	4.05	1
4	0.63	1.31	2.5	4.09	1
5	0.69	1.19	2.54	4.02	2
6	0.67	1.28	2.52	4.03	2
7	0.67	1.26	2.43	4.05	2
8	0.67	1.27	2.48	3.96	3
9	0.65	1.29	2.46	4.02	3
10	0.68	1.26	2.48	4	3
11	0.68	1.19	2.47	4.02	4
12	0.67	1.29	2.53	3.95	4
13	0.68	1.32	2.5	3.99	4
14	0.64	1.23	2.45	3.97	5

Rys. 1. Dane wejściowe do analiz – arkusz STATISTICA

Zmienna	Statystyki opisowe (Siarka.sta)	
	Średnia	Odch. std
Poziom 1	0.681481	0.027970
Poziom 2	1.231481	0.048493
Poziom 3	2.493333	0.039419
Poziom 4	4.011111	0.045938

Rys. 2. Statystyki opisowe otrzymane w STATISTICA

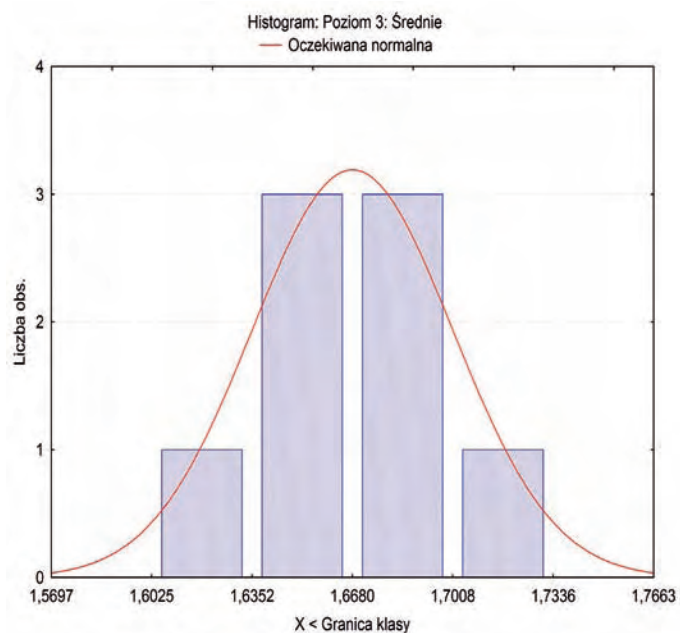
Przykładowe dane pochodzą z oznaczania zawartości siarki w węglu i zostały pokazane w tabeli. Wykorzystane wartości wzorców dla kolejnych poziomów wynoszą odpowiednio: 0.7; 1.255; 2.45; 4.07. Dla potrzeb przykładu przyjmujemy, że wartości te zostały wyznaczone z niepewnością rozszerzoną 0.04 (niepewność rozszerzona jest wielkością, która mówi, w jakim zakresie znajduje się większość otrzymywanych pomiarów). Dane zostały zaczerpnięte z [1].

Opisowa analiza danych

Pierwszym krokiem jest opis zebranych danych. Oczywiście w końcowym raporcie z walidacji muszą znaleźć się dane surowe w postaci

Id grupy	Poziom 1		Poziom 2		Poziom 3		Poziom 4	
	Średnie	Odch. std	Średnie	Odch. std	Średnie	Odch. std	Średnie	Odch. std
1	0.692500	0.050580	1.247500	0.051881	2.517500	0.067515	4.027500	0.059090
2	0.676667	0.011547	1.243333	0.047258	2.496667	0.058595	4.033333	0.015275
3	0.666667	0.015275	1.273333	0.015275	2.473333	0.011547	3.993333	0.030551
4	0.676667	0.005774	1.266667	0.068069	2.500000	0.030000	3.986667	0.035119
5	0.670000	0.022361	1.226000	0.011402	2.494000	0.032863	4.024000	0.048785
6	0.666667	0.015275	1.210000	0.060828	2.480000	0.034641	3.993333	0.035119
7	0.720000	0.010000	1.223333	0.020817	2.483333	0.051316	4.020000	0.096437
8	0.686667	0.032146	1.160000	0.036056	2.493333	0.025166	3.996667	0.023094

Rys. 3. Statystyki opisowe dla klas otrzymane w STATISTICA

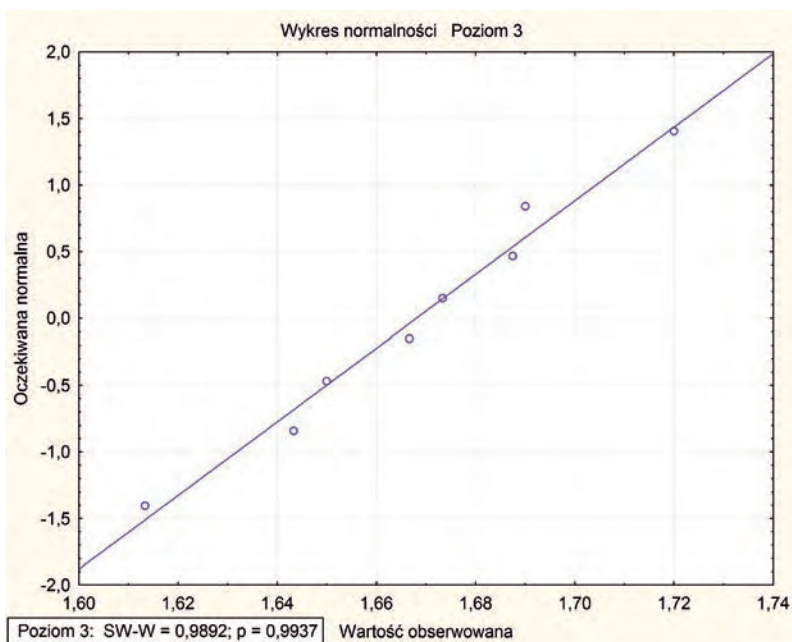


Rys. 4. Histogram dla poziomu 3 otrzymany w STATISTICA

tabeli, ale ludzkie oko będzie miało problem ze wstępnym oszacowaniem zbioru surowych liczb. Dlatego na początku zazwyczaj obliczamy takie statystyki, jak wartość średnia i odchylenie standardowe, które mówią nam, gdzie jest „środek” danych i jak bardzo się one zmieniają. Poniżej znajduje się zestawienie w postaci tabeli (rys. 2.)

Powyżej zostały obliczone statystyki dla kolejnych poziomów w całości. Zauważmy jednak, że podczas zbierania danych pomiarowych zostały one podzielone na grupy (klasy), o czym mówi zmienna grupująca „Id grupy” w pliku danych. Dane w obrębie grupy mogły być zebrane np. w różnym czasie, co mogło wprowadzić dodatkową zmienność (niepewność) do pomiarów i warto przyrzeć się statystykom w kolejnych grupach (Rys. 3.).

W tabeli widać, że w grupach kolejne wartości średnie i odchylenia standardowe różnią się nieznacznie od siebie. Powstaje w związku z tym pytanie, czy różnice te są istotne. Aby w sposób formalny odpowiedzieć sobie na to pytanie, należy przeprowadzić odpowiednie testy, o których będzie mowa w kolejnym paragrafie. Zanim jednak będzie można wykorzystać większość testów statystycznych, należy zweryfikować, czy zebrane dane są zgodne z rozkładem normalnym. Na początek dobrze jest stworzyć histogram, w tym przypadku 4 histogramy osobno dla każdego poziomu. Poniżej przedstawiony jest przykładowy wykres dla poziomu 3 (Rys. 4.).



Rys. 5. Wykres normalności poziomu 3 otrzymany w STATISTICA

Histogram pokazuje empiryczny rozkład wartości, do którego można dopasować teoretyczną krzywą gęstości rozkładu normalnego. Daje to ogólny pogląd na „zgodność” zbioru danych z rozkładem normalnym. Należy zauważyć, że w przypadku małej liczby danych wyciąganie wniosków o ich normalności często będzie błędne. Dużo lepszym podejściem jest zastosowanie wykresu normalności, który jest tak utworzony, że dla danych normalnych zależność pomiędzy danymi a oczekiwaną normalnością jest liniowa. Wykres normalności dla danych z histogramu przedstawiony jest poniżej.

Widać, że pokazana niżej zależność jest liniowa, zatem można uznać, że normalność została zweryfikowana pozytywnie. W raporcie z walidacji powinien również znaleźć się „liczbowy” test weryfikujący normalność, np. test Shapiro-Wilka (Rys. 5.).

Wartości odstające

Testy wykrywające wartości odstające w zbiorach danych są poniekąd związane z rozkładem normalnym, i poszukując wartości odstających, należy wziąć pod uwagę normalność, i badając normalność wziąć pod uwagę wartości odstające. Zaproponowane poniżej testy są najbardziej odpowiednie dla danych o rozkładzie normalnym, w których wystąpiła wartość odstająca. W przypadku, kiedy spodziewamy się, że dane będą mieć rozkład inny niż normalny, testy te mogą doprowadzić do błędnych wniosków.

Korzystając z danych przykładowych, możemy ocenić, czy wśród wartości średnich i wśród odchyłeń standardowych w klasach nie ma wartości odstających. W celu sprawdzenia, czy w zbiorze danych jest wartość odstająca ze względu na wartość średnią, w pierwszej kolejności musimy znaleźć kandydata na wartość odstającą, czyli wartość, która jest najbardziej oddalona od ogólnej średniej. Następnie wartość bezwzględną różnicy pomiędzy wartością średnią a wartością kandydata należy podzielić przez odchylenie standardowe. Otrzymujemy w ten sposób wartość statystyki zwanej statystyką Grubbsa:

$$G = \frac{|\bar{x} - x_{\max}|}{s}$$

Pozostaje nam otrzymaną wartość G porównać z wartością krytyczną, którą możemy odczytać z odpowiednich tablic. Jeżeli G będzie zbyt wysokie, wtedy należy uznać kandydata za wartość odstającą.

Analogicznie należy postąpić podczas wykrywania, czy największe odchylenie standardowe w klasach nie odstaje zbyt od pozostałych. Oczywiście inaczej jest konstruowana statystyka testowa, którą nazywamy statystyką Cochran:

$$C = \frac{s_{\max}^2}{\sum_{i=1}^n s_i^2}$$

Obliczone C tak jak wyżej porównujemy z odpowiednią wartością krytyczną z tablic testu Cochran.

Należy zwrócić uwagę, że przeprowadzając test na wartość odstającą, możemy odrzucić tylko jedną wartość i po odrzuceniu tej wartości należy powtórzyć test kolejny raz.

Normalność a wartości odstające

Zauważmy, że wystąpienie wartości odstającej „niszczy” zbiór danych tym bardziej, im bardziej wartość jest oddalona od pozostałych. Należy również pamiętać, że wartości odstające mają duży wpływ na badanie normalności, a konkretnie wystąpienie wartości odstającej spowoduje najprawdopodobniej odrzucenie założenia o normalnym rozkładzie zbioru danych. Często problem ten można zaobserwować na wykresie normalności, który pozwala w sposób graficzny ocenić podobieństwo zbioru do rozkładu normalnego.

Podczas graficznej oceny normalności danych, jak również podczas obliczania liczbowych testów istotna jest liczba danych. Im mniej elementów w zbiorze danych, tym trudniej ocenić normalność w sposób graficzny i jednocześnie zaobserwować wartości odstające. Im mniej danych, tym bardziej przydatne są liczbowe testy normalności, takie jak test Shapiro-Wilka (w skrócie test S-W).

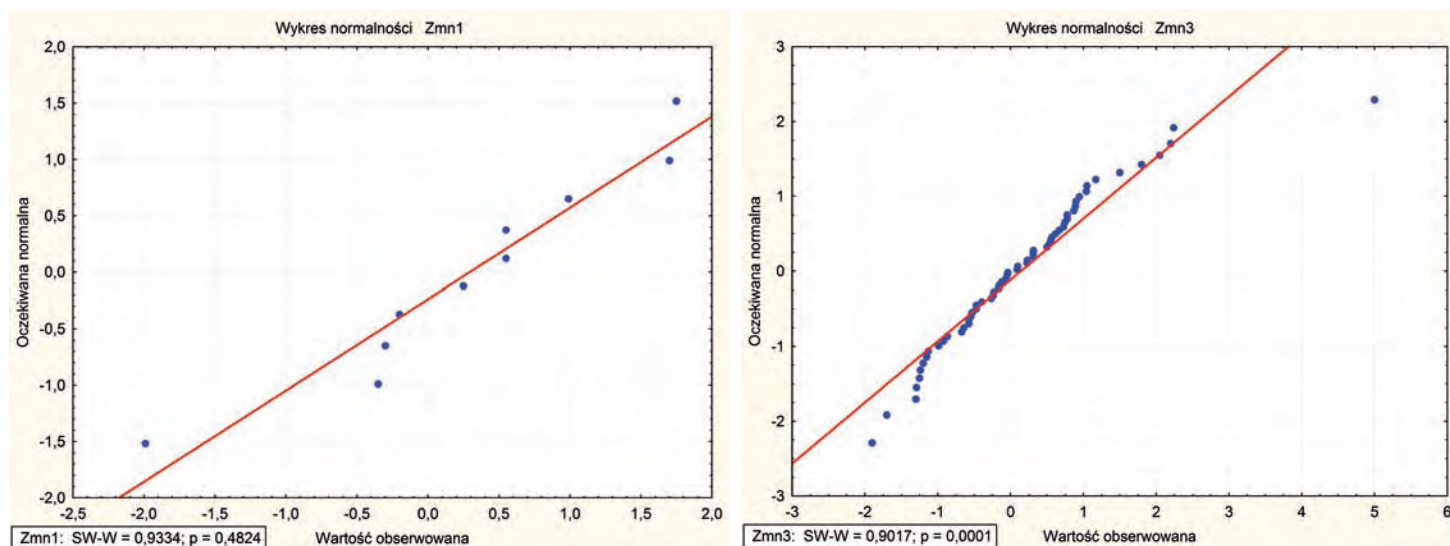
Rozważmy przykład pokazany na rys. 6, w którym na pierwszym wykresie przedstawiony jest mały zbiór danych, którego rozkład jest podobny do rozkładu normalnego. Drugi wykres sporządzono na podstawie zbioru o większej liczbie danych, w którym stwierdzono jedną wartość odstającą (punkt w zielonej obwódce). W pierwszym przypadku prawdopodobieństwo testowe p testu S-W wynosi 0,48, a w drugim jest bliskie 0. Oznacza to, że dane na wykresie po lewej wykazują podobieństwo do rozkładu normalnego ($p > 0,05$), natomiast na wykresie po prawej nie ($p < 0,05$ i obserwujemy tam wartość odstającą).

Dlatego, zanim przejdziemy do kolejnych analiz na zbiorze danych, powinniśmy przyjrzeć się wartości odstającej, jeżeli nie był to błąd powstały podczas wpisywania danych, powinniśmy ją usunąć ze zbioru, gdyż najprawdopodobniej jest obciążona błędem grubym i będzie istotnie zaburzać następne analizy.

Precyzja metody pomiarowej.

Ocena powtarzalności i odtwarzalności

Do tej pory omówiona została wstępna analiza danych, która z jednej strony pomaga lepiej zrozumieć dane, a z drugiej usunąć podejrzane wartości, które mogą zafałszować obliczenia. Po sprawdzeniu normal-



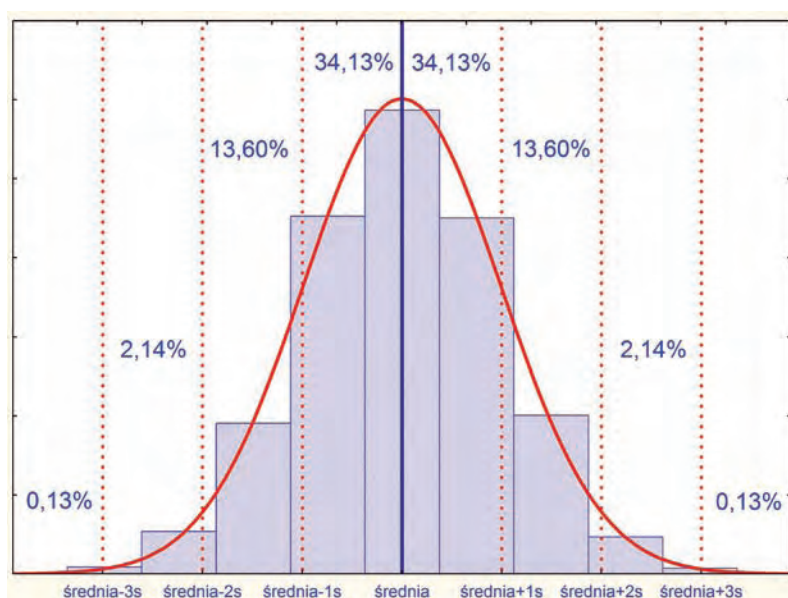
Rys. 6. Przykładowe wykresy normalności zbiorów danych o zróżnicowanej liczebności

ności danych i odrzuceniu ewentualnych wartości odstających możemy oszacować to, co jest najbardziej krytyczne dla systemu pomiarowego, czyli jego zmienność. Precyzja metody analitycznej jest jednym z najważniejszych parametrów charakteryzujących metodę pomiarową. Jest pojęciem ogólnym opisującym zmienność wyników powtarzanych badań, a potrzeba jej stosowania wynika z niemożliwości uniknięcia błędów różnego rodzaju podczas wykonywania analiz.

Pojęcie zmienności (sigmy) w statystyce jest pojęciem wymagającym dokładniejszego omówienia. W najprostszym przypadku dla zbioru danych możemy obliczyć wartość średnią i odchylenie standardowe. Wartość średnia mówi, gdzie jest najprawdopodobniej środek (centrum) danych, natomiast trudniej jest zrozumieć, o czym mówi odchylenie standardowe, będące miarą zmienności zbioru danych. Odchylenie standardowe informuje, jaka jest średnia odległość poszczególnych danych od wartości średniej. W przypadku danych pomiarowych należy pojęcie odchylenia standardowego powiązać z rozkładem danych, który jest podobny do rozkładu normalnego. Odejmując i dodając odchylenie standardowe od wartości średniej, otrzymujemy zakres, w którym wartość pomiaru może znaleźć się z pewnym prawdopodobieństwem. Przykładowo w odległości jednego odchylenia standardowego od średniej znajduje się około 1 pomiarów, a w odległości trzech odchyleń standardowych od wartości średniej 99,73% pomiarów, czyli „prawie wszystkie”. Rozkład prawdopodobieństw występowania pomiarów w danym zakresie jest oczywiście związany z krzywą gęstości rozkładu normalnego i został przedstawiony na rys. 7.

Przejdźmy teraz do sedna, czyli do obliczenia zmienności dla systemu pomiarowego, za pomocą którego otrzymujemy wyniki analiz. Okazuje się, że w całościowym zaobserwowanym wskaźniku zmienności możemy wyodrębnić pewne jego składowe, np. powtarzalność i odtwarzalność. Zdarza się, że w zależności od kontekstu użycia tych terminów ich interpretacja różni się, my natomiast postaramy się wytłumaczyć ich sens w sposób ogólny.

Żeby oddzielić od siebie te dwa składniki zmienności, należy w odpowiedni sposób pobrać dane. Wyobraźmy sobie, że mierzymy



Rys. 7. Krzywa gęstości rozkładu normalnego.

wartość wzorcową i pomiar jest powtarzany kilka razy przez laboranta w jednakowych warunkach pomiarowych, które obejmują:

- taką samą procedurę pomiarową,
- ten sam przyrząd pomiarowy,
- to samo miejsce oraz
- powtarzanie w krótkich odstępach czasu.

Otrzymana na podstawie tych kilku pomiarów miara zmienności, np. odchylenie standardowe (s_r), mówi o tym, jak dobrze w stałych warunkach jesteśmy w stanie powtórzyć pomiar, czyli jest to miara powtarzalności. Następnie inny laborant mierzy ten sam wzorec kilkakrotnie i znów możemy otrzymać pewne oszacowanie powtarzalności (s_2). Możemy teraz wyliczyć powtarzalność dla systemu pomiarowego, która będzie sumą wariancji otrzymaną przez dwóch laborantów: $s_r = \sqrt{s_1^2 + s_2^2}$. Otrzymana wartość jest miarą zmienności mówiącą o najlepszej, czyli najmniejszej zmienności systemu pomiarowego. Ale może być gorsza, czyli wynikająca z odtworzenia wyniku pomiaru przez

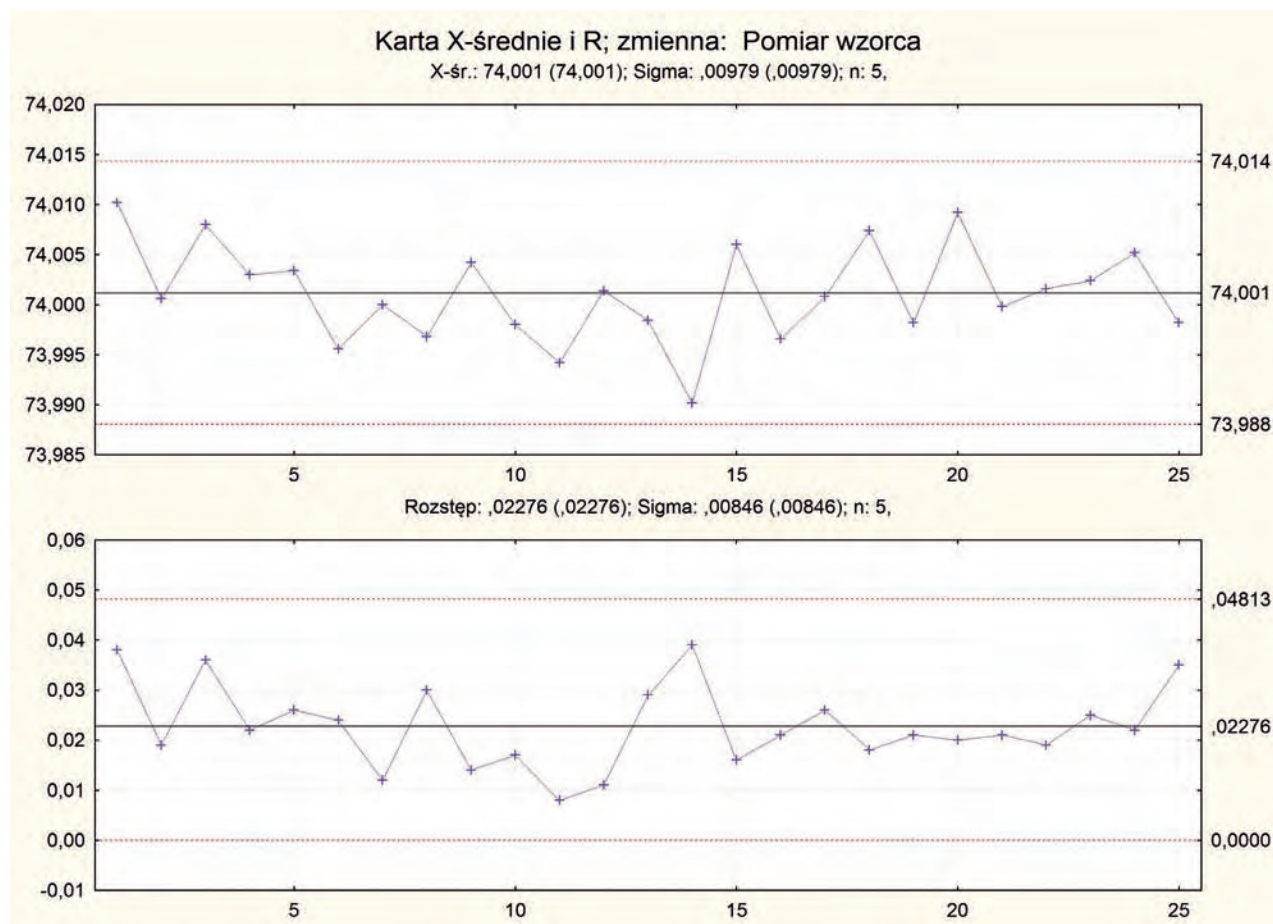
różnych operatorów w tym samym laboratorium, za pomocą różnej aparatury, w ciągu dłuższego przedziału czasowego, czyli wyznaczona w warunkach odtwarzalności. Warunki odtwarzalności obejmują również sytuację, gdy pomiar wykonywany jest w różnych laboratoriach. Wtedy procedura nieznacznie komplikuje się obliczeniowo i jest dokładnie opisana w normie PN-ISO 5725-2:2002 na str. 25. Może się okazać (i tak byłoby najlepiej), że odtwarzalność nie była istotna, a mówiąc bardziej precyzyjnie nie była istotnie większa od powtarzalności. Oznaczałoby to, że jakość pomiaru nie zależy od tego, który z laborantów go wykonuje.

Doszliliśmy tutaj niejako do problemu testowania istotności różnic pomiędzy wartościami zmienności otrzymanymi z różnych zbiorów danych. Wyobraźmy sobie sytuację, kiedy chcemy porównać ze sobą dwie metody pomiarowe. Należy wykonać wówczas równoległe oznaczenie badanego analitu metodą badaną oraz inną metodą uznaną za referencyjną, dającą wiarygodne i dokładne wyniki. Przygotowujemy więc odpowiedni wzorec CRM czy RM i wykonujemy odpowiednią do zastosowanego materiału liczbę powtórzeń pomiaru dla każdej metody. To, czy zmienność wyników nie różni się istotnie, możemy zweryfikować, wykorzystując prosty w obliczeniu test statystyczny oparty na rozkładzie χ^2 .

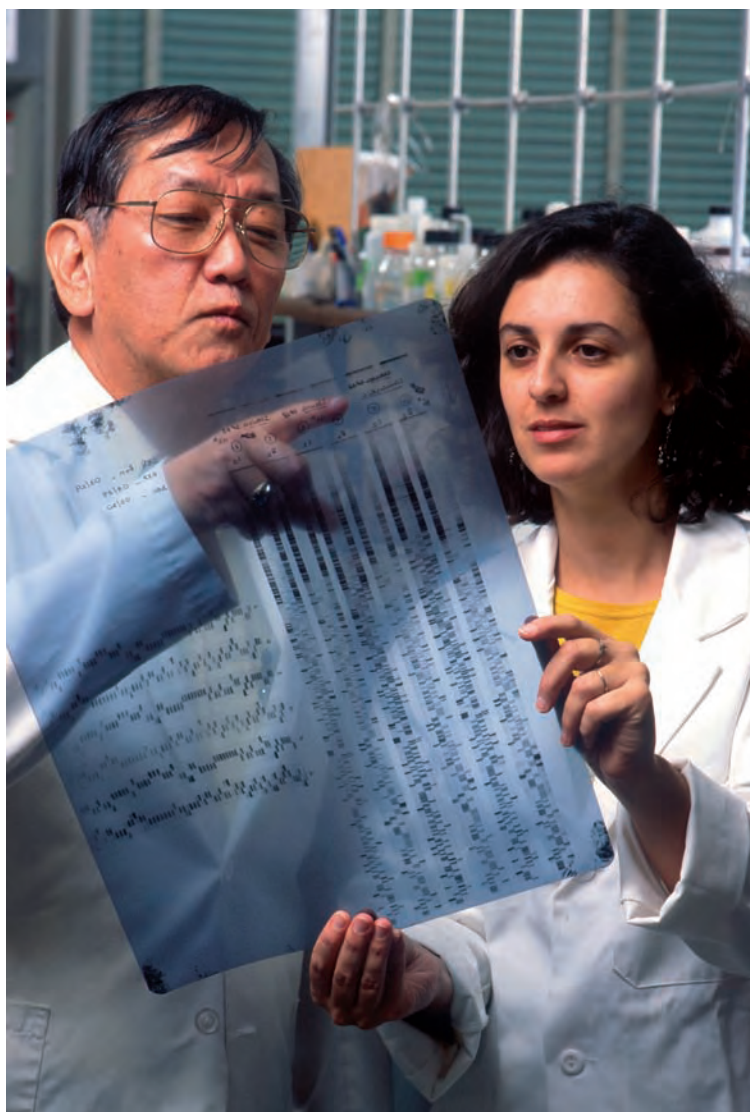
Poprawność metody pomiarowej

Podczas oceny metody pomiarowej mówimy o jej dokładności. Według międzynarodowej normy ISO 5725 na dokładność składają się poprawność oraz precyzja. Poprawność związana

jest ze zgodnością pomiędzy średnią arytmetyczną z dużej liczby pojedynczych wyników badania a wartością prawdziwą lub przyjętą wartością odniesienia, natomiast precyzja jest miarą zgodności pomiędzy wynikami badania. Składnikami precyzji są powtarzalność i odtwarzalność, o których była mowa wyżej. Żeby w pełni ocenić dokładność, musimy oszacować również poprawność. O poprawności mówić można wtedy, gdy jest sens mówić o prawdziwej wartości mierzonej wielkości. Niestety wartość prawdziwa (rzeczywista) jest zazwyczaj nieznaną. W wielu przypadkach możemy ją wyznaczyć, wykorzystując wzorec, certyfikowany materiał referencyjny (CRM) lub materiał referencyjny (RM), który został przygotowany z odpowiednią dla naszej metody dokładnością. W przypadku stosowania materiału CRM (ze znaną zawartością analitu - μ) dokładność metody wyznacza się, wykonując analizę z sześciokrotnym powtórzeniem, następnie na podstawie otrzymanych wyników oblicza się wartość średnią \bar{x} oraz odchylenie standardowe s . Jeżeli wyznaczona wartość średnia mieści się w zakresie $\mu - 2s < \bar{x} < \mu + 2s$, to oznacza, że badana metoda jest dokładna, a poprawność i precyzja są zadowalające. W przypadku gdy stosowany jest materiał referencyjny (RM), wykonuje się 10 powtórzeń, wyznacza się wartość średnią \bar{x} oraz odchylenie standardowe s . Na podstawie uzyskanych wyników za pomocą statystyki t-Studenta oblicza się wartość t , następnie porównuje jej wartość z wartością krytyczną odczytaną z tablic dla określonej liczby stopni swobody i oczekiwanego poziomu ufności. Jeśli obliczona wartość statystyki $t < t_{kryt}$, oznacza to, że metoda



Rys. 8. Karta kontrolna otrzymana w STATISTICA



jest dokładna i nie ma istotnej różnicy pomiędzy wartością średnią a wartością prawdziwą badanego analitu w materiale referencyjnym (RM). Oprócz omówionych sposobów wyznaczenie dokładności metody można przeprowadzić również na drodze międzylaboratoryjnych badań porównawczych. W tym miejscu warto zauważyć, że jest to metoda kosztowna i rzadko stosowana. Jednak dla laboratorium uczestnictwo w programach badań międzylaboratoryjnych, zwłaszcza dla metod, gdzie trudno jest uzyskać odpowiedni wzorzec czy materiał certyfikowany, może dostarczyć wielu cennych informacji na temat metody pomiarowej oraz ocenić własną biegłość w danym zakresie.

Jak więc ocenić, czy metoda pomiarowa jest wystarczająco poprawna? Najbardziej ogólnie, należy wykonać kilka pomiarów znanej wielkości (wzorca) i obliczyć różnicę pomiędzy wartością średnią z pomiarów i wartością wzorcową. Łatwo zauważyć, że zazwyczaj różnica ta będzie różna od zera. Weźmy jednak pod uwagę fakt, że jednocześnie mamy do czynienia ze zmiennością wartości pomiarowych, która została wcześniej oszacowana jako precyzja, czyli suma powtarzalności i odtwarzalności (oznaczymy ją przez σ). Dlatego jeżeli teraz otrzymaliśmy różnicę np. -0,13, to jeżeli wykonamy ponownie kilkanaście pomiarów tego wzorca, to możemy dostać

różnicę np. 0,03. Wynika to stąd, że σ (zmienność) jest większa od zera. Powinniśmy, więc przeprowadzić test statystyczny, który odpowie na pytanie, czy otrzymana różnica jest przypadkowa, czy też powinniśmy się nią przejmować. Tworzymy więc statystykę testową, która będzie miała w przybliżeniu rozkład normalny. Im większa różnica pomiędzy wartością wzorca a wartością średnią z pomiarów tym gorzej, czyli moduł różnicy znajdzie się w liczniku wzoru. Z drugiej strony im większa jest σ , tym trudniej określić, czy różnica jest istotna, gdyż „ginie” ona w zmienności. Czyli wartość σ znajduje się w mianowniku naszej statystyki testowej, która ostatecznie wygląda następująco:

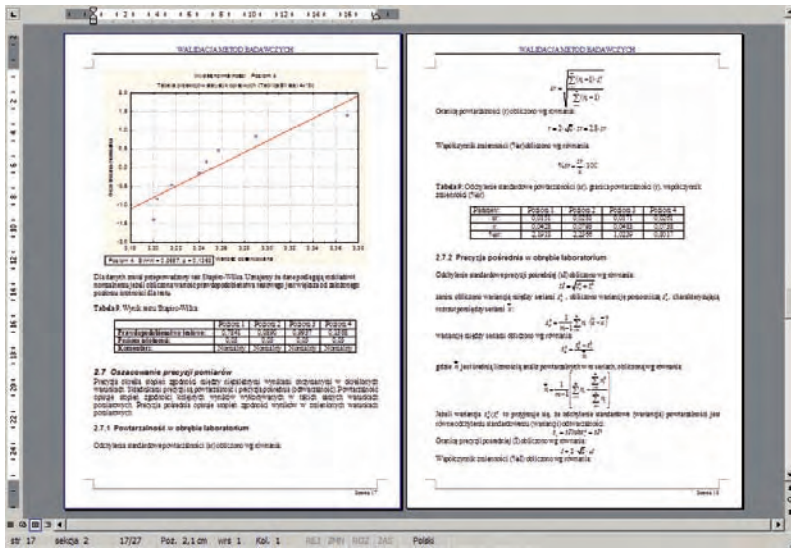
$$N = \frac{|x_w - \bar{x}|}{\sigma}$$

Wartość statystyki N zawsze będzie różna od zera, pozostaje więc pytanie, jaka musi być ta wartość, aby móc powiedzieć, że różnica jest już zbyt duża, aby była przypadkowa. Na pytanie to nie odpowiemy wprost, ale poprzez prawdopodobieństwo dla rozkładu normalnego, odpowiadające wartości N. Podobnie jak w przypadku testu Shapiro-Wilka dopóki wartość prawdopodobieństwa jest większa od 0,05, dopóty nie ma podstaw aby twierdzić że różnica jest istotna statystycznie.

Jeżeli jednak wykażemy z dużym prawdopodobieństwem, że różnica jest istotna, wtedy warto zastanowić się nad stałą korektą dla wykonywanych pomiarów.

Liniowość

W przypadku metod ilościowych szczególnie ważnym parametrem opisującym metodę analityczną jest liniowość. Oznacza ona zdolność metody do uzyskiwania wyników badań bezpośrednio (lub pośrednio po odpowiednich przekształceniach matematycznych) proporcjonalnych do stężenia analitu. Razem z liniowością powinien być wyznaczony zakres roboczy metody analitycznej, czyli przedział, w obrębie którego metoda daje wyniki o akceptowanej precyzji, dokładności i liniowości. Najprościej liniowość określa się poprzez kolejne rozcieńczenia substancji wzorcowej wykonywane w taki sposób, aby pokryły 40-100% oczekiwanego zakresu roboczego. Najkorzystniej jest wykonać 6-10 roztworów (kolejnych rozcieńczeń wzorca) pokrywających oczekiwany przedział roboczy z dodatkowym pomiarem w ślepej próbce. Każdy roztwór powinien być analizowany, co najmniej trzykrotnie, a wynik dla danego stężenia jest średnią arytmetyczną ze wszystkich powtórzeń. Następnie, po otrzymaniu wyników i przeprowadzeniu ich analizy wraz z uwzględnieniem rachunku błędów wynikającego z procedury rozcieńczania wzorca wykreśla się wykres. Sprawdza on korelację pomiędzy wartościami stężeń oczekiwanych a wartościami stężeń otrzymanych dla poszczególnych próbek (oś X stężenia oczekiwane, Y stężenia uzyskiwane). Otrzymane punkty powinny układać się w linię prostą $y=ax+b$. Kolejnym bardzo ważnym etapem jest analiza otrzymanej prostej i układu punktów. Gdy punkty układają się



Rys. 9. Przykład automatycznie wygenerowanego raportu

w sposób liniowy, co potwierdza współczynnik korelacji liniowej Pearsona r^2 - im bliższy jedności (najlepiej 0,97-1,00), tym lepsza proporcjonalność wyników otrzymanych ze stężeniami założonymi - to możemy wyciągnąć wniosek, że metoda w badanym zakresie jest liniowa. Problemy powstają wówczas, gdy obserwujemy rozrzuty punktów od liniowego przebiegu, co może wskazywać na błędy w rozcieńczeniach lub na brak powtarzalności w obrębie jednego stężenia, jeśli widoczne są pojedyncze odstające od wykresu punkty. Artykuł na ten temat pod tytułem „Kalibracja liniowa w zagadnieniu walidacji metod pomiarowych” można znaleźć w [5].

Stabilność

Jakość procesów pomiarowych, podobnie jak to jest w przypadku procesów produkcyjnych, nie jest w te procesy wpisana i ciągle należy ją kontrolować. Kontrolę taką możemy przeprowadzić w obrębie własnego laboratorium, stosując karty kontrolne Shewharta lub (oraz) uczestnicząc regularnie w badaniach biegłości. Karty kontrolne Shewharta lub ich modyfikacje można stosować do monitorowania poprawności i precyzji na jednym poziomie lub też można monitorować jakość samej funkcji kalibracji.

W obu przypadkach należy regularnie przeprowadzać pomiary obiektów o znanych właściwościach i po wykonaniu niezbędnych obliczeń umieszczać wyniki na karcie lub kartach kontrolnych. Wynik uznajemy za stabilny, jeżeli mieści się w wyznaczonych wcześniej granicach kontrolnych. Przykład karty kontrolnej wartości średniej i rozstępu znajduje się na rys. 8.

Sposób wyznaczenia granic kontrolnych zależy od rodzaju karty, w związku z czym zostanie pominięty w tym artykule. Więcej informacji na ten temat można znaleźć w [5].

Automatyzacja

Jak to przedstawiono powyżej, przeprowadzenie walidacji metody pomiarowej niezwykle jest czasochłonne i kosztowne. Czas przeprowadzenia walidacji zależy od czasu potrzebnego na opisanie metody, od ilości i rodzaju pomiarów (oznaczeń), jakie należy

wykonać, i wreszcie od biegłości osoby odpowiedzialnej za walidację w zagadnieniach statystycznych wykorzystywanych do obróbki zebranych danych. W przypadku kiedy zachodzi potrzeba walidacji niejednokrotnie dziesiątek metod pomiarowych, usprawnienie któregośkolwiek z tych etapów przyniesie znaczący zysk. Opis metody, czy też wykonywanie pomiarów zazwyczaj nie pozostawia wiele pola do manewru, w przeciwieństwie do statystycznej obróbki danych, którą można usystematyzować i zautomatyzować. Oczywiście wiedza statystyczna u zaangażowanych w walidację osób jest czynnikiem nieodzownym, niemniej jednak odpowiednie jej usystematyzowanie może przynieść znaczącą poprawę jakości przeprowadzanych analiz. Celem stworzonego przez StatSoft systemu było przede wszystkim zaproponowanie odpowiednich szablonów analiz, które pozwolą usystematyzować statystyki oraz pozwolą zautomatyzować wykonywanie analiz. Dzięki takiemu podejściu udało się pogodzić dwa przeciwstawne cele: wykonać bezbłędnie odpowiedni zestaw analiz w krótkim czasie.

System *STATISTICA Walidacja Metod Pomiarowych* (www.StatSoft.pl) daje możliwość automatycznego tworzenia raportów w oparciu o zdefiniowane wcześniej szablony dokumentów MS Word w zależności od struktury danych wejściowych. To, co musi zrobić użytkownik, to wprowadzenie danych w odpowiednim formacie do arkusza *STATISTICA*, wybór zmiennych do analizy i naciśnięcie przycisku. Program stworzy gotowy do wydruku raport, który można dalej dostosowywać do własnych potrzeb.

Warto zauważyć, że poniższy raport jest kompletny, tzn. że oprócz wyników analiz zawiera również (o ile jest to możliwe i wskazane) wzory, na podstawie których zostały przeprowadzone obliczenia, oraz w przypadku testów statystycznych automatycznie zapisywane są wnioski, co ułatwia interpretację całej analizy.

Mówiąc o rozwiązaniach informatycznych, trudno nie wspomnieć o walidacji samego systemu komputerowego zwłaszcza w odniesieniu do [4]. StatSoft może przeprowadzić walidację poprawności działania systemu komputerowego w trakcie jego wdrożenia.

Dodatkowe informacje dotyczące powyższego rozwiązania można znaleźć w [6].

Literatura

1. LAB – Laboratoria, Aparatura, Badania, Teresa Topolnicka, Michał Iwaniec, „Automatyczna walidacja metod badawczych i pomiarowych cz. 1”, rok 13, nr 6.
2. 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, grudzień 2002.
3. PN-ISO 11095, Kalibracja liniowa z zastosowaniem materiałów odniesienia, sierpień 2001.
4. Wyrażanie niepewności pomiaru, Przewodnik, Główny Urząd Miar, 1999.
5. Biuletyn Informacyjny Klubu POLLAB, 1/50/2008, „Wytczne do zarządzania komputerami i oprogramowaniem w laboratoriach w odniesieniu do normy ISO/IEC 17025: 2005”, EUROLAB TR 2/2006.
6. Czytelnia internetowa StatSoft Polska: www.statsoft.pl/czytelnia/czytelnia.html.
7. <http://www.statsoft.pl/industries/Laboratorium.htm>.