

OCENA SYSTEMU POMIAROWEGO DLA CECH NIEMIARZALNYCH Z ZASTOSOWANIEM WSPÓLCZYNNIKÓW KAPPA COHEN'A I AC_1 GWET'A

Magdalena DIERING

Streszczenie: W artykule opisano sposoby szacowania poziomu zgodności decyzji oceniających w ramach analizy przemysłowych systemów pomiarowych dla cech niemierzalnych. Omówiono współczynniki Kappa Cohen'a i AC_1 Gwet'a. Na przykładzie procesu produkcji cewników diagnostycznych oszacowano poziom zgodności operatorów i eksperta z zastosowaniem obu rozwiązań. Dalej w pracy dokonano porównania Kappa Cohen'a i AC_1 ze względu na kryteria ważne w kontekście możliwości stosowania tych współczynników w praktyce. Artykuł kończy rekomendacja autorki dotycząca sposobu szacowania poziomu zgodności decyzji oceniających w ramach analizy przemysłowych systemów pomiarowych dla cech niemierzalnych.

Słowa kluczowe: cecha niemierzalna, analiza systemów pomiarowych (MSA), poziom zgodności oceniających, współczynnik Kappa Cohen'a, współczynnik AC_1 Gwet'a

1. Wprowadzenie

Nadrzędnym celem prowadzenia w przedsiębiorstwach produkcyjnych analizy systemów pomiarowych (ang. Measurement System Analysis), w skrócie MSA, jest doskonalenie elementów tego systemu, przede wszystkim procesu pomiarowego i kontroli jakości produkcji. Metody i procedury do analizy systemów pomiarowych umożliwiają stwierdzenie czy badany system jest odpowiedni do stawianych mu zadań i celów, tj. czy stosowany system pomiarowy pozwala na pozyskanie z procesu wytwarzania wiarygodnych danych o stanie tego procesu. System pomiarowy spełnia więc ważną funkcję w procesie produkcyjnym, ponieważ każda decyzja operatora (oceniającego) pociąga za sobą konkretną korzyść lub stratę dla organizacji [1].

MSA dla cech niemierzalnych znajduje zastosowanie gdy wyrób jest oceniany (a nie mierzony) czy jest zgodny czy nie w odniesieniu do stawianych mu wymagań (najczęściej występują dwie kategorie w ocenie: 0 – wyrób wadliwy, 1 – wyrób dobry). Taka ocena jest charakterystyczna przede wszystkim w kontroli sprawdzianem lub przy ocenie sensorycznej, także w przypadku cech mierzalnych, których pomiar jest utrudniony lub ekonomicznie nieuzasadniony. W przypadku orzekaniu o zgodności, ważnym celem MSA jest poznanie umiejętności rozpoznawania przez oceniających wad i niezgodności wyrobu.

W badaniu systemów pomiarowych stosowane znajdują metody statystyczne i modele matematyczne. W każdym przypadku powinny one być odpowiednio dobierane, tj. z uwzględnieniem specyfiki procesu wytwarzania i stosowanych w nim technik i technologii pomiaru. Mimo, że literatura przedmiotu proponuje wiele rozwiązań (np. w [2,3]), to w praktyce przedsiębiorstw przemysłowych zastosowanie znajdują najczęściej te opisane w przewodniku AIAG [4], stowarzyszenia zrzeszającego ekspertów z obszaru nauki, przedstawicieli polityki, przemysłu i biznesu. AIAG tworzy dla największych

producentów samochodów na świecie i ich poddostawców standardy w zakresie dobrych praktyk produkcyjnych, w tym także dla analiz z obszaru MSA.

2. Ocena poziomu zgodności z zastosowaniem współczynników Kappa Cohen'a i AC₁

2.1. Metodyka i kryteria oceny

Do analizy systemów pomiarowych dla cech niemierzalnych AIAG proponuje metodę tabel krzyżowych (ang. cross-tab method), która pozwala na oszacowanie poziomu zgodności par oceniających.

W ocenie poziomu zgodności decyzji oceniających w praktyce przyjęły się kryteria, które opracowano w oparciu o wytyczne Fleiss'a [5] i rekomendowane są przez AIAG [4]:

- wartość współczynnika typu Kappa wynosi poniżej 0,4 – zgodność oceniających jest niska (wymagane są działania naprawcze),
- wartość współczynnika w zakresie od 0,4 do 0,75 – poziom zgodności oceniających jest dobry (zalecane jest podjęcie działań zapobiegawczych i doskonalących),
- gdy wartość współczynnika Kappa przekracza 0,75 – poziom zgodności oceniających badanego systemu pomiarowego jest bardzo dobry (działania nie są wymagane, ale zaleca się podjęcie działań doskonalących).

Więcej o tej metodzie znaleźć można w [4,6,7]. Do oszacowania zgodności decyzji oceniających w metodzie tabel krzyżowych, zgodnie z [4], stosuje się współczynnik Kappa Cohen'a. Autorka proponuje rozważyć zastosowanie w tym miejscu innego współczynnika – AC₁, opracowanego przez Gwet'a [2].

Z analizy wyników, w ramach oceny poziomu zgodności decyzji oceniających, osoba odpowiedzialna za badanie powinna także dążyć do uzyskania odpowiedzi na pytania: czy oceniający mają odpowiednie kwalifikacje?, czy otoczenie (środowisko) przeprowadzania oceny jest odpowiednie?, czy ryzyko błędów jest akceptowalne dla organizacji i/lub klienta?

2.2. Współczynnik Kappa Cohen'a

Współczynnik Kappa Cohen'a można zapisać (1) [8]:

$$\kappa_C = \frac{p_a - p_e}{1 - p_e}, \text{ gdzie } p_a = \sum_{k=1}^q p_{kk} \text{ i } p_e = \sum_{k=1}^q p_{k+} p_{+k}, \quad (1)$$

gdzie:

- κ_C – współczynnik zgodności decyzji Kappa Cohen'a,
- q – liczba kategorii (najczęściej dwie: 0 – wyrób wadliwy, 1 – wyrób dobry),
- k – kategoria,
- p_n – suma zaobserwowanych zgodnych decyzji typu 0-0 i 1-1 w odniesieniu do liczby możliwych par decyzji,
- p_e – szansa (prawdopodobieństwo) na zgodność przypadkową dla decyzji typu 0-0 i 1-1 w odniesieniu do liczby możliwych par decyzji typu 0-0 i 1-1,
- p_{kk} – udział części, dla których para oceniających wskazała kategorię k ,
- p_{k+} – udział części, dla których pierwszy oceniający wskazał kategorię k ,
- p_{+k} – udział części, dla których drugi oceniający wskazał kategorię k .

O problemach w wyznaczaniu i interpretacji tego współczynnika autorka pisała w [9]. Więcej o Kappa Cohen'a znaleźć można także w [3,8,10,11].

2.3. Współczynnik AC_1 Gwet'a

Współczynnik AC_1 Gwet'a jest rozwinięciem Kappy Fleiss'a [5], może być stosowany dla wielu oceniających i wielu kategorii ocen.

Współczynnik Gwet'a można zapisać (2) [2]:

$$AC_1 = \frac{p_a - p_e}{1 - p_e}, \quad (2)$$

gdzie:

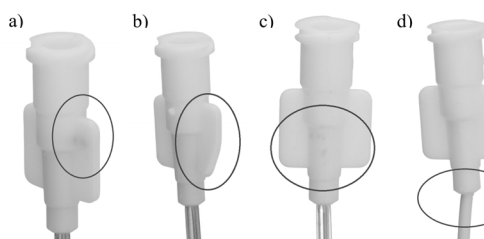
$$p_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}-1)}{r_i(r_i-1)}, \quad p_e = \frac{1}{q-1} \sum_{k=1}^q \pi_k(1 - \pi_k) \quad \text{i} \quad \pi_k = \frac{1}{n} \sum_{i=1}^n \frac{r_{ik}}{r_i},$$

- AC_1 – współczynnik zgodności decyzji Gwet'a,
 - r – liczba oceniających,
 - q – liczba kategorii,
 - k – kategoria,
 - n – liczba wszystkich obiektów w badaniu,
 - n' – liczba obiektów w badaniu, które oceniło co najmniej dwóch oceniających,
 - p_a – udział zaobserwowanych zgodnych decyzji w odniesieniu do liczby możliwych decyzji zespołu oceniających,
 - p_e – szansa (prawdopodobieństwo) na zgodność przypadkową,
 - r_{ik} – liczba oceniających, którzy zaklasyfikowali obiekt i do kategorii k ,
 - r_i – liczba oceniających, którzy ocenili obiekt i ,
 - π_k – prawdopodobieństwo wskazania kategorii k przez dowolnego oceniającego dla dowolnej części.
- Więcej o AC_1 Gwet'a znaleźć można także w [2,9,12].

2.4. MSA dla cech niemierzalnych z zastosowaniem Kappa Cohen'a i AC_1 – studium przypadku

Dla przykładu przeprowadzono badanie, mające na celu ocenę systemu pomiarowego dla cech niemierzalnych przy produkcji cewników diagnostycznych. Ocenie poddano przyłącze na cewniku zwane hubem, oznaczane w obliczeniach jako HUB. Na podstawie instrukcji stanowiskowych, katalogu błędów i wad oraz specyfikacji technicznych zdefiniowano listę możliwych wad, które mogą pojawić się w produkcji, a które wykrywać powinni operatorzy tych stanowisk, na których elementy te są wykonywane. Wybrane z wad przedstawiono na rysunku 1.

Do badania przygotowano próbkę liczącą 30 sztuk przyłączy (zgodnie z metodyką opisaną w [6]). Badanie wykonano z udziałem trzech oceniających (operatorów) i eksperta (inżyniera ds. jakości). Wykonano dwie serie ocen. Arkusz z wynikami przedstawiono w tabeli 1.



Rysunek 1. Przyłącze na cewniku diagnostycznym (HUB) – przykłady wad: a) wtopienia na hubie, b) zagięty „motylek”, c) przebarwienie na hubie, d) zakrzywiony hub.

Źródło: opracowanie własne

Tabela 1. Arkusz wyników z badania systemu pomiarowego dla cech niemierzalnych przy produkcji cewników diagnostycznych: ocena przyłączy na cewniku. Źródło: opracowanie własne

nr części	Operator 1		Operator 2		Operator 3		Ekspert	
	seria 1	seria 2	seria 1	seria 2	seria 1	seria 2	seria 1	seria 2
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1
9	0	0	1	1	0	0	0	0
10	0	0	1	1	0	0	0	0
11	1	1	0	0	0	1	0	0
12	1	1	1	1	1	1	1	1
13	1	1	0	0	0	0	0	1
14	1	1	0	0	0	0	0	1
15	1	1	1	1	1	1	1	1
16	1	1	0	1	1	1	1	1
17	1	1	1	1	1	1	1	0
18	0	0	1	1	1	1	0	0
19	0	0	1	1	0	1	0	0
20	1	1	1	1	1	0	1	0
21	0	0	0	0	0	0	0	0
22	0	1	1	1	0	1	0	0
23	0	0	1	0	0	0	0	0
24	0	0	0	0	0	0	0	0
25	0	0	0	0	0	1	0	0
26	1	1	1	1	1	1	1	1
27	1	1	0	0	1	1	1	1
28	0	0	1	1	0	1	0	1
29	1	1	1	1	1	1	0	0
30	1	1	1	1	1	1	0	0

Do analizy przygotowano tabele krzyżowe z zestawieniem wskazań:

- osobno dla każdego z oceniających (dla operatorów i dla eksperta) (w pracy zamieszczono przykład jednej z tabel krzyżowych – tabela 2a),
- dla każdej z par operatorów (Operator 1-Operator 2, Operator 1-Operator 3 i Operator 1-Operator 3) (patrz: tabela 2b) oraz
- dla każdej z par operator-ekspert (Operator 1-Ekspert, Operator 2-Ekspert, Operator 3-Ekspert) (patrz: tabela 2c).

Poziom zgodności decyzji oceniających oszacowano z zastosowaniem współczynników Kappa Cohen'a oraz AC_1 . W pierwszej kolejności oszacowano zgodność własną (wewnętrzną) eksperta i poszczególnych oceniających (zgodnie z rekomendacją autorki zawartą w [13]). W tym celu w obliczeniach, zamiast liczby oceniających, przyjęto liczbę serii powtórzeń (patrz wzory (1) i (2)). Uzyskane wyniki zestawiono w tabeli 3. Zgodność

wewnętrzna dostarcza informacji o tym czy osoby badane oceniają tak samo te same elementy w kolejnych seriach ocen (badanie powtarzalności), jednak nie mówi o tym czy te oceny są wiarygodne (czy są zgodne z wartością ustaloną, „prawdziwą” dla danej części). Wyjątkiem jest wynik eksperta, jeśli to on ustalał wartość „prawdziwą” dla badanych części.

Jak wynika z przeprowadzonych obliczeń, w oparciu o kryteria opisane w 2.1, w obu przypadkach, tj. zarówno przy użyciu Kappa Cohen’a jak i AC_1 , dla Operatora 1 i Operatora 2 uzyskano wysoką zgodność wewnętrzną (od 0,83 do 0,94), dla Operatora 3 i dla Eksperta – zgodność dobrą (od 0,56 do 0,67). Dla osoby odpowiedzialnej za badanie już ten etap daje wskazówki do doskonalenia – należałoby ustalić ocenę ekspercką w odniesieniu do badanych części oraz omówić z ekspertem aktualny katalog błędów i wad (szerzej różnice w ocenach eksperta i operatorów autorka omawia w [9]).

Tabela 3. Zgodność własna (wewnętrzna) eksperta i poszczególnych oceniających z zastosowaniem współczynników Kappa Cohen’a i AC_1 Gwet’a. Źródło: opracowanie własne

	Zgodność wewnątrz (własna)	
	Kappa Cohen'a	AC_1
Operator 1	0,92	0,94
Operator 2	0,83	0,89
Operator 3	0,56	0,64
Ekspert	0,67	0,67

Tabela 2. Zgodność własna (wewnętrzna) eksperta i poszczególnych oceniających z zastosowaniem współczynników Kappa Cohen’a i AC_1 Gwet’a. Źródło: opracowanie własne

a) Rozkład decyzji Operatora 3

		Seria 2		Suma
		0	1	
Seria 1	0	7	5	12
	1	1	17	18
Suma		8	22	30

b) Rozkład decyzji pary Operator 1-Operator 2

		Operator 2		Suma
		0	1	
Operator 1	0	7	12	19
	1	9	32	41
Suma		16	44	60

c) Rozkład decyzji pary Operator 3-Ekspert

		Ekspert		Suma
		0	1	
Operator 3	0	19	2	21
	1	10	29	39
Suma		29	31	60

Mając na uwadze cel podjętych badań, przystąpiono do dalszego etapu analizy. Oszacowano, w oparciu o wzory (1) i (2), poziom zgodności decyzji operatorów i eksperta (par operator-operator i par operator-ekspert). Uzyskane wyniki analizy zestawiono w tabeli 4.

Tabela 4. Poziom zgodności decyzji operatorów i eksperta z zastosowaniem współczynników Kappa Cohen'a i AC_1 Gwet'a. Źródło: opracowanie własne

Kappa Cohen'a	Operator 1	Operator 2	Operator 3	Ekspert
Operator 1	x	0,16	0,55	0,60
Operator 2	0,16	x	0,42	0,22
Operator 3	0,55	0,42	x	0,60
Oceniający łącznie	brak możliwości oszacowania*			

* Współczynnik Kappa Cohen'a ma zastosowanie dla dwóch oceniających, tj. do analizy parami (każdy z każdym) [3,12]

AC_1	Operator 1	Operator 2	Operator 3	Ekspert
Operator 1	x	0,40	0,64	0,62
Operator 2	0,40	x	0,56	0,28
Operator 3	0,64	0,56	x	0,61
Oceniający łącznie	0,56			x
	0,50			

Jak wynika z przeprowadzonej analizy, uzyskane wartości odzwierciedlające poziomy zgodności decyzji operatorów i eksperta różnią się od siebie. Oszacowane poziomy zgodności decyzji par operatorów i par operator-ekspert z zastosowaniem Kappa Cohen'a są niższe (czasem równe) niż oszacowania z zastosowaniem współczynnika Gwet'a.

W oparciu o kryteria opisane w 2.1, w oszacowaniu za pomocą Kappa Cohen'a dla pary Operator 1-Operator 2 uzyskano zgodność niską. Również na nieakceptowalnym poziomie jest zgodność Operator 2 z Ekspertem. W pozostałych przypadkach uzyskano zgodność dobrą. W ocenie z użyciem AC_1 para Operator 2-Ekspert również uzyskała zgodność niską, a w pozostałych przypadkach – dobrą.

Obliczono także łączną zgodność oceniających – łączną dla trzech operatorów oraz zgodność łączną dla operatorów i eksperta (tabela 4). W tym przypadku oszacowanie możliwe było tylko dla AC_1 , ponieważ Kappa Cohen'a zastosować można jedynie dla dwóch oceniających. Dla AC_1 uzyskano dobrą zgodność łączną.

Oba współczynniki różni aparat matematyczny – inaczej szacowana jest zgodność przypadkowa [3]. Różnica w rezultatach wynika także stąd, że Kappa Cohen'a jest wrażliwy na zmiany rozkładu wartości dla par decyzji typu 0-0 i 1-1. Im większa koncentracja ocen dla jednej z tych par decyzji, tym mniejsza wartość współczynnika [3,12]. Jeśli liczba decyzji o częściach dobrych od decyzji o częściach złych będzie się znacznie różniła, wartość Kappa Cohen'a zostanie „zaburzona”, a interpretacja wyniku i uzyskanie informacji o faktycznej zgodności oceniających będą trudne. Te różnice wystąpiły w omawianym przykładzie (patrz: tabela 2) i, mając na uwadze, że współczynnik AC_1 jest niewrażliwy na rozkład wskazań, stąd różnice w wynikach. Reasumując, lepszym oszacowaniem wydają się być wyniki uzyskane z zastosowaniem AC_1 .

2.5. Współczynniki typu Kappa – porównanie Kappa Cohen'a i AC_1

W oparciu o wnioski z przeprowadzonych badań i analizy wyników, dokonano porównania obu współczynników typu Kappa, tj. Kappa Cohen'a oraz AC_1 . W porównaniu uwzględniono kryteria, które, zdaniem autorki, są ważne w kontekście możliwości

stosowania tych współczynników w analizie przemysłowych systemów pomiarowych dla cech niemierzalnych (kryteria ważne ze względu na przygotowanie i organizację badania oraz możliwość dostosowania do istniejących w danej organizacji warunków produkcyjnych) (tabela 5):

- liczba obiektów (części) niezbędnych do przeprowadzenia badania (oceny systemu

Tabela 5. Współczynniki typu Kappa – porównanie współczynników Kappa Cohen’a i AC₁ Gwet’a. Źródło: opracowanie własne

I.p.	Kryterium porównania	Współczynnik Kappa Cohen’a (metoda tabel krzyżowych)	Współczynnik AC ₁ Gwet’a
1	Liczba obiektów (części) w badaniu	im więcej tym lepsze oszacowanie; najczęściej 30-50 sztuk	im więcej tym lepsze oszacowanie; najczęściej 30-50 sztuk
2	Udział obiektów (części) wadliwych w całkowitej liczbie elementów w badaniu	50%	dowolny (zaleca się około 50% lub więcej)
3	Skala pomiaru dla ocenianej cechy	nominalna (najczęściej dychotomiczna; 0 jako wyrób wadliwy i 1 jako wyrób dobry)	nominalna (przykładowo, poza „dobry” i „zły”, możliwa jest np. kategoria „do naprawy”); porządkowa
4	Liczba oceniających	w jednym porównaniu – dwóch oceniających, tj. analizy parami (każdy z każdym)	dowolna (porównania parami lub łącznie)
5	Wrażliwość na rozkład wartości par obserwacji typu 0-0 i 1-1	duża (nierównomierny rozkład zaburza /zaniża/ wynik)	brak
6	Zakres możliwych wyników analizy	wartości z przedziału od -1 do 1 (w praktyce, tj. w metodzie tabel krzyżowych – od 0 do 1)	wartość z przedziału od 0 do 1
7	Interpretacja wyników	interpretacja wyników wymaga znajomości założeń metody szacowania zgodności oceniających (patrz: kryterium 2 i 5)	interpretacja wyników intuicyjna – im bliżej wartości 1 tym lepsza zgodność oceniających
8	Poziom trudności w oszacowaniu wskaźnika bez użycia wspomaganie komputerowego	średni	wysoki (trudny)
9	Stosowanie współczynnika jest w zaleceniach / wymaganiach branżowych	tak (np. w podręczniku AIAG, MSA, wyd. 4) [4]	nie

- pomiarowego dla cech niemierzalnych) (więcej o liczbie części w [15]),
- udział części wadliwych w całkowitej liczbie elementów przygotowanych do badania,
- skala pomiaru dla ocenianej cechy (więcej o możliwych skalach pomiarowych można znaleźć w [13,16]),
- liczba oceniających,
- wrażliwość na rozkład wartości par obserwacji typu 0-0 i 1-1 (szczegółowo omawia to Gwet w [12]),
- zakres możliwych wyników analizy,
- interpretacja wyników,
- poziom trudności w oszacowaniu wskaźnika bez użycia wspomaganie komputerowego,
- oraz informacja o tym czy stosowanie danego współczynnika jest w zaleceniach / wymaganiach branżowych (przedsiębiorstwa chętniej stosują te rozwiązania, które wprost „dyktuje” im klient).

W tabeli 5 wyróżniono kryteria 2, 5 i 7 jako te, które najbardziej wskazują na różnice między porównywanymi współczynnikami typu Kappa. Są to jednocześnie kryteria, które, zdaniem autorki, stanowią o przewadze współczynnika AC_1 nad Kappa Cohen'a w odniesieniu do możliwości ich zastosowania w praktyce inżynierskiej.

Kryterium nr 2 jest ważne dla organizacji badania. Do oszacowania zgodności oceniających z zastosowaniem współczynnika Kappa Cohen'a do badania należy tak dobrać części, aby połowa z nich należała do kategorii „wyrób wadliwy”. Niespełnienie tego wymagania spowoduje nierównomierny rozkład wartości dla par zaobserwowanych zgodnych decyzji typu 1-1 i 0-0 (skutek tego omówiono przy kryterium 5) [3,14]. W przypadku współczynnika AC_1 udział części z poszczególnych kategorii nie ma znaczenia dla analizy (nie wpływa na wyniki analiz). Jednakże, z punktu widzenia jednego z celów badania, tj. dla poznania umiejętności rozpoznawania przez oceniających wad i niezgodności wyrobu zaleca się, aby udział części wadliwych był znaczny. Ponadto, należy zapewnić, aby w obu grupach znalazły się takie części, dla których decyzja jest trudna do podjęcia (części na granicy akceptowalności/odrzućcia). Uzyskanie wysokiej zgodności na tych elementach zwykle jest najtrudniejsze.

Porównanie według kryterium nr 5 daje szczególnie ważną informację dla osoby wykonującej analizę i ocenę systemu pomiarowego dla cech niemierzalnych – Kappa Cohen'a jest wrażliwy na zmiany rozkładu wartości dla par decyzji typu 1-1 i 0-0 (współczynnik Gwet'a jest od niego niezależny).

Kryterium 7 ma szczególne znaczenie dla zastosowań inżynierskich – umiejętność interpretacji wyników z analiz statystycznych to podstawa świadomości ich przydatności w doskonaleniu procesów wytwarzania i/lub procesów kontroli. Z obserwacji autorki wynika, że świadomość korzyści wynikających z MSA czy umiejętność przeprowadzenia badania systemu pomiarowego w przedsiębiorstwach produkcyjnych są na niskim poziomie. Stąd, im mniej wymagająca interpretacja wyników, tym większa szansa na zaangażowanie zespołu biorącego udział w badaniu i na podjęcie odpowiednich działań doskonalących. W przypadku współczynnika Kappa Cohen'a przy interpretacji wyników należy uwzględnić, między innymi, równomierność rozkładu wartości dla poszczególnych par obserwacji (patrz: kryterium 5) oraz sposób przygotowania części do badania (patrz: kryterium 2). Interpretacja wartości uzyskanych w analizie i ocenie poziomu zgodności oceniających z zastosowaniem współczynnika Gwet'a jest intuicyjna – im bliżej wartości 1 tym lepsza zgodność oceniających.

3. Wnioski i kierunek dalszych prac

Na podstawie przeglądu literatury przedmiotu oraz przeprowadzonej analizy wyników pozyskanych z badań własnych autorka zaleca, by w ocenie przemysłowych systemów pomiarowych dla cech niemierzalnych stosować współczynnik AC_1 Gwet'a. Współczynnik AC_1 jest wolny od ograniczeń Kappa Cohen'a, co zwiększa zakres możliwości dla nadzorującego badanie na etapie przygotowania części i ułatwia analizę wyników.

Nauki matematyczne dostarczają wiele różnych modeli i podejść do oceny poziomu zgodności oceniających, jednakże niewiele z nich zaimplementowano do zastosowań inżynierskich. Stąd, w oparciu o dotychczasowe wyniki badań i obserwacji analizowanych procesów wytwarzania i procesów pomiarowych, autorka obecnie pracuje nad nową metodyką analizy i oceny poziomu zgodności oceniających (kontrolerów jakości i ekspertów) w przemysłowych systemach pomiarowych dla oceny niemierzalnej z uwzględnieniem różnych warunków produkcyjnych.

Prezentowane wyniki badań, zrealizowane w ramach tematu nr 02/23/DSPB/7640, zostały sfinansowane z dotacji na naukę przyznanej przez Ministerstwo Nauki i Szkolnictwa Wyższego.

Bibliografia

1. Bożek M., Rogalewicz M.: Nieskuteczność kontroli końcowej przyczyną niskiej efektywności procesu wytwarzania, *Inżynieria Maszyn*, 2013, rocznik 18, zeszyt 1; s. 84-96.
2. Gwet K. L.: *Handbook of Inter-Rater Reliability. The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters*, 3rd ed., 2012.
3. Jarosz-Nowak J.: Modele oceny stopnia zgody pomiędzy dwoma ekspertami z wykorzystaniem współczynników Kappa, *Matematyka Stosowana*, nr 8, Wrocław 2007, s. 126-154.
4. *Measurement Systems Analysis*, 4th ed., Reference manual, AIAG-Work Group, Daimler Chrysler Corporation, Ford Motor Company, General Motors Corporation, 2010.
5. Fleiss J., Cohen J.: The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability, *Educational and Psychological Measurement* 1973, vol. 33, s. 613-619.
6. Diering M., Kujawińska A.: *MSA – Analiza Systemów Pomiarowych. Przewodnik po procedurach*, Poznań 2012.
7. Bower Keith M.: *Measurement system analysis with attribute data*, Trainer's Corner, Minitab Inc., 2002, s. 10-11.
8. Cohen J.: A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* 1960, vol. 20, s. 37-46, doi: 10.1177/001316446002000104.
9. Diering M., Dyczkowski K., Hamrol A.: Szacowanie poziomu zgodności ocen w kontroli wizualnej – problemy w wyznaczaniu współczynników typu Kappa, w: *Innowacje w zarządzaniu i inżynierii produkcji*, Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, Opole 2015, tom 2, s. 257-368.
10. Carletta J.: Assessing agreement on classification tasks: the Kappa statistic, *Computational Linguistics*, vol. 22/2, s. 1-6.

11. Scott W.A.: Reliability of Content Analysis: The Case of Nominal Scale Coding, *The Public Opinion Quarterly*, 19(3), s. 321-325, 1955.
12. Wongpakaran N., Wongpakaran T., Wedding D., Gwet K.L.: A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples, *BMC Medical Research Methodology*, 2013, dostęp online: doi: 10.1186/1471-2288-13-61.
13. Diering M., Dyczkowski K., Hamrol A.: New method for assessment of raters agreement based on fuzzy similarity, 10th International Conference on Soft Computing Models in Industrial and Environmental Applications, *Advances in Intelligent Systems and Computing*, © Springer International Publishing Switzerland, Volume 368, 2015, s. 415-425, doi: 10.1007/978-3-319-19719-7_36.
14. Viera Anthony J., Garrett Joanne M.: Understanding Interobserver Agreement: The Kappa Statistic, *Family Medicine*, Vol. 37, No. 5, s. 360-363, 0742-3225.
15. Feliks J., Lichota A.: Wspomaganie analizy systemów pomiarowych (MSA), *Archives of Foundry Engineering*, Volume 10, Special Issue 3/2012, s. 169-174.
16. Stachowiak, A., Żywica, P., Dyczkowski, K., Wójtowicz, A.: An interval-valued fuzzy classifier based on an uncertainty-aware similarity measure, In: *Intelligent Systems' 2014*, Springer, 2015, s.741–751.

Dr inż. Magdalena DIERING
Katedra Zarządzania i Inżynierii Produkcji
Wydział Budowy Maszyn i Zarządzania
Politechnika Poznańska
Pl. M. Skłodowskiej-Curie 5, 60-965 Poznań
tel./fax.: (061) 665 2738 / (061) 665 2774
e-mail: magdalena.diering@put.poznan.pl