



BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
VEGYÉSZMÉRNÖKI ÉS BIOMÉRNÖKI KAR
OLÁH GYÖRGY DOKTORI ISKOLA

Minősítéses mérőeszközök képességvizsgálata

Tézisfüzet

Szerző:

Vágó Emese
okleveles vegyészmérnök

Témavezető:

Dr. Kemény Sándor
egyetemi tanár

Kémiai és Környezeti Folyamatmérnöki Tanszék

2011

1. Bevezetés

A mérőeszközök képességvizsgálatának célja a mérőrendszer bizonytalanságának jellemzése, és annak eldöntése, hogy az eszköz alkalmas-e az adott folyamat-jellemző mérésére. A méréses mérőeszközök képességvizsgálata jól kidolgozott és széles körben alkalmazott módszer. Statisztikai alapja az egyes bizonytalanság-források varianciáinak becslése, és e varianciák összehasonlítása. Minősítéses mérőeszközök esetében a mérés eredménye nem folytonos, hanem diszkrét (az általunk vizsgált esetekben bináris) változó, így a varianciaanalízis nem alkalmazható, más megközelítésre van szükség.

A legelterjedtebben használt módszerek a minősítéses mérőeszközök képességvizsgálatára az autóiipari beszállító cégek számára kidolgozott QS-9000 kvázi-szabvány kézikönyvében¹ található. Az itt bemutatott módszer feltételezi a referencia-értékek ismeretét. Ez azt jelenti, hogy annak ellenére, hogy a minősítés során két kategóriába soroljuk a darabokat (megfelelt/nem felelt meg), létezik egy folytonos skálán mérhető tulajdonság a döntés háttérében. Ezt a tulajdonságot a mérés idő és/vagy költségigénye miatt a gyakorlatban ismeretlennek tekintik, helyette csak megfelelt/nem felelt meg típusú döntést hoznak. Disszertációmban az ilyen típusú minősítéses mérőrendszerek elemzésével foglalkozom.

2. Irodalmi háttér, célkitűzések

A bináris kategorikus döntés mögött álló folytonos változó lehetséges értékeit a szakirodalom három különböző részre osztja: elfogadási-, elutasítási, és ún. szürke tartományra. Ha a folytonos változó értéke az elfogadási (vagy elutasítási) tartományban van, akkor a mérés eredménye mindig elfogadás (illetve elutasítás) lesz. Ha a vizsgált jellemző tényleges értéke a szürke tartományba esik, akkor a mérés eredménye véletlen változó, itt az elfogadás valószínűsége a folytonos változó nagyságától függ. A modellt egy S-alakú görbe szemlélteti, ami az angol nyelvű szakirodalomban „Gauge Performance Curve” (GPC) néven ismer. Magyarul mérőeszköz-képesség-görbének, vagy mérőeszköz-jelleggörbének szokás nevezni. Az ideális mérőeszköz képesség-görbéje egy ugrásfüggvény lenne, ahol az ugrás a specifikációs határnál van. Ez azt jelentené, hogy a specifikációs határ alatti darabokat mindig elutasítjuk, a specifikációs határ felettieket mindig elfogadjuk (alsó specifikációs határ esetén). A valóságban nem ugrás-, hanem S-alakú a görbe, az S-alak inflexiós pontja feleltethető meg az ugrás helyének. Ha ez eltolódott a specifikációs

¹ Automotive Industry Action Group. Measurement System Analysis; Reference Manual, 3rd ed. Detroit, MI: Automotive Industry Action Group, 2002; 125-140.

határtól, a mérőeszköz torzít. Az inflexiós pont annál a referencia-értéknél van, ahol az elfogadás valószínűsége 50%.

Az AIAG előiratban javasolt képességvizsgálati eljárás két teljesen elkülönülő részből áll. Az egyik a keresztosztályozásos-, a másik az analitikus módszer. A keresztosztályozásos vizsgálat elvégzéséhez 50 elemű véletlen mintát vesznek a folyamatból. Három operátor háromszor minősíti mindegyik munkadarabot. A minősítés háttérében álló folytonos változót mérik, ez alapján minden darabról megállapítják, hogy az ténylegesen megfelelő-e. Az előirat a Cohen-féle kappát javasolja az operátorok páronkénti és referencia-értékkel történő összehasonlítására. Az operátorok referenciával való egyezését a kapp mellett különböző relatív gyakoriságokkal jellemzi az előirat: a hatékonyság alatt a helyes döntések arányát érti, és számolja külön-külön is a téves elfogadások és elutasítások arányát. Ismételhetőségnek azon darabok arányát nevezi, melyekre az ismételt minősítések megegyező eredményt hoztak.

Az AIAG előirat javasolja még a $P(\text{rossz}|\text{elutasít})$ és $P(\text{jó}|\text{elfogad})$ feltételes valószínűség használatát is. Ezek a véletlen mintából relatív gyakoriságokkal becsülhetők. $P(\text{rossz}|\text{elutasít})$ a ténylegesen rossz darabok arányát adja meg az elutasított darabok között, $P(\text{jó}|\text{elfogad})$ pedig az elfogadott darabok között lévő ténylegesen jó darabok arányát jelöli. A minősítési mérőeszközök sajátossága, hogy ezek a valószínűségek függenek attól, milyen darabokat minősítünk (pontosabban a vizsgált darabok referencia-értékének eloszlásától). Abban az esetben, ha a minősített darabok eloszlása változik (ahhoz képest, ami a mintavételkor volt), az előirat a Bayes-tételt javasolja $P(\text{rossz}|\text{elutasít})$ számolására.

Az előiratban javasolt analitikus módszer alkalmazásához nyolc elemű mintára van szükség. A mintaelemeket úgy válogatják ki, hogy azok minél jobban lefedjék a mérőeszköz szürke zónáját. Minden darabnak mérik a referencia-értékét, és mindegyiket többszöri ismétléssel minősíti egyetlen operátor. Az ismételt minősítésekből relatív gyakoriságokkal becsülik az elfogadás valószínűségét az egyes darabokra. Az összetartozó referencia-érték-elfogadási valószínűség adatpárokat egy Gauss-háló jellegű ábrán ábrázolják, majd a pontokra egyenest illesztnek, ezzel a GPC egy linearizált (grafikus) becslését kapják. Az ábráról leolvasható a mérőrendszer torzítása és ismételhetősége.

A mérési mérőrendszerek képességvizsgálatánál az ismételhetőség, reprodukálhatóság és torzítás jelentése egyértelműen meghatározott. A jó ismételhetőség azt jelenti, hogy ha egy adott operátor egy adott darabot többször egymás után megmér, akkor az eredmények közel lesznek egymáshoz. A reprodukálhatóság az operátorok közötti egyezés mértéke. A tökéletes reprodukálhatóság azt jelenti, hogy ugyanazon darabon mért értékek várható értékei megegyeznek, függetlenül attól melyik operátor végzi a mérést. A torzítás a mérési esetben a mérés várható értéke és a tényleges érték közti

eltérést jelenti. Minősítéses esetben e fogalmak definiálásával, és még inkább jellemzésükkel kapcsolatban bőven akadnak nehézségek.

Az ismételhetőség függ a kiválasztott darabtól. Értelmezésére az analitikus és a keresztosztályozásos módszer két teljesen különböző megközelítést javasol. Két operátor közti rossz reprodukálhatóság azt jelenti, hogy az egyik operátor elfogadóbb/elutasítóbb döntéseket hoz, mint a másik. Az, hogy éppen elutasítóbb vagy elfogadóbb az operátor, darabfüggő lehet. Véletlen minta többszöri minősítésénél a reprodukálhatóság az eltérően minősített darabok arányával hozható kapcsolatba. Ez az arány azonban nemcsak a rossz reprodukálhatóság, hanem a rossz ismételhetőség miatt is lehet magas. Tehát ez a mutató a mérőrendszer e két tulajdonságát együttesen jellemzi. Ugyanez igaz a Cohen-kappára is. A torzítást az AIAG előirat analitikus módszere a GPC inflexiós pontja és a specifikációs határ közti távolságnak tekinti. Más források szerint a torzítás a véletlen mintából számolható, és a téves elfogadások és elutasítások arányával jellemezhető.

A fentiek alapján megállapíthatjuk, hogy a minősítéses mérőeszközöknél az ismételhetőség, reprodukálhatóság és torzítás jellemzésére és definiálására az irodalom különböző megközelítéseket használ. Nincs egységes módszer a mérőrendszer tulajdonságainak jellemzésére. Disszertációmban elvégeztem az irodalomban bemutatott képességvizsgálati módszerek kritikai elemzését (különös tekintettel az AIAG előíratra), eredményként rámutatva ezek számos súlyos elvi hibáira. Céлом egy olyan képességvizsgálati módszer kidolgozása volt, amely megoldást nyújt a vázolt problémákra és használatával átfogó, egységes képet kap a felhasználó a mérőeszköz tulajdonságairól, és alkalmasságáról az adott minősítési feladat elvégzésére.

3. Számítási módszerek

A minősítéses mérőeszközök képességvizsgálatára az irodalom főként leszámhlálási mutatókon alapuló módszereket alkalmazását javasolja. Disszertációmban a logit modell használatát javaslom a feladat megoldására. A logit modell az általánosított lineáris modellek (generalized linear model) családjába tartozik, illesztése maximum likelihood módszerrel történik. Az elemzési módszerek hatékonyságát részben analitikusan, részben szimulációval vizsgáltam.

A doktori munkám során végzett statisztikai számítások - ahol ezt külön nem jeleztem - a kutatócsoportban általánosan használt STATISTICA szoftver 8.0 verziójával² készültek. A szimulációs vizsgálatokhoz használtam a szoftverbe épített Visual Basic programnyelvet.

A képességvizsgálat véletlen faktoros megközelítésekor általánosított vegyes modell (generalized linear mixed model) illesztésére volt szükség. Ehhez a nyílt

² StatSoft, Inc. (2007). STATISTICA (data analysis software system), version 8.0. www.statsoft.com

forráskódú R szoftver lme4 modulját³ használtam. A disszertációmban javasolt numerikus integrálást alkalmazó számításokat a MATHEMATICA szoftverben⁴ írt programmal végeztem.

4. Eredmények

A disszertációmban bemutatott kutatási munka két fő részre osztható. Az első témakörben a jelenlegi képességvizsgálati módszerek hibáira mutatok rá, a másodikban egy új módszert javaslok a problémák megoldására.

4.1. Az irodalomban alkalmazott minősítési mérőeszköz képességvizsgálati módszerek kritikája

Az AIAG előiratban javasolt keresztosztályozásos módszer többek között a Cohen-féle kapa mutatót használja a minősítést végző személyek egyezésének jellemzésére. Kapa számolásához az egyazon darabon végzett ismételt minősítéseket operátoronként párba rendezi. A párosítás alapja a mérés sorrendje. Mivel az ismételt minősítések egymástól függetlenek egy önkényes párosítás elméleti szempontból megalapozatlan, kapa helytelen becslését eredményezi.

$P(\text{rossz}|\text{elutasít})$ és $P(\text{jó}|\text{elfogad})$ feltételes valószínűségek Bayes-tétellel történő becslésekor nem teljesülnek a tétel alkalmazási feltételei, mivel az egyenlet jobboldalán szereplő $P(\text{elutasít}|\text{rossz})$ és $P(\text{elfogad}|\text{jó})$ feltételes valószínűségek nem állandóak. Ez abból adódik, hogy az elfogadás/elutasítás valószínűsége függ a darabok referencia-értékétől, így ha ennek eloszlása változik, $P(\text{elutasít}|\text{rossz})$ és $P(\text{elfogad}|\text{jó})$ is változni fog.

Az AIAG előirat keresztosztályozásos módszere csak azt vizsgálja, hogy az operátor elfogadta vagy elutasította az adott darabot. A feladatban azonban az alsó és felső tűréshatár adott, így különbséget lehet tenni „túl nagy” és „túl kicsi” darabok között. Tehát az „elutasított” kategória az előbbi két lehetséges kimenet egyesítéséből származik. Ez az egyszerűsítés információvesztéshez vezet. A példában valójában nem kettő, hanem három döntési kategória létezik, ilyen esetekben más megközelítés szükséges a mérőrendszer vizsgálatához.

A keresztosztályozásos módszer keretén belül többféle relatív gyakorisággal jellemzik a mérőrendszer hatékonyságát. A használt mutatók becsléséhez az 50 elemű minta háromszori ismétléssel kapott minősítéseinek eredményeit használják. A becslés hatásosabb lenne úgy, ha nagyobb (a konkrét példában 150) elemszámú mintát használnának, ismétlés nélküli minősítésekkel.

³ D. Bates, M. Maechler lme4 Version 0.999375-35

⁴ Mathematica, version 6.0.0 (2007) Wolfram Research, Inc.

Az előirat két teljesen különálló képességvizsgálati módszert javasol, melyek között semmilyen kapcsolat nincs. Van olyan fogalom (ismételhetőség), melyet a két módszer két különböző módon jellemez.

A szakirodalommal kapcsolatos kritikámat és javaslataimat az *1.* és a *2. tézisben* foglaltam össze.

4.2. Modell-alapú megközelítés minősítéses mérőeszközök képességvizsgálatára

Új eljárást javasoltam a minősítéses mérőeszközök képességvizsgálatára. A módszer lényege, hogy a mérőeszköz-képesség-görbét nem grafikusán becsülöm, ahogy azt az AIAG előiratban teszik, hanem matematikai modellel írom le. A javasolt modell logit transzformációval linearizálja a referencia-érték és az elfogadási valószínűség közötti kapcsolatot:

$$\text{logit}(p_i) = \ln \frac{p_i}{1-p_i} = \alpha + \beta_i^O + \beta^P x + \gamma_i^{O*P} x$$

ahol p_i annak valószínűsége, hogy az i -edik operátor elfogadja az x referencia-értékkel jellemezhető darabot

x referencia-érték

α modellkonstans

β_i^O i -edik operátor „hatása”

β^P darab „hatása”

γ_i^{O*P} operátor-darab „kölcsonhatás” az i -edik operátornál

A fenti modell segítségével operátoronként számolható:

- a mérőeszköz torzítása, azaz a görbe inflexiós pontja és a specifikációs határ közötti távolság;

- a szürke zóna szélessége: az a referencia-érték tartomány, amelyben az elfogadás valószínűsége tetszőleges mértékben különbözik nullától és egytől (a görbe S-alakú része). Például az a referencia-érték intervallum, amiben az elfogadási valószínűség 5 és 95% között van.

A javasolt logit modell paramétereinek becslése a gyakorlatban úgy valósítható meg, hogy az AIAG-előiratban is követett módon ismert referencia-értékű darabokat többszöri ismétléssel minősítenek az operátorok. Az eredmények alapján maximum likelihood módszerrel becsülhetők a logit modell paramétere. Vizsgáltam, hogy a becslés javítható-e a ridge-módszer alkalmazásával. A ridge becslést általánosan azzal a céllal alkalmazzák, hogy csökkentsék a független változók korreláltsága miatt megnövekedett varianciát. Igaz, hogy a mérőeszköz-képesség-görbe paramétereinek becslésekor a független változók nem korreláltak, de a logit modell alakja miatt feltételeztem, hogy a ridge módszer javíthatja a becslést. A feltételezést szimulációval vizsgáltam. Az eredmények azt mutatták, hogy a ridge alkalmazásával a

linearizált modell becsült paramétereinek hibája jelentősen csökken. Ez a javulás azonban sokkal kisebb mértékű, ha az inflexiós pont helyének és a szürke zóna szélességének becslését vizsgálom. Tehát a képességvizsgálat gyakorlati szempontból érdekes eredményeit a ridge nem javítja számottevően, ezért alkalmazása nem indokolt.

A mérőeszköz és az operátor tökéletesen jellemezhető az inflexiós pont helyével és a szürke zóna szélességével, arról azonban e tulajdonságok nem adnak információt, hogy a mérőrendszer alkalmas-e az adott minősítési feladat elvégzésére. E kérdés megválaszolásakor az érdekes a felhasználó számára, hogy mekkora a rossz darabok aránya az elfogadottak között, és mekkora a jó darabok aránya az elutasítottak között, mivel ezek a valószínűségek hozhatók közvetlenül kapcsolatba a hibás döntés okozta veszteséggel. Disszertációmban ezért $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ feltételes valószínűségeket javaslom a mérőrendszer képességének jellemzésére. A mérőeszköz-képesség-görbe matematikai modelljének segítségével e valószínűségek számolhatók, ha ismert a referencia-értékek eloszlása. Az i -edik operátorra:

$$P(\text{jó}|\text{elutasít})_i = \frac{\int_{x=SH}^{x=+\infty} f(x)[1-p_i(x)]dx}{\int_{x=0}^{x=+\infty} f(x)[1-p_i(x)]dx}$$

$$P(\text{rossz}|\text{elfogad})_i = \frac{\int_{x=0}^{x=SH} f(x)[p_i(x)]dx}{\int_{x=0}^{x=+\infty} f(x)[p_i(x)]dx}$$

Ahol $f(x)$ a referencia-értékek eloszlását jelöli.

A javasolt számolási módszer nagy előnye, hogy elkülönül benne a mérőeszköz jellemzése és a referencia-értékek eloszlása. Ezért nem lép fel az a probléma, mint az előiratban használt Bayes-tételes megközelítés esetén. Így a modell-alapú számolás helyes akkor is, ha a folyamat megváltozik ahhoz képest, mint amilyen a mérőeszköz-képesség-görbe becslésekor volt.

$P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ feltételes valószínűségek becsülhetők modell illesztése nélkül is, relatív gyakoriságokkal, az AIAG előirat keresztosztályozásos módszerét követve. Disszertációmban részben analitikusan, részben szimulációval vizsgáltam a két különböző becslés hatékonyságát. A közepes négyzetes hiba drasztikusan kisebb a modell-alapú becslés esetén a vizsgált minta-elemszám-tartomány egészében. Ezen felül megállapítható, hogy a GPC-alapú becslés hibája egy gyakorlatban is megvalósítható számú minősítés fölött már alig függ a minősítések számától. A

vizsgált mintaelemszám-tartomány alsó részében az AIAG-típusú számolás nagy valószínűséggel ad nulla becsült értéket. Ilyenkor az AIAG-típusú becslés különösen gyengén teljesít a modell-alapúhoz képest.

A méréses mérőeszközök képességvizsgálatánál a mérési eredmények ingadozását két okra vezetik vissza: a darabok különbözőségére és a mérőrendszer bizonytalanságára. Ez utóbbi ingadozásforrás (a mérőrendszer bizonytalansága) további két fő részből áll: a mérőeszköz operátorok miatti (reprodukálhatóság) és operátoroktól független (ismételhetőség) bizonytalanságából. Minden ingadozásforrást egy-egy varianciával jellemeznek. Mivel a véletlen faktoros modell paraméterei a különböző hatások varianciái, felmerülhet, hogy alkalmazásával a minősítéses mérőeszközöket is a mérésesekhez hasonló módon tudjuk jellemezni.

Dolgozatomban szimulációval vizsgáltam, hogy véletlen faktoros modellt alkalmazó megközelítés esetén hogyan becsülhetők a $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ feltételes valószínűségek.

Az első esetben (véletlen tengelymetszet) csak az inflexiós pont helye különbözött operátoronként, a szürke zóna szélessége állandó volt. Becsültem $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ operátorok szerinti várható értékét és varianciáját rögzített és vegyes modell használatával egyaránt. A vegyes modell illesztésével a linearizált modell paraméterei kisebb hibával becsülhetők, de ez a különbség $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ várható értékének és varianciájának becslésekor nem minden esetben jelentkezik. Rögzített modell illesztése esetén $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ operátorok szerinti várható értékét és varianciáját kétféleképpen becsültem: egy egyszerűbb módon, átlag és szórás számolásával, és egy bonyolultabb, integrált alkalmazó megközelítés szerint. Az integrált alkalmazó becslés hibája kisebb, de a különbség nem jelentős.

Ezután azt az esetet vizsgáltam, amikor az inflexiós pont helye mellett a szürke zóna szélessége is különbözik operátoronként (véletlen tengelymetszet és meredekség). Itt csak a rögzített modellt illesztettem, és azt találtam, hogy többségében itt is a bonyolultabb, integrált alkalmazó számolási mód hibája kisebb, de az elért javulás nem jelentős.

Összességében a szimulációs eredményekből az állapítható meg, hogy a vizsgált becslési módszerek mindegyike nagyságrendileg jó tájékoztatást nyújt a felhasználónak $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ várható értékéről és varianciájáról. A rögzített modell illesztése azonban könnyebben kivitelezhető az alkalmazott modell szélesebb körű szoftveres implementációja miatt (a vegyes modellhez képest). Ezért az operátoronként különböző tengelymetszetet feltételező esetekben a rögzített modell illesztését javaslom. A becsült modell-

paraméterekből $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ várható értékének és varianciájának becslésére úgyszintén az egyszerűbb eljárást, az átlag és szórás számolását javaslom, mivel a bonyolultabb, integrált is alkalmazó módszerrel becsült értékek hibája nem annyival kisebb, hogy az indokolja a jóval bonyolultabb, gyakorlatban nehezebben kivitelezhető számolás alkalmazását. Az operátoronként különböző tengelymetszet és meredekség esetén csak a rögzített modellt illesztettem. A becsült modell-paraméterekből $P(\text{jó}|\text{elutasít})$ és $P(\text{rossz}|\text{elfogad})$ várható értékének és varianciájának becslésére úgyszintén az egyszerűbb eljárást, az átlag és szórás számolását javaslom, ez előzővel megegyező okokból.

A modell-alapú képességvizsgálat eredményeit a **3. 4. és 5. tézisben** foglaltam össze.

5. Tézisek

1. tézis

Feltártam a gyakorlatban széles körben alkalmazott AIAG előirat minősítéses mérőeszköz képesség vizsgálati módszerének alábbi elvi hibáit és hiányosságait [2].

(a) Az előirat szerint 50 mintaelemet háromszori ismétléssel minősítenek a képességvizsgálat során, majd az eredményeket úgy kezelik, mintha azok egy 150 elemű minta egyszeri minősítéséből származnának. E hibás megközelítés a kappa mutató helytelen becslését eredményezi.

(b) Ha a minősítendő darabok eloszlása megváltozik (a folyamat javul), a jó és rossz darabok elfogadási valószínűsége is változik. Ezért a mérőrendszert jellemző pozitív és negatív prediktív érték Bayes-tétellel történő becslésekor nem alkalmazhatók az eredeti eloszlást jellemző feltételes valószínűségek. Tehát a tételt helytelenül használják.

(c) Az előirat két teljesen különálló képességvizsgálati módszert javasol: a keresztosztályozásos és analitikus módszert. A kettőhöz külön minták szükségesek, a matematikai kezelés sem közös, pedig ugyanarról a jelenségről adnak számot. Ennek például az a következménye, hogy az ismételtetésre két különböző mutatót is javasolnak.

2. tézis

Kimutattam, hogy binomiális és multinomiális eloszlásnál a túlszóródás jelensége csak ismételt megfigyelések esetén jelentkezik [4]. Ez a minősítéses mérőeszközök képességvizsgálatára vonatkozóan azzal a következménnyel jár, hogy leszámplálásos mutatók használatakor a becslés hatásosabb lesz, ha nagyobb mintákat egyszer minősítenek, mintha kisebb mintákat ismételtelen minősítenek, feltéve hogy a minősítések száma megegyezik a két esetben.

3. tézis

Az általánosan elfogadott AIAG módszer helyett új eljárást javasoltam a minősítéses mérőeszközök képességvizsgálatára [3]. A módszer alapvető újdonsága, hogy matematikai modellt használ a mérőeszköz-képesség-görbe leírására. A modell függő változója az elfogadás valószínűségének logit transzformáltja, független változói a vizsgált darab referencia-értéke (mérete) és az operátor.

(a) Egyértelműen interpretálható jellemzőket javasoltam a mérőeszköz-képesség-görbe karakterizálására, nevezetesen a szürke zóna szélességét és a torzítást. Megmutattam, hogy ezek a javasolt módszerrel hogyan számíthatók.

(b) A GPC-modell-alapú becsléssel hatásosabban (ugyanannyi mérésből kisebb bizonytalansággal) becsülhetők a hibás döntést jellemző feltételes valószínűségek, mint az AIAG-típusú (leszámlálási jellegű) becslést követve.

(c) Az új módszer elkülönítve (külön mintákból) jellemzi és becsli a referencia-értékek eloszlását és a mérőeszköz-képesség-görbét. Ebből következően, ha a folyamat eloszlása változik, nem szükséges a mérőeszköz-képesség-görbe újbóli becslése egy új mintából, csak a referencia-értékek új eloszlását kell felvenni, és ezzel lényeges megtakarítást érhetünk el.

4. tézis

A mérőeszköz-képesség-görbe modelljének becslésére a ridge módszert [1] is alkalmaztam. A javasolt becslési módot szimulációval vizsgálva azt találtam, hogy a ridge alkalmazásával a linearizált modell becsült paramétereinek hibája jelentősen csökken. Ez a javulás azonban sokkal kisebb mértékű, ha az inflexiós pont helyének és a szürke zóna szélességének becslését vizsgálom. A képességvizsgálat gyakorlati szempontból érdekes eredményeit a ridge módszer tehát nem javítja számottevően, alkalmazása ezért nem indokolt.

5. tézis

A modell-alapú képességvizsgálati módszerre a véletlen faktoros megközelítést is alkalmaztam [5] két változatban: az egyikben csak a tengelymetszet (vagyis a GPC görbe helyzete), a másikban a tengelymetszet és a meredekség (vagyis a GPC görbe helyzete és alakja) is véletlen változó volt. Különböző megközelítéseket (különböző egyszerűsítő feltételeket) hasonlítottam össze a hibás döntés valószínűsége operátorok szerinti varianciájának és várható értékének becslésére. Szimulációval vizsgáltam, hogy ezek az egyszerűsítések milyen hatással vannak a becslés hibájára. Azt találtam, hogy mindegyik megközelítéssel nagyságrendileg jó becslést kapunk, és nincs jelentős eltérés a becslések hibájában. A gyakorlatban tehát elfogadható a legegyszerűbb számolási mód alkalmazása.

6. Alkalmazási lehetőségek

A minősítéses mérőrendszereket az ipar különböző ágaiban sokféle célra sokan használják. Gyakran a minősítés során hozott kategorikus döntés háttérében egy mérhető folytonos változó áll. Az ilyen típusú mérőeszközök vizsgálata tehát gyakorlati szempontból érdekes, széles felhasználói kört érintő feladat.

Disszertációmban rámutattam, hogy a minősítéses mérőeszközök képességvizsgálatára jelenleg legszélesebb körben használt AIAG előiratban közölt módszerek statisztikailag több szempontból is hibásak. A szakterület egyéb irodalmait is figyelembe véve megállapítottam, hogy az alkalmazott fogalmak jelentése esetenként különböző. A rendelkezésre álló módszerek a képességvizsgálat során felmerülő részkérdésekre adnak csak választ, nincs olyan megközelítés, mellyel egyszerre lehetne jellemezni a mérőrendszer tulajdonságait, az operátorok közti különbséget, és választ kapnánk arra a kérdésre is, hogy a mérőeszköz alkalmas-e az adott folyamatból származó darabok minősítésére.

Az általam javasolt modell-alapú megközelítés nem szenved az AIAG előirat elvi hibáitól, és egyértelműen definiált fogalmakat használ a mérőrendszer jellemzésére. A javasolt módszer segítségével a képességvizsgálat során felmerülő összes részkérdés megválaszolható. Az új megközelítés további előnye, hogy az AIAG előiratban javasolthoz képest kevesebb mérést igényel, így joggal számíthatok arra, hogy elterjed gyakorlati alkalmazása.

7. Közlemények

Az értekezés témájában megjelent publikációk

Idegen nyelven megjelent publikációk

1. **E. Vágó**, S. Kemény: Logistic ridge regression for clinical data analysis (a case study). *Applied Ecology and Environmental Research* 4(2), 2006.
2. **E. Vágó**, S. Kemény: Critique of the AIAG Cross-Tabulation Procedure for Attribute Gauge R&R Study, *International Journal of Quality Engineering and Technology*, 2(1), p. 75-93, 2011, DOI: 10.1504/IJQET.2011.038724
3. **E. Vágó**, S. Kemény: A model based approach for attribute R&R Analysis, *Quality and Reliability Engineering International*, **IF: 0,452**, 2010, DOI: 10.1002/qre.1154
4. **E. Vágó**, Zs. Lang, S. Kemény: Overdispersion at the Binomial and Multinomial Distribution, *Periodica Polytechnica*, **IF: 0,042**, 55(1), p.17-20, 2011, DOI:10.3311/pp.ch.2011-1.03
5. **E. Vágó**, S. Kemény: Random effects model for attribute gauge R&R *Quality and Reliability Engineering International*, **IF: 0,452**, (elfogadva)

Előadások, posztterek nemzetközi konferenciákon

1. **E. Vágó**, S. Kemény: Analysis of attribute measurement systems. *Conferentia Chemometrica VII*, Hajdúszoboszló (2005)
2. **E. Vágó**, S. Kemény: Analysis of attribute measurement systems. *Symposium on Computer Applications and Chemometrics in Analytical Chemistry (SCAC)*, Tihany (2006)
3. **E. Vágó**, S. Kemény: Overdispersion for multinomial distribution: a Case Study for Attribute Gage Analysis. *Symposium on Computer Applications and Chemometrics in Analytical Chemistry (SCAC)*, Balatonalmádi (2008)

Előadások hazai konferenciákon, szakmai fórumokon

1. **E. Vágó**, S. Kemény: Minősítéses mérőeszközök képességvizsgálata *Műszaki Kémiai Napok*, Veszprém (2005)
2. **E. Vágó**, S. Kemény: A minősítés minősítése. *Műszaki Kémiai Napok*, Veszprém (2006)
3. **E. Vágó**, S. Kemény: Overdispersion. *Klinikai Biostatistikai Társaság ülése*, Budapest (2008)

A doktori tanulmányokhoz kapcsolódó egyéb publikációk

Könyv társszerzője

Kemény Sándor - Deák András - Lakné Komka Kinga - **Vágó Emese**: Statisztikai elemzés a STATISTICA programmal, Műegyetemi Kiadó, Budapest, 2004

Egyéb előadások, posztterek

1. **E. Vágó**: Páros t-próba és varianciaanalízis. *Műszaki Kémiai Napok*, Veszprém (2002)
2. **E. Vágó**: Páros t-próba és varianciaanalízis. *EOQ Six Sigma Szakbizottság ülése*, Budapest (2003)
3. **E. Vágó**: Ridge regresszió alkalmazása logit modell esetén. BME, *Vegyész és Biomérnöki Kar Doktoráns Konferencia*, Budapest (2004)
4. **E. Vágó**, S. Kemény: Paired t-test and analysis of variance. *Modern Methods of Data Analysis Oroszország*, Belokurikha, (2003)
5. **E. Vágó**, S. Kemény: Ridge regression method to logit model. *Symposium on Computer Applications and Chemometrics in Analytical Chemistry (SCAC)*, Balatonfüred (2004)
6. **E. Vágó**, Kemény Sándor: Logistic ridge regression for clinical data analysis (a case study) VII. *Magyar Biometriai és Biomatematikai Konferencia*, Budapest (2005)
7. **E. Vágó**, S. Kemény: Ridge regresszió alkalmazása logit modell esetén. *Klinikai Biostatistikai Társaság ülése*, Budapest (2005)