

Panelcheck szoftver statisztikai lehetőségei az érzékszervi bírálócsoporthoz teljesítményének monitorozásában

*Gere Attila, Szabó Dániel, Franku Tamás,
Györey Annamária, Kókai Zoltán és Sipos László*

Budapesti Corvinus Egyetem, Élelmiszertudományi Kar,
Érzékszervi Minősítő Laboratórium

Érkezett: 2012. április 12.

Az érzékszervi vizsgálatok kulcskérdése az egyes bírálók és a bírálói csoport (panel) eredmények megbízhatósága és reprodukálhatósága. Ezek a szempontok kiemelkedően fontosak az akkreditált laboratóriumok eljárásai (ISO/IEC, 2005), az érzékszervi és a műszeres analitikai módszerek párhuzamba állítása során, vagy olyan tesztek (pl. preferencia térképezés) esetében, ahol a bemenő adatokat a szakértői panel átlagai jelentik (SIPOS et al., 2008).

A bírálói panel – a bíráló kiválasztó tesztek, majd több hónapos termékspecifikus sikeres elvégzése után – képzés egyfajta mérőműszerként használható az érzékszervi tulajdonságok vizsgálatához. Az ipari gyakorlatban általában az érzékszervi panel a felelős azért, hogy meghatározza és ellenőrizze az érzékszervi specifikációkat, ezért teljesítményük nyomonkövetése kulcskérdés az eredmények értelmezésénél.

A teljesítmény ellenőrzése során tulajdonképpen azt vizsgáljuk, hogy a panel vagy az egyes bíráló rendelkezik-e azokkal a képességekkel, melyekkel valós eredményeket tud adni egy tulajdonság vizsgálatakor. A panelek vagy bírálók célra való alkalmasságát jellemzik az alábbi paraméterek: ingerek között különbségtételi képessége, egyén és a panel ismétlőképessége, reprodukálhatóság (állandóság, hibamentesség). A teljesítménymutatók monitorozásával folyamatosan értékelni lehet a panel vagy a bírálók egyéni teljesítményét így többek között lehetőség nyílik azonosítani:

- a gyengébben teljesítő bírálókat,
- a gyengébb teljesítés gyökereit (skálahasználat, megváltozott érzékenység, mentális túlterheltség, megbetegedés, nem kidolgozott referenciaminta használata, hibás kérdésfeltevés, motiváció hiánya),

- az egyén vagy a panel tagjai számára személyre szabott fejlesztési területeket (ISO/DIS, 2011).

Anyag és módszer

Kutatásunk célja három különböző országban előállított (Bulgária, Csehország, Magyarország) kereskedelmi forgalomban kapható sör minta tesztelésén keresztül bemutatni az érzékszervi panelteljesítmény-mutatók alkalmazását. A 10 fős képzett bírálócsoport profilanalitikus érzékszervi módszerrel (ISO, 2003) elemezte a vizsgálati mintákat, amelyet két ismétlésben, nemzetközi szabványok alapján kialakított érzékszervi laboratóriumban (ISO, 2007) végzett. A panel teljesítményének mérését a norvég Nofima (Norwegian Food Research Institute, Osloveien 1, 1430 Ås, Norway) által fejlesztett PanelCheck 1.4. programmal végeztük. Az elemzés során a program fejlesztői által javasolt „Workflow”-t követtük, mivel ennek relevanciáját több nemzetközi cikkben is bizonyították (TOMIC et al., 2007; TOMIC et al., 2010; DAHL et al., 2008). Így ezek alapján mutatjuk be az általunk alkalmazott módszereket (1. ábra).

A többváltozós elemzések megkezdése előtt egy kevert ANOVA-t (varianciaanalízist) futtattunk le, melynek során az analízis első lépéseként a minták közti szignifikáns különbségeket vizsgáljuk meg mixed modell ANOVA segítségével. A módszer a minták, a bírálók és azok interakcióinak két szempontos ANOVA modellezésén, illetve a minták, bírálók, ismétlések és azok interakcióinak három szempontos ANOVA modellezésén, majd egy általános F-próba mintahatáson történő alkalmazását jelenti. Minden esetben (2- vagy 3-szempontos), a kölcsönhatásokat és a bíráló hatásokat is véletlennek tekintik. Csak azokat a tulajdonságokat kell vizsgálni a továbbiakban, amelyek bizonyos szinten szignifikánsak (5%-os szignifikancia szint) a termékhatással (NAES et al., 2010). Azokat a tulajdonságokat, amelyek nincs szignifikáns termékhatása, a további elemzésnél nem vettük figyelembe.

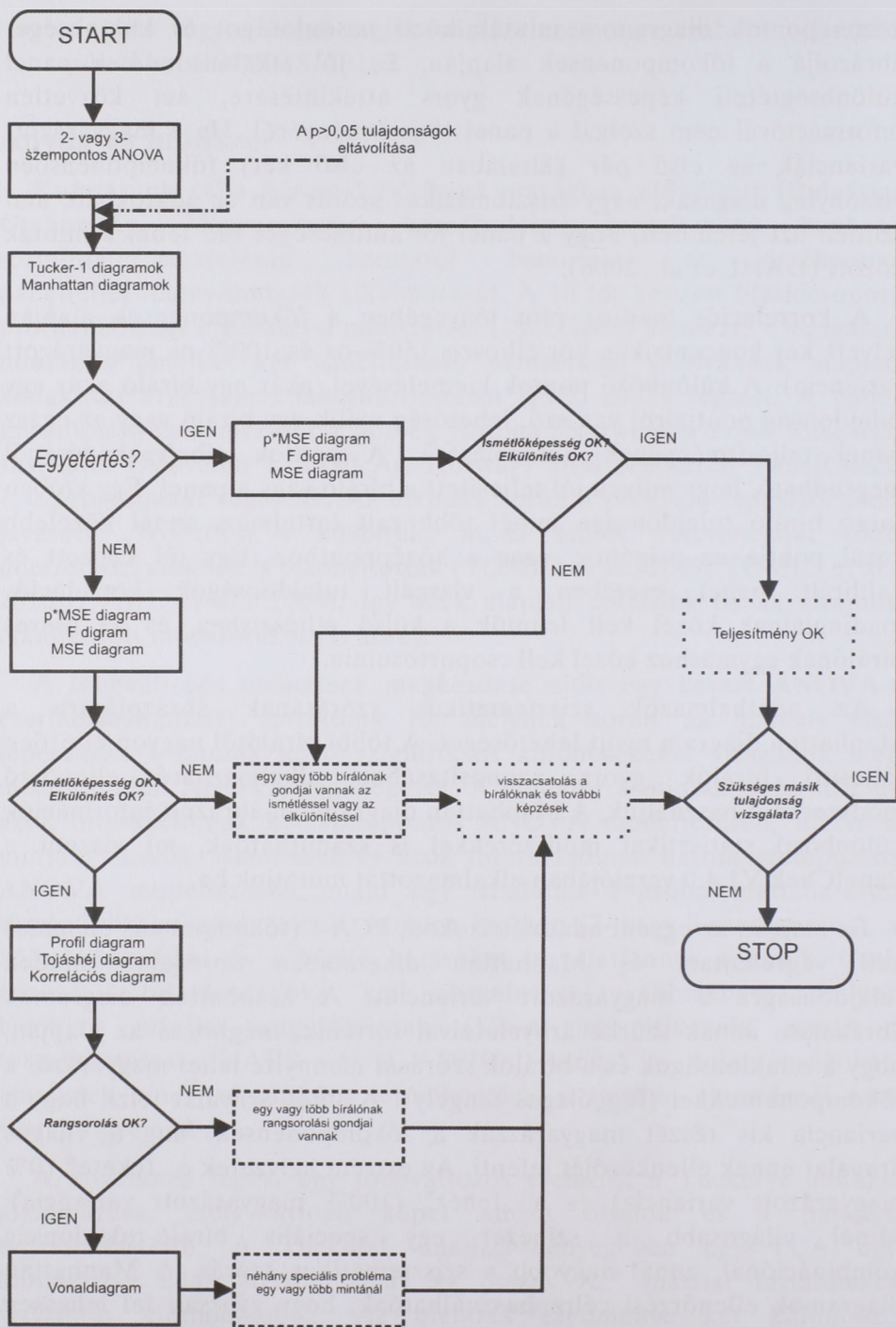
A következő lépés, egy többváltozós elemzés, a Tucker-1 analízis alkalmazása, mely átfogó képet ad a bírálók és a panelek teljesítményéről. A Tucker-1 analízis lényegében egy PCA egy kibontatlan adatmátrixon, mely az összes $X_i^{bá}$ mátrixt tartalmazza vízszintes elrendezésben. Az elemzés eredménye két különböző diagram: a közös pontok diagram és a korrelációs loading diagram. A

közös pontok diagram a minták közti hasonlóságot és különbséget ábrázolja a főkomponensek alapján. Ez jól alkalmazható a panel különbségtételi képességének gyors áttekintésére, ám közvetlen információval nem szolgál a panel teljesítményéről. Ha a magyarázott varianciák az első pár (általában az első két) főkomponensben viszonylag magasak, nagy szisztematikus szórás van az adatokban, ami szintén azt jelentheti, hogy a panel jól különbséget tud tenni a minták között (DAHL et al., 2006).

A korrelációs loading plot lényegében a főkomponensek alapján felvett két koncentrikus kör/ellipszis (50%-os és 100%-os magyarázott variancia). A különböző pontok kiemelésével, akár egy bíráló akár egy tulajdonság pontjairól van szó, lehetőség nyílik egy bíráló vagy az egész panel teljesítményének ábrázolására. A pontok elhelyezkedéséből megtudható, hogy milyen jól teljesített a bíráló vagy a panel. Egy szóban forgó bíráló tulajdonsága minél több zajt tartalmaz, annál közelebb kerül pontja az origóhoz, azaz a középponthoz. Egy jól képzett és kalibrált panel esetében a vizsgált tulajdonságok korrelációs loadingjainak közel kell lenniük a külső ellipszishoz, és az összes bírálóknak egymáshoz közel kell csoportosulnia.

Az adathalmazok szisztematikus szórásának ábrázolására a Manhattan diagram nyújt lehetőséget. A többi bírálótól nagyon eltérően teljesítő bírálók gyors azonosításához alkalmazható ellenőrző módszerként használjuk. A Manhattan diagramban ábrázolt információk különböző statisztikai módszerekkel is számíthatóak, mi viszont a PanelChek V1.4.0 verziójában alkalmazottat mutatjuk be.

Ez alapján az egyéni adatmátrixokon, PCA-t (főkomponens elemzés) kell végrehajtani és Manhattan diagramban ábrázolni minden tulajdonságra a magyarázott varianciát. A Manhattan diagramok ábrázolása annak szürke árnyalataival történik, méghozzá az alapján, hogy a tulajdonságok és a bírálók szórását mennyire lehet magyarázni a főkomponensekkel (függőleges tengely). A sötét árnyalat jelzi, hogy a variancia kis részét magyarázzák a főkomponensek, míg a világos árnyalat ennek ellenkezőjét jelenti. Az extrém helyzetek a „fekete” (0% magyarázott variancia) és a „fehér” (100% magyarázott variancia). Minél világosabb a színezet egy speciális bíráló-tulajdonság kombinációnál, annál nagyobb a szisztematikus szórás. A Manhattan diagramok ellenőrzési célra használhatóak, hogy gyorsan fel lehessen deríteni azokat a bírálókat, akik nagyon elérően viselkednek, vagy azokat a tulajdonságokat, melyek nem jól magyarázhatóak a másikkal.



1. ábra: PanelCheck Workflow (Forrás: PanelCheck Manual)

A többváltozós elemzések (Tucker-1 és Manhattan elemzések) után a Workflow alapján az elkülönítő képességet (F és p^* MSE diagram) és az ismétlőképességet (MSE és p^* MSE diagram) vizsgáljuk.

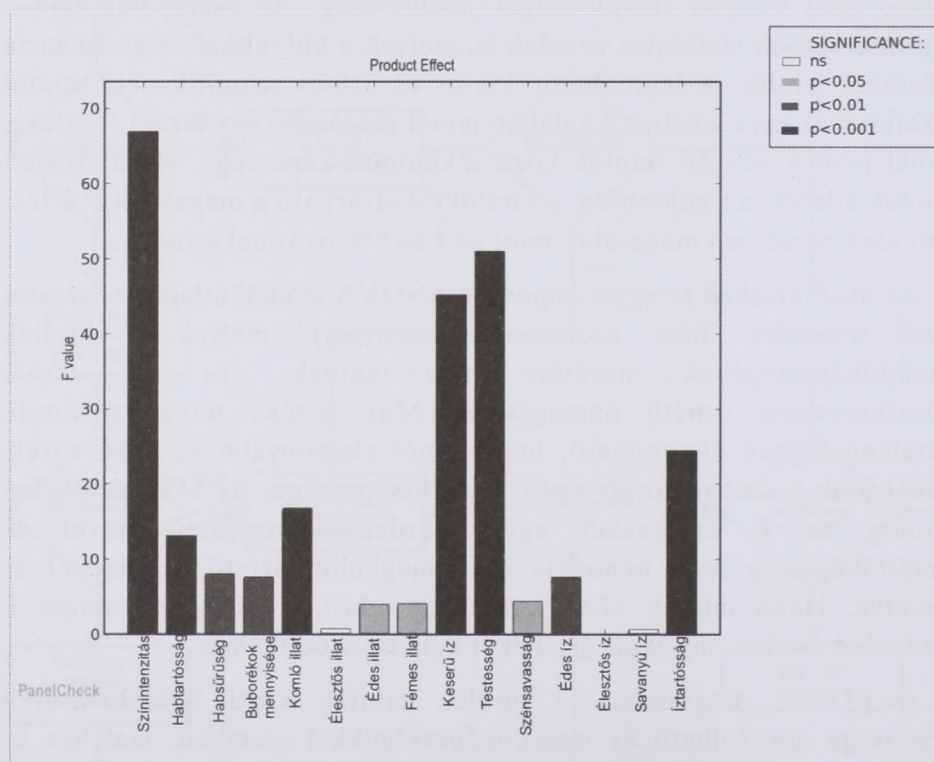
Az F érték a bírálók azon képességének ellenőrzésére használható, hogy a megadott tulajdonságban képesek-e a minták között különbséget tenni. Az elemzés során az összes F érték kiszámításra kerül és ábrázolható egy oszlopdiagramban, melyben minden oszlop egy bizonyos bíráló által vizsgált tulajdonságot jelenít meg. Az oszlopdiagramban feltüntethetőek vízszintes vonalak is, melyek a különböző szignifikancia szinteket jelölik. A legtöbbször 1%-os és 5%-os szignifikancia szintet alkalmazzák erre a célra. Általában minél magasabb egy bíráló F értéke, annál jobb a vizsgált minták közti elkülönítő képessége. Ha a vizsgált minták között van különbség, a bírálótól elvárható a magasabb F érték, ami ideális esetben magasabb, mint az 1 és 5%-os vonal szintje.

Az MSE értékek az egyszempontos ANOVA modell átlagos négyzetes hibái (véletlen hiba szórásának becslései), melyek a bírálók ismétlőképességének mérésére használhatóak. Ha egy bíráló következetesen ismétli önmagát, az MSE értéke nullához közeli. Általánosságban elmondható, hogy minél alacsonyabb az MSE érték, annál jobb a szóban forgó bíráló ismétlőképessége. Az MSE értékeket mindig az F értékekkel együtt érdemes vizsgálni, mivel az ismétlőképesség nem biztosítja a jó megkülönböztető képességet és fordítva. Ha a minták közti különbség adott, a bírálónak magas F értékeket és alacsony MSE értékeket kellene produkálnia.

A p^* MSE diagramban a bírálók minták közti különbségtételi képessége ábrázolható az ismétlőképességükkel szemben, melyhez az egyszempontos ANOVA számítások p és MSE értékeit használják fel. Ideális esetben minden bírálónak alacsony p és alacsony MSE értékeket kell elérnie az összes tulajdonságra, ha a termékek közti különbség valóban jelen van, és emiatt a diagram bal alsó sarkában kell csoportosulniuk. A p^* MSE diagrammal egyszerűen és gyorsan ki lehet deríteni, melyik bíráló teljesít gyengén egy adott tulajdonság vizsgálatánál. A p^* MSE diagram nagy előnye, hogy az összes bírálót és tulajdonságot egy diagramban ábrázolva egy lépésben gyorsan felfedi a megkülönböztető képesség és ismétlőképesség hiányosságait.

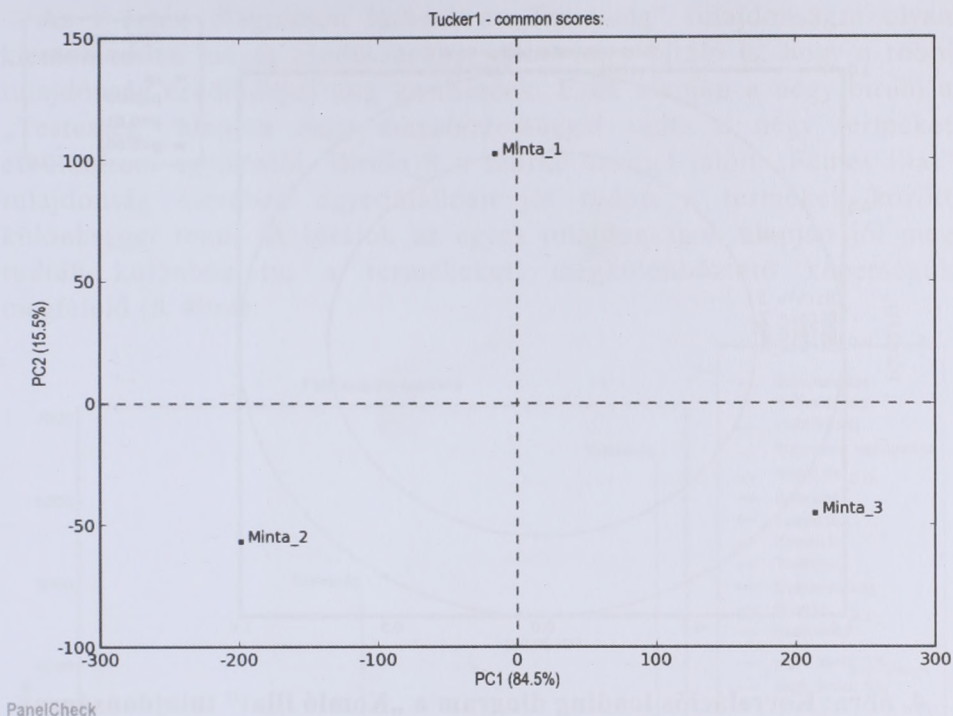
Eredmények

A Workflow első lépése, hogy azonosítsuk azokat a termék jellemzőket, melyeknek nincsen szignifikáns termék hatása. A fehér színnel jelzett tulajdonságok ilyen szempontból nem szignifikánsak, így a további adatelemzésből kizárandók az „Élesztős illat”, az „Élesztős íz” és a „Savanyú íz” tulajdonságok. A következő 2. ábrán látható a Workflow első lépésének, a kevert ANOVA eredménye.



2. ábra: Az ANOVA elemzés eredményei

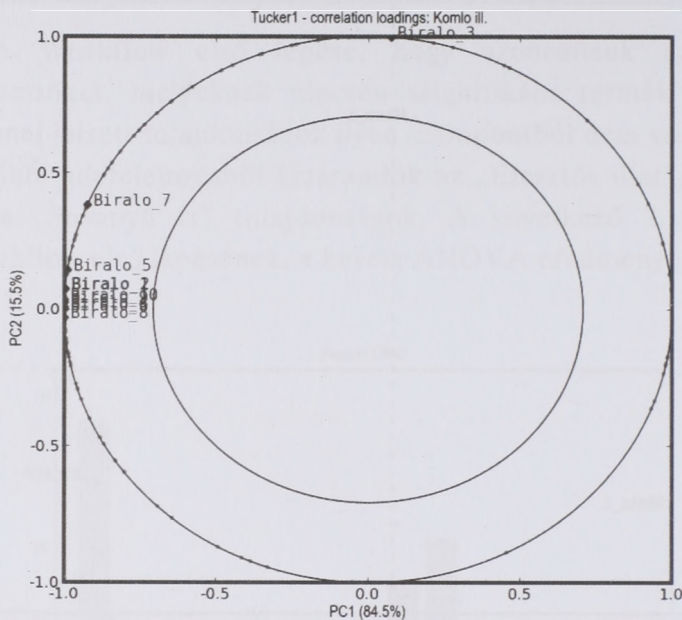
A közös pontok diagramon az egyes minták egymástól való távolsága adja meg, mennyire hasonlóak vagy különbözőek egymáshoz képest. A panel eredményei alapján a három vizsgált sör minta teljes mértékben elkülöníthető és egymástól különböző. Minél közelebb helyezkednek el egymáshoz a minták, annál hasonlóbbnak mondhatóak. Esetünkben a három minta közel egyforma távolságban van egymástól, ami alapján kijelenthető, hogy teljes mértékben különböznek egymástól (3. ábra).



3. ábra: Közös pontok diagram.

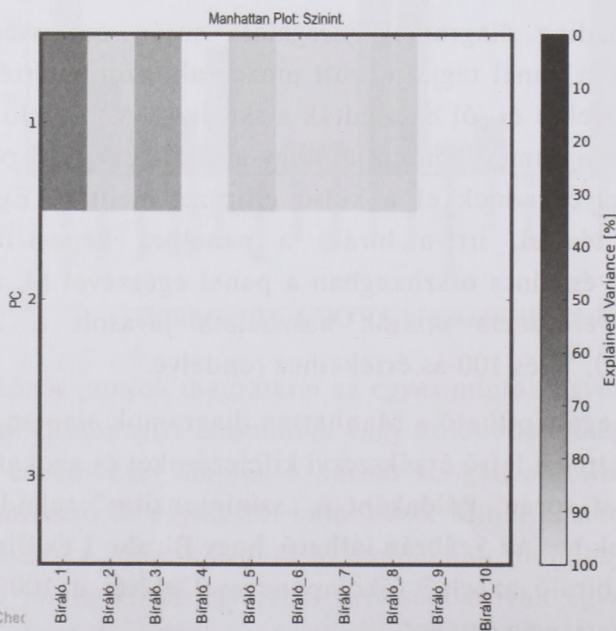
A korrelációs loading diagramok vizsgálata során összességében megállapítható, hogy a panel tagjai együtt mozogtak, azonos értékeket adtak az azonos ingerekre és jól használták a skálákat. A „Komló illat” korrelációs loading diagramja bemutatja, hogy a bírálókat jelző pontok egymáshoz közel helyezkednek el a külső ellipszis mentén. Egyedül Biraló_3 pozíciója tér el, itt a bíráló a panelhez képest eltérő eredményeket adott, és nincs összhangban a panel egészével (4. ábra). Ennek kivédésére referencia minták használata javasolt a 100-as beosztású skála 25, 50, 75 és 100-as értékeihez rendelve.

Összefoglalóan megállapítható a Manhattan diagramok alapján, hogy a bírálók jól megértették a leíró érzékszervi kifejezéseket és azokat jól is alkalmazták a bírálat során. Példaként a „színintenzitás” tulajdonság eredményeit mutatjuk be. Az 5. ábrán látható, hogy Biraló_1 és Biraló_3 kivételével minden bíráló az első főkomponenssel elérte a 100%-hoz közeli magyarázott variancia értéket.



PanelCheck

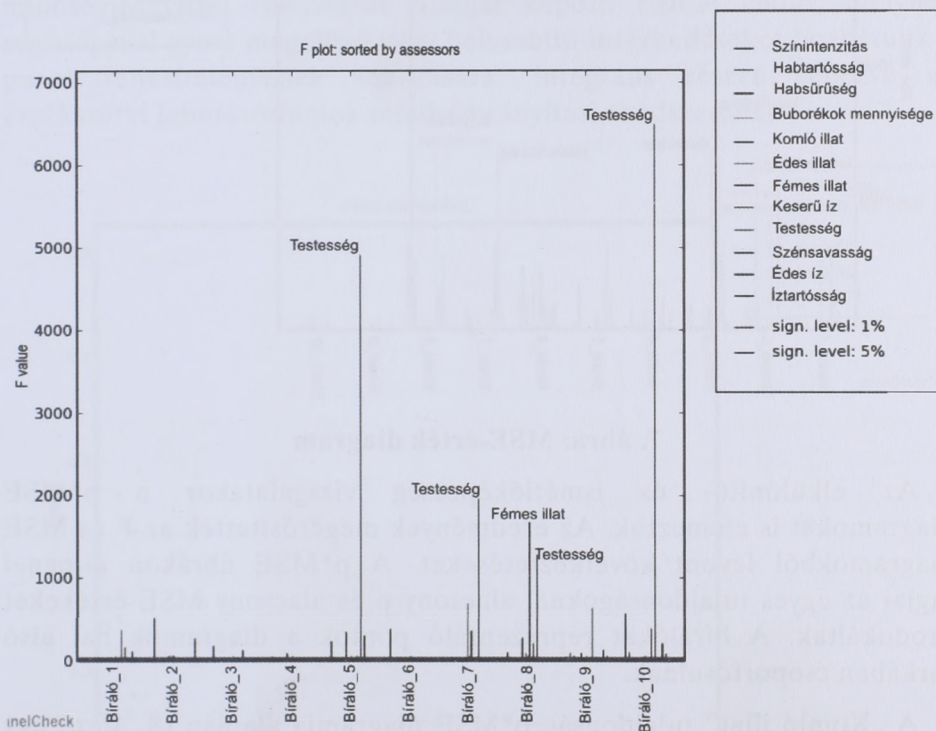
4. ábra: Korrelációs loading diagram a „Komló illat” tulajdonságra



PanelCheck

5. ábra: Manhattan diagram a „Színintenzitás” tulajdonságra

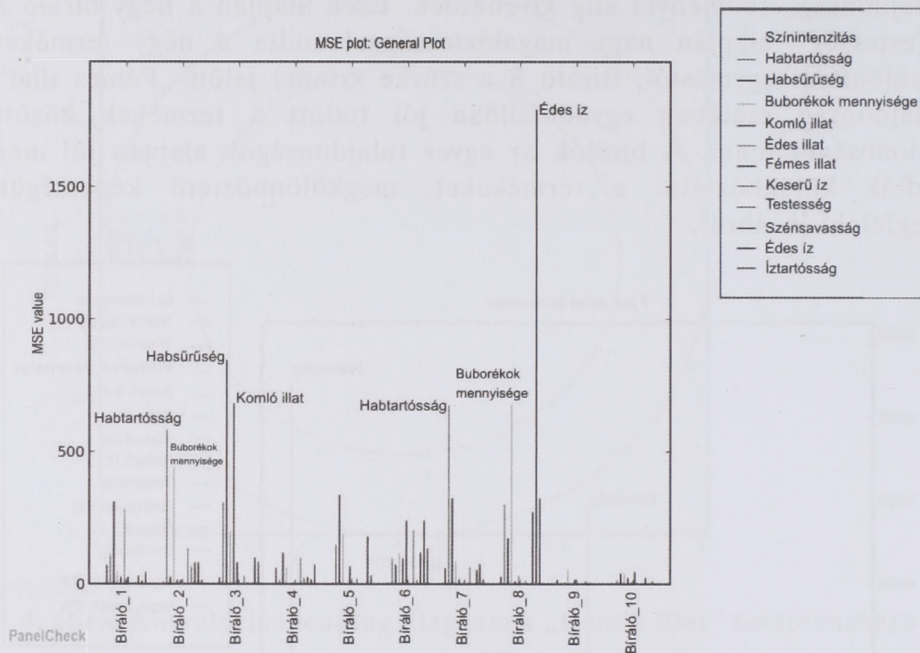
Az F-érték diagramon látható a „Testesség” tulajdonságra olyan kiemelkedően magas eredményeket adott négy bíráló is, hogy a többi tulajdonság eredményei alig kivehetőek. Ezek alapján a négy bíráló a „Testesség” alapján nagy magabiztossággal tudta a négy terméket elkülöníteni egymástól. Biraló_8 a szürke színnel jelölt „Fémes illat” tulajdonság esetében egyedülállóan jól tudott a termékek között különbséget tenni. A bírálók az egyes tulajdonságok alapján jól meg tudták különböztetni a termékeket, megkülönböztető képességük megfelelő (6. ábra).



6. ábra: F-érték diagram

A laboratóriumi tapasztalatunk azt mutatja, hogy az 500-as MSE érték alatt teljesítő bírálók ismétlőképessége elfogadható egy nem termék-specifikus panel esetében. Néhány esetben előfordult, hogy a panel egyes tagjai magasabb értékeket produkáltak, ezek közül különösen kiemelkedő a Biraló_8 „Komló illat” tulajdonságra adott értékei közti eltérés. Itt nagyon magas értéket kaptunk. Ebben az esetben szükséges lenne a bírálónak további ismételt érzékszervi gyakorlásra. Kiemelkedő a Biraló_4, Biraló_9 és Biraló_10 eredménye, ami alapján nagyon precízen és következetesen ismételt eredményeket

kaptunk (7. ábra). Összességében elmondható, hogy a panel ismétlőképessége megfelelő.



7. ábra: MSE-érték diagram

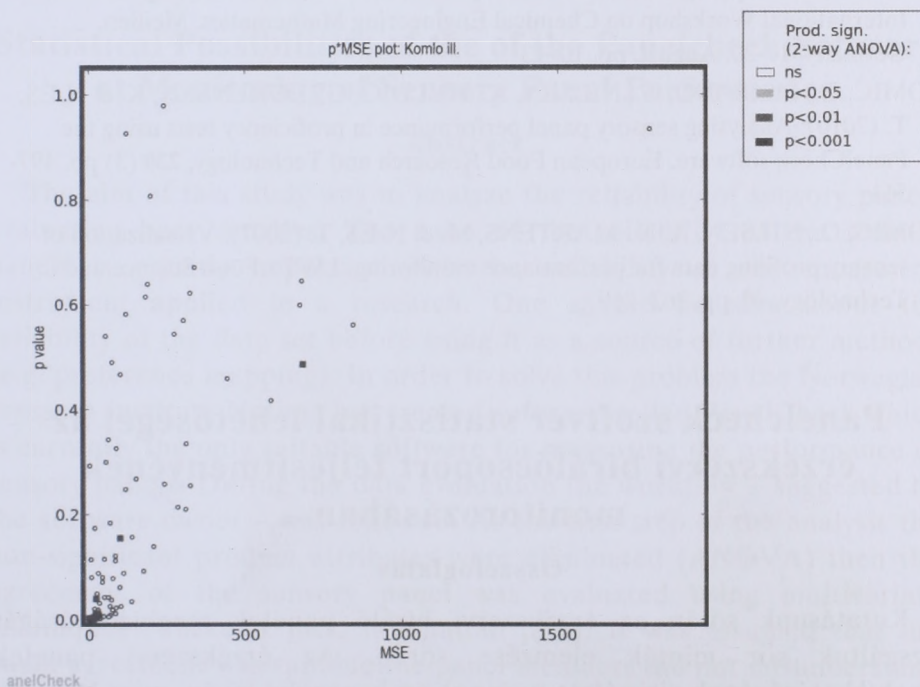
Az elkülönítő- és ismétlőképesség vizsgálatok a p^* MSE diagramokat is elemeztük. Az eredmények megerősítették az F és MSE diagramokból levont következtetéseket. A p^* MSE ábrákon a panel tagjai az egyes tulajdonságoknál alacsony p és alacsony MSE értékeket produkáltak. A bírálókat reprezentáló pontok a diagramok bal alsó sarkában csoportosulnak.

A „Komló illat” tulajdonság p^* MSE diagramja alapján (8. ábra) egy bíráló különül el a panel többi tagjától. Az összes diagram vizsgálata alapján összességében megállapítható, hogy a panel jó megkülönböztető és ismétlőképességgel rendelkezett a vizsgált tulajdonságok tekintetében.

Következtetések

A nemzetközi szabványokon (ISO) alapuló, informatikai támogatással megvalósított érzékszervi tesztek hatékonyan segítik a bírálatok tervezését, végrehajtását és elemzését. A bírálók teljesítményének monitoring rendszerei azonban sok esetben csak

néhány területre fókuszálnak, így nem teszik lehetővé a panelek összetett teljesítményértékelését (FIZZ, Compusense, Senstools, XL-Stat Sensory, Profisense). Jelenleg a legsokoldalúbban az egyén és a panel teljesítményének monitorozását a norvég Nofima által fejlesztett PanelCheck támogatja. Kutatásunkban bizonyítottuk, hogy a Workflow statisztikai módszereit követve megfelelően azonosíthatók a fejlesztésre váró területek (kopló illat – panel összhang, testesség és fémes illat – elkülönítő képesség, kopló illat – ismétlőképesség). Az érzékszervi bírálók folyamatos nyomonkövetésével így hatékonyan valósítható meg a minőségirányítási rendszerek alapját képező PDCA ciklus, melynek segítségével gyors megelőző vagy helyesbítő intézkedéseket hozhatunk a panel teljesítményének fokozására, integráns részét képezve az érzékszervi laboratóriumok minőségirányítási rendszerének.



8. ábra: A „Kopló illat” p*MSE diagramja

Hivatkozások

- DAHL, T. & NAES, T. (2006). A bridge between Tucker-1 and Carroll's generalised canonical analysis. *Computational statistics and data analysis*, **50** (11), pp. 3086-3098
- DAHL, T., TOMIC, O., WOLD, J.P. & NÆS T. (2008): Some new tools for visualising multi-way sensory data. *Food Quality and Preference*, **19** (1), pp. 103-113

- ISO 2003: Sensory analysis – Methodology – General guidance for establishing a sensory profile. No. 13299:2003
- ISO 2007: Sensory analysis – General guidance for the design of test rooms. No. 8589:2007
- ISO/DIS 2011: Sensory analysis – Methodology – General guidance for monitoring the performance of a quantitative sensory panel. No. 11132:2011
- ISO/IEC 2005: General requirements for the competence of testing and calibration laboratories. No. 17025:2005
- NAES, T., BROCKHOFF, P. B. & TOMIC, O. (2010): Statistics for sensory and consumer science. Wiley, Chicester. pp. 11–34, 193–245
- NAES. T., BROCKHOFF P.B. & TOMIC, O. (2010): Statistics for sensory and consumer science. Wiley, Chichester, pp. 126-154, 193-206, 249-262
- SIPOS, L., KOLLÁR–HUNEK, K., HESZBERGER, J. & KÓKAI Z. (2008): Preference Mapping, Panel Consistency and PCA in Food Marketing. 28th International Workshop on Chemical Engineering Mathematics. Meißen, Germany, 17-20. August, pp. 10–13
- TOMIC, O., LUCIANO, G., NILSEN, A., HYLDIG, G., LORENSEN, K. & NÆS, T. (2010): Analysing sensory panel performance in proficiency tests using the PanelCheck software. European Food Research and Technology, **230** (3) pp. 497-511
- TOMIC, O., NILSEN, A. N., MARTENS, M. & NÆS, T. (2007): Visualization of sensory profiling data for performance monitoring, LWT - Food Science and Technology, **40**, pp. 262–269

Panelcheck szoftver statisztikai lehetőségei az érzékszervi bírálócsoport teljesítményének monitorozásában

Összefoglalás

Kutatásunk során az érzékszervi bíráló panelek megbízhatóságát vizsgáltuk sör minták elemzése során. Az érzékszervi panelek megbízhatóságának vizsgálata ugyanolyan hangsúlyt kíván, mint bármely más mérőműszer. A további módszerekben (pl. preferencia-térképezés) történő felhasználás előtt meg kell győződni az adatok megbízhatóságáról. Ennek a problémának a megoldására nyújt lehetőséget a norvég Nofima által kifejlesztett PanelCheck program, mely jelenleg egyedüli az érzékszervi panelteljesítmény mérésére alkalmazható statisztikai módszereket tartalmazó szoftverek közül. Kutatásunkban a szoftver fejlesztői által javasolt Workflow-t követtük.

Az elemzések során először a nem szignifikáns termék tulajdonságokat zártuk ki (ANOVA), majd a többváltozós statisztikai elemzések segítségével (Tucker-1 plot, Manhattan plot) a panel összhangját vizsgáltuk. A vizsgálat során megállapítottuk, hogy a panel összhang megfelelő volt, nem voltak félreértések a skálahasználattal kapcsolatban. Az egyváltozós elemzések során (F plot, MSE plot, p^* MSE plot) a panel tagjait egyenként vizsgáltuk, melynek során sikerült azonosítani, hogy az egyes vizsgált tulajdonságok esetében melyik paneltag szorul további gyakorlásra. Emellett sikerült azonosítani a nagyon jó megkülönböztető- és ismétlőképességgel rendelkező bírálókat is. Az érzékszervi panelek egészének, illetve a panel tagjainak elemzése segítséget nyújt, hogy akár személyre szabott teljesítményjavító feladatokat lehessen összeállítani a panel vagy egy bíráló részére.

Statistical Possibilities of the of the Panelcheck Software at Monitoring of Sensory Panel Performance

Abstract

The aim of this study was to analyze the reliability of sensory panels evaluating beer samples. The issue of the reliability of the sensory panels is equally important with the reliability of any measuring instrument applied in a research. One should be sure about the reliability of the data set before using it as a source of further methods (e.g. preference mapping). In order to solve this problem the Norwegian research institute Nofima has created software called PanelCheck which is currently the only suitable software for measuring the performance of sensory panels. During the data evaluation the workflow – suggested by the software owner – was followed. As the first step of the analysis the non-significant product attributes were eliminated (ANOVA) then the agreement of the sensory panel was evaluated using multivariate techniques (Tucker-1 plot, Manhattan plot). It was assumed that the panel agreement was suitable the panel members did not misunderstand the use of the scale. The panel members were evaluated one by one applying the univariate statistical techniques (F plot, MSE plot, p^* MSE plot) and panel members who needed more practice were identified. In addition panel members having good discrimination ability and repeatability were identified, too. Analyzing the performance of the whole panel or an individual assessor can help to create aimed tasks to improve the performance of the panel members and also the whole panel.