

TÖBBVÁLTOZÓS FOLYAMATSZABÁLYOZÁS VIZSGÁLATA R KÖRNYEZETBEN

INVESTIGATION OF MULTIVARIATE STATISTICAL PROCESS CONTROL IN R ENVIRONMENT

József MIHALKÓ, Róbert RAJKÓ

*Szegedi Tudományegyetem, Mérnöki Kar, Folyamatmérnöki Intézet, 6725 Szeged, Moszkvai
krt. 5-7.*

*Institute of Process Engineering, Faculty of Engineering, University of Szeged,
5-7. Moszkvai krt., Szeged, Hungary, H-6725*

Abstract

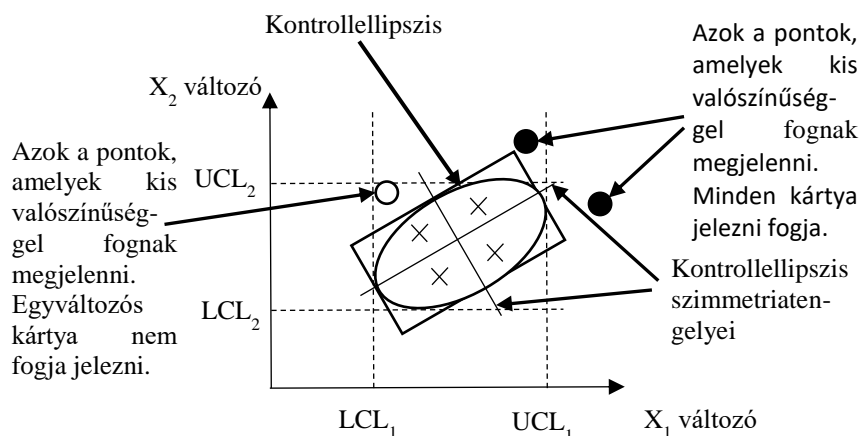
At the first stage of our work, the theoretical knowledge needed to use the multivariate statistical process control (MSPC) was explored. Last year, we clarified the sometimes confused concepts, equations and formulas (Mihalkó and Rajkó, 2016). At the second stage, R project simulation studies and some food industrial practical model investigations are carried out for confirming the MSPC advantages compared with the univariate ones. Furthermore, we analyze, using principal component analysis (PCA), what could cause the outlying values, moreover we will demonstrate how to use the MYT-decomposition.

Bevezetés

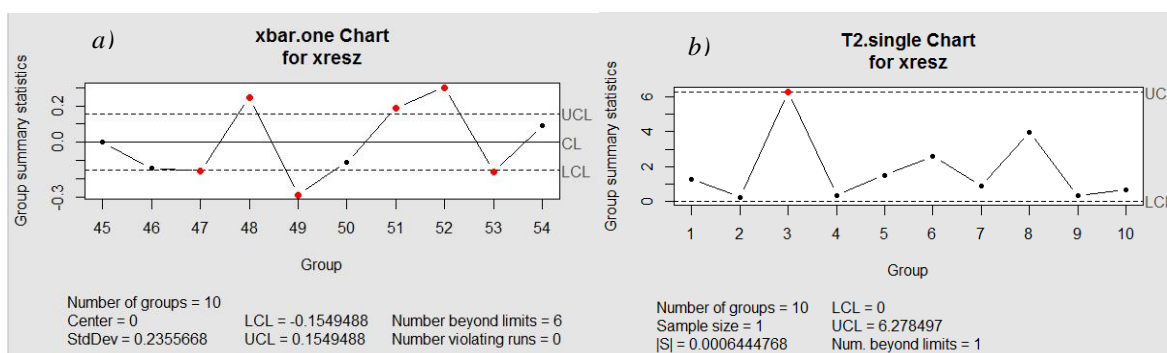
A statisztikai folyamatszabályozás (Statistical Process Control, SPC) során akkor történik beavatkozás a termékgyártás adott folyamatába, ha ismerünk olyan okot, amelynek hatására a minőségjellemző (pl. tömeg) értéke megváltozik. Az SPC fő eszközei közé az ellenőrző (más néven szabályozó) kártyák sorolhatók (Kemény et al., 1998).

Módszerek

A statisztikai folyamatszabályozáson belül elkülöníthetők az egy-, illetve a többváltozós folyamatszabályozási módszerek. A két módszer közötti fő különbség, hogy az egyváltozós módszereknél (Univariate SPC, USPC) egy ismert változónak az értelmezése történik egy vagy több ismert – nem mesterséges – változóval, míg a többváltozós módszereknél (Multivariate SPC, MSPC) a több ismert változót kevesebb számú mesterséges változóval értelmezzük (Sváb, 1979). A két módszer közötti különbséget mutatja az 1. ábra.



4. ábra: Egy- és többváltozós módszerek összehasonlítása. Forrás: Rogalewicz (2012).



2. ábra: Véletlen számok generálásával kapott a) átlag-kártya és b) T²-kártya (95%-os megbízhatósági szinten).

A T²-statisztika pontos eloszlása két szemponttól függ (Ittész, 1999):

- egyrészt attól, hogy egyedi vagy csoportosított adatokkal dolgozunk;
- másrészt attól, hogy visszatekintő elemzést végzünk (I. fázis) vagy az aktuális folyamatot felügyeljük (II. fázis).

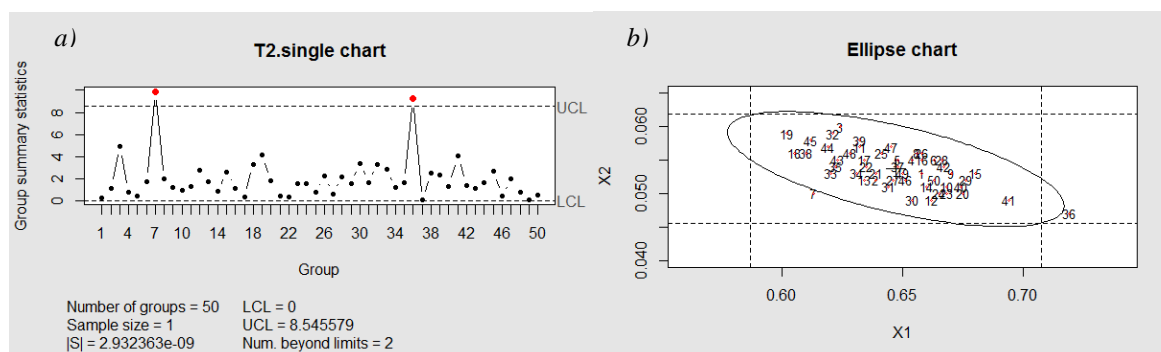
A II. fázisban azonban nehéz értelmezni, hogy mi okozhatta a jel tartományán kívülre kerülését. Lehet, hogy azt az egyik minőségjellemző, esetleg kettő vagy több változó együttműködése, vagy a kovariancia megváltozása váltja ki. Néhány módszert már kidolgoztak erre a problémára, pl. főkomponens-analízist, MYT-felbontást (Mason et al., 1997, Rogalewicz, 2012). Mason et al. (1997) az MYT-felbontás konkrét számítási sémáját írja le.

Eredmények és értékelésük

A tavalyi, 22nd International Symposium on Analytical and Environmental Problems elnevezésű konferencián poszterelőadás keretében bemutattuk az MSPC használatához szükséges elméleti ismereteket, tisztáztuk az alkalmazható összefüggések alakjait és szerepüket, ill. összegyűjtöttük az MSPC alkalmazásának főbb előnyeit és hátrányait az USPC-vel szemben (Mihalkó és Rajkó, 2016).

A munkánk második fázisában elsőként R project-ben megvalósított szimulációs vizsgálatokat végeztünk el, amelyet a következőkben részletezünk. Az algoritmusok egyike a

100 db véletlen szám generálásával („runif” parancs segítségével történt) kapott adatsor felhasználásával elkészített egyváltozós folyamatszabályozás eszközeként használható átlag-kártya látható a 2.a) ábrán. A CL a középvonalat, az UCL az elfogadási tartomány felső határát, az LCL az elfogadási tartomány alsó határát jelenti. A 2.b) ábrán az MSPC során alkalmazható T^2 -kártya látható, ahol a véletlen számok generálásáért az „rmvnorm” parancs a felelős.



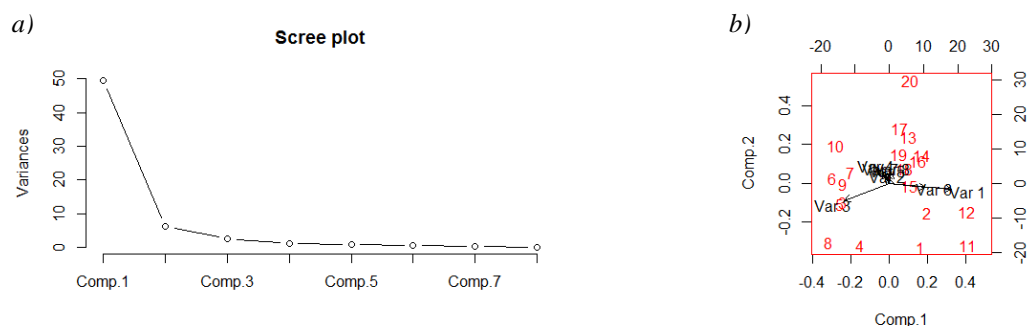
3. ábra: Pácolt sonka víz- és sótartalma alapján készített a) T^2 -kártya és b) kontrollellipszist tartalmazó ellenőrző kártya Ittész és Zukál példáján (99%-os megbízhatósági szinten).

Az algoritmusok megvalósítása után Ittész András és Zukál Endre mérési adatsorának R-ben történő ellenőrzése következett abból a célból, hogy az Ittész és Zukál (1999) által publikált Hotelling-féle T^2 -kártyát és a kontrollellipszist tartalmazó kártyát visszkapjuk-e (3.a) ábra).

A T^2 -kártyát visszakaptuk, azonban a kontrollellipszist tartalmazó szabályozó kártyát egy hibakód („kisminták számának nagyobbak kell lennie, mint 1”) miatt nem kaptuk vissza a szakirodalomban szereplővel. Az „ellipseChart” nevű parancs forráskódjának átalakítása révén meg tudtuk jeleníteni az ellenőrző kártyát (3.b) ábra).

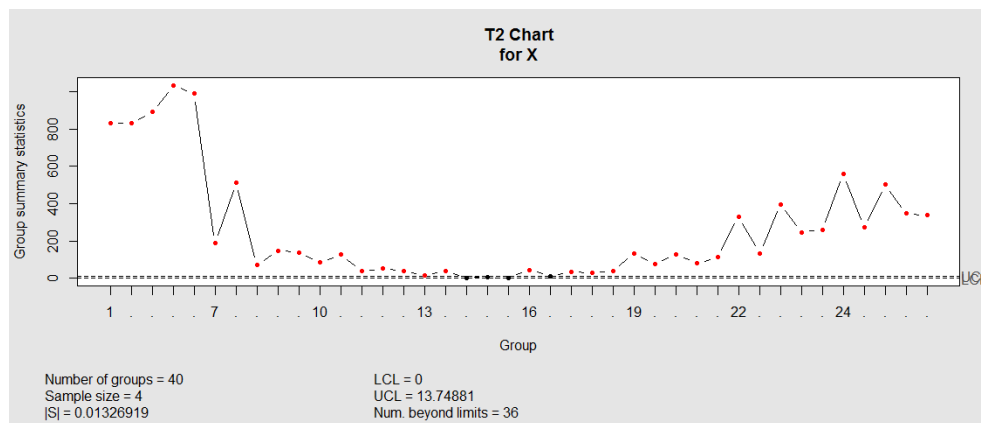
Továbbiakban gyakorlati alkalmazási illusztrációkat hajtottunk végre annak érdekében, hogy bizonyítsuk az MSPC előnyeit az USPC-vel szemben. A gyakorlati alkalmazási illusztrációk egyike a Koppenhágai Egyetem által elérhetővé tett Pármai sonka érlelésének mérési eredményeinek felhasználásával bemutatott folyamatszabályozás (models.life.ku.dk, 2016). Többek között azt sikerült megállapítani, hogy nem szükséges a sonkákat az előállításától kezdődően 15-18 hónapig érlelni, hanem elegendő csak a szűk esztendő (11-12 hónapos) érlelés, legalábbis csak az analitikai mérési eredményeket figyelembe véve.

Ezt követően a főkomponens-analízis módszerét felhasználva az R-ben megvizsgáltuk azt, hogy a 8 megmért változó – nedvesség-, só-, fehérjetartalom, 2 db érzékszervi bírálat, valamint a színérésekre tartozó L, a és b jellemzők – közül melyeknek lesz faktorhatása. A hegyomlás ábrán (4.a) ábra) látható, hogy 3 főkomponens van. A kettős diagram (4.b) ábra) segítségével belátható, hogy a 3. változó (fehérjetartalom, ennek faktorhatása is van), az 1. és a 6. változó (nedvességtartalom és az L értékek) lényegesen eltér a másik öt változótól.



4. ábra: a) Hegyomlás ábra és b) kettős diagram a Pármai sonka érlelésének változóira nézve.

A másik gyakorlati alkalmazási illusztráció egy saját márkás szalámi érlelésének folyamatkövetése volt, amelynek mérési adatait Rostás Csabának és témavezetőjének, Dr. Eszes Ferencnek köszönünk.

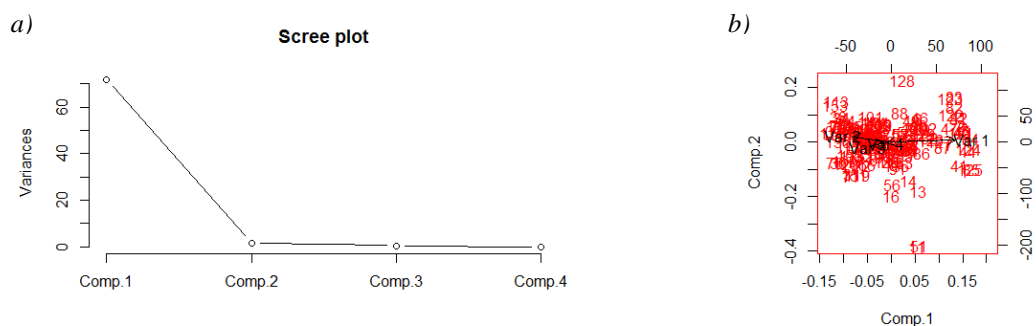


5. ábra: Folyamatkövetés a szalámi érlelésének változóira nézve (95%-os megbízhatósági szinten).

A 5. ábrán jól elkülöníthetőek a szalámi érlelésének főbb szakaszai:

- a füstölés utáni első napokban visszanedvesítés, a kéreg felpuhítása történik,
- a 7. naptól kezdődik a termék szárítása,
- a 10. naptól a 19. nap között a szalámi érlelése stabilizálódik,
- ezután pedig újfent vízelvonás történik, megszűnik a stabil folyamat.

Emellett a hegyomlás ábrán (6.a) ábra) látható, hogy 2 főkomponens van, illetve a kettős diagramról (6.b) ábra) leolvasható, hogy az első változó, illetve a második változó egymástól és a többi minőségjellemzőtől lényegesen eltér.



6. ábra: a) Hegyomlás ábra és b) kettős diagram a szalámi érlelésének változóira nézve.

Végezetül pedig egy példát mutatunk be az MYT-felbontás alkalmazására, vagyis arra keressük a kérdést, hogy mi okozhatja egy jel tartományon kívülre kerülését. Ehhez Ittész és Zukál (1999) mérési adataiból kapott T^2 -statisztika eredményeit használtuk fel (3.a) ábrán látható a T^2 -kártya). Ezen az ábrán látható, hogy 2 zavart jelző pont van, ez a 7. és a 36. minta. A részletezett számítást a 7. mintán keresztül mutatjuk be. Mivel kétváltozós folyamatszabályozásról van szó, ezért kétféleképpen lehet felbontani a 7. mintából kapott T^2 -értéket:

$$T^2 = T_1^2 + T_{2.1}^2 \quad (1)$$

$$T^2 = T_2^2 + T_{1.2}^2 \quad (2)$$

A 7. minta T^2 -értéke 9,888, a felső beavatkozási határ (angol rövidítése: UCL) pedig 8,546, vagyis látható az, hogy a T^2 -érték túllépi az UCL-t, ezért ki kell számolni a feltétel nélküli T^2 -értékeket (T_1^2 és T_2^2 értékeket) és egy új felső beavatkozási határértéket, ezen értékeket az 1. táblázatban összegezzük.

1. táblázat: Feltétel nélküli T^2 -értékek és a hozzá tartozó UCL-érték.

T_1^2	T_2^2	UCL
2,038	1,544	7,326

Az 1. táblázat értékei azt mutatják, hogy a feltétel nélküli T^2 -értékek az elfogadási tartományba esnek (kisebb az értékük, mint az UCL értéke), így még nem találtuk meg a zavarjelzés okát, emiatt ki kell számolnunk a feltételes T^2 -értékeket (az (1) és (2) egyenlet segítségével), illetve a jellemző felső beavatkozási határértéket, amelyeket a 2. táblázatban összegzünk.

2. táblázat: Feltételes T^2 -értékek és a hozzá tartozó UCL-érték.

$T_{2.1}^2$	$T_{1.2}^2$	UCL
7,850	8,344	7,491

Mivel a feltételes T^2 -értékek túllépik a felső beavatkozási határértéket, ezért a 7. minta esetében a jel tartományon kívülre kerülését az okozza, hogy ennél a mintánál a víz és a sótartalom korrelációja nem követi a többi mérésnél tapasztaltat.

Következtetések, javaslatok

Ebben a cikkben bemutatunk az R project-ben megvalósított szimulációs vizsgálatokat és gyakorlati alkalmazási illusztrációkat, amelyekkel igazoltuk az MSPC előnyeit az USPC-vel szemben. Azonban fontos megjegyezni, hogy ezen adatok nem a folyamatszabályozásra megtervezett mérési eredmények, csupán illusztrációkként szolgálnak (mind a Pármai sonka, mind a saját márkás szalámi folyamatkövetése esetében). Ezeknél az adatoknál később megnéztük, hogy mi okozta a jel tartományon kívülre kerülését a II. fázisban, amelyhez vagy főkomponens-analízist vagy az ún. MYT-felbontást használhatjuk.

A későbbiekben elsődlegesen a célunk az, hogy ipari partner segítségével egy adott technológiai folyamat gyakorlatban megvalósuló szabályozását hajtsuk végre, amelyhez pl. optimális kísérlettervet kell kidolgozni.

Köszönetnyilvánítás



Az Emberi Erőforrások Minisztériuma UNKP-17-2 kódszámú Új Nemzeti Kiválóság Programjának támogatásával készült.

Irodalom

http://www.models.life.ku.dk/ParmaHam_Fluor (2016. 10. 25.)

Ittész A. (1999): Többváltozós statisztikai folyamatszabályozás, *Minőség és Megbízhatóság*, 33(5), pp. 226-231.

Ittész A., Zukál E. (1999): Többváltozós folyamatszabályozás alkalmazási lehetőségei a húsiparban, *A Hús*, 9(3), pp. 179-183.

Kemény S., Papp L., Deák A. (1998): *Statisztikai minőség- (megfelelőség-) szabályozás*, Műszaki Könyvkiadó – Magyar Minőség Társaság, Budapest, pp. 56-57., pp. 81-83.

Mason R.L., Tracy N.D., Young J.C. (1997): A practical approach for interpreting multivariate T^2 control chart signals, *Journal of Quality Technology*, 29(4), pp. 399., pp. 404-405.

Mihalkó J., Rajkó R. (2016): Advantages and drawbacks of using multivariate statistical process control in food industry, In: *Proceedings of the 22nd International Symposium on Analytical and Environmental Problems*, Szeged, pp. 385-389., http://www2.sci.uszeged.hu/isaep/index_htm_files/Proceedings%20of_ISAEP_2016.pdf

Rogalewicz M. (2012): Some Notes on Multivariate Statistical Process Control, *Management and Production Engineering Review*, 3(4), pp. 81.

Sváb J. (1979): *Többváltozós módszerek a biometriában*, Mezőgazdasági Kiadó, Budapest, pp. 16.