

**PARADICSOMSÚRÍTMÉNYEK
MIKROBIOLÓGIAI ELLENŐRZÉSÉNÉL FELLEPŐ
STATISZTIKAI PROBLÉMÁKRÓL**

ÉLTETŐ ÖDÖN és SARKADI KÁROLY

A Konzerv-, Hús- és Hűtőipari Kutató Intézet kérésére foglalkoztunk a következő problémával.

Paradicsomsűrítmények átvételi minőségellenőrzése úgy történik, hogy a tétel valamelyik dobozából vesznek egy mintát; ezt szabvány szerint hígítják, majd belőle néhány cseppet kivéve, mikroszkóppal megvizsgálják, hogy 50—100 mikroszkópi látótér közül hányban található penészfonal. A penészfonalas, azaz pozitív részek százalékos számaránya a minta ún. Howard-száma. Ha ez a Howard-szám egy bizonyos korlátnál nagyobb, akkor visszautasítják az egész tételt, ha kisebb, akkor átveszik. Külföldi átvételnél a szereplő korlát 40. A kérdés az, hogy milyen korlátot kell megadni hazai előzetes ellenőrzésnél a Howard-számra ahhoz, hogy megfelelő minőségű tételek kerüljenek exportra, továbbá hogyan módosul ez a hazai átvételi korlát, ha tételenként nem egy, hanem két vagy esetleg több előzetes vizsgálat történik.

A problémát több szempontból és különböző módokon lehet tárgyalni, azonban mindegyiknél több-kevesebb egyszerűsítő feltételezéssel kell élnünk. Feltesszük, hogy a megvizsgált látóterek száma rögzített (a gyakorlatban rendszerint 100 látóteret vizsgálnak meg). Ha a látóterek pozitív vagy negatív volta egymástól független lenne, akkor a mintában talált Howard-szám binomiális eloszlású lenne. Azonban egyrészt a penészfonalak nem egyenletesen elszórva jelentkeznek, hanem kisebb-nagyobb csomókban, másrészt egy-egy penészfonal esetleg nem csak egy látótérben észlelhető, hanem átnyúlhat két-három szomszédos látótérbe is. Ily módon a szomszédos látóterek nem függetlenek egymástól, az eloszlás nem lesz binomiális, ami a fenti okok folytán elsősorban pl. abban jelentkezik, hogy a tapasztalati szórás elég jelentősen nagyobb, mint a megfelelő binomiálisnál lenne. A tárgyalás során ezért feltesszük, hogy a Howard-számok eloszlására a normális eloszlás a megfelelő közelítés. E feltételezést a tapasztalat is megerősíti. A megbízó intézet ugyanis előzőleg több tételre vonatkozóan 500—500 mérésből álló részletes vizsgálatot végzett, s a nyert tapasztalati eloszlások általában elég jó egyezést mutattak a normális eloszlással. Feltesszük továbbá, hogy a Howard-szám szórása független a várható értéktől, s ennek folytán a szórás állandónak és a tapasztalati adatokból ismertnek tekintjük. Ez a feltételezés csak bizonyos határok között nem jelent nagy eltérést a valóságtól. Ha a vizs-

gálatra kerülő tételek Howard-számának várható értékei között nagy eltérések vannak, akkor az állandó szórással való számolás lényeges hibát okozhat. Ezt a hibát az [1] 7.3. pontja szerint végzett homoscedasztikus transzformációval küszöbölhetjük ki. A transzformációt a megbízó intézet részletes vizsgálataiból nyert adatok alapján az alább példaként ismertetett konkrét esetben elvégeztük.

A probléma megoldására lényegében három módszer kínálkozik.

1. Az első tárgyalásmódnál még azt a további feltételezést tesszük, hogy az egyes tételek minősége egymástól független, s így az ellenőrző vizsgálat alapján történő döntés alapja csak a konkrét mintára vonatkozó vizsgálat lehet.

Ezen feltevések alapján a probléma a következő feladat megoldására vezethető vissza:

Legyen adva egy normális eloszlású ξ valószínűségi változó ismeretlen μ várható értékkel és ismert σ szórással. Ezen változó értékére nézve két megfigyelést végzünk, és az első ismeretében akarunk a másodikra következtetni.

Pontosabban a következő feladattal állunk szemben:

Azt kívánjuk, hogy bármekkora is a μ értéke, legfeljebb $1-\beta$ legyen annak együttes valószínűsége, hogy ξ az első esetben ne haladjon meg egy korlátot — jelöljük ezt x -szel, a második megfigyelésnél viszont egy másik korlát, t fölé essék; β a biztonsági együttható. t és β adottak; keressük x azon értékét, amely ennek a feltételnek megfelel. Könnyű belátni, hogy e feltételek teljesítése szempontjából legkedvezőtlenebb az az eset, amelynél μ

éppen x és t számtani közepe. Jelölje $\Phi\left(\frac{u-\mu}{\sigma}\right)$ a ξ eloszlásfüggvényét — $\Phi(u)$ a standard normális eloszlásfüggvény —; akkor a fentiek szerint a

$$(1) \quad \Phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\mu-t}{\sigma}\right)$$

kifejezést kell maximalizálni μ szerint. Deriválás után kapjuk az (1)-et maximálissá tevő μ_0 -ra a következő egyenletet:

$$(2) \quad Z\left(\frac{x-\mu_0}{\sigma}\right) = Z\left(\frac{\mu_0-t}{\sigma}\right)$$

ahol $Z(u) = \frac{\varphi(u)}{\Phi(u)}$ az ún. Mill-féle hányados, a standard normális sűrűség- és eloszlásfüggvény hányadosa. Mivel $Z(u)$ monoton csökkenő, (2)-ből következik az állítás: $\mu_0 = \frac{x+t}{2}$.

(1)-nek μ ezen legkedvezőtlenebb értéke mellett is $(1-\beta)$ -val kell egyenlőnek lennie, azaz

$$\Phi\left(\frac{x-\frac{t+x}{2}}{\sigma}\right) = \sqrt{1-\beta},$$

s így a keresett korlátra a következő kifejezést kapjuk:

$$(3) \quad x = t + 2\sigma \Phi^{-1}(\sqrt{1-\beta}).$$

Ha a hazai ellenőrzés során nem egy, hanem több, mondjuk n darab vizsgálatot végeznek, akkor a talált n darab Howard-szám átlagára a korlátot a

$$(4) \quad \sqrt{n} Z\left(\frac{x-\mu}{\sigma} \sqrt{n}\right) = Z\left(\frac{\mu-t}{\sigma}\right)$$

$$\Phi\left(\frac{x-\mu}{\sigma} \sqrt{n}\right) \Phi\left(\frac{\mu-t}{\sigma}\right) = 1-\beta$$

egyenletrendszer x -re való megoldásával kapjuk meg. A (4) egyenletrendszer a Mill-féle hányados táblázata alapján minden különösebb nehézség nélkül megoldható.

Az alábbi táblázatban adjuk az előzetes vizsgálatnál alkalmazandó x korlátot $n=1, 2, 3, 4$ -re $t=40$ és $\sigma=7,43$ mellett és néhány valószínűségi szintre. (A szórás a megbízó intézet által végzett nagyszámú vizsgálat révén kapott empirikus adat.)

$\beta \backslash n$	0,85	0,90	0,95	0,99
1	36	33	29	21
2	36	33,5	30	23
3	36	33,67	30,33	24
4	35,75	33,5	30,25	24,25

A táblázatból látható, hogy az előzetes vizsgálatok számát növelve a kapott korlát eleinte egy kicsit növekszik, majd lassan csökken. Ez természetesen nem jelenti azt, hogy ha van rá lehetőség, nem érdemes minél több előzetes vizsgálatot végezni, hiszen a feltüntetett korlát a kapott Howard-számok átlagára vonatkozik.

Ha a transzformált értékekkel számítjuk ki az előzetes minőségellenőrzési határokat, akkor azt tapasztaljuk, hogy a 85 és 90 %-os szintek esetén a transzformáció nem változtatja meg a határokat. A magasabb 95, illetve 99 %-os szintek esetén a transzformáció általában emeli a határokat. Az utóbbi két szintre a transzformációval nyert határokat az alábbi táblázat adja:

$\beta \backslash n$	0,95	0,99
1	29	23
2	30	25
3	30,67	25,33
4	30,75	25,5

2. Az előző pontban úgy kerestük az előzetes vizsgálatnál alkalmazandó korlátot, hogy a másodfajú hiba, $1-\beta$, előre rögzítve volt. A valószínűségi szint megválasztása azonban önkényes. Ez az önkényesség elkerülhető, ha nem rögzítjük előre a valószínűségi szintet, hanem gazdaságossági számítások alapján állapítjuk meg.

Legyen az a kár, amit az a helytelen döntés okoz, hogy az előzetes vizsgálat során visszautasítottuk a tételt, bár az a második vizsgálat alkalmával átment volna, w_1 , az a kár pedig, amit az okoz, hogy elfogadtunk egy olyan tételt, amelyiket a második vizsgálat során visszautasítanak, w_2 (az oda-vissza szállítás költsége, esetleg még más okozott költség). Ekkor a helytelen döntés által okozott várható kár

$$(5) \quad w_1 \Phi\left(\frac{\mu-x}{\sigma}\right) \Phi\left(\frac{t-\mu}{\sigma}\right) + w_2 \Phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\mu-t}{\sigma}\right).$$

x -et úgy akarjuk megválasztani, hogy ez a kár μ legkedvezőtlenebb értéke mellett a lehető legkisebb legyen. Vagyis az (5) kifejezés μ szerinti maximumának x szerinti minimum-helyét kívánjuk meghatározni.

Az első részben láttuk, hogy (5) első tagjának μ szerinti maximuma a $\mu = \frac{x+t}{2}$ helyen van. Hasonló módon belátható, hogy a második tagban szereplő kifejezés is μ -ben ugyanezen a helyen veszi fel a maximumát. Így azt kapjuk, hogy (5) μ szerinti maximuma a $\mu = \frac{x+t}{2}$ helyen van. Tehát a

$$(6) \quad w_1 \left[\Phi\left(\frac{t-x}{2\sigma}\right) \right]^2 + w_2 \left[\Phi\left(\frac{x-t}{2\sigma}\right) \right]^2$$

kifejezést kell x -ben minimálissá tenni. Felhasználva a $\Phi(-x) = 1 - \Phi(x)$ azonosságot, (6) a következő alakot ölti:

$$(7) \quad \left[\Phi\left(\frac{x-t}{2\sigma}\right) \right]^2 (w_1 + w_2) - 2 \Phi\left(\frac{x-t}{2\sigma}\right) w_1 + w_1$$

Ebből deriválás után azt kapjuk, hogy az x szerinti minimum az

$$(8) \quad x = t + 2\sigma \Phi^{-1}\left(\frac{w_1}{w_1 + w_2}\right)$$

helyen van. Ezt a határt kell tehát alkalmaznunk az előzetes vizsgálatnál. (7)-ből látható, hogy az előzetes korlát csak akkor lesz az átvevő által megszabott t korlát alatt, ha a második tag negatív, azaz $w_2 > w_1$. A módszer alkalmazásához szükség van minden esetben w_1 és w_2 előzetes megbecslésére, de észszerű feltenni, hogy a szállítás csak növelheti a visszautasítás által okozott kárt ($w_2 > w_1$). Ennél az eljárásnál tehát $1 - \beta$ értéke (az ún. másodfajú hiba valószínűsége) nincs előre meghatározva, azt az (1) kifejezés adja, x helyébe (8)-at, μ helyébe $\frac{x+t}{2}$ -t téve. Ha n számú előzetes vizsgálat történik, akkor (5) helyett a

$$w_1 \Phi\left(\frac{\mu-x}{\sigma} \sqrt{n}\right) \Phi\left(\frac{t-\mu}{\sigma} \sqrt{n}\right) + w_2 \Phi\left(\frac{x-\mu}{\sigma} \sqrt{n}\right) \Phi\left(\frac{\mu-t}{\sigma} \sqrt{n}\right)$$

kifejezés μ szerinti maximumát kell x -ben minimalizálni.

3. Az eddig tárgyalt két módszernél feltételezzük, hogy az előzetes vizsgálat alapján történő döntés alapja csak a konkrét mintára vonatkozó vizsgálat lehet. Mivel az egyes tételek minősége egymástól nem független

— ugyanazon gyárból kikerülő tételeknél, vagy legalább is egy részüknél nagyjából azonos volt a nyersanyag, azonos a feldolgozás minősége, stb. — hatásosabb eljárás nyerhető, ha az egyes tételekre vonatkozó döntést ugyanattól a gyártól származó egyéb minták eredményétől is függővé tesszük.

Ebben a tárgyalásban tehát azt tesszük fel, hogy egy mintában talált Howard-szám olyan normális eloszlású valószínűségi változó, amelynek várható értéke tételtől tételre változva maga is valószínűségi változó, amely egy-egy tételben meghatározott értéket vesz fel. Erről a változóról feltesszük, hogy egy gyáron belül az eloszlása normális. A mintabeli szórásnál most is feltesszük, hogy független a várható értéktől és előre ismert. Vagyis feltesszük, hogy $\xi = \eta + \zeta$, ahol η és ζ függetlenek, η normális eloszlású μ várható értékkel és σ_0 szórással, ζ szintén normális eloszlású 0 és σ paraméterekkel.

Vegyünk most egy mintát az adott tételből. Ennek eredménye legyen $\xi_1 = \eta + \zeta_1$. Feladatunk ennek alapján az x korlátot úgy meghatározni, hogy a $\xi_1 \leq x$ feltétel mellett az adott β mennyiséggel legyen egyenlő annak valószínűsége, hogy a második mintában $\xi_2 = \eta + \zeta_2 \leq t$ legyen, azaz

$$(9) \quad P\{\xi_2 \leq t | \xi_1 \leq x\} = \frac{P\{\xi_1 \leq x, \xi_2 \leq t\}}{P\{\xi_1 \leq x\}} = \beta$$

álljon. Itt a számlálóban egy kétdimenziós normális eloszlásnak az (x, t) helyen felvett értéke áll. Ennek a kétváltozós normális eloszlásnak a paramétereit könnyen kiszámíthatjuk:

$$M\{\xi_1\} = M\{\xi_2\} = \mu$$

$$D^2\{\xi_1\} = D^2\{\xi_2\} = \sigma_0^2 + \sigma^2$$

és

$$r(\xi_1, \xi_2) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} = \frac{\sigma_0^2}{\sigma_1^2}, \quad \text{ahol } \sigma_1^2 = \sigma_0^2 + \sigma^2.$$

Ezen paraméterek ismeretében a kétdimenziós normális eloszlás táblázatának [2] igénybevételével feladatunk numerikusan megoldható.

Egy másik elv lehet az x kiszámításánál az, hogy a $P\{\xi_2 \leq t | \xi_1 \leq x\}$ helyett a $P\{\xi_2 \leq t | \xi_1 = x\}$ valószínűség értékét rögzítjük. Megjegyezzük, hogy ez utóbbi elv szigorúbb, mert

$$(10) \quad P\{\xi_2 \leq t | \xi_1 = x\} = \beta$$

esetén az elküldött tételeket átlagosan β -nál nagyobb arányban fogják átvenni: csak az elküldöttek közül a legrosszabbakat, az előzetes ellenőrzésen éppen átcúsóztakat veszik át β arányban. Ezen elv alapján a határ kiszámítása valamivel egyszerűbb, a megoldás a normális eloszlás inverz függvénye segítségével zárt alakban megadható. Mégpedig, mivel ξ_1 és ξ_2 kétdimenziós normális eloszlású fentebb meghatározott paraméterekkel, együttes sűrűségfüggvényük

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1^2\sqrt{\sigma_1^4 - \sigma_0^4}} e^{-\frac{1}{2(1-\sigma_0^4/\sigma_1^4)}\left(\frac{(x-\mu)^2}{\sigma_1^2} - \frac{2\sigma_0^2(x-\mu)(y-\mu)}{\sigma_1^4} + \frac{(y-\mu)^2}{\sigma_1^2}\right)} = \\ &= \frac{1}{2\pi\sigma\sigma_1^2\sqrt{\sigma_1^2 + \sigma_0^2}} e^{-\frac{\sigma_1^2}{2\sigma^2(\sigma_1^2 + \sigma_0^2)}\left((x-\mu)^2 - 2\frac{\sigma_0^2}{\sigma_1^2}(x-\mu)(y-\mu) + (y-\mu)^2\right)} \end{aligned}$$

tehát ξ_2 feltételes sűrűségfüggvénye a $\xi_1 = x$ feltétel mellett

$$f(y|x) = \frac{e^{-\frac{\sigma_1^2}{2\sigma^2(\sigma_1^2+\sigma_0^2)}\left((x-\mu)^2-2\frac{\sigma_0^2}{\sigma_1^2}(x-\mu)(y-\mu)+(y-\mu)^2\right)}}{\int_{-\infty}^{\infty} e^{-\frac{\sigma_1^2}{2\sigma^2(\sigma_1^2+\sigma_0^2)}\left((x-\mu)^2-2\frac{\sigma_0^2}{\sigma_1^2}(x-\mu)(t-\mu)+(t-\mu)^2\right)} dt} =$$

$$= \frac{1}{\sigma \sqrt{1+\sigma_0^2/\sigma_1^2}} \varphi\left(\frac{y - \frac{\sigma_0^2 x + \sigma^2 \mu}{\sigma}}{\sigma \sqrt{1+\sigma_0^2/\sigma_1^2}}\right)$$

Igy x -re (10) alapján a következő egyenletet kapjuk:

$$P\{\xi_2 \leq t | \xi_1 = x\} = \Phi\left(\frac{t - \frac{\sigma_0^2 x + \sigma^2 \mu}{\sigma}}{\sigma \sqrt{1+\sigma_0^2/\sigma_1^2}}\right) = \beta$$

Ezt x -re megoldva adódik a keresett határ:

$$(11) \quad x = \frac{\sigma_1^2}{\sigma_0^2} \left(t - \frac{\sigma^2}{\sigma_1^2} \mu - \sigma \sqrt{1 + \frac{\sigma_0^2}{\sigma_1^2}} \Phi^{-1}(\beta) \right);$$

n számú előzetes vizsgálat esetén a (10) formula — mint könnyen belátható — a következőképp módosul:

$$x = \frac{\sigma_1^2}{\sigma_0^2} \left(t - \frac{\sigma^2}{n\sigma_1^2} \mu - \sigma \sqrt{1 + \frac{\sigma_0^2}{n\sigma_1^2}} \Phi^{-1}(\beta) \right);$$

itt $\sigma_1^2 = \sigma_0^2 + \frac{\sigma^2}{n}$.

A fentiekből látható, hogy ezen számítási mód esetén az ismertnek feltételezett σ -n kívül szükség van az egyes gyárak termékeit jellemző μ és σ_0 értékekre is. Ezeket előzetes adatokból becslés útján határozhatjuk meg. Ha σ értékét pontosnak vesszük, akkor $\sigma_0^2 = \sigma_1^2 - \sigma^2$ mennyiség becslésének relatív

szórása $\frac{\sigma^2}{2n(\sigma_1^2 - \sigma^2)}$. Ha ez a mennyiség nagy, — pl. az 5–10 %-ot meghaladja —, akkor a becslés bizonytalan és meghamisítja az eredményt. Ilyenkor kevésbé bízhatunk meg a homogénebb anyagra vonatkozó kevés adatból mint a bár kevésbé homogén de több adatból származó becslésben, azért célszerű ilyen esetben σ_1 helyébe az összes — nem csak az illető gyárra vonatkozó — adatokból számított értéket írni.

A megoldásnál a fő nehézséget μ és σ_0 becslése okozza. A megbízó intézet vizsgálatai szerint ugyanis a paradicsompüré minősége nemcsak gyárról-gyárra változik, nemcsak évről-évre változik még ugyanazon gyárnál is, hanem elsősorban a beérkezett nyersanyagtól függően ugyanabban az évben is egészen különböző lehet az egymást követő időszakokban. Ehhez járul még az a kísérletek során megfigyelt tény is, hogy a talált Howard-szám függ a megfigyelő személytől is, ugyanazon tételből vett mintákban különböző megfigyelők szignifikánsan eltérő Howard-számokat figyeltek meg. Ezért

ezen paramétereket csak akkor lehet akár az előző évi adatokból, akár az abban az évben már leszállított tételek adataiból eredményesen becsülni, ha feltehető, hogy a nyersanyag minőségében, a gyártási viszonyokban nem következett be időközben lényeges változás.

IRODALOM

- [1] HALD, A: *Statistical theory with engineering applications*. Wiley, New York, 1952.
[2] PEARSON, K.: *Tables for statisticians and biometricians, II*. Cambridge University Press, Cambridge, 1931.

О СТАТИСТИЧЕСКИХ ПРОБЛЕМАХ, ВОЗНИКАЮЩИХ ПРИ МИКРО-БИОЛОГИЧЕСКОМ КОНТРОЛЕ ТОМАТНЫХ КОНЦЕНТРАТОВ

Ö. ÉLTETŐ и K. SARKADI

Резюме

Авторы занимаются с вопросами, связанными с следующей проблемой:

Некоторая партия томатного пюре только тогда принимается покупателем, если так называемое „число Howard-a“, найденное с помощью выборки, взятой из партии, не превосходит некоторое определённое число.

Спрашивается, как приходится проделать предварительный контроль качества, чтобы нехорошие партии задерживались. Предположено, что распределение „числа Howard-a“ приблизительно нормальному. Предполагается, что дисперсия известна из эмпирических данных. Ошибка, возникающая из гетероскедастичности, может быть элиминирована гомоскедастичным преобразованием.

Трактовка проблемы проделана авторами на основе следующих трёх принципов:

1. Предписано, что вероятность следующего события: при первом контроле партия принимается и при втором контроле отклоняется, не превосходит $1-\beta$. Решение даётся формулой (3).

2. Пусть w_1 означает вред, подорванный при первом контроле, если задержится одна партия, которая могла бы приниматься, и w_2 вред, подорванный в том случае, когда передаётся одна партия, которая не будет приниматься. В этом случае минимаксное решение даётся формулой (8).

3. Предполагается, что среднее значение „числа Howard-a“ одной партии тоже является случайной величиной, распределение которой в случае одной фабрики есть нормальное с параметрами μ и σ . Эта проблема может быть численно решена с помощью таблицы двухмерного нормального распределения [2] (см. формулу (9)).

ON THE STATISTICAL PROBLEMS ARISING IN MICROBIOLOGICAL EXAMINATION OF TOMATO PURÉES

Ö. ÉLTETŐ and K. SARKADI

Summary

Authors deal with the questions connected with the following problem.

A lot of tomato purée will be accepted by the consumer if the Howard-count found in the sample taken out from the lot in question doesn't exceed a fixed limit. What limit is to be used by the producer on the occasion of a previous quality control for the purpose of retaining the bad lots? It is assumed to be legitimate to approach the distribution of the Howard-count by the normal one. The variance is assumed to be known from empirical data. The error arising from the heteroscedasticity can be eliminated by a homoscedastic transformation.

The problem is discussed on the grounds of the following three principles:

1. It will be prescribed that the joint probability of accepting in the first and rejecting in the second examination would be not greater than $1-\beta$. The solution is given by formula (3).

2. Let w_1 be the loss caused by the wrong decision in the first examination if we retain a lot which would be accepted, and w_2 the loss if we dispatch a lot which will be rejected. In this case the minimax solution is given by formula (8).

3. It is assumed that the expected value of the Howard count of a lot is itself a random variable varying in the lots transported from the same producer according to $N(\mu, \sigma_i)$ normal probability law. The problem can be solved numerically by the aid of the table of the two-dimensional normal distribution [2] (see formula (9)).