

A kubai nyers grapefruitlevek minősítésének fejlesztése

I. A nyers levek minőség szerinti csoportosítása

TORRICELLA, R.,* ÖRSI, F.** és PINO, J.*

* Központi Élelmiszeripari Kutató Intézet Kuba, Havanna

** BME Biokémiai és Élelmiszertchnológiai Tanszék

Érkezett: 1984. május 16.

Kubában a forradalom győzelme után a citrusfélék termőterületét az 1959. évi 10 000 ha-ról 100 000 ha-ra növelték (és még ma is növelik), így a következő években a világ nagy citrustermelő országává válik (1).

A termelés több citrusfajt ölel fel: a narancsot, a grapefruitot, a citromot, a mandarint és másokat. Ezek közül a grapefruit a termelt mennyiséget tekintve, a narancs után, a második helyet foglalja el.

A grapefruit gyümölcs termelésének növekedésével egyre fontosabbá válik a feldolgozási technológia fejlesztése. Mivel a nyerslé minősége változik a kampány során, a légyártásnál a minőséget a különböző minőségű nyerslevek keverésével kell biztosítani. E feladat megoldásához a lé minősítése, és minőség szerinti csoportosítása szükséges.

A nyerslé minősítésére jelenleg a szárazanyag-tartalom és a savtartalom hányadosát alkalmazzák. Mivel ebben az indexben különbözőképpen változhat a szárazanyag-tartalom és savtartalom, a minősítés szempontjából sokkal teljesebb képet kaphatunk, ha mind a négy minőség mutatót (szárazanyag, sav, szárazanyag/sav és az idő) összefüggéseiben vizsgáljuk. Ilyen vizsgálatra különösen alkalmasak a többváltozós biometriai módszerek, mint pl. a *főkomponens*, a *cluster* és a *díszkri-minancia* analízis (2).

Ennek a munkának célja az „Isla de la Joventud” új típusú citrus kombinátban az 1980/81., 1981/82. és 1982/83-as kampányokban termelt nyers grapefruit-
lé vizsgálati adatainak feldolgozása, a minőség szerinti csoportosítás megoldása.

A nyerslé szárazanyag-tartalmát refraktometriás módszerrel, savtartalmát fenoftalein indikátor jelenlétében végzett titrálással határoztuk meg (3). A feldolgozás szempontjából az egyedek (minták) az előbbi adatok napi átlagait jelentik. A kampányi időszakát úgy vettük figyelembe, hogy minden egyes mintához egy idővel arányos számot rendelünk. Ezt a számot az 1. táblázatban definiáltuk.

1. táblázat

A kampány időszakának a definíciója

| Kampány időszak | | T értéke |
|-------------------|------------------------|----------|
| – | szeptember 15-ig | 1 |
| szeptember 16-tól | október 15-ig | 2 |
| október 16-tól | november 15-ig | 3 |
| november 16-tól | december 15-ig | 4 |
| december 16-tól | – | 5 |

Főkomponens analízis

Ezzel a módszerrel az eredeti adat mátrixát olyan koordináta transzformációnak vetjük alá, amelynél az új koordináták egymásra merőlegesek (ortogonálisak) és az első tengely a legnagyobb változás irányába, és minden következő a maradék legnagyobb változás irányába mutat. Így érhető el, hogy pontjaikat az első két (esetleg három) főkomponens síkjában (térben) ábrázolva az eredeti struktúra változásait leíró ábrát nyerünk.

Matematikai megoldása a sajátérték számításon alapszik. Leggyakrabban a korrelációs mátrixból indulnak ki és ennek sajátértékeit a következő vektor egyenlet megoldásai adják:

$$|R - \lambda I = 0|$$

ahol:

- R : a korrelációs koeficiensok mátrixa
- λ : a sajátérték
- I : egység mátrix

Az alkalmazott számítógépes program az R mátrixból a sajátértéket a Jordan-féle eliminációs eljárással iterációs módszerrel számítja.

Minden sajátértékhez tartozik egy sajátvektor, amelyet a következő vektor egyenlet megoldásából nyerünk:

$$(R - \lambda_i I) \cdot U_i = 0$$

Ahol U_i a λ_i sajátértékhez tartozó sajátvektor. A Jordán-féle eliminációs eljárás egyidejűleg adja meg a sajátértékeket és a hozzájuk tartozó sajátvektorokat.

A főkomponenseket az eredeti változókból az alábbi vektor egyenlettel számíthatjuk ki:

$$C_i = U_i \cdot X^1$$

ahol:

- C_i : az i főkomponens
- U_i : az i sajátértékhez tartozó sajátvektor
- X^1 : a normált eredeti változó vektor
(a változóból kivonva a változók átlaga és osztva a szórással)

Az így kapott főkomponenseket ábrázoltuk és alkalmaztuk a *cluster* analízisben.

Cluster analízis

A cluster analízis az adatok csoportosításával foglalkozik az egyedek hasonlósága alapján.

A módszer alkalmazásánál a csoportosítandó egyedeket, mint vektorokat tekintjük és a vektorok hasonlóságát euklédesi távolságukból a következőképpen számítjuk ki:

$$D_{12}^2 = \sum_{i=1}^n (x_{1i} - x_{2i})^2$$

$$S_{12} \% = 100 \cdot (D_{\max} - D_{12}) / D_{\max}$$

ahol:

X_{1i} : az X_1 vektor i -ik koordinátája

X_{2i} : az X_2 vektor i -ik koordinátája

D_{12} : a két vektor távolsága

D_{\max} : az összes lehetséges távolságok közül a legnagyobb

n : a vektorok koordinátáinak száma

S_{12} : a két vektor hasonlósági indexe.

Az a két vektor hasonló a legjobban, amelyek távolsága az összes lehetséges vektorpár távolságából a legkisebb. A számítás a következő lépésekből áll:

- Az eredeti adatokból kiszámított főkomponenseket tekinti a vektorok koordinátáinak, ezekből meghatározza az összes lehetséges távolságokat $[k \cdot (k-1)/2]$, és kiválasztjuk a legkisebbhez tartozó két vektort. Ezeket egyesítjük, vagyis helyettesítjük egy olyan vektorral, amelynek koordinátái az eredeti koordináták számtani átlaga. Az egyesítés súlyozottan történik, amely figyelembe veszi, hogy előzőleg már hány vektort egyesítettünk a kiválasztott vektorba. Ez megakadályozza, hogy egy eltérő vektor „elhúzza” a már egyesített átlag vektort.
- Annak eldöntésére, hogy elértünk-e a számítás befejezéséhez, vizsgáljuk meg az utolsó minimális távolságot. Ha ez túllépte az átlagos távolság 50%-át, akkor a számítást befejezettnek tekintjük, ellenkező esetben ismétlődik az előző lépés.
- Az így kapott adatpár-hasonlósági adatok dendogram, vagy Venn diagram formájában ábrázolhatók.
- Az ily módon kapott csoportok a *diszkriminancia* analízissel vizsgálhatók, hogy valóban elkülönülő csoportokról, vagy egymást átfedő azonos eloszláshoz tartozó adatokról van szó.

Diszkriminancia analízis

A *diszkriminancia* analízis olyan számítási módszer, amelynek a segítségével a többváltozós egyedek csoportba osztása, illetve a csoportok elkülönítésére szükséges súlyozó faktorok megállapítása történhet meg. Mivel az eredeti adatok eloszlása eltért a normálistól és a szórások különbözőek voltak; a Fisher által kidolgozott lineáris diszkriminancia számítás nem alkalmazható.

Ezért alkalmaztuk a Welsh által kidolgozott nem lineáris eljárást, amely a P paraméterrel jellemzett egyedek sűrűség függvényére állít fel becslést, és valamely terméket abba a csoportba sorol, amelynek sűrűség függvényébe helyettesítve a legnagyobb függvény értékeket kapjuk.

A számítások kiinduló pontja az egyesített kovariancia mátrix, melyet a következő képlet alapján számítunk:

$$D_{m, mi} = \frac{\sum_{k=1}^k \sum_{i=1}^{n \cdot k} (x_{kmi} - \bar{x}_{km}) \cdot (x_{kmi} - \bar{x}_{km})}{\sum_{k=1}^k nk - 1}$$

ahol:

X_{kmi} : az m paraméter értéke a k -ik csoportban az i -ik egyednél.

X_{km} : az m paraméter átlaga a k -ik csoportban.

k : csoportok száma

$n \cdot k$: egyedek száma a k csoportban.

A sűrűség függvényt a következő egyszerű függvény formájában állítjuk elő:

$$f_{k(x)} = \text{EXP} \left(B_{K0} - \sum_{m=1}^m B_{km} \cdot x_m \right)$$

ahol:

B_{k0} és B_{km} : a diszkriminancia egyenlet együtthatói a k -ik csoportban.

$f_{k(x)}$: a sűrűség függvény a k -ik csoportban.

x_m : az egyed paraméterei

m : a paraméterek száma.

A B_{k0} és B_{km} értékek a következő egyenletekkel számíthatók ki:

$$B_{k0} = -1/2 \cdot \sum_{m=1}^m \sum_{ml=1}^m D(m, ml) \bar{x}_{km} \cdot \bar{x}_{kml},$$

$$B_{km} = \sum_{m=1}^m D(m, ml) \bar{x}_{km}.$$

A döntésnél az adott egyed paramétereit minden csoport sűrűség függvényébe behelyettesítjük és az egyedet abba a csoportba soroljuk, amelyben a legnagyobb $f_{(x)}$ értéket kapjuk.

A szétválaszthatóságot χ^2 próbával ellenőrizhetjük.

A χ^2 a következő összefüggéssel számítható:

$$\chi^2 = \sum_{m=1}^m \sum_{ml=1}^m D(m, ml) \cdot n \cdot k (\bar{x}_{km} - \bar{x}_m) \cdot (\bar{x}_{kml} - \bar{x}_{ml})$$

a szabadsági fok $(k-1) \cdot m$.

Az összefüggést alkalmazva azt is megadhatjuk, hogy az egyedek milyen valószínűséggel sorolhatók az adott alapsokaságba, mely az $f_{(x)}$ értékből a következő összefüggéssel számítható:

$$P = 1 \sum_{k=1}^k f_{k(x)} \Bigg/ \text{MAX } f_{k(x)},$$

ahol

P : annak a valószínűsége, hogy az egyed a kiválasztott osztályba tartozik.

$f_{k(x)}$: az egyed sűrűség függvénye a k csoportban.

$\text{MAX } f_{k(x)}$: a legnagyobb sűrűség függvény értéke.

A szükséges számításokat a BME Biokémiai és Élelmiszertechnológiai Tanszék által kidolgozott T. B.* programcsomag felhasználásával az egyetemi számítógépes központ R-32-es számítógépén végeztük.

* Többváltozós Biometriai Programcsomag.

Eredmények és kiértékelés

Az 1. ábra a nyerslé szárazanyagtartalmát és összes titrálható savtartalmát mutatja az idő függvényében az 1980/81, 1981/82 és 1982/83-as kampányok során. A $T = 0$ az augusztus 15-ét jelenti az ábrán.

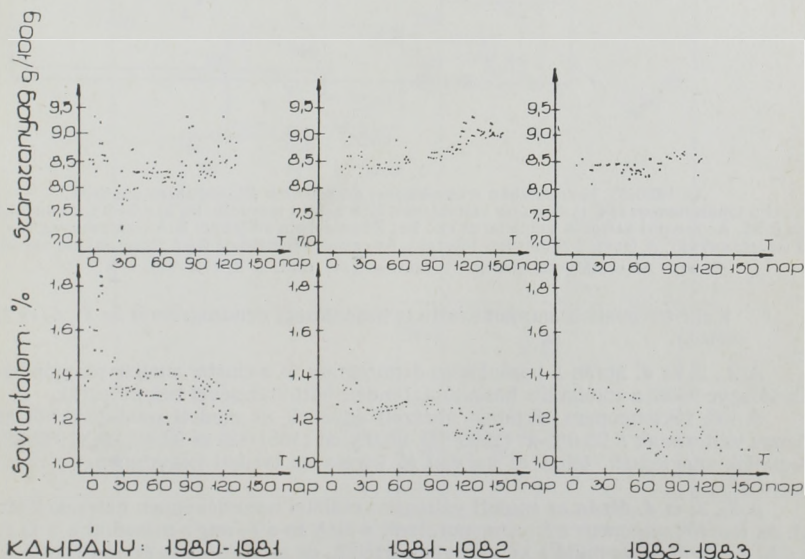
Mind a három kampányban a szárazanyagtartalom növekedő, a savtartalom pedig csökkenő tendenciát mutatott. A nyerslé-minőség szempontjából a nagy szárazanyagtartalom és az alacsony savtartalom kívánatos (bizonyos határokon belül). Gyakorlati szempontból a nyerslevet három minőségi osztályba szokták sorolni: az I. osztályba a nagy szárazanyag- és kis savtartalom, a II. osztályba a közepes szárazanyag- és savtartalom jellemző, míg a III. osztályba a kis szárazanyag- és nagy savtartalmú nyersleveket sorolják.

Az osztályozásban a szárazanyag/savtartalom (B/A) indexet alkalmazzák. A jobb csoportosítás érdekében nemcsak a B/A indexet, hanem a szárazanyag- (B), a savtartalom (A) és a kampány időszakát (T) vettük figyelembe együttesen.

A négy indexes (dimenziós) egyedek (vektorok) síkbeli ábrázolását főkomponens analízis segítségével, a csoportosítását hasonlóságuk alapján cluster analízissel végeztük; végül a csoportosítás felülvizsgálatát és a súlyozó faktorok meghatározását diszkriminancia analízissel hajtottuk végre.

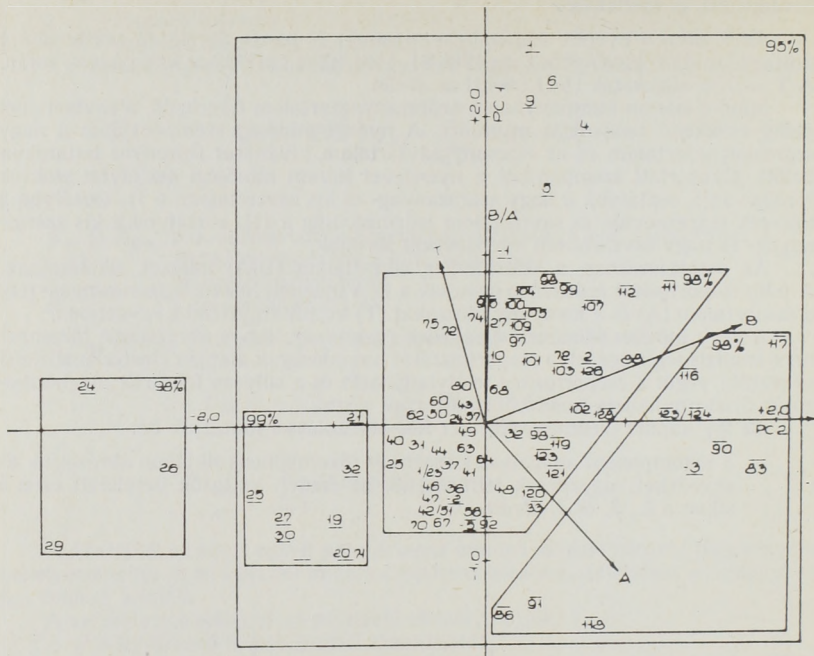
Az így kapott eredményeket két féle ábrázolással mutatjuk be:

- a főkomponens analízissel kapott két főkomponens síkjában ábrázoltuk az egyedeket, ugyancsak feltüntettük az eredeti változók vetületeit ezen a síkon a 2., 3. és 4. ábrán,



1. ábra

A kubai nyers grapefruitlé szárazanyag (Brix) és összes titrálható savtartalma (%) a termelési idő függvényében (T napokban) az 1980/81, 1981/82 és 1982/83-as kampányok során. Az ábrában a $T = 0$ az augusztus 15-ét jelenti



1980 - 1981

2. ábra

Az 1980/81. kampányban termelt nyers grapefruitlé főkomponens analízise.

Az első főkomponens (PC 1) az összes változásnak 62,8%-át, a második (PC 2) a 24,9%-át foglalja magába. Az eredeti változók vetületei az első két főkomponens síkjában: B/A (szárazanyag/sav), B (szárazanyag), A (sav), T (kampány időszak). Az egyedek kódszáma (k) a termelési idő napokban. Az osztályozás jelölése: I. oszt. = k, II. oszt. = k, III. oszt. = k.

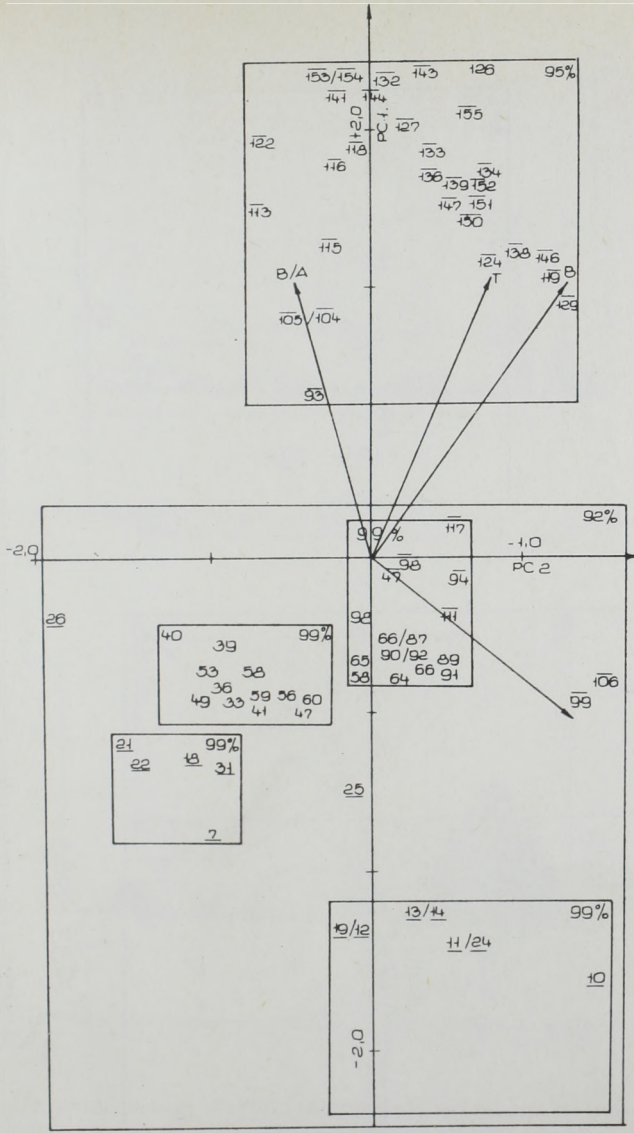
– a cluster analízis alapján készített hasonlósági dendrogramot az 5., 6. és 7. ábrán.

A 2., 3. és 4. ábrán a hasonlósági dendrogram és a cluster analízis alapján hasonló egyedeket a minimális hasonlósági index feltüntetésével bekereteztük.

A két főkomponens síkjában ábrázolt egyedek az eredeti rendszerekből az összes változónak a 96,0%-át (1980/81), 96,0%-át (1981/82) és 92,4%-át (1982/83) foglalják magukban, tehát ez nagyon jó leírása az eredeti négydimenziós rendszernek.

A 2., 3. és 4. ábrán az eredeti változók vetületei hasonlóképpen helyezkednek el. Az első főkomponens irányába mutatnak a B/A és a T, míg a másodikba a B és az A. Az elválasztás mind a két irányba történik, de az 1980/81-es és az 1982/83-as kampánynál a második, míg 1981/82-ben az első főkomponens irányába jobb az elválasztás. Ezenkívül az elkülönítés az 1981/82-es kampányban a legélesebb.

A hasonlósági dendrogramban (5., 6., 7. ábra) a diszkriminancia analízissel megerősített csoportokat fekete színnel bejelöltük. Egyértelműen látszik az, hogy

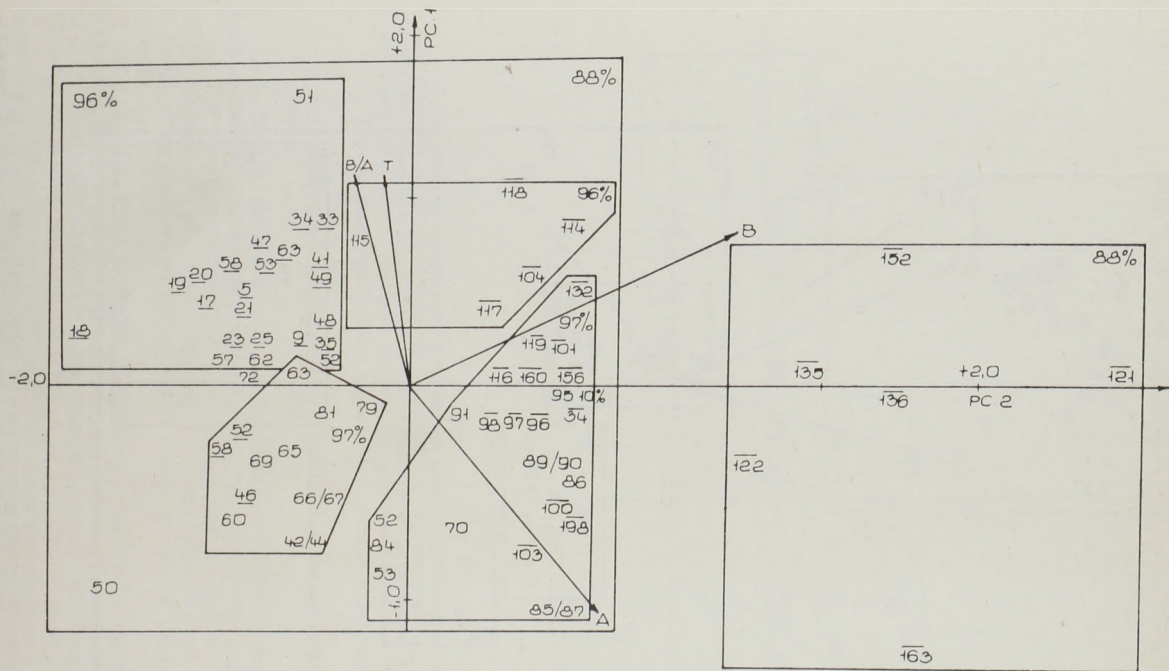


1981 - 1982

3. ábra

Az 1981/82-es kampányban termelt nyers grapefruitlé főkomponens analízise. Az első főkomponens (PC 1) az összes változásnak a 84,5%-át, a második (PC 2) 11,5%-át foglalja magába.

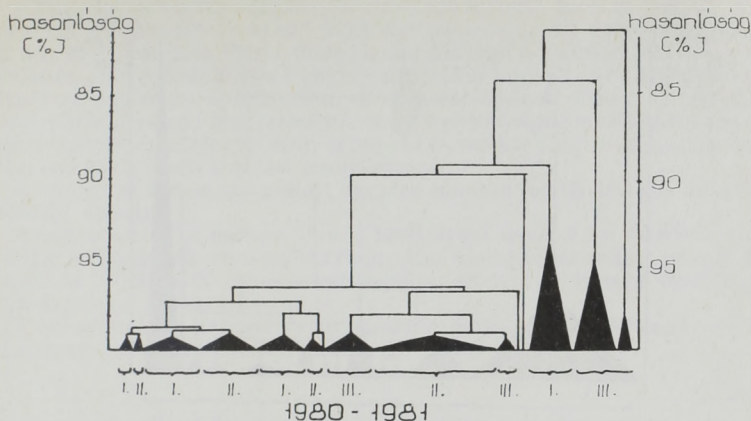
A többi jelölések azonosak a 2. ábrán alkalmazottakkal



1982 - 1983

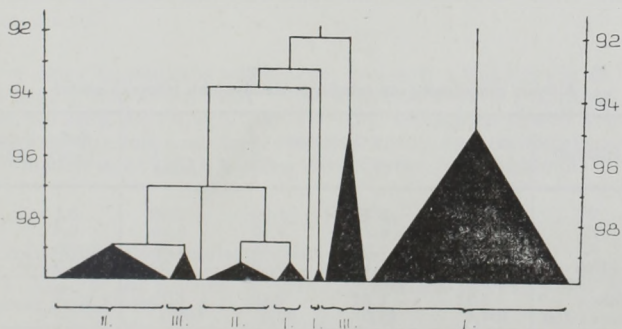
4. ábra

Az 1982/83-as kampányban termelt nyers grapefruitlevek főkomponens analízise. Az első főkomponens (PC 1) az összes változásnak a 76,5%-át, a második (PC 2) a 15,9%-át foglalja magába. A többi jelölések azonosak a 2. ábrán alkalmazottakkal



5. ábra

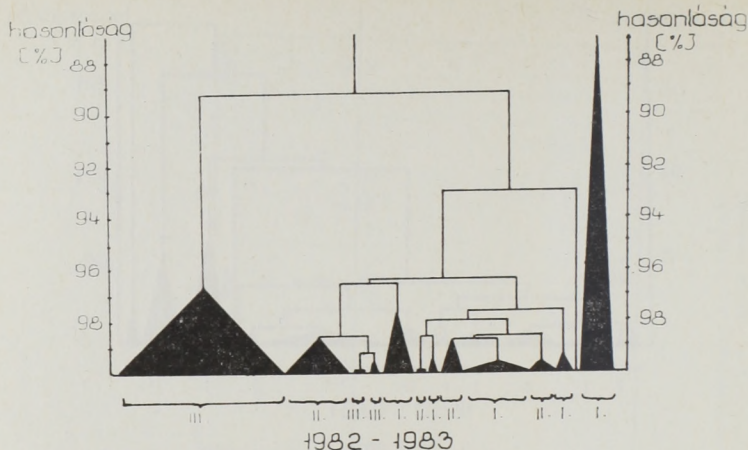
Az 1980/81-es kampányban termelt grapefruitlevek cluster analízise. Az ordinátában az egyedek kódszáma (k). A koordinátában a hasonlósági index (%)



6. ábra

Az 1981/82-es kampányban termelt grapefruitlevek cluster analízise. Az ordinátában az egyedek kódszáma (k). A koordinátában a hasonlósági index (%)

az I., ill. III. osztályok nagyobb része jól elkülönült, miközben a II. osztályba sorolt egyedek többé-kevésbé elkülönültek az előző osztályoktól. Az itt tapasztalható különbségek abból eredhetnek, hogy a diszkriminancia analízis az eredeti adatokkal dolgozik, míg a főkomponens és a cluster analízis a standardizált adatokkal. Ezért az állítható, hogy a csoportosítás megfelelőnek bizonyult. Ezenkívül mind a három kampány esetében a diszkriminancia analízissel végzett csoportosítások szignifikánsak voltak ($p = 0,01$).



7. ábra

Az 1982/83-as kampányban termelt grapefruitlevelek cluster analízise. Az ordinátában az egyedek kódszáma (k). A koordinátában a hasonlósági index (%)

2. táblázat

A nyers grapefruitlevelek minőségi osztályainak jellemző adatai

| Kampany | Jellemző | Osztályok | | | | Súlyozó faktorok | |
|---------|------------|-----------|-------|------|-------|------------------|-----------|
| | | I. | II. | III. | Átlag | Index | Index-100 |
| 1980/81 | T | 3,97 | 2,39 | 1,00 | 2,55 | 0,1 | 1,49 |
| | B | 8,57 | 8,27 | 8,19 | 8,35 | 0,4 | 5,96 |
| | A | 1,27 | 1,39 | 1,48 | 1,35 | 0,0 | 0,00 |
| | B/A | 6,77 | 6,19 | 5,60 | 6,22 | 0,5 | 7,45 |
| | Index | 7,21 | 6,67 | 6,18 | 6,70 | — | — |
| | Index-100 | 107,4 | 99,0 | 92,0 | 100,0 | — | — |
| 1981/82 | T | 4,59 | 2,38 | 1,00 | 3,18 | 0,2 | 3,00 |
| | B | 8,94 | 8,50 | 8,50 | 8,71 | 0,2 | 3,00 |
| | A | 1,17 | 1,25 | 1,30 | 1,22 | 0,0 | 0,00 |
| | B/A | 7,64 | 6,79 | 6,59 | 7,14 | 0,6 | 9,00 |
| | Index | 7,29 | 7,26 | 5,82 | 6,66 | — | — |
| | Index-1000 | 109,3 | 93,8 | 87,3 | 100,0 | — | — |
| 1982/83 | T | 4,27 | 2,86 | 1,67 | 2,87 | 0,1 | 1,40 |
| | B | 8,77 | 8,50 | 8,42 | 8,56 | 0,2 | 2,80 |
| | A | 1,11 | 1,11 | 1,29 | 1,18 | 0,0 | 0,00 |
| | B/A | 7,95 | 7,70 | 6,55 | 7,34 | 0,7 | 9,81 |
| | Index | 7,75 | 7,38 | 6,43 | 7,14 | — | — |
| | Index-100 | 108,5 | 103,3 | 90,2 | 100,0 | — | — |

A kampányok közötti különbségek abból adódhatnak, hogy az 1980/81. kampány a citrus kombinátnak az első teljes kampánya volt. Itt a nyersanyagellátás még nem stabilizálódott. Már a második kampányban a nyersléellátás jelentősen állandósult, amit a jobb minőség szerinti elválasztás bizonyít. Az 1981/82-es kampánynál az elválasztás kissé romlott, de még mindig jobb maradt az 1980/81-es kampányénál. Ez azzal magyarázható, hogy nyersanyaghiány miatt a kampány végén a nyerslé termelése csökkent, éppen akkor, amikor a minőség gyorsan emelkedne, ami a jobb szétválasztást eredményezné.

Az előbbi módon csoportosított egyedek minőségi osztályonkénti átlagait a 2. táblázat mutatja.

A megszokott B/A indexen kívül a táblázatban szerepel két összetett index. Az INDEX figyelembe veszi a diszkriminancia analízissel számított súlyozó faktorokat. Az INDEX-100 olyan összetett index, amelyet az átlaggal normáltunk, vagyis az átlagot 100-ra állítottuk be.

A három indexet kétszemponos variancia analízissel vizsgáltuk. Az így kapott eredmények a 3. táblázatban találhatóak.

3. táblázat

■ A nyers grapefruitlé minősítő indexek összehasonlítása a variancia analízis alapján

| Variancia forrása | Számított F értéke | | Index - 100 |
|--------------------------|--------------------|--------|-------------|
| | B/A | Index | |
| Kampányok | 167,7* | 113,0* | 15,0* |
| Minőségi osztályok | 165,4* | 321,6* | 325,3* |

* Szignifikáns $p = 0,01$

A B/A index szignifikáns különbséget mutatott a kampányok és a minőségi osztályok között, míg az összetett INDEX jobban kiemeli a minőségi osztályok közötti különbséget. Az INDEX-100 értékében lényegesen kisebb különbség van a kampányok között, míg a minőségi osztályok között a különbség kiemelten jelentkezett. Ily módon ez az INDEX-100 a minősítésben felhasználható.

I R O D A L O M

- (1) KGST XXXV. ülése, (1980) Szófia, Bulgária.
- (2) Sváb, J.: Többváltozós módszerek a biometriában. Mezőgazdasági Kiadó, Budapest, 1979.
- (3) Nagy, S., Shaw, P. and Veldhuis, M.: Citrus Sciences and Technology. The AVI Pub. Co, Connecticut, USA, (1977) pp. 497 - 504.

A KUBAI NYERS GRAPEFRUITLEVEK MINŐSÍTÉSÉNEK FEJLESZTÉSE I. A NYERSLEVEK MINŐSÉG SZERINTI CSOPORTOSÍTÁSA

R. Torricella, F. Örsi és J. Pino

A kubai grapefruitlé előállítására alkalmazott nyerslevet vizsgáltuk minősítés céljából az 1980/81, 1981/82 és 1982/83-as kampányok során. A nyerslevet a szárazanyag- és a savtartalommal, a szárazanyag/savtartalom-mal és a kampány időszakkal jellemeztük. Az adatok feldolgozását főkomponens, cluster és diszkriminancia analízissel végeztük, így három minőségi osztályba soroltuk a nyerslevet és egy olyan összetett indexet definiáltunk, amely nagy segítséget nyújt a kubai grapefruitlé minőségének a biztosításában.

СОВЕРШЕНСТВОВАНИЕ КАЧЕСТВА КУБИНСКИХ
СОКОВ-СЫРЦОВ ИЗ ГРЕЙПФРУТОВ
I. ГРУППИРОВКА СОКОВ ПО КАЧЕСТВУ

Р. Торичелла, Ф. Ёрши и Я. Пино

Авторы, с целью оценки качества, провели анализы соков-сырцов, применяемых в период компаний 1980/81, 1981/82 и 1982/83 гг. в производстве напитков из грейпфрутов.

Соки-сырцы характеризовались содержанием сухих веществ и содержанием кислоты, содержание сухих веществ/содержание кислоты и периодом компаний.

Обработка результатов испытаний проводилась методом анализа главного компонента, кластера и дискриминации.

Таким образом, соки-сырцы причислялись к трем классам качества и был дефинирован такой составной индекс, который окажет большую помощь в обеспечении качества напитков из кубинских грейпфрутов.

IMPROVEMENT OF THE EVALUATION OF CUBAN RAW GRAPE-FRUIT
JUICE QUALITY

I. CLASSIFICATION OF RAW JUICES ACCORDING TO THEIR QUALITY

R. Torricella, F. Örsi and J. Pino

The raw juice used for the preparation of Cuban grape-fruit juice was examined for the purpose of qualification during the campaigns in 1980/81, 1981/82 and 1982/83. The raw juice was characterised by the dry material and acid contents, by the dry material/acid content rate and by the campaign period. The data were processed by principal component, cluster and discriminant analysis, and in this way the raw juice was classified into three quality groups and a complex quality index was defined, which promotes the assurance of the quality of Cuban grape-fruit juice.

ENTWICKLUNG DER BEWERTUNG DER QUALITÄT VON KUBANISCHEN
GRAPEFRUITROHSÄFTEN

I. KLASSIFIZIERUNG DER ROHSÄFTE NACH IHRER QUALITÄT

Torricella, R., F. Örsi und J. Pino

Während der Kampagnen 1980/81, 1981/82 und 1982/83 wurde der zur Herstellung von kubanischen Grapefruit-säften verwendeten Rohsaft untersucht und die Qualität beurteilt. Der Rohsaft wurde mittels Gehalt an Trockensubstanz und Säure sowie anhand des Verhältnisses Trockensubstanz/Säuregehalt charakterisiert und die Kampagneperiode angegeben. Die Daten wurden mit des Hauptkomponenten-, Cluster- und Diskriminanzanalyse ausgewertet, wobei der Rohsaft in drei Qualitätsklassen eingeordnet wurde. So konnte eine solche komplexe Kennziffer definiert werden, die zur Sicherung der Qualität des kubanischen Grapefruitsaftes mit Erfolg genutzt werden kann.