

**ПРИЛОЖЕНИЕ НА КЛЪСТЕР АНАЛИЗА ЗА ГРУПИРАНЕ НА НОВИ
ГЕНОТИПОВЕ ТЮТЮН ВИРЖИНИЯ СПОРЕД НЯКОИ ХИМИЧНИ ПОКАЗАТЕЛИ**

Нели Керанова^{1*}, Марина Друмева-Йончева², Йонко Йончев²

¹*Аграрен университет-Пловдив, Факултет по икономика,
4000, България, Пловдив, бул. „Менделеев“, № 12*

²*Институт по тютюна и тютюневите изделия-Марково,
4108, България, Марково*

**nelikeranova@abv.bg*

**APPLICATION OF CLUSTER ANALYSIS FOR GROUPING NEW GENOTYPES
VIRGINIA TOBACCO ACCORDING TO SOME CHEMICAL INDICATORS**

**Neli Todorova Keranova^{1*}, Marina Drumeva Drumeva-Yoncheva² and Yonko Georgiev
Yonchev²**

¹*Agricultural University-Plovdiv, Faculty of Economics,
4000, Bulgaria, Plovdiv, 12 "Mendeleev" Blvd.*

²*Tobacco and Tobacco products Institute-Markovo, 4108, Bulgaria, Markovo*

**Corresponding author: nelikeranova@abv.bg*

Abstract

The study was carried out during 2013-2015 at the Tobacco and Tobacco Products Institute, Markovo. The object of analysis in the present work were newly selected hybrid combinations of Virginia tobacco: H 27, H 33, H 51 and V 0514 Ø. Changes in chemical indicators of genotypes were reported - nicotine, sugars, total nitrogen. The aim of the study is to achieve grouping of mentioned hybrids according to similarity to the given chemical composition. The hybrid combinations of Virginia tobacco are classified and grouped through hierarchical cluster analysis based on the chemical indicators. It was found that in 2013 studied genotypes were grouped into three clusters. There is a similarity between the B 0514 Ø, Hybrid 33 and Hybrid 51. The most remote is Hybrid 27. For 2014, three clusters have been obtained. Nearest to B 0514 Ø is Hybrid 51, and the most distant - again Hybrid 27. This year it is with the most balanced chemical composition. In 2015, similarity to B 0514 Ø was established with Hybrid 33 and greatest difference with Hybrid 51. Hybrid 51 and Hybrid 27 are with most balanced chemical composition in this year of research.

Keywords: Virginia tobacco, chemical components, cluster analysis

Въведение

Тютюнът (*Nicotiana tabacum* L.) е важна техническа култура за икономиката на много страни [1,5,14], включително и за България. Тютюн Виржиния (Flue-cured) е един от основните търговско-технически типове тютюни. Той се отличава с най-високо съдържание на разтворими въглехидрати (12-28%) [3] и представлява суровина, служеща за вкусова основа при съставяне на композиционни тютюневи смеси [7].

Качеството на тютюневата суровина зависи от химичния ѝ състав [6], което обуславя проучването му. Редица автори изучават и анализират химичните показатели при тютюна [4,7,9,15,23].

Съдържанието на основни химични компоненти в тютюна е специфично и зависи от сортовия състав, почвено-климатичните условия на района и прилаганата агротехника на отглеждане [8].

Терминът „кълъстерен анализ“ е въведен за първи път от Труон през 1939 г. При този метод се постига разделяне на дадено множество от обекти на непресичащи се еднородни множества, наречени кълъстери, така че всеки кълъстер се състои от подобни обекти, а обектите от различните кълъстери съществено се различават. Съществува два вида кълъстерен анализ: йерархичен и нейерархичен. При йерархичната кълъстеризация в зависимост от избора на метод, последователно се обединяват малки кълъстери в големи или се разделят големи кълъстери на по-малки. Този метод не изисква предварително да се знае броят на кълъстерите. Нейерархичната кълъстеризация се основава на итеративни методи за разделяне на началната съвкупност от данни. Тук се изисква предварително определяне на броя на кълъстерите.

Много учени са приложили възможностите, които предоставя кълъстерният анализ, за да изследват различни сортове тютюн [11,16,21,24,25]. В [12] са анализирани 48 хибридни линии тютюн и са класифицирани въз основа на осем химични показателя чрез кълъстерен анализ. Разликите между елементите от отделните кълъстери са оценени чрез Т-тест. Резултатите показват, че пробите могат да бъдат класифицирани в четири групи, като съществуват значителни разлики между тях по повечето от признаците.

Feng, [17] използва кълъстер анализ за сравнение на химичните показатели и качеството на flue-cured тютюна.

Kalivas [20] прилага кълъстерен анализ на 34 различни сорта ориенталски тютюн. Те са класифицирани в три основни кълъстера. На базата на това проучване прави заключение, че има основание изследваните сортове да бъдат използвани в селекционни програми.

Целта на настоящата работа е да се направи групиране на новоселекционирани хибридни комбинации тютюн: Хибрид 27, Хибрид 51, Хибрид 33 и стандарт В 0514 на базата на някои химични показатели (процентно съдържание на никотин, захари, общ азот), както и да се открият генотипове с най-благоприятен химичен състав.

Материали и методи

Обект на изследване в настоящата работа са четири хибридни форми тютюн тип Виржиния: Хибрид 27, Хибрид 51, Хибрид 33 и стандарт В 0514. В Института по тютюна и тютюневите изделия, Марково за периода 2013-2015 г. са направени изследвания, свързани с химичните показатели - никотин, захари и общ азот (%). Експерименталните данни са получени чрез автоматичен анализатор в поток AutoAnalyzer II C, Technicon с методи, съответно ISO 15152, ISO 15154, БДС 15836.

За постигането на поставената в началото на изследването цел е приложен йерархичен кълъстерен анализ по метода на междугруповото свързване и мярка за сходство квадратичното евклидово разстояние. При този метод разстоянието между два кълъстера А и В и се дефинира като средната стойност на $n_A \cdot n_B$ на брой разстояния между n_A точки от А и n_B точки от В чрез формулата:

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(x_i, x_j) \quad (1)$$

където сумата се изменя по всички x_i и x_j от А и В. Чрез

$$d(x_i, x_j) = \sum_{m=1}^p (x_{im} - x_{jm})^2, \quad i, j = \overline{1, n} \quad (2)$$

означаваме квадратичното евклидовото разстояние между два вектора $x_i(x_{i1}, x_{i2}, \dots, x_{ip})$, $x_j(x_{j1}, x_{j2}, \dots, x_{jp})$.

С цел по-комплексно изследване в сравнителен аспект на разглежданите показатели, е приложен факторен анализ. Чрез него множество от корелиращи признаци се редуцира до друго с некорелиращи, но обясняващи в най-висока степен общата дисперсия на изходните

данни. Корелиращите променливи се обединяват в общ фактор, а некорелиращите – в различни [18,19]. Съвместното прилагане на двата анализа дава възможност за получаване на по-пълна информация за значението на признаците в групирането на хибридите линии в клъстерите. Методът, чрез който се извличат факторите, е методът на главните компоненти (PCA – Principal Component Analysis). Направено е допълнително преобразуване на факторите чрез метода на въртене Varimax.

Обработката на експерименталните данни е извършена чрез статистическия програмен продукт IBM Statistics SPSS 24 [13,22].

Резултати и обсъждане

Химичните показатели на всеки генотип тютюн, и в частност на тютюн тип Виржиния, са важни за характеризиране качеството и пушателните свойства на генотиповете.

Типичната Виржиния флю кюрд съдържа никотин около 2-2,5 %, а тази, произвеждана в нетипични райони – 0,6-1,2 % [3]. Анализът на резултатите за съдържанието на никотина в изследвания тютюн показва, че той се колебае в пределите както на генотипа, така и на годините на отглеждане. Максимално количество никотин и при трите хибрида е отчетено през 2013 г. През 2013 г. и през 2014 г. съдържанието на общ азот е значително. През 2015 г. с изключение на Хибрид 33 то е в оптимални граници. Табакова [10] посочва характерни стойности за тютюн Виржиния - от 1,7 до 2 (2,3 %).

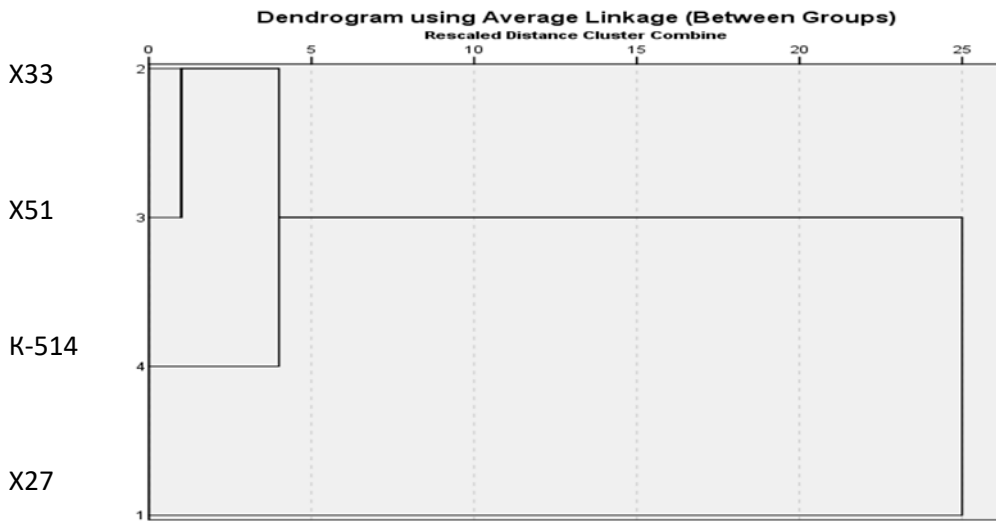
Нивото за захари варира в широки граници през трите години на изследване. Количеството им през 2013 г.е минимално, което не е характерно за тютюн Виржиния. При този тип тютюн съдържанието на разтворими въглехидрати е (12-28%) [10]. Хибрид 27 и Хибрид 51 са с най-благоприятни стойности и на трите изследвани показателя през 2015 г. – Хибрид 27 и Хибрид 51.

В таблица 1 са дадени последователността на получаването на отделните клъстери при всяка стъпка и междугруповите разстояния при групиране според степен на сходство по химични показатели за 2013 г. Четвъртата колона на таблицата (Coefficients) показва разстоянието между клъстерите, които са обединени на тази стъпка.

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	3	0,931	0	0	2
2	2	4	2,707	1	0	3
3	1	2	15,399	0	2	0

Таблица 1. Комбиниране на клъстерите и междугруповите разстояния по химични показатели за 2013 г.

На дендрограмата на фигура 1 е представен резултатът от клъстеризацията в графичен вид. Изследваните генотипове са групирани в три клъстера. Най-отдалечен е Хибрид 27, който е със значително по-малко съдържание на захари в сравнение с останалите хибриди. Разстоянието, на което се присъединява към останалите хибридни форми, е 25 евклидови единици. Хибридна форма В 0514 формира самостоятелен клъстер, поради минималното съдържание на общ азот и максимално – на захари. Хибрид 33 и Хибрид 51 формират общ клъстер на минимално евклидово разстояние.

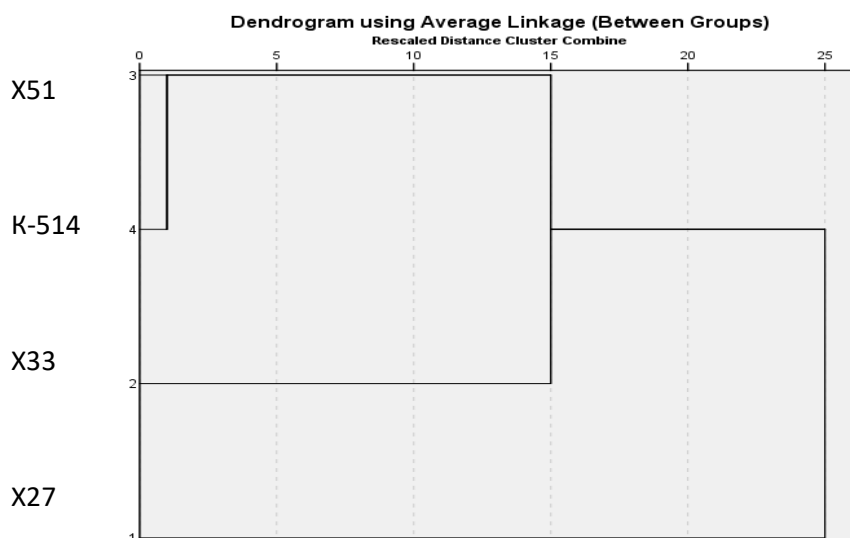


Фигура 1. Дендрограма на клъстеризационната процедура по химични показатели за 2013 г.

Основавайки се на данните за химичните характеристики на изследваните хибриди за 2014 г., са получени три клъстера (таблица 2, фигура 2).

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	4	0,253	0	0	2
2	2	3	1,429	0	1	3
3	1	2	2,244	0	2	0

Таблица 2. Комбиниране на клъстерите и междугруповите разстояния по химични показатели за 2014 г.



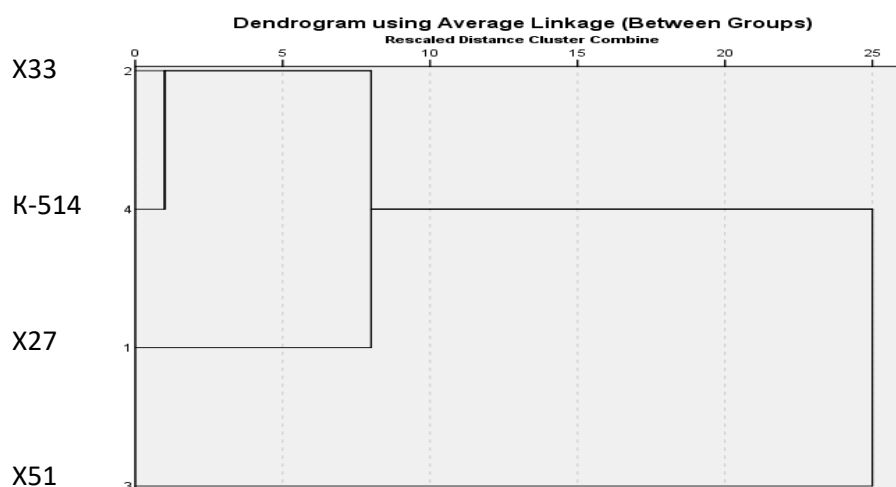
Фигура 2. Дендрограма на клъстеризационната процедура по химични показатели за 2014 г.

Това се обяснява със следните факти: Хибрид 51 и В 0514 са с максимално съдържание на общ азот, чувствително надвишаващо това при другите две форми, същевременно Хибрид 27 е с минимално съдържание на никотин и общ азот и максимално – на захари (значително надвишаващо това на останалите). Хибрид 27 се откроява като хибриден с най-благоприятен химичен състав.

По-различен е начинът на групиране на хибридите по химични показатели за 2015 г. Стандартът В 0514 е най-близък с Хибрид 33, поради сходство в съдържанието на азот (максимално) и на захари (минимално). Това обуславя обединението им в общ клъстер на минимално евклидово разстояние. Хибрид 51 формира самостоятелен клъстер, най-отдалечен от останалите. Това се обяснява с минималното съдържание на никотин и максималното – на захари (таблица 3, фигура 3). А Хибрид 27 е със средни размери в сравнение с останалите изследвани хибриди, което го превръща в сравнително балансиран по съответните показатели.

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	4	0,279	0	0	2
2	1	2	3,737	0	1	3
3	1	3	11,663	2	0	0

Таблица 3. Комбиниране на клъстерите и междугруповите разстояния по химични показатели за 2015 г.



Фигура 3. Дендрограма на клъстеризационната процедура по химични показатели за 2015 г.

При клъстерния анализ не се формулират статистически хипотези и затова няма проверка на статистическа и на практическа значимост. Все още не е разработена процедура за оценка на надеждността и валидността на клъстеризацията [2]. Това обуславя необходимостта от комбинирането му с друг метод, при който се провежда тест за достоверност на получените резултати. Тук е използван факторен анализ. Установи се, че всички предпоставки за неговото прилагане, са изпълнени за всяка от годините на изследване. За трите години детерминантите на корелационните матрици са положителни числа, удовлетворени са условията на КМО-теста (Kaiser-Meyer-Olkin>0.5) и теста за сферичност на Bartlett (Sig.<0.05). Доказано е, че изследваните променливи корелират помежду си.

Оказа се, че за 2013 г. трите показателя се трансформират до един компонент, чрез който се обяснява 67% от общата дисперсия. Тук най-силно е влиянието на процентното съдържание на никотин. За 2014 г. отново факторите са редуцирани до един, обясняващ 56% от общата дисперсия и отново никотинът е с доминиращи функции. За 2016 г. имаме редукция до два фактора. Първият обяснява 60% от общата дисперсия и с водещо значение е никотинът. Заедно с втория компонент се обясняват 99,8% от общата дисперсия. Според втория компонент определящо е значението на съдържанието на общ азот.

Изводи

След проведените клъстеризации биха могли да се направят изводи за степента на сходство на новополучените хибридни форми с контролата В 0514 със сходни характеристики по химични показатели. За 2013 г. се установи сходство между В 0514, Хибрид 33 и Хибрид 51. Най – отдалечен е Хибрид 27. За 2014 г. най-близък до стандарта В 0514 е Хибрид 51, а най-отдалечен отново е Хибрид 27. За 2015 г. се установи сходство на В 0514 с Хибрид 33 и най-голямо различие с Хибрид 51. Хибрид 27 и Хибрид 52 са с най-балансиран химичен състав през тази година на изследване. Всички тези резултати биха могли да се имат предвид при бъдещи научни разработки, целящи подобрения на хибридните форми. По този начин могат да се създадат нови линии тютюн с характеристики, максимално близки до контролната линия, а дори и превъзхождащи я.

Използвана литература

1. Бозуков Х. История, настояще и бъдеще на ориенталския тютюн в България, Юбилейна международна научна конференция „70 години Институт по тютюна и тютюневите изделия“, 2014
2. Ганева, З. преоткрием статистиката с IBM SPSS Statistics, 2016, Елестра, София
3. Гюзелев, Л. Стокознание на тютюна, Христо Г. Данов, Пловдив, 1983
4. Кочоска К., Илија Ристески, И., Смоквоски, М. Физичко-хемиски карактеристики на признати и новосоздадени виржински сорти тутун, Тутун/Тобасо, 2008, 58, (3-4), 63-68
5. Машева В., Тахсин, Н., Мутафчиева, М. Изследване на биологичните и морфологични качества на чуждестранни сортове ориенталски тютюн, Юбилейна международна научна конференция „70 години институт по тютюна и тютюневите изделия“, 2014
6. Николаев Е.В. Особенности табачного сырья, 2013, [www. po-teme.com.ua](http://www.po-teme.com.ua)
7. Писклов В.П., С. К. Кочеткова, Н.А. Дурунча, Л.В. Кокорина, С.А. Медведева, И.А. Резниченко, Сравнительный анализ химического состава табачного сырья, Извешкия вузов. Пищевая технология, 2012, 5-6, 20-24.
8. Стоилова А., Станкев, Г. Химико-технологични показатели на линии тютюн от региона на Дупница, Български тютюн, 2003, 6, 20-23
9. Стоилова А., Даньо, С., Станкев, Г. Линии тютюн от произход Дупница с подобрен химичен състав и пушателни свойства, Научни трудове на Национален център за аграрни науки, София, 2005, 3, 163-166.
10. Табакова Е. Основни показатели за оценка на тютюн Виржиния, Български тютюн, 1992, 5, 18-21.
11. Chaves, M., Berebtsen P., Oude Lansink, A., Creating a typology of tobacco farms according to determinants of diversification in Valle de Lerma (Salta-Argentina), Spanish Journal of Agricultural Research, 2010, 8 (2), 460-471

12. Chen Xueping, Zhang Liang, Gui Jiaming, Wei Li, Yan Zheng. A study on the cluster analysis of 48 leaf-tobacco samples with 8 chemical compositions, *Acta Tabacaria Sinica*. 2002, 4.
13. Cronk, B, *How to Use SPSS: A Step-By-Step Guide to Analysis and Interpretation*, 2016, Routledge
14. Davalieva, K., Maleva, I., Filipovski, K., Spirovski, O., Efremov, G., Genetic Variability of Macedonian Tobacco Varieties Determined by Microsatellite Marker Analysis, *Diversity*, 2010 (2), 439-449,
15. Dyulgierski Y., Docheva, M.. Production characteristics and chemical indicators of perspectiva lines of Burley tobacco, *Тютун/Tobacco*, 2017, 67, (1-6), 41-47.
16. Dzalilov, Z., Bagirov, A., Cluster Analysis of A Tobacco Control Data Set, *International Journal of Lean Thinking*, 2010, 1, 40-50
17. Feng, G., Li L., Li, D., Li, N., Dai., Y., Grey Equal Weight Cluster Analysis of Flue-Cured Tobacco Based on Chemical Compositions, *Advances Materials Research*, 2011, 339, 299-309
18. Gocheva-Ilieva, S. G., I. P. Iliev. Statistical models of characteristics of metal vapor lasers, Nova Science Publishers, 2009, ISBN 978-1-61324-293-3.
19. Iliev, I. P., Gocheva-Ilieva, S., Astadjov, D., Denev, N., Sabotinov. N. Statistical approach in planning experiments with a copper bromide vapor laser, *Quantum Electron.*, 2008a, 38, (5), 436-440
20. Kalivas, A., Ganopoulos, I., Bosmali, I., Tsaliki, E., Osathanunkul, M., Xanthopoulou, A., Moysiadis, T., Avramidou, E., Grigoriadis, I., Zambounis, A., Tsaftaris, A., Nianiou-Obeidat, I., Madesis, P., Genetic Diversity and Structure of Tobacco in Greece on the Basis of Morphological and Microsatellite Markers, *Alliance of Crop, Soil and Environmental Science Societies*, 2016, 56 (5), 2652-2662
21. Li Dong-liang, Xu Zi-cheng, Analysis of Chemical Components in Flue-cured Tobacco Leaves Based on Grey Relational Degree Cluster Method, *System Sciences and Comprehensive Studies in Agriculture*, 2007, 4,
22. Meyers L., Gamst G., Guarino A., *Performing Data Analysis Using IBM SPSS*, Wiley, 2013
23. Trajkoski J., Mitreski, M., Pelivanoska, V., Zdraveska, N., Mavroski, R.. Chemical properties of tobacco in some oriental varieties from the type Prilep, *Тютун/Tobacco*, 2015, 65 (7-12), 59-66.
24. Xiao Bingguang, Zhang Yanchun, Lu Xiuping, Wnag Shaokun, The Principal Component Analysis and Cluster Analysis of the Flue-cured Tobacco Varieties, 2000-02.
25. Yu Na, Sheng Ke, Sun Xiekun, Tao Liqi, Principal component cluster analysis of aroma components in cut tobacco strips under different drying intensities, *Tobacco Chemistry*, 2016, 49 (2), 29-35