

**Н. В. Коханюк**, кандидат сільськогосподарських наук  
**І. В. Темченко, Т. М. Штуць, А. А. Лехман, С. В. Барвінченко**  
*Інститут кормів та сільського господарства Поділля НААН*

## **КЛАСТЕРНИЙ АНАЛІЗ У СЕЛЕКЦІЇ ЗЕРНОБОБОВИХ КУЛЬТУР**

*За результатами кластерного аналізу зроблено розподіл досліджуваних 423 генотипів (56 комбінацій)  $F_2$  сої, 122 генотипів (17 комбінацій)  $F_2$  бобів кормових та 27 генотипів (4 комбінації)  $F_2$  квасолі, створених методом гібридизації, на п'ять кластерів за основними господарсько-цінними ознаками. Встановлено, що найбільший інтерес викликає розподіл рослин  $F_2$  сої на 5 кластерів ( $K_5$ ), за якого в кращій групі  $K_5$ -IV виділено 12 (2,8 %) рослин, вони за основними господарсько-цінними ознаками перевищували відповідні показники як інших груп  $K_5$ , так і кращих груп  $K_2$ ;  $K_3$  і  $K_4$ . При розподілі рослин  $F_2$  бобів кормових на 5 кластерів ( $K_5$ ) крацою виявилася група IV ( $K_5$ -IV), в якій концентрувалися 17 (13,9 %) рослин з ознаками, які перевищували середній показник усіх груп даного кластеру. Розподіл рослин  $F_2$  квасолі на 5 кластерів ( $K_5$ ) дав змогу виділити групу  $K_5$ -II, в якій розміщені рослини з найвищими показниками.*

**Ключові слова:** *соя, квасоля, боби кормові, кластерний аналіз, селекційний процес, групи кластерів.*

Термін «кластерний аналіз» вперше вжив англійський вчений Р. Тріон у 1939 році, він охоплює 100 різних алгоритмів [1]. Більшість сучасних методів кластеризації запропоновано в 60-ті роки ХХ століття. Цей час характеризується величезною кількістю публікацій. Як найважливіші можна виділити роботи таких авторів: Г. Болла і Д. Холла, Дж. Мак-Кіна – щодо методу К-середніх; Р. Сокала і Дж. Снитсья, Г. Ланса і У. Вільямса, Н. Джардайн і Р. Сібсон – щодо ієрархічних процедур; Дж. Роджесу і Т. Танімото, Е. М. Бравермана, А. А. Дорофеюк, І. Б. Мучника – щодо процедур типу послідовного формування кластерів і діагоналізації; В. М. Йолкіної, Н. Г. Загоруйко – щодо еталонних алгоритмів й деякі інші дослідження [2]. Ці та багато інших авторів сформували математичну базу для застосування кластерних методів у різних галузях науки.

Кластерний аналіз (англ. *Data clustering*) – це задача розподілу заданої вибірки об'єктів (ситуацій) на підмножини, які називаються кластерами, так, щоб кожен кластер складався зі схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися. Кластерний аналіз – це багатовимірна статистична процедура, що виконує збір даних, які містять інформацію про вибірку

об'єктів, і потім впорядковує об'єкти в порівняно однорідні групи (кластери). Кластер – група елементів, які характеризуються загальною властивістю, головна мета кластерного аналізу – знаходження груп схожих об'єктів у вибірці [3].

Використання багатовимірних методів аналізу у селекційному процесі допомагає скоротити тривалість селекційного процесу через більш повну і комплексну оцінку факторів, що впливають на реалізацію генетичного потенціалу рослин [4—7].

Оцінка гібридних рослин зернобобових культур у самому ранньому поколінні  $F_2$ , що розщеплюється, і виділення тих, які мають селекційне значення за комплексом кількісних ознак і індексів, за допомогою кластерного аналізу представляється нам досить цікавою і багатообіцяючою.

З використанням цього методу добору ми можемо розвантажити селекційний процес і звільнитися на ранніх етапах селекції від неперспективного матеріалу шляхом бракування рослин, що потрапили в гірші групи кластерів.

Головним передуманням цьому аналізу є візуальна оцінка в полі і бракування в  $F_2$  по вегетуючих рослинах, що не представляють селекційної цінності, не відповідають вимогам селекціонера за архітектонікою, стійкістю до шкідників і хвороб, забарвленням листя, скоростиглістю і т. д, і лише після лабораторного аналізу популяцій  $F_2$  за основними ознаками і індексами використовуємо кластерний аналіз з подальшим бракуванням гірших рослин по групах кластера. Потім посів у  $F_{3-6}$ , аналіз за фенотипічними ознаками і продуктивністю, структурний аналіз (встановлення генотипічної варіанси за середніми значеннями ознак і індексів не менше, ніж по 25 рослинах) і повторне комп'ютерне бракування в старших поколіннях у міру встановлення гомозиготності ознак.

Таким чином, запропонований селекційно-технологічний цикл аналізу популяцій  $F_2$  зернобобових культур, що розщеплюються, дає змогу, використовуючи багатовимірний аналіз, відбирати кращі за господарсько-цінними ознаками рослини, звільняючись при цьому від неперспективних, і в подальшому працювати з малими об'ємами, надаючи більшу увагу гомеостазу, адаптивним властивостям і якості насіння.

**Мета.** Використовуючи кластерний аналіз, скоротити тривалість селекційного процесу шляхом групування гібридних комбінацій у кластерні класи і відібрати для подальшої роботи елітні рослини з груп кластерів, які включають ознаки і індекси, що мають найтісніші кореляційні зв'язки з продуктивністю.

**Методи.** Нами було проведено оцінку 423 генотипів (56 комбінацій)  $F_2$  сої, 122 генотипів (17 комбінацій)  $F_2$  бобів кормових та 27 генотипів (4 комбінації)  $F_2$  квасолі, створених методом гібридизації. У кластерному аналізі в основу групування по всіх рослинах були взяті основні господарсько-цінні ознаки: висота рослин, кількість продуктивних вузлів,

кількість бобів на рослині, кількість насінин на рослині та маса насіння з рослини.

Кластерний аналіз виконувався в модулі Cluster Analysis пакету програм Statistica, 10. Для побудови кластерів використовувався метод К-середніх кластерного аналізу.

**Результати.** За результатами кластерного аналізу нами зроблено розподіл досліджуваних зразків, створених методом гібридизації, на п'ять основних кластерів за основними господарсько-цінними ознаками.

При розподілі на 2 кластери ( $K_2$ ) за середньою величиною всіх 5 аналізованих ознак кращою виявилася група I ( $K_2$ -I), в якій концентрувалися 91 рослина сої з ознаками, які значно перевищували середній показник групи II (табл. 1).

### 1. Розподіл рослин гібридів $F_2$ сої по кластерах і групах за основними господарсько-цінними ознаками (2018 р.)

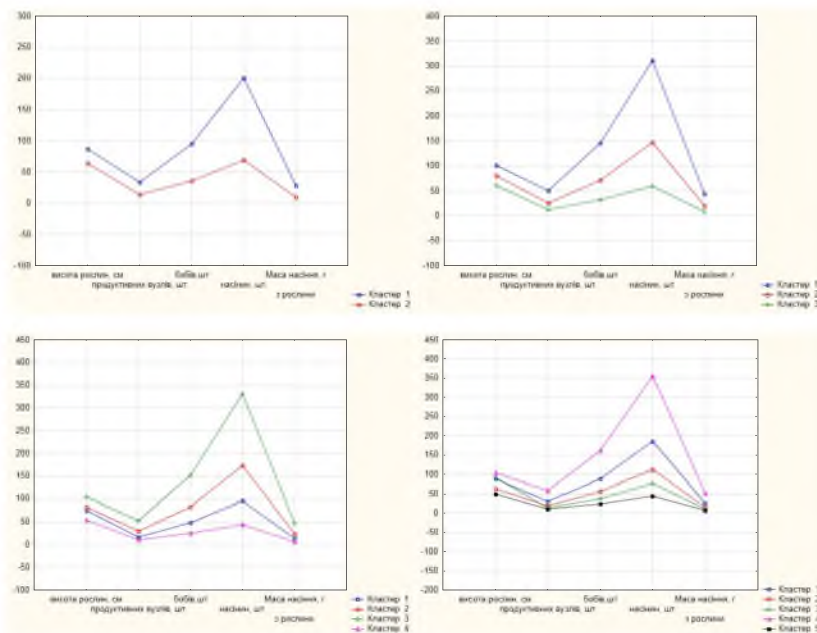
Кластери		Число рослин, шт.	Висота рослин, см	Кількість продуктивних вузлів, шт.	Кількість бобів на рослині, шт.	Кількість насінин на рослині, шт.	Маса насіння з рослини, г
$K_2$	I	91	87,5±26,1	33,5±13,7	95,3±31,7	201,1±70,4	27,9±10,7
	II	332	63,9±22,6	13,5±5,3	36,1±15,1	69,7±31,6	10,2±4,5
$X_{\text{сєр.}}$			75,7	23,5	65,7	135,4	19,1
$K_3$	I	20	101,3±18,2	50,8±16,9	145,0±29,2	310,9±70,3	43,7±11,8
	II	130	80,8±25,1	24,9±7,4	70,5±15,8	146,2±32,9	20,3±5,0
	III	273	61,1±21,5	12,1±4,2	31,4±11,9	59,4±24,3	8,9±3,7
$X_{\text{сєр.}}$			81,1	29,3	82,3	172,2	24,3
$K_4$	I	165	74,7±23,1	16,9±4,7	47,6±10,3	94,9±19,6	13,5±3,2
	II	77	83,1±27,0	29,3±8,1	82,8±15,2	172,6±28,2	23,9±4,5
	III	16	103,4±15,2	52,3±18,6	151,5±28,6	330,4±64,9	46,8±11,0
	IV	165	53,5±16,4	9,9±3,0	24,3±8,5	43,7±16,1	6,8±2,6
$X_{\text{сєр.}}$			78,7	27,1	76,6	160,4	22,8
$K_5$	I	65	88,8±26,3	30,9±8,3	88,6±16,8	185,6±31,3	25,3±5,2
	II	104	63,6±19,8	19,2±5,2	55,9±10,1	112,5±18,9	16,3±3,1
	III	84	91,2±16,2	14,7±4,3	38,4±9,2	76,8±16,4	10,5±2,8
	IV	12	103,8±14,6	58,1±17,2	161,1±26,2	354,0±56,3	50,6±9,9
	V	158	50,1±11,3	10,1±3,2	24,8±9,3	44,2±17,3	7,0±2,8
$X_{\text{сєр.}}$			79,5	26,6	73,7	154,6	21,7
По всьому масиву		423	78,7	26,6	74,6	155,6	22,0

Розподіл на 3 кластери дав змогу виділити найкращу групу I (20 рослин з 423 або 4,7 %), в якій розміщені рослини з найвищими показниками: висота рослин – 101,3 ± 18,2 см., кількість продуктивних вузлів – 50,8 ± 16,9 шт., кількість бобів на рослині – 145,0 ± 29,2 шт., кількість насінин на рослині – 310,9 ± 70,3 шт. та маса насіння з рослини – 43,7 ± 11,8 г.

При розподілі на 4 кластери ( $K_4$ ) за середньою величиною всіх 5 аналізованих ознак кращою виявилася група III ( $K_4$ -III), в якій концентрувалися 16 (3,8 %) рослин з ознаками, які значно перевищували середній показник інших груп даного кластеру.

Найбільший інтерес викликає розподіл рослин  $F_2$  на 5 кластерів ( $K_5$ ), за якого в кращій групі  $K_5$ -IV концентрувалося 12 (2,8 %) рослин сої, які за основними господарсько-цінними ознаками перевищували відповідні показники як інших груп  $K_5$ , так і кращих груп  $K_2$ ,  $K_3$  і  $K_4$ .

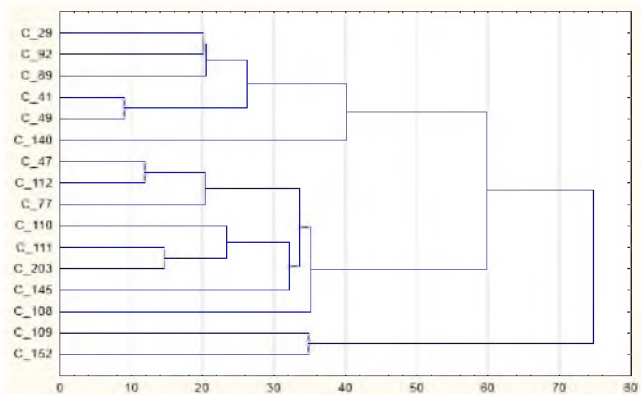
Графік середніх показників рослин  $F_2$  сої для кожного кластеру зображений на рис. 1.



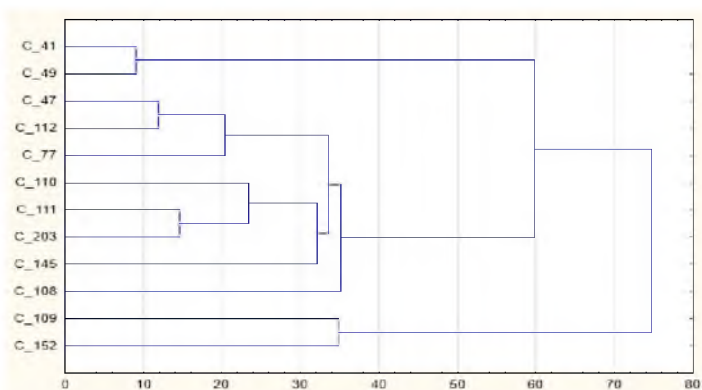
**Рис. 1. Графік середніх показників рослин  $F_2$  сої для кожного кластеру**

Як показано на дендрограмах (рис. 2, 3) всі 12 кращих рослин сої з кластера  $K_5$ -IV увійшли в число 16 кращих в кластері  $K_4$ -III.

Таким чином, у процесі відбору з використанням кластерного аналізу з 423 аналізованих рослин  $F_2$  сої, з використанням групуючих ознак, у кластері  $K_5$ -IV виділено 12 кращих рослин, що мають максимальні значення основних показників, які визначають їх продуктивність при оптимальному поєднанні ознак генеративних і вегетативних органів рослини. Потомства  $F_2$ , відібрані із кращих кластерів елітних рослин, будуть висіватись для подальшого їх вивчення і використання в практичній селекції.



**Рис. 2.** Дендрограма розподілу 16 кращих рослин сої F<sub>2</sub> в K<sub>4</sub>-III



**Рис. 3.** Дендрограма розподілу 12 кращих рослин сої F<sub>2</sub> в K<sub>5</sub>-IV

Розподіл рослин гібридів F<sub>2</sub> бобів кормових по кластерах і групах за основними господарсько-цінними ознаками наведений в табл. 2.

При розподілі на 2 кластери (K<sub>2</sub>) за середньою величиною всіх 5 аналізованих ознак кращою виявилася група I (K<sub>2</sub>-I), в якій концентрувалися 68 рослин бобів кормових з ознаками, які перевищували середній показник групи II.

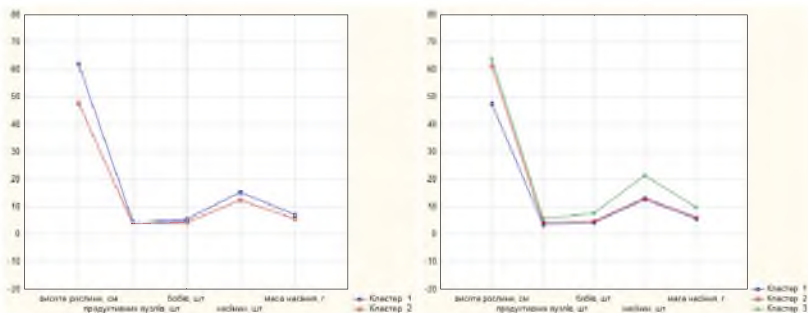
Розподіл на 3 кластери дав можливість виділити найкращу групу III (17 рослин з 122 або 13,9%), в якій розміщені рослини з найвищими показниками: висота рослин –  $63,7 \pm 6,0$  см, кількість продуктивних вузлів –  $5,8 \pm 1,5$  шт., кількість бобів на рослині –  $7,6 \pm 1,3$  шт., кількість насіння на рослині –  $21,4 \pm 3,6$  шт. та маса насіння з рослини –  $9,9 \pm 2,6$  г.

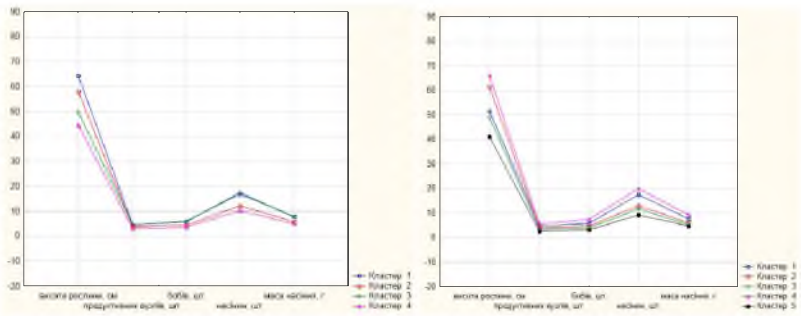
Розподіл на 4 кластери дав змогу виділити групу К<sub>4</sub>-III (18 рослин (14.8 %), в якій розміщені рослини з найвищими показниками.

## 2. Розподіл рослин гібридів F<sub>2</sub> бобів кормових по кластерах і групах за основними господарсько-цінними ознаками

Кластери		Число рослин, шт.	Висота рослин, см	Кількість продуктивних вузлів, шт.	Кількість бобів на рослині, шт.	Кількість насінин на рослині, шт.	Маса насіння з рослини, г
K <sub>2</sub>	I	68	62,2±4,5	4,5±1,3	5,5±1,6	15,3±4,5	7,2±2,4
	II	54	47,7±4,7	3,6±1,1	4,2±1,2	12,6±3,9	5,7±1,7
X <sub>сер.</sub>			55,0	4,1	4,9	14,0	6,5
K <sub>3</sub>	I	52	47,4±4,5	3,6±1,1	4,2±1,2	12,6±3,9	5,8±1,6
	II	53	61,4±4,0	4,1±0,9	4,8±0,9	13,2±2,6	6,2±1,5
	III	17	63,7±6,0	5,8±1,5	7,6±1,3	21,4±3,6	9,9±2,6
X <sub>сер.</sub>			57,5	4,5	5,5	15,7	7,3
K <sub>4</sub>	I	38	64,5±4,2	4,9±1,4	6,1±1,6	17,1±4,1	8,0±2,4
	II	38	58,0±3,3	4,0±0,8	4,4±0,9	12,2±2,4	5,8±1,3
	III	18	49,7±2,8	4,7±1,2	5,7±1,2	17,6±3,3	7,7±1,9
	IV	28	44,8±3,9	3,1±1,0	3,6±1,1	10,4±3,2	4,9±1,3
X <sub>сер.</sub>			54,3	4,2	5,0	14,3	6,6
K <sub>5</sub>	I	20	51,7±3,6	4,8±1,1	5,8±1,2	17,7±3,2	7,7±1,8
	II	46	61,4±3,2	4,0±0,9	4,7±0,9	13,0±2,6	6,2±1,6
	III	27	49,1±3,0	3,5±0,8	4,0±0,8	11,7±2,1	5,2±1,2
	IV	17	66,1±4,6	5,6±1,5	7,4±1,4	20,0±4,1	9,3±2,7
	V	12	41,2±3,1	2,8±1,0	3,2±1,2	9,3±4,3	4,8±1,7
X <sub>сер.</sub>			53,9	4,1	5,0	14,3	6,6
По всьому масиву		122	55,2	4,2	5,1	14,6	6,7

При розподілі на 5 кластерів (K<sub>5</sub>) за середньою величиною всіх 5 аналізованих ознак кращою виявилася група IV (K<sub>5</sub>-IV), в якій концентрувалися 17 рослин бобів кормових з ознаками, які перевищували середній показник усіх груп даного кластеру (рис. 4).





**Рис. 4. Графік середніх показників рослин F<sub>2</sub> бобів кормових для кожного кластера**

Розподіл рослин гібридів F<sub>2</sub> квасолі по кластерах і групах за основними господарсько-цінними ознаками наведений в табл. 3.

### **3. Розподіл рослин гібридів F<sub>2</sub> квасолі по кластерах і групах за основними господарсько-цінними ознаками**

Кластери		Число рослин, шт.	Висота рослин, см	Кількість продуктивних вузлів, шт.	Кількість бобів на рослині, шт.	Кількість насіння на рослині, шт.	Маса насіння з рослини, г
K <sub>2</sub>	I	14	33,9±4,6	6,8±1,3	7,6±1,3	25,8±6,6	7,9±2,1
	II	13	38,8±5,4	10,0±1,7	13,2±2,5	43,0±6,6	12,9±2,8
X <sub>сер.</sub>			36,4	8,4	10,4±	34,4	10,4
K <sub>3</sub>	I	8	37,1±5,0	10,3±1,6	13,8±2,2	46,9±5,0	14,2±2,8
	II	12	37,3±5,4	8,3±2,1	10,0±3,0	33,8±3,7	10,1±1,5
	III	7	33,4±5,8	6,3±0,5	7,0±0,6	20,0±2,9	6,2±0,6
X <sub>сер.</sub>			35,9	8,3	10,3	33,6	10,2
K <sub>4</sub>	I	7	42,0±4,5	9,8±1,9	13,1±3,2	38,1±3,5	11,3±0,9
	II	7	34,4±3,4	7,3±1,6	8,3±1,6	31,6±2,5	9,6±1,7
	III	6	35,0±3,7	10,2±1,6	13,3±1,6	48,7±4,3	14,8±3,1
	IV	7	33,4±5,8	6,3±0,5	7,0±0,6	20,0±2,9	6,2±0,7
X <sub>сер.</sub>			36,2	8,4	10,4	34,6	10,5
K <sub>5</sub>	I	7	33,4±3,4	7,3±1,6	8,3±1,6	31,6±2,5	9,6±1,7
	II	6	35,0±3,7	10,2±1,6	13,3±1,6	48,7±4,3	14,8±3,1
	III	7	42,0±4,5	9,8±1,9	13,1±3,2	38,1±3,5	11,3±0,9
	IV	3	39,3±0,6	6,7±0,6	6,7±0,6	18,0±2,6	6,4±0,9
	V	4	29,0±2,6	6,0±0,1	7,2±2,4	21,5±2,4	6,1±0,5
X <sub>сер.</sub>			35,7	8,0	9,7	31,6	9,6
По всьому масиву		27	36,1	8,3	10,2	33,5	10,2

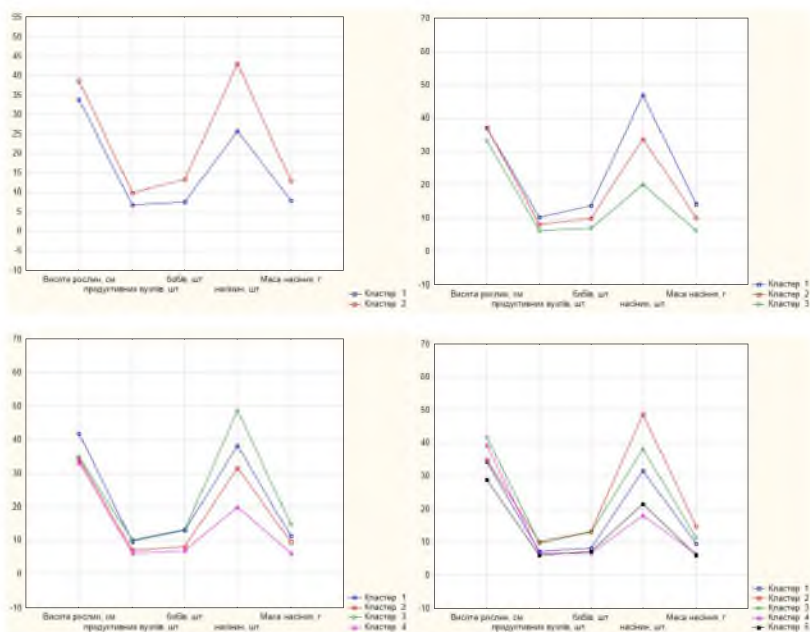
При розподілі на 2 кластери (K<sub>2</sub>) за середньою величиною всіх 5 аналізованих ознак кращою виявилася група II (K<sub>2</sub>-II), в якій

концентрувалися 13 (48,1 %) рослин квасолі з ознаками, які перевищували середній показник групи I.

Розподіл на 3 кластери дав можливість виділити групу К<sub>3</sub>-I (8 рослин з 27 або 29,6%), в якій розміщені рослини з найвищими показниками.

Найбільший інтерес викликає розподіл рослин F<sub>2</sub> на 4 кластери (К<sub>4</sub>), за якого в кращій групі К<sub>4</sub>-III концентрувалося 6 (22,2 %) рослин, які за основними господарсько-цінними ознаками перевищували відповідні показники інших груп даного кластеру.

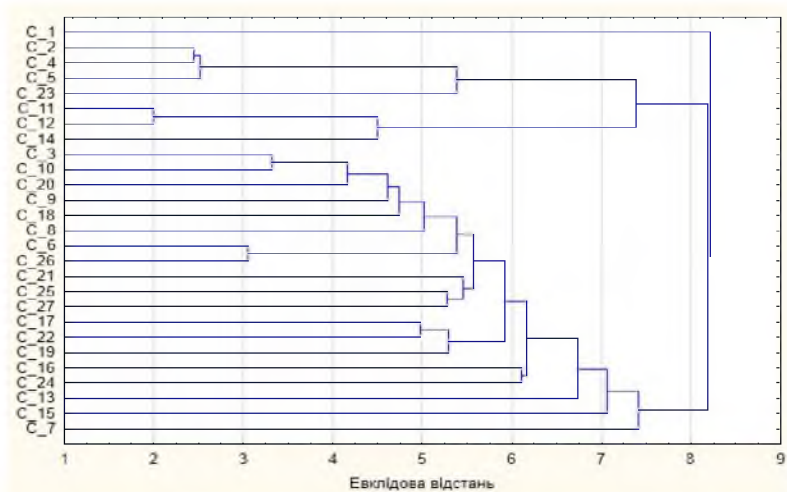
Розподіл рослин F<sub>2</sub> на 5 кластерів (К<sub>5</sub>) дав змогу виділити групу К<sub>5</sub>-II, в якій розміщені рослини з найвищими показниками: висота рослин – 35,0 ± 3,7 см, кількість продуктивних вузлів – 10,2 ± 1,6 шт., кількість бобів на рослині – 13,3 ± 1,6 шт., кількість насінин на рослині – 48,7 ± 4,3 шт. та маса насіння з рослини – 14,8 ± 3,1 г. (рис. 5).



**Рис. 5. Графік середніх показників рослин F<sub>2</sub> квасолі для кожного кластеру**

Дендрограма розподілу 27 рослин квасолі F<sub>2</sub> зображена на рис. 6.





**Рис. 6. Дендрограма розподілу 27 рослин квасолі F<sub>2</sub>**

**Висновки.** Використання кластерного аналізу в селекції зернобобових культур дає можливість на ранніх етапах селекційного процесу провести швидку оцінку, розподіл та підбір вихідного матеріалу.

1. У процесі відбору з використанням кластерного аналізу з 423 аналізованих рослин F<sub>2</sub> сої, з використанням групуючих ознак, в кластері K<sub>5</sub>-IV виділено 12 кращих рослин, що мають максимальні значення основних показників, які визначають його продуктивність при оптимальному поєднанні ознак генеративних і вегетативних органів рослини. Потомства F<sub>2</sub>, відібрані із кращих кластерів елітних рослин, будуть висіватись для подальшого їх вивчення і використання в практичній селекції.

2. При розподілі рослин F<sub>2</sub> бобів кормових на 5 кластерів (K<sub>5</sub>) за середньою величиною всіх 5 аналізованих ознак кращою виявилася група IV (K<sub>5</sub>-IV), в якій концентрувалися 17 рослин з ознаками, які перевищували середній показник усіх груп даного кластеру.

3. Розподіл рослин F<sub>2</sub> квасолі на 5 кластерів (K<sub>5</sub>) дав змогу виділити групу K<sub>5</sub>-II, в якій розміщені рослини з найвищими показниками: висота рослин –  $35,0 \pm 3,7$  см, кількість продуктивних вузлів –  $10,2 \pm 1,6$  шт., кількість бобів на рослині –  $13,3 \pm 1,6$  шт., кількість насінин на рослині –  $48,7 \pm 4,3$  шт. та маса насіння з рослини –  $14,8 \pm 3,1$  г.

### **Бібліографічний список**

1. Tryon R. C. Cluster analysis / R. C. Tryon L, D. E. Bailey. – Mc Graw-Hill Inc., US – 1970.
2. Мандель И. Д. Кластерный анализ / И. Д. Мандель – М.: Финансы и статистика. – 1988.

3. Чубукова И. А. Data Mining: учебное пособие / И. А. Чубукова. — М.: Интернет-университет информационных технологий : БИНОМ: Лаборатория знаний, 2006.

4. Тищенко В. Н. Использование кластерного анализа для идентификации и отбора высокопродуктивных генотипов озимой пшеницы на ранних этапах селекции / В. Н. Тищенко, Н. М. Чекалин, М. Е. Зюков // Фактори експериментальної еволюції організмів. Збірник наукових праць. — Том 2. — Київ, Аграрна наука, 2004. — С. 270—278.

5. Мельник А. В. Використання кластерного аналізу за підбору сортів і гібридів ріпаку ярого для вирощування в Лівобережному Лісостепу України / А. В. Мельник // Вісник Полтавської державної аграрної академії. — 2013. — № 4. — С. 6—11.

6. Вільчинська Л. А. Кластерний аналіз в селекції гречки / Л. А. Вільчинська, О. П. Городинська, М. В. Диянчук, О. О. Камінна // Вісник Українського товариства генетиків і селекціонерів. — 2017. — Том 15. — № 2. — С. 145—149.

7. Чекалин Н. М. Использование кластерного анализа как метода индивидуального отбора у проса (*Panicum miliaceum* L.) / Н. М. Чекалин, В. Н. Тищенко, П. М. Панченко, В. С. Сидоренко // Вісник Полтавської державної аграрної академії. — 2009. — № 2. — С. 10—17.

*Надійшла до редколегії 25. 06. 2019 року  
Рецензенти В. Д. Бугайов, кандидат сільськогосподарських наук*