

УДК 624.15:631.431.6

Г.Б. Гуменюк, канд.біол.наук, доцент,

В.О. Хоменчук, канд.біол.наук, доцент,

Б.Б. Гавришок, канд.геогр.наук, викладач,

Г.В. Станіславчук, завідувач лабораторії екобіотехнології та основ здоров'я

Тернопільський національний педагогічний університет імені Володимира Гнатюка,
Україна

ФАКТОРНИЙ АНАЛІЗ: ВИКОРИСТАННЯ У БІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

H. V. Humeniuk, PhD, Assoc. Prof

V. O. Khomenchuk, PhD, Assoc. Prof.

B. B. Havryshok, Ph.D

A. V. Stanislavchuk, PhD, Head of the laboratory of ecobiotechnology and basics of health

FACTOR ANALYSIS: USE IN BIOLOGICAL RESEARCH

Факторний аналіз, як сукупність математичних методів і моделей з латентними показниками покликаний вирішувати наступні задачі: зменшити кількість змінних, які описують досліджувані об'єкти; опосередковано кількісно оцінити латентні показники; класифікувати змінні разом із уведеним більш загальних змінних (головних компонент) шляхом агрегування первинних ознак.

Для оцінки ефективності запропонованої моделі та узагальнення результатів ми використали підрозділ факторного аналізу – метод головних компонент.

Побудова моделі головних компонент здійснюється в три етапи: розрахунок кореляційної матриці, знаходження головних компонент і розрахунок факторних навантажень; - ідентифікація головних компонент.

На основі первинних даних з допомогою модуля STATISTICA 10 [1,2] будемо матрицю кореляцій (табл. 1)

Таблиця 1

Матриця кореляцій агрохімічного стану ґрунтів Шепетівського району Хмельницької області та основних показників родючості

Variable	Correlations (Spreadsheet1) Casewise deletion of MD N=9						
	K	NO3	pH	Ca	NH4	Гумус	P
K	1,00	-0,37	0,45	-0,31	0,11	0,12	-0,01
NO3	-0,37	1,00	-0,48	0,09	0,45	0,27	0,43
pH	0,45	-0,48	1,00	0,52	0,38	-0,06	-0,05
Ca	-0,31	0,09	0,52	1,00	0,56	0,04	0,30
NH4	0,11	0,45	0,38	0,56	1,00	-0,09	0,64
Гумус	0,12	0,27	-0,06	0,04	-0,09	1,00	-0,11
P	-0,01	0,43	-0,05	0,30	0,64	-0,11	1,00

Для знаходження коефіцієнтів значення факторів використаємо метод головних компонент. Аналіз головних компонент (Principal Components Analysis) заснований на визначенні мінімального числа факторів, які вносять найбільший вклад в дисперсію даних. За критерієм кам'янистого насипу залишаємо 2 фактори (рис 1).

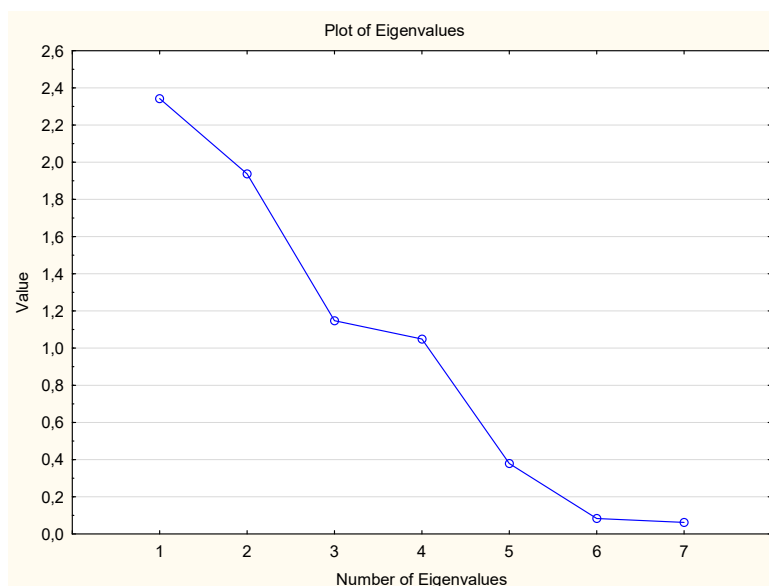


Рис. 1. Критерій кам'янистого насипу

Значення характеристичних чисел кореляційної матриці представлено на осі ординат. Як бачимо, ці значення стрімко зменшуються і лише два більші за одиницю.

За установкою Eigenvalues система видає таблицю значень характеристичних чисел, які є дисперсіями головних компонент, а також внесок кожної з них у сумарну варіацію ознакової множини - % total Variance. Внесок першої компоненти в сумарну дисперсію ознакової множини становить 33,46%, другої - 27,67%. Разом (Cumul.%) дві компоненти пояснюють 61,14% сумарної варіації, що свідчить про високий ступінь факторизації (табл.1). Власні значення факторів становлять 2.34 та 1,93 та включають 61,14% сумарної варіації. Накопичений відсоток сумарної варіації факторів визначає, наскільки повно вдалося описати нашу сукупність даних з допомогою виділених факторів (табл. 2).

Таблиця 1

Характеристичні числа та їх внесок у загальну варіацію

Value	Eigenvalues (Spreadsheet1) Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,342497	33,46424	2,342497	33,46424
2	1,937184	27,67405	4,279681	61,13830

Після цього проаналізуємо факторні навантаження – коефіцієнти кореляції кожної із аналізованих змінних із кожним з виділених факторів (табл. 2). Застосовуємо метод обертання факторів методом нормалізованого варімакса (Varimax normalized). Після обертання кожна змінна має велике навантаження тільки по одному фактору, що дозволяє його інтерпретувати через змінні, що входять до нього.

Серед процедур обертання факторів - Factor rotation вибираємо Varimax normalized (Варімакс нормалізований). За опцією Factor loadings маємо таблицю факторних навантажень, значення яких наближаються до 1 або до 0. Ознаки, які навантажує кожна компонента, виділено (табл. 3)

Результати обертання факторів (факторні ваги)

Variable	Factor Loadings (Varimax raw) (Spreadsheet1) Extraction: Principal components (Marked loadings are >,700000)	
	Factor 1	Factor 2
K	-0,080813	-0,650016
NO3	0,445194	0,791074
pH	0,382589	-0,876223
Ca	0,758344	-0,185351
NH4	0,927767	-0,058288
Гумус	-0,041828	0,168209
P	0,739687	0,248437
Expl.Var	2,335826	1,943855
Prp.Totl	0,333689	0,277694

Перша компонента зв'язана з ознаками 4,5 і 7, що характеризують вміст обмінного кальцію в ґрунтах, аміачного азоту та обмінного фосфору; друга компонента навантажує ознаки 2 і 3, які характеризують вміст нітратного азоту та водневий показник. Наведені в останніх рядках таблиці характеристичні числа і внесок окремих компонент у сумарну дисперсію визначені за трансформованими факторними навантаженнями, а тому відрізняються від первинних, проте сумарний їх внесок процедура обертання не змінює: $Prp.Total. = 0,346 + 0,267 = 0,781$.

Представляємо факторні навантаження у факторному полі (рис.2). Аналіз графіка дозволяє виявити латентні фактори та інтерпретувати фактори за навантаженнями (рис.1).

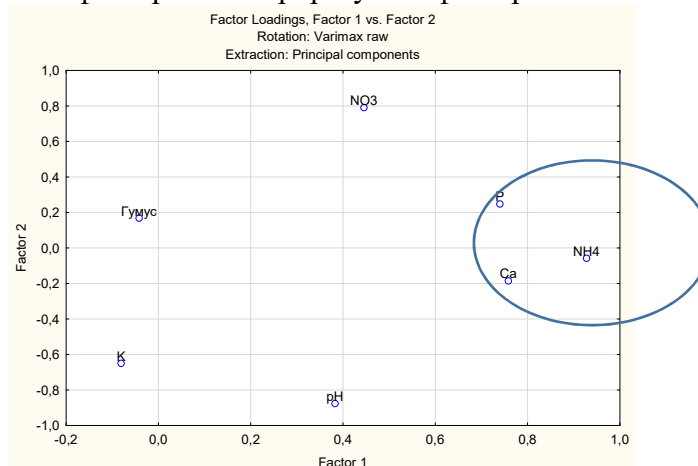


Рис 2. Графічне представлення факторних навантажень (факторне поле)

Чим вищий показник факторного навантаження, тим більшу частину масиву даних вдалося факторизувати і тим достовірніша факторна модель.

Нам вдалося факторизувати 3 ознаки, а саме, вміст рухомого фосфору, обмінного кальцію та аміачного азоту. Ймовірно ці показники найбільшою мірою впливають на агрохімічний стан досліджуваних ґрунтів Шепетівського району Хмельницької області.

Література:

1. Гойко О.В. Практичне використання пакета STATISTICA для аналізу медико-біологічних даних: навчальний посібник для студентів вищих навчальних закладів (Рекомендовано МОН України, ISBN 966-8326-31-8) / О.В. Гойко. - Київ, 2004. - 76 с.

2. Янковой А.Г. Многомерный статистический анализ в системе STATISTICA. – Одесса, Оптимум. – 2001. Вып.1. – 216 с.