

УДК: 330.4: 332.1

Олена Піскунова  
Ольга ОсиповаOlena Piskunova  
Olha Osypova**КЛАСТЕРНИЙ АНАЛІЗ РЕГІОНАЛЬНИХ ОСОБЛИВОСТЕЙ  
У СПОЖИВАННІ НАСЕЛЕННЯМ ПРОДУКТІВ ХАРЧУВАННЯ****CLUSTER ANALYSIS OF REGIONAL FEATURES  
IN PUBLIC FOODSTUFFS CONSUMPTION**

*З метою виявлення та аналізу регіональних особливостей у споживання населенням України продуктів харчування проведено кластеризацію регіонів за ієрархічним агломеративним методом повних зв'язків та за методом *k*-середніх на основі статистичних даних щодо середньодушових обсягів споживання населенням регіонів базових продуктів харчування. Проведений порівняльний аналіз центрів кластерів та якості розбиття для варіантів двох, трьох і чотирьох кластерів показав, що найбільш прийнятним є поділ регіонів на три групи. Отримані кластери регіонів мають тісний зв'язок з географією та рівнем їх соціально-економічного розвитку: перша група – це промислово розвинені східні регіони та Київська область, другу групу склали центральні регіони з потужним агропромисловим комплексом, в третю групу потрапили західні регіони з менш розвинутою промисловістю. Результати одновимірного та багатовимірного дисперсійного аналізу свідчать про добру якість розбиття регіонів на три кластери. Проаналізовано середньодушові обсяги споживання нутрієнтів за регіонами України.*

**Ключові слова:** споживання базових продуктів харчування, споживання нутрієнтів, регіональні особливості, кластерний аналіз, дисперсійний аналіз.

**Актуальність проблеми.** Розробка ефективних заходів щодо забезпечення продовольчої безпеки на національному та регіональному рівнях передбачає врахування регіональних особливостей у споживанні населенням продуктів харчування, обумовлених соціально-економічними особливостями розвитку регіонів та звичками у харчуванні. Звідси витікає важливість дослідження цих особливостей, зокрема їх аналізу на підґрунті статистичних даних з використанням методів багатовимірної статистики.

**Аналіз останніх джерел досліджень і публікацій.** Над дослідженням проблем продовольчої безпеки працювали такі українські вчені як: О. І. Гойчук, А. С. Лисецький, Б. М. Данилишин, М. М. Одінцов, П. Т. Саблук, але в їх працях недостатньо висвітленими залишаються питання дослідження проблеми продовольчої безпеки на підґрунті методів математичного моделювання. У цьому контексті слід звернути увагу на роботу [1, с. 424–437], де авторами пропонується підхід щодо визначення рівня продовольчої безпеки домогосподарств із застосуванням методів кластерного аналізу. Деякі аспекти класифікації регіонів України на підґрунті методів кластерного аналізу розглянуто у праці [2, с. 82–89], проте питання виявлення регіональних особливостей у споживанні населенням України продуктів харчування на основі статистичних даних із застосуванням методів багатовимірної статистики, аналізу статистичної значимості цих особливостей залишається невирішеним.

**Метою статті** є виявлення і аналіз регіональних особливостей у споживанні населенням України продуктів харчування за допомогою методів кластерного аналізу.

**Викладення основного матеріалу дослідження.** У даному дослідженні аналізується споживання населенням базових продуктів харчування, набір яких визначається діючими нормативними актами України [3, с. 2; 4, с. 2] та включає наступні продукти: м'ясо та м'ясопродукти у перерахунку на м'ясо, включаючи сало і субпродукти (надалі м'ясо); молоко і молочні продукти у перерахунку на молоко (надалі молоко); яйця; рибу і рибопродукти (надалі рибу); цукор; олію; картоплю; овочі та баштанні продовольчі культури (надалі овочі); плоди, ягоди та виноград без переробки на вино (надалі фрукти); хлібні продукти – хліб, макаронні вироби у перерахунку на борошно, борошно, крупи, бобові (надалі хліб або хлібні продукти). Розглядаються обсяги споживання вказаних продуктів харчування за регіонами України у розрахунку на 1 особу. Джерело інформації – дані Державної служби статистики України [5, с. 13–31].

Виявлення регіональних особливостей у споживанні населенням України харчових продуктів, обумовлених різними соціально-економічними передумовами розвитку регіонів та звичками населення у харчуванні, здійснювалось за допомогою методів кластерного аналізу. Ці методи надають можливість дослідникові розподіляти багатовимірні сукупності вхідних даних на однорідні групи так, щоб об'єкти всередині групи були подібними між собою згідно з деяким критерієм, а об'єкти із різних груп відрізнялися один від одного. При цьому класифікація об'єктів проводиться одночасно за декількома ознаками на основі введення певної міри сумарної близькості за всіма ознаками класифікації [6, с. 7]. Попередній аналіз здійснювався із використанням ієрархічних агломеративних методів кластерного аналізу, сутність яких полягає в тому, що процес об'єднання кластерів відбувається послідовно на основі матриці відстаней: на першому кроці кожний об'єкт вибірки розглядається як окремий кластер, на наступному кроці об'єднуються найближчі об'єкти, і т.д., у кінцевому підсумку всі об'єкти будуть об'єднані в один кластер. Послідовність об'єднання легко піддається геометричній інтерпретації і може бути подана у вигляді графа-дерева (дендрограми), на якій вказуються об'єднувальні об'єкти і відстані, за якої відбулось об'єднання.

Дендрограма об'єднання регіонів України за середньодушовими обсягами споживання базових продуктів харчування за 2013 р. наведена на рис. 1. Як алгоритм класифікації тут використовувався метод повних зв'язків, за якого відстані між кластерами визначаються найбільшими відстанями між будь-якими двома об'єктами в різних кластерах (тобто між найбільш віддаленими сусідами). Цей метод зазвичай працює дуже добре, коли об'єкти в реальності поділяються на декілька різних груп [7, с. 106].

Аналіз дендрограми показує, що найбільш чітко виокремлюються три кластери регіонів: один кластер – Київська, Донецька, Дніпропетровська області; другий кластер – Івано-Франківська, Сумська, Рівненська, Хмельницька, Тернопільська, Львівська, Чернівецька і Закарпатська області; третій кластер – усі інші області. Проте можна також сказати, що регіони поділяються на чотири кластери (виокремлюється ще одна група – Полтавська, Житомирська, Чернігівська, Волинська, Черкаська, Миколаївська, Херсонська і Вінницька області) або на два кластери.

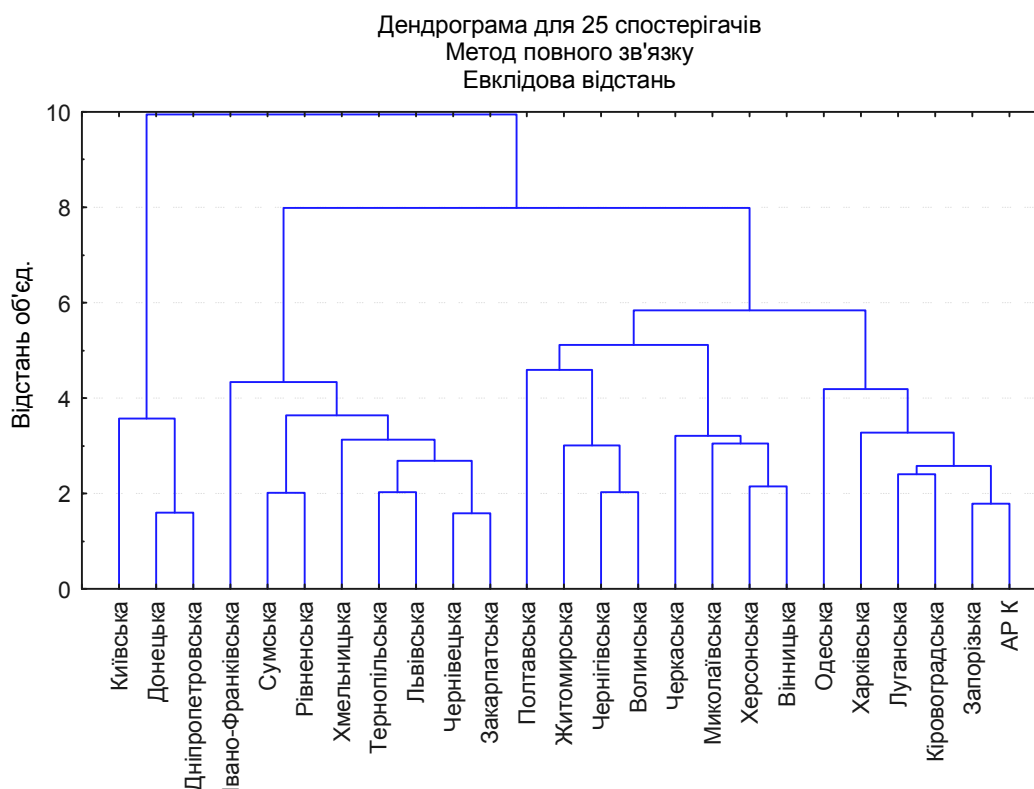


Рис. 1. Дендрограма об'єднання регіонів України за середньодушовими обсягами споживання базових продуктів харчування

Для поглибленого дослідження міжрегіональних відмінностей у споживанні населенням України продуктів харчування застосовувався метод кластерного аналізу  $k$ -середніх. Сутність методу полягає у побудові наперед заданої кількості кластерів  $k$ , таких, що середні значення всіх змінних за об'єктами одного з кластерів максимально відрізняються від середніх значень цих змінних за об'єктами інших кластерів [6, с. 125]. Результати розбиття регіонів України на 2, 3 і 4 кластери представлені у табл. 1. Крім того, у табл. 1 наведено відстані регіонів до центрів кластерів, в які вони потрапили. Для розрахунків використовувались статистичні дані за 2013 р., які було попередньо нормалізовано.

Таблиця 1

Результати розбиття регіонів на кластери за методом  $k$ -середніх

Регіон	Номер кластеру			Відстань до центра кластеру		
	2 кластери	3 кластери	4 кластери	2 кластери	3 кластери	4 кластери
АР Крим	1	1	1	0,64	0,44	0,64
Дніпропетровська	1	1	1	0,41	0,52	0,36
Донецька	1	1	1	0,51	0,60	0,45
Запорізька	1	1	1	0,67	0,51	0,65
Київська	1	1	1	1,34	1,46	1,23
Харківська	2	1	1	0,89	0,76	0,82
Одеська	1	1	2	0,76	0,84	0,47
Луганська	2	1	2	0,73	0,75	0,62
Миколаївська	1	2	2	1,19	0,72	0,76
Кіровоградська	2	1	3	0,65	0,77	0,72
Вінницька	2	2	3	0,83	0,42	0,51
Полтавська	2	2	3	1,04	0,93	0,92
Херсонська	2	2	3	0,91	0,47	0,60
Черкаська	2	2	3	1,27	0,71	0,87
Волинська	2	3	3	0,65	0,67	0,69
Житомирська	2	3	3	0,71	0,76	0,71
Чернігівська	2	3	3	0,71	0,80	0,66
Закарпатська	2	3	4	0,70	0,62	0,59
Івано-Франківська	2	3	4	1,38	1,06	1,00
Львівська	2	3	4	0,66	0,56	0,56
Рівненська	2	3	4	0,51	0,54	0,60
Сумська	2	3	4	0,78	0,71	0,67
Тернопільська	2	3	4	0,79	0,61	0,49
Хмельницька	2	3	4	0,85	0,75	0,66
Чернівецька	2	3	4	0,52	0,56	0,57

З даних, представлених у табл. 1, можна бачити, що регіони з першої «гілки» ієрархічного дерева (Київська, Донецька та Дніпропетровська області) при розбитті на 2, 3 і 4 кластери за методом  $k$ -середніх були віднесені до кластеру 1. Регіони, що утворили другу «гілку» дендрограми (Івано-Франківська, Сумська, Рівненська, Хмельницька, Тернопільська, Львівська, Чернівецька і Закарпатська області), за методом  $k$ -середніх для всіх трьох варіантів розрахунків були віднесені до останнього кластеру. Тобто тут ми спостерігаємо досить стійкі групи регіонів. Трохи інша картина має місце для регіонів з третьої та четвертої гілок ієрархічного дерева. Регіони з третьої гілки дендрограми за розрахунками для першого варіанту двох кластерів були віднесені до другої групи (за винятком Миколаївської області), за розрахунками для чотирьох кластерів – до третьої групи (знову ж за винятком Миколаївської області), а за розрахунками для трьох кластерів перше «гіроно» регіонів (Житомирська, Чернігівська та Волинська області) було віднесено до третьої групи, а друге «гіроно» (Черкаська, Миколаївська, Херсонська і Вінницька області) – до другої групи. Регіони з четвертої гілки у випадку їх розбиття за методом  $k$ -середніх на три кластери потрапили до першої групи, а у випадку ро-

збиття на два і чотири кластери потрапили в різні групи. Тобто об'єднання у кластери регіонів третьої та четвертої гілки менш стійкі.

Для економічного аналізу отриманих груп розглянемо середні за кластерами середньодушові обсяги споживання базових продуктів харчування, які показано на рис. 2 (графіки побудовано для нормалізованих даних).

Аналізуючи графіки, представлені на рис. 2, можна зробити такі висновки:

- у першому випадку, коли регіони розбивались на два кластери, перша отримана група характеризувалась більшими ніж друга середньодушовими обсягами споживання м'яса, риби, фруктів, яєць, овочів і олії, та меншими – молока, хліба, картоплі та цукру;
- у другому випадку, коли регіони розбивались на три кластери, перша та остання групи мали такі ж характеристики, як і у попередньому випадку двох кластерів (за винятком обсягів споживання олії, які виявились однаковими в цих двох кластерах), водночас друга група регіонів характеризувалась найвищими обсягами споживання риби, яєць, овочів, цукру, хліба, та середніми серед трьох груп обсягами споживання м'яса, фруктів, картоплі та молока;
- у третьому випадку, коли регіони розбивались на чотири групи, взаємне розташування графіків центрів першого, останнього та третього кластерів схоже на попередній випадок. Розташування графіку центрів другого кластеру нагадує розташування графіку центрів третього кластеру, проте відрізняється за обсягами споживання риби, які в другому кластері максимальні, та обсягами споживання картоплі та молока, які тут мінімальні.

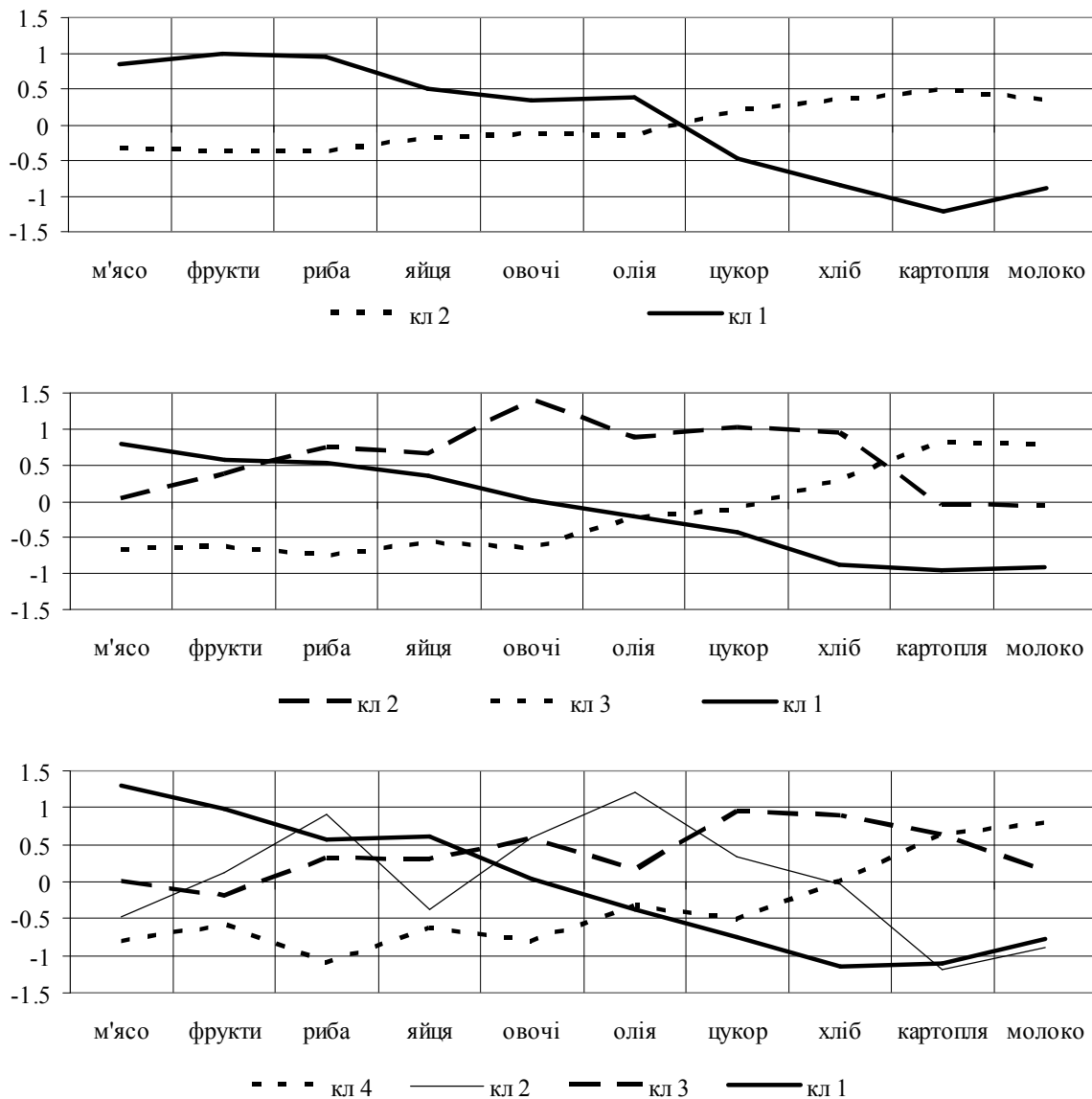


Рис. 2. Центри кластерів регіонів

Для отримання більш повної картини, дослідимо різниці в середніх за кластерами середньодушових обсягах споживання та їх значимість, для чого побудуємо рівняння регресії з фіктивними змінними. Зокрема, для варіанту чотирьох кластерів такі рівняння матимуть вигляд:

$$y_i = a_1 + a_2 \cdot x_{ki} + a_3 \cdot x_{li} + a_4 \cdot x_{ri} + u_i,$$

де  $y_i$  – значення середньодушових обсягів споживання певного продукту харчування в  $i$ -му регіоні;  $x_{ki}$ ,  $x_{li}$  і  $x_{ri}$  – фіктивні змінні, що набувають таких значень:

$$x_{ki} = \begin{cases} 1, & \text{якщо регіон } i \text{ входить до кластеру } k; \\ 0, & \text{в усіх інших випадках,} \end{cases}$$

$$x_{li} = \begin{cases} 1, & \text{якщо регіон } i \text{ входить до кластеру } l; \\ 0, & \text{в усіх інших випадках,} \end{cases}$$

$$x_{ri} = \begin{cases} 1, & \text{якщо регіон } i \text{ входить до кластеру } r; \\ 0, & \text{в усіх інших випадках,} \end{cases}$$

$a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$  – параметри моделі, які необхідно оцінити на основі статистичних даних;  $u_i$  – відхилення фактичного значення показника в  $i$ -му регіоні від середнього по групі. За вказаного вибору фіктивних змінних параметр  $a_1$  є середнім значенням показника у кластері  $j$ , параметр  $a_2$  є різницею між середніми значеннями показника у кластері  $j$  та кластері  $k$ , параметр  $a_3$  – між середніми значеннями показника у кластері  $j$  та кластері  $l$ , а параметр  $a_4$  – між середніми значеннями показника у кластері  $j$  та кластері  $r$ . Якщо, наприклад, оцінка параметра  $a_2$  є значимою, то середнє значення досліджуваної величини у кластері  $k$  значимо відрізняється від середнього значення цієї величини у кластері  $j$ , тобто ці відмінності не є випадковими, а обумовлені факторами, притаманними регіонам, які потрапили у кластер  $k$  [8].

У табл. 2–4 наведено середні за кластерами обсяги споживання базових продуктів харчування та рівні значимості їх різниць у кластері  $i$  та кластері  $j$  при розбитті регіонів на два кластери (табл. 2), три кластери (табл. 3) та чотири кластери (табл. 4).

Таблиця 2

**Середні за кластерами обсяги споживання базових продуктів харчування  
(розбиття регіонів на два кластери)**

Базові продукти харчування	Середні за кластерами середньодушові обсяги споживання		Різниця в середніх за кластерами середньодушових обсягах споживання	$t$ -статистика	$P$ -рівень
	кластер 1	кластер 2			
м'ясо	58,9	51,1	7,82	3,10	0,005
фрукти	61,3	49,7	11,7	3,91	0,001
яйця	314	303	11,3	1,58	<b>0,127</b>
риба	16,5	12,8	3,72	3,68	0,001
овочі	170	162	7,73	1,03	<b>0,315</b>
олія	13,6	13,1	0,480	1,20	<b>0,242</b>
цукор	35,9	38,7	-2,77	1,55	<b>0,135</b>
хлібні продукти	105	115	-10,5	3,07	0,005
картопля	112	156	-44,6	5,81	0,000
молоко	205	233	-27,6	3,30	0,003

Таблиця 3

**Середні за кластерами обсяги споживання базових продуктів харчування  
(розбиття регіонів на три кластери)**

Базові продукти харчування	Середні за кластерами середньодушові обсяги споживання			Різниця в середніх обсягах споживання у кластері <i>i</i> та кластері <i>j</i>			Рівень значимості різниці в середніх обсягах споживання у кластері <i>i</i> та кластері <i>j</i>		
	кл. 1	кл. 2	кл. 3	<i>i</i> =3, <i>j</i> =1	<i>i</i> =3, <i>j</i> =2	<i>i</i> =2, <i>j</i> =1	<i>i</i> =3, <i>j</i> =1	<i>i</i> =3, <i>j</i> =2	<i>i</i> =2, <i>j</i> =1
м'ясо	58,6	53,5	48,8	9,77	4,71	5,06	0,000	<b>0,101</b>	0,089
фрукти	57,9	56,0	47,5	10,4	8,51	1,86	0,004	0,038	<b>0,646</b>
яйця	311	317	296	15,3	20,4	-5,2	0,028	0,016	<b>0,529</b>
риба	15,3	15,9	11,7	3,60	4,16	-0,567	0,001	0,001	<b>0,635</b>
овочі	165	188	154	11,2	34,5	-23,4	0,037	0,000	0,001
олія	13,0	14,0	13,0	0,008	1,00	-1,00	<b>0,983</b>	0,038	0,046
цукор	36,2	42,1	37,5	-1,34	4,62	-5,96	<b>0,421</b>	0,028	0,008
хлібні продукти	104	121	115	-10,5	5,75	-16,2	0,002	<b>0,117</b>	0,000
картопля	118	142	165	-46,8	-22,8	-24,0	0,000	0,018	0,016
молоко	205	224	242	-37,4	-18,5	-18,9	0,000	0,032	0,035

Таблиця 4

**Середні за кластерами обсяги споживання базових продуктів харчування  
(розбиття регіонів на чотири кластери)**

Базові продукти харчування	Середні за кластерами середньодушові обсяги споживання				Рівень значимості різниці в середніх обсягах споживання у кластері <i>i</i> та кластері <i>j</i>					
	кл. 1	кл. 2	кл. 3	кл. 4	<i>i</i> =4, <i>j</i> =2	<i>i</i> =4, <i>j</i> =3	<i>i</i> =4, <i>j</i> =1	<i>i</i> =2, <i>j</i> =3	<i>i</i> =2, <i>j</i> =1	<i>i</i> =1, <i>j</i> =3
м'ясо	61,9	50,1	53,3	48,0	<b>0,453</b>	0,017	0,000	<b>0,260</b>	0,001	0,001
фрукти	61,4	53,9	51,3	48,0	<b>0,233</b>	<b>0,361</b>	0,002	<b>0,597</b>	<b>0,151</b>	0,016
риба	15,5	16,4	14,8	10,8	0,000	0,000	0,000	<b>0,192</b>	<b>0,471</b>	<b>0,474</b>
яйця	316	300	311	296	0,020	0,055	<b>0,685</b>	<b>0,287</b>	<b>0,142</b>	<b>0,537</b>
овочі	165	174	174	151	0,023	0,004	0,078	<b>0,968</b>	<b>0,363</b>	<b>0,257</b>
олія	12,9	14,3	13,4	12,9	0,022	<b>0,291</b>	<b>0,934</b>	<b>0,108</b>	0,024	<b>0,290</b>
цукор	34,8	39,3	41,9	35,8	0,001	0,102	<b>0,534</b>	<b>0,217</b>	0,046	0,000
хлібні продукти	102	112	120	112	<b>0,961</b>	0,017	0,006	0,063	0,032	0,000
картопля	114	113	161	161	0,000	<b>0,997</b>	0,000	0,000	<b>0,852</b>	0,000
молоко	208	205	228	242	0,005	<b>0,112</b>	0,001	0,067	<b>0,834</b>	0,044

Як бачимо з табл. 2–4, найкраща якість розбиття спостерігається для варіанта трьох кластерів. Тут різниці в середніх обсягах споживання у першому та третьому кластерах значимі майже для всіх продуктів за винятком олії та цукру, – у другому та третьому кластерах значимі майже для всіх продуктів за винятком м'яса і хлібних продуктів, – у першому та другому кластерах значимі майже для всіх продуктів за винятком фруктів, яєць і риби. Для варіанта чотирьох кластерів різниці в середніх обсягах споживання у другому та третьому кластерах майже для всіх продуктів є не значимими за винятком картоплі, молока і хлібних продуктів, що відповідає висновкам, зробленим при аналізі графіків центрів кластерів.

Про загальну якість розбиття об'єктів на групи дозволяють судити методи дисперсійного аналізу. При проведенні одновимірного дисперсійного аналізу обсяги споживання кожного продукту харчування розглядаються окремо. Для досліджуваних показників розраховується співвідношення між груповою  $MS_{BG}$  та внутрішньо груповою  $MS_{WG}$  дисперсій:  $F_{\delta i \zeta \delta} = MS_{BG} / MS_{WG}$ , яке має  $F$ -розподіл.

Якщо розрахункове значення  $F$ -статистики перевищує критичне:  $F_{\delta\zeta\delta} > F_{\alpha, m-1, n-m}$  ( $\alpha$  – обраний рівень значимості,  $n$  – кількість регіонів,  $m$  – кількість кластерів), то відмінності у середніх значеннях обсягів споживання в отриманих кластерах від середнього значення за всією вибіркою можна вважати значимими [9, с. 646]. Розраховані значення  $F$ -статистики для розглянутих продуктів харчування для варіантів розбиття регіонів на 2, 3 і 4 кластери наведено у табл. 5.

Таблиця 5

## Результати одновимірного дисперсійного аналізу

Базові продукти харчування	2 кластери		3 кластери		4 кластери	
	$F$ -статистика	$p$ -рівень	$F$ -статистика	$p$ -рівень	$F$ -статистика	$p$ -рівень
м'ясо	9,6	0,005	9,1	0,001	13,9	0,000
молоко	10,9	0,003	15,4	0,000	6,0	0,004
яйця	2,5	<b>0,127</b>	4,5	0,023	2,6	0,078
хлібні продукти	9,4	0,005	11,5	0,000	10,0	0,000
картопля	33,8	0,000	20,0	0,000	20,6	0,000
овочі	1,1	<b>0,315</b>	16,3	0,000	4,1	0,019
фрукти	15,3	0,001	5,8	0,010	4,3	0,017
риба	13,6	0,001	10,1	0,001	11,7	0,000
цукор	2,4	<b>0,135</b>	4,4	0,024	8,3	0,001
олія	1,4	<b>0,242</b>	2,8	0,082	2,5	0,087

З табл. 5 ми бачимо, що у випадку двох кластерів відмінності у середніх значеннях обсягів споживання в кластерах від середнього значення за всією вибіркою не можна вважати значимими для таких продуктів харчування як яйця, овочі, цукор, олія. У випадку трьох і чотирьох кластерів відмінності у середніх значеннях обсягів споживання в отриманих групах від середнього значення за всіма регіонами можна вважати значимими при рівні значимості, меншим, ніж 0,1.

Зазначимо, що при наявності кореляційного зв'язку між залежними змінними застосовують багатовимірний дисперсійний аналіз, в якому досліджувані характеристики об'єктів розглядають одночасно [10, с. 2]. Для аналізу кореляційного зв'язку між обсягами споживання базових продуктів харчування побудовано матрицю частинних коефіцієнтів кореляції (див. табл. 6).

Таблиця 6

## Матриця частинних коефіцієнтів кореляції

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	
$R^* =$	1,00	-0,08	-0,14	0,50	-0,06	-0,16	-0,58	0,04	-0,47	0,36	$X_1$
	-0,08	1,00	0,12	-0,15	-0,48	-0,02	-0,30	0,39	0,11	-0,16	$X_2$
	-0,14	0,12	1,00	0,33	-0,54	-0,51	-0,34	-0,14	-0,02	0,25	$X_3$
	0,50	-0,15	0,33	1,00	-0,39	-0,63	-0,46	0,24	-0,62	0,22	$X_4$
	-0,06	-0,48	-0,54	-0,39	1,00	0,62	0,50	-0,13	-0,02	-0,09	$X_5$
	-0,16	-0,02	-0,51	-0,63	0,62	1,00	0,34	-0,20	0,10	-0,14	$X_6$
	-0,58	-0,30	-0,34	-0,46	0,50	0,34	1,00	-0,40	0,53	-0,38	$X_7$
	0,04	0,39	-0,14	0,24	-0,13	-0,20	-0,40	1,00	-0,39	-0,28	$X_8$
	-0,47	0,11	-0,02	-0,62	-0,02	0,10	0,53	-0,39	1,00	-0,36	$X_9$
	0,36	-0,16	0,25	0,22	-0,09	-0,14	-0,38	-0,28	-0,36	1,00	$X_{10}$

Перевірка значимості наведених у табл. 6 частинних коефіцієнтів кореляції за допомогою  $t$ -критерію показала наявність значимого зв'язку між такими парами змінних: хліб і м'ясо; фрукти і м'ясо; картопля і яйця; овочі і яйця; овочі і хліб; цукор і хліб; овочі і картопля; фрукти і картопля; цукор і фрукти. Тому для аналізу якості отриманих кластерів було застосовано багатовимірні критерії, які ґрунтуються на обчисленні визначника добутку двох матриць – матриці сум квадратів і коефіцієнтів коваріації  $\mathbf{H}$  і оберненої матриці  $\mathbf{E}^{-1}$  залишкових сум квадратів і коефіцієнтів коваріації:  $\mathbf{H} \times \mathbf{E}^{-1}$ . Як відомо, визначник матриці є мірою загальної варіабельності даних або дисперсії усїєї матриці. Визначник матриці може бути обчислений як добуток власних значень матриці  $\lambda_i$  ( $i = \overline{1, s}$ ,  $s$  – кількість ненульових власних значень матриці), оскільки кожне власне значення  $\lambda_i$  є, в свою чергу, часткою загальної дисперсії. Для багатовимірної оцінки відмінностей між середніми розраховано такі статистичні показники [10, с. 7]:

- критерій Піллая (Pillai's trace):  $V = \sum_{i=1}^s \frac{1}{1 + \lambda_i}$ ;
- критерій Уїлкса (Wilks's lambda):  $W = \prod_{i=1}^s \frac{1}{1 + \lambda_i}$ ;
- критерій Хотеллінга (Hotelling trace):  $V = \sum_{i=1}^s \lambda_i$ ;
- критерій Роя (Roy's largest root):  $R = \frac{\lambda_{max}}{1 + \lambda_{max}}$ .

Для кожного з цих багатовимірних критеріїв побудовані власні розподіли, проте усі вони можуть бути перетворені у статистики, які добре апроксимуються  $F$ -розподілом [10, с. 8]. Розраховані за допомогою програмного пакету Statistica значення багатовимірних критеріїв дисперсійного аналізу наведено у табл. 7. За всіма розглянутими критеріями можна зробити висновок про добру якість розбиття регіонів на кластери.

Таблиця 7

## Результати багатовимірної дисперсійної аналізу

Критерій	2 кластери			3 кластери			4 кластери		
	значення	$F$ -статистика	$P$ -рівень	значення	$F$ -статистика	$P$ -рівень	значення	$F$ -статистика	$P$ -рівень
Уїлкса	0,29	3,48	0,017	0,03	6,32	0,000	0,00	6,28	0,000
Піллая	0,71	3,48	0,017	1,63	6,11	0,000	2,44	6,16	0,000
Хотеллінга	2,49	3,48	0,017	10,80	6,48	0,000	16,73	5,95	0,000
Роя	2,49	3,48	0,017	7,98	11,18	0,000	8,28	11,60	0,000

Таким чином, проведений порівняльний аналіз результатів кластеризації регіонів України за обсягами споживання базових продуктів харчування показав, що найбільш прийнятним є розбиття регіонів на три кластери: розбиття на два кластери не вловлює деякі суттєві регіональні особливості у споживанні базових продуктів харчування, водночас розбиття регіонів на чотири кластери нагромаджує подробиці у характеристиці кластерів, які затушовують загальну картину. Крім того, аналіз значимості різниць в середніх обсягах споживання показав, що для варіанту розбиття регіонів на 4 кластери дані різниці у другому та третьому кластерах майже для всіх продуктів є не значимими, а за результатами одновимірної дисперсійної аналізу найгірша якість розбиття відповідає варіанту двох кластерів.

За результатами дослідження можна сказати, що регіони України поділяються на три групи, які мають тісний зв'язок з географією та рівнем їх соціально-економічного розвитку: перша група – це



промислово розвинені східні регіони та Київська область, в третю групу потрапили західні регіони з менш розвинутою промисловістю, а другу групу склали центральні регіони з потужним агропромисловим комплексом. Перша група регіонів характеризується найвищими обсягами споживання м'яса та фруктів, та найнижчими обсягами споживання цукру, хліба, картоплі та молока; друга група – найвищими обсягами споживання риби, яєць, овочів, олії, цукру та хлібних продуктів, третя група – найвищими обсягами споживання картоплі та молока, і найнижчими – м'яса, фруктів, риби, яєць, овочів. Можна зробити припущення, що дані специфічні особливості обумовлені в основному двома чинниками: спеціалізацією регіонів у виробництві певних видів продуктів харчування та величиною доходів населення. Так, 1-й кластер утворений регіонами з найвищим рівнем доходів по Україні, що і відображається на структурі споживання: високі обсяги споживання дорогих продуктів та низькі – дешевих. Регіони, що формують 2-й кластер, вирізняються сільськогосподарським напрямом своєї діяльності, результатом чого і є високі або середні (але не найнижчі) обсяги споживання продуктів харчування. І в 3-му кластері, що, як правило, представлений економічно найменш розвиненими регіонами, обсяги споживання дорогих продуктів харчування знаходяться на найнижчому рівні, а помітно високі обсяги споживання молока та картоплі пояснюються тим, що в силу природно-кліматичних умов даних регіонів ці продукти харчування є традиційними видами виробництва.

Важливою характеристикою споживання населенням харчових продуктів, є обсяги споживання нутрієнтів (основних поживних речовин), що являють собою складові частини харчових продуктів, які організм використовує для побудови, оновлення та нормального функціонування органів, тканин і клітин, а також як джерело енергії для виконання роботи і забезпечення життєдіяльності організму в період спокою [11]. Перерахунок продуктів харчування у нутрієнти було здійснено на основі даних щодо вмісту основних поживних речовин (нутрієнтів) у продуктах [12, с. 12–104]. У табл. 8 наведено середні за кластерами обсяги споживання нутрієнтів та рівні значимості їх різниць у кластері  $i$  та кластері  $j$  при розбитті регіонів на три кластери.

Таблиця 8

**Середні за кластерами обсяги споживання нутрієнтів  
(розбиття регіонів на три кластери)**

Нутрієнт	Середні за кластерами середньодушові обсяги споживання нутрієнтів			Рівень значимості різниці в середніх обсягах споживання нутрієнтів у кластері $i$ та кластері $j$		
	кл. 1	кл. 2	кл. 3	$i=3, j=1$	$i=3, j=2$	$i=2, j=1$
Білки, г	29761	31620	30107	<b>0,489</b>	0,018	0,006
Жири, г	34358	35249	33360	<b>0,166</b>	0,034	<b>0,314</b>
Вуглеводи, г	132549	152444	145410	0,000	0,069	0,000
Енерг, ккал	992462	1092245	1040900	0,004	0,010	0,000
Натрій, мг	508983	565954	542611	0,000	0,017	0,000
Калій, мг	1631668	1825375	1834982	0,000	<b>0,803</b>	0,000
Кальцій, мг	364018	400632	405970	0,000	<b>0,579</b>	0,001
Магній, мг	178102	199818	190950	0,000	0,009	0,000
Фосфор, мг	563951	617473	613866	0,000	<b>0,740</b>	0,000
Залізо, мг	7406	8088	7622	0,057	0,002	0,000
Вітамін А, мг	314	308	298	0,097	<b>0,369</b>	<b>0,614</b>
Каротин, мг	1526	1740	1434	0,051	0,000	0,001
Вітамін В1, мг	616	681	673	0,000	<b>0,521</b>	0,000
Вітамін В2, мг	840	907	879	0,019	<b>0,145</b>	0,002
Вітамін РР, мг	6874	7361	7042	<b>0,192</b>	0,043	0,005
Вітамін С, мг	52852	53961	47292	0,012	0,012	<b>0,664</b>

Як бачимо з табл. 8, перший кластер характеризується найнижчими обсягами споживання майже усіх груп нутрієнтів (за винятком жирів, каротину, та вітамінів А і С), а найвищі обсяги спожи-

вання нутрієнтів мають місце у кластері 2 (за винятком калію, кальцію та вітаміну А). Практично для всіх груп нутрієнтів відмінності в середніх обсягах їх споживання в різних кластерах є значимими (на рівні значимості меншим за 0,1). Зокрема, різниця у середніх рівнях споживання в першому та третьому кластерах є значимою для всіх груп нутрієнтів за виключенням білків, жирів та вітаміну РР. Значимі відмінності між першим та другим кластерами було отримано для всіх груп нутрієнтів, окрім жирів, вітамінів А та С. Середні обсяги споживання нутрієнтів у другому та третьому кластерах виявились більш близькими, їх різниці є незначимими для калію, кальцію, фосфору, вітамінів А, В1 та В2. Таким чином ми бачимо, що виявлені міжрегіональні відмінності у споживанні населенням базових продуктів харчування суттєвим чином позначаються на структурі споживання нутрієнтів.

**Висновки.** З метою виявлення та аналізу регіональних особливостей у споживання населенням України продуктів харчування на підґрунті статистичних даних за 2013 рік проведено кластеризацію регіонів за ієрархічним агломеративним методом кластерного аналізу повних зв'язків, який показав що у структурі регіонів найбільш чітко виокремлюються три кластери, проте можна казати, також, про групування у два або чотири кластери. Проведений порівняльний аналіз центрів кластерів, отриманих за методом k-середніх, та якості розбиття для варіантів двох, трьох і чотирьох кластерів показав, що найбільш прийнятним є поділ регіонів на три групи. Отримані кластери регіонів мають тісний зв'язок з географією та рівнем їх соціально-економічного розвитку: перша група – це промислово розвинені східні регіони та Київська область, другу групу склали центральні регіони з потужним агропромисловим комплексом, в третю групу потрапили західні регіони з менш розвинутою промисловістю. Перша група регіонів характеризується найвищими обсягами споживання м'яса та фруктів, та найнижчими обсягами споживання цукру, хліба, картоплі та молока; друга група – найвищими обсягами споживання риби, яєць, овочів, олії, цукру та хлібних продуктів, третя група – найвищими обсягами споживання картоплі та молока, і найнижчими – м'яса, фруктів, риби, яєць, овочів. Результати одновимірного та багатовимірного дисперсійного аналізу свідчать про добру якість розбиття регіонів на три кластери. Аналіз середньодушових обсягів споживання нутрієнтів за регіонами України показав, що виявлені міжрегіональні відмінності у споживанні населенням базових продуктів харчування суттєвим чином позначаються на структурі споживання нутрієнтів.

Подальше дослідження проблем, висвітлених у даній статті, вбачаємо передусім у проведенні процедури кластерного аналізу регіонів України в динаміці та у визначенні факторів, що обумовлюють регіональні особливості у споживанні населенням продуктів харчування, на підґрунті методів економетричного моделювання.

#### Список використаної літератури

1. Suresh C. B. Food security, poverty and nutrition policy analysis: Statistical methods and applications / C. B. Suresh, N. G. Shailendra, S. Prabuddha. – 2<sup>nd</sup> ed. – Academic Press, 2014.
2. Піскунова О. В. Кластерний аналіз у моделюванні продовольчої безпеки на регіональному рівні / О. В. Піскунова, О. І. Осипова // Вісник Київського національного університету імені Тараса Шевченка. – К. : Київський університет, 2014. – № 9 (162). – С. 82–89. – (Економіка).
3. Закон України «Про продовольчу безпеку» № 4227 – VI від 22.12.2011 року.
4. Постанова «Про затвердження наборів продуктів харчування, наборів непродовольчих товарів та наборів послуг для основних соціальних і демографічних груп населення» № 656 від 14.04.2000 року.
5. Баланси та споживання основних продуктів харчування населенням України. Статистичний збірник / [Відп. за вип. О. М. Прокопенко]. – К. : Державна служба статистики України, 2014. – 57 с.
6. Everitt B. (2011) Cluster analysis / B. Everitt, S. Landau, M. Leese, D. Stahl. – 5th ed. – Chichester : Wiley, 2011.
7. Noor R. R. A comparison between single linkage and complete linkage in agglomerative hierarchical cluster analysis for identifying tourists segment / R. R. Noor, A. Sabri, M. Safiek // Engineering Journal. – Vol. 12. – 2011. – № 6. – Pp. 105–116.
8. Осмирко Я. О. Исследование сезонных колебаний экспорта Российской Федерации природного газа в страны дальнего зарубежья за 2008-2013 годы с применением метода фиктивных переменных в эконометрическом анализе / Я. О. Осмирко, Л. О. Бабешко // Материали VII Международной студенческой электронной научной конференции «Студенческий научный форум» [Электронный ресурс]. – Режим доступа : <http://www.scienceforum.ru/2015/802/8606>. – Назва з екрану (13.04.2015).
9. Левин Д. М. и др. Статистика для менеджеров с использованием Microsoft Excel / Левин, Дэвид М., Стефан, Дэвид, Кребийл, Тимоти С, Беренсон, Марк Л ; пер. с англ. – 4-е изд. – М. : «Вильямс», 2004. – 1312 с.

10. French A. A. Multivariate analysis of variance (MANOVA) / A. French, M. Macedo, J. Poulsen, T. Yu. Waterson, [Електронний ресурс]. – Режим доступу : <http://www.pdflibrary.org/pdf/multivariate-analysis-of-variance-manova.html>. – Назва з екрану (13.04.2015).
11. Фармацевтична енциклопедія України [Електронний ресурс]. – Режим доступу : <http://www.pharmencyclopedia.com.ua/article/1285/nutriciologiya>. – Назва з екрану (13.04.2015).
12. Химический состав пищевых продуктов: Книга 1: Справочные таблицы содержания основных пищевых веществ и энергетической ценности пищевых продуктов / Под ред. проф., д-ра техн. наук И. М. Скурихина, проф., д-ра мед. наук М. Н. Волгарева. – 2-е изд., перераб. и доп. – М. : Агропромиздат, 1987. – 224 с.

### References

1. Suresh, C. B., Shailendra, N. G. and Prabuddha, S. (2014) Food security, poverty and nutrition policy analysis: Statistical methods and applications. 2<sup>nd</sup> ed. Academic Press.
2. Piskunova, O. V. and Osypova, O. I. (2014) Cluster analysis in modeling food security at the regional level. *Visnyk Kyivskoho natsionalnoho universytetu imeni Tarasa Shevchenka*. Kyiv: Kyivskiy universitet, 9 (162), pp. 82–89. (Economika).
3. Law of Ukraine «On food security» (2011) № 4227–VI. Kyiv: Holos Ukrayiny.
4. Resolution «On approval of food sets, sets of sets of non-food products and services for major social and demographic groups» (2000) № 656.
5. Prokopenko, O. M. (2014). Balances and consumption of basic foodstuffs population of Ukraine. Kyiv: Derzhavna sluzhba statystyky Ukrainy, p. 57.
6. Everitt, B., Landau, S., Leese, M. and Stahl, D. (2011) Cluster analysis. 5th ed. Chichester: Wiley, p. 330.
7. Noor, R. R., Sabri, A. and Safiek, M. (2011) A comparison between single linkage and complete linkage in agglomerative hierarchical cluster analysis for identifying tourists segment. *Engineering Journal*, 12 (6), pp. 105–116.
8. Osmyrko, Ya. O. and Babeshko, L. O. (2015) The study of seasonal fluctuations of exports of Russian Federation of natural gas in foreign countries in 2008–2013, using the method of dummy variables in econometric analysis [Internet]. Available from: <<http://www.scienceforum.ru/2015/802/8606>> [accessed 13 April 2015].
9. Levine, D., Stephan, D., Krehbiel, T. and Berenson, M. (2005) Statistics for managers using Microsoft Excel. 4<sup>th</sup> ed. Moscow: Williams, 1312 p.
10. French, A., Macedo, M., Poulsen, J. and Waterson, T. Yu (2008) Multivariate analysis of variance (MANOVA) [Internet]. Available from: <<http://www.pdflibrary.org/pdf/multivariate-analysis-of-variance-manova.html>> [accessed 13 April 2015]
11. Pharmaceutical encyclopedia of Ukraine [Internet]. Available: from: <<http://www.pharmencyclopedia.com.ua/article/1285/nutriciologiya>> [accessed 13 April 2015].
12. Chemical composition of foods (1987). In : Y. M. Skurykhyn, M. N. Volharev (Eds.). 2<sup>nd</sup> ed. Moscow: Ahropro-myzdats, 224 p.

**Olena Piskunova, Olha Osypova**

### CLUSTER ANALYSIS OF REGIONAL FEATURES IN PUBLIC FOODSTUFFS CONSUMPTION

*One of the points in the formation of an effective strategy to ensure food security at the national and regional levels is to pay attention to the regional differences in the consumption of food due to socio-economic characteristics of regions and nutritional habits. Studies of regional features can be performed using the methods of multivariate statistics, in particular cluster analysis.*

*The purpose of this study is the identification and analysis of regional features in Ukraine consumption of food by means of cluster analysis.*

*The initial data for the study are the State Statistics Service of Ukraine data on average per basic consumption of food the population of Ukraine in 2013.*

*First hierarchical clustering procedure for complete connection method was conducted. Visual analysis of the results led to the assumption that the most likely options for partitioning regions will be split into 2, 3 or 4 clusters. Under this assumption partitioning regions was performed at 2, 3 and 4 clusters by using k-means. The comparative analysis of cluster centers that were derived by the method of k-means and quality partitioning options for two, three and four clusters showed that the most appropriate was the division of regions into three groups. The resulting clusters of regions have a close relationship with their geography*

*and the level of socio-economic development: the first group is the group of industrialized eastern regions and Kyiv region, the second group consisted of central regions with developed agriculture, the third group includes western regions with less developed industry. The division into 2 clusters does not reflect some significant regional differences in consumption of basic foodstuffs, while division into 4 clusters of regions cannot give a clear economic interpretation of the clusters. In addition, the results of univariate and multivariate analysis of variance indicate good quality partition regions into three clusters.*

*Analysis of average consumption of nutrients by regions of Ukraine has showed that revealed regional differences in consumption of basic foodstuffs materially affected the structure of consumption of nutrients.*

**Keywords:** *consumption of basic foodstuffs, consumption of nutrients, regional features, cluster analysis, analysis of variance.*

*Статтю представляє д.е.н. О. В. Піскунова, професор кафедри економіко-математичного моделювання, ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана».*