

УДК 338:515.1

МЕТОДИКА ЗАСТОСУВАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ІДЕНТИФІКАЦІЇ  
ОБРАЗІВ СПОЖИВАЧІВ ОСВІТНІХ ПОСЛУГ

CLUSTER ANALYSIS APPLICATION FOR IDENTIFICATION IMAGES OF  
EDUCATION SERVICES USERS

МЕТОДИКА ПРИМЕНЕНИЯ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ  
ИДЕНТИФИКАЦИИ ОБРАЗОВ ПОТРЕБИТЕЛЕЙ ОБРАЗОВАТЕЛЬНЫХ УСЛУГ

Войтович С.Я., к.е.н., доцент; Букало Н.А.

Serhiy Y. Voityovych -Doctor of Philosophy (Economics), Associate Professor  
Nadia A. Bukalo – Teacher

Войтович С.Я., к.э.н., доцент; Букало Н.А.

*Анотація:* З метою формування образів споживачів освітніх послуг та виявлення їх споживацької поведінки в роботі представлений теоретичний огляд різноманітних методів розбиття сукупностей об'єктів на однорідно-стійкі групи. Досліджено сутність і практично реалізовано послідовність етапів проведення кластерного аналізу ідентифікації образів споживачів освітніх послуг.

With the purpose of forming of educational services consumers' images and exposure of their consumer conduct, the theoretical review of various methods of division of objects aggregates into homogeneously proof groups is presented in the work. The essence and the sequence of the stages of cluster analysis conducting of identification of educational services consumers' images is practically realized and investigated.

*Аннотация:* С целью формирования образов потребителей образовательных услуг и выявления их потребительского поведения в работе представлен теоретический обзор разнообразных методов разбивки совокупности объектов на однородно стойкие группы. Исследовано сущность и практически реализована последовательность этапов проведения кластерного анализа идентификации образов потребителей образовательных услуг.

*Ключові слова:* кластери, кластерний аналіз, метод деревовидної кластеризації.

*Key Notions:* clusters, cluster analysis, method of of derevopodibnoy clusterization.

*Ключевые слова:* кластеры, кластерный анализ, метод деревовидной кластеризации.

*Постановка наукової проблеми та її значення.* Для вищого навчального закладу освіти одним із основних напрямів маркетингової діяльності є вивчення ринку освітніх послуг, відповідних потреб і настроїв потенційних споживачів, а також факторів, що впливають на попит освітянських послуг та вибір освітнього закладу, спеціальності, форми навчання.

Перспективи змін місткості ринку освітніх послуг потребує особливої уваги з боку вищого менеджменту закладів освіти [1, с. 48].

Нині вищі навчальні заклади повинні завойовувати прихильність у свідомості споживачів освітніх послуг. Вони повинні знайти свій шлях до повного розуміння споживача. «Молоді» вузи мають створювати власні бренди, ефективно позиціонувати себе. Як свідчить практика маркетингу послуг, вдала назва та правильне позиціонування суттєво вплинуть на рівень популярності товарів і послуг [2, с. 256], тому актуальним на сьогодні є для дослідження ринкових факторів, що впливають на формування потреб у задоволенні освітніх послуг. Споживання освітніх послуг знаходиться в залежності від типу споживацької поведінки, яка в свою чергу визначається окремим образом споживача. Виявлення образів споживачів освітніх послуг та їх споживацької поведінки доцільно проводити із використанням математичних методів, серед яких доцільно виокремити методи кластерного аналізу.

*Аналіз останніх досліджень із цієї проблеми.* Дослідженню проблем ринку освітніх послуг та діяльності ВНЗ присвячені роботи Д.Макдугалла, Д.Маккорта, А.Б.Дайновського, В.Н.Зотова, С.М.Ілляшенка, О.А.Кратта, Т.Є.Оболенської, А.П.Панкрухіна, В.С.Сотнікова, Д.А.Штефанича та ін. Проте в працях згаданих науковців дослідження ринку освітніх послуг в більшій мірі розглядається в контексті елементів комплексу маркетингу і на цих засадах здійснюється аналіз діяльності ВНЗ. Однак, недостатньо уваги приділено питанням прогнозування обсягів споживання освітніх послуг на засадах виявлення і формування образів споживачів освітніх послуг та їх споживацької поведінки.

*Формулювання мети та завдань статті.* Для досягнення поставлених у дослідженні цілей необхідно вирішити наступні завдання:

- провести теоретичний розгляд різних типів розбиття сукупностей об'єктів на однакові класи, вибір найбільш доцільного способу розрахунку відстаней між ними, з точки зору, практичної реалізації;

- дослідити послідовність та сутність етапів проведення кластерного аналізу з метою ідентифікації образів споживачів освітніх послуг;

- провести практичну реалізацію розбиття споживачів освітніх послуг на стійкі кластери із застосуванням методу розрахунку Евклідової відстані.

*Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.* Методика кластерного аналізу базується на поняттях подібності об'єктів або їх ознак. З допомогою підбору найбільших «подібних» одиниць виконується розподіл сукупності на кластери (групи). Кластерний аналіз дозволяє сегментувати споживачів, об'єднуючи їх у групи з однаковим врахуванням вибраних для класифікації ознак, або групувати ознаки, тобто характеристики споживачів, в окремі кластери.

За твердженням авторів, Лапач С. Н., Чубенко А. В., Бабич П. Н. кластерний аналіз це – сукупність багатомірних статистичних процедур, які дозволяють впорядкувати об'єкти на однорідні групи [3, с. 213].

У ієрархічній кластеризації об'єкти об'єднуються один з одним через послідовні кроки, що у результаті приводить до побудови «дерева», яке містить усі об'єкти. Підсумкова деревовидна діаграма виразно представляє гілки або групи об'єктів, які більш схожі один з одним, ніж інші. Таким чином, деякі природні структури об'єктів можуть бути виразно представлені групами в результаті проведення такого аналізу.

Методи кластерного аналізу дозволяють проводити різноманітні маніпуляції із сукупностями об'єктів, залежно від поставлених у дослідженні цілей. Так, метод середніх дозволяє заздалегідь визначити необхідну кількість кластерів, а математичний апарат виконує функцію пошуку якнайкращого розбиття об'єктів на необхідне число кластерів. В ході обчислень пропонується застосування різноманітних статистичних оцінок і досліджень з метою з'ясування рівня

досягнення відповідної кластеризації, тобто чи дійсно елементи усередині кожного кластера схожі між собою більше, ніж елементи з інших кластерів.

При двовхідному об'єднанні одночасно кластеризуються і спостереження, і змінні. Математичний апарат виконує побудову кластерів з однаковими даними (значеннями).

Для проведення кластерного аналізу можна використати метод дендритів. Дендрит – це ламана лінія, якою об'єднанні два будь-яких об'єкти з обраної множини (або дві будь-які ознаки об'єкта, якщо групуються ознаки об'єктів). Дендрит може розгладжуватися, але не може містити замкнених ламаних ліній [4, с.119].

Кластерний аналіз дозволяє обчислювати різні типи відстаней між об'єктами, формуючи матриці відстаней незалежно і використовувати їх безпосередньо в процедурі об'єднання [3, с. 213– 218; 5].

Процедура об'єднання визначає необхідність з'ясування правила об'єднання або зв'язку для двох кластерів. Використовують наступні варіанти: односторонній зв'язок (метод найближчого сусіда); повний зв'язок (метод найбільш віддалених сусідів); незважене попарне середнє (відстань розраховується, як середня відстань між усіма парами об'єктів); зважене попарне середнє (метод незваженого попарного середнього, за винятком того, при якому розмір відповідних кластерів використовується як ваговий коефіцієнт); незважений центроїдний метод (відстань між двома кластерами визначається як відстань між їх центрами тяжкості); зважений центроїдний метод (ідентичний попередньому, за винятком того, при якому використовуються питомі ваги для обліку різниці між розмірами кластерів; метод Варда (відрізняється від усіх інших методів, оскільки він використовує методи дисперсійного аналізу для оцінки відстаней між кластерами).

Як було раніше зазначено, визначення подібності між об'єктами здійснюється шляхом розрахунку відстаней між ними. Чим менша відстань, тим більше є подібними ці об'єкти. Щоб бути придатною для визначення відстані, міра повинна характеризуватися такими властивостями: симетрією, мінімальною відстанню

самого об'єкта до центру ваги, монотонними змінами в описовому просторі та змістовною інтерпретацією міри.

В кластерному аналізі використовують такі міри відстаней:

*Квадрат евклідової відстані.* Зводить до квадрату стандартну евклідову відстань, щоб надати більшої ваги віддаленішим один від одного об'єктам. Ця відстань розраховується таким чином :

$$(x,y) = \sum_i (x_i - y_i)^2 \quad (1)$$

*Відстань міських кварталів (манхеттенська відстань).* Ця відстань є просто середньою різниць по координатах. В більшості випадків ця міра відстані спричиняє такі ж результати, як і звичайна Евкліда відстань. Проте відзначимо, що для цієї міри вплив окремих великих різниць (викидів) зменшується. Манхеттенська відстань розраховується за формулою:

$$(x,y) = \sum_i |x_i - y_i| \quad (2)$$

*Відстань Чебишева.* Ця відстань застосовується у випадку, коли необхідно визначити два об'єкти як "різні", якщо вони розрізняються за якою-небудь однією координатою (яким-небудь одним вимірюванням). Відстань Чебишева розраховується за формулою:

$$(x,y) = \text{Максимум}|x_i - y_i| \quad (3)$$

*Степенева відстань.* Дозволяє прогресивно збільшити або зменшити вагу, що відноситься до розмірності, стосовно якої відповідні об'єкти сильно відрізняються. Степенева відстань розраховується за формулою:

$$(x,y) = (\sum_i |x_i - y_i|^p)^{1/r} \quad (4)$$

де  $r$  і  $p$  параметри, що визначаються користувачем.

*Відсоток незгоди.* Цей вид відстаней використовується в тих випадках, коли дані є категоріальними. Ця відстань розраховується за формулою:

$$(x,y) = (\text{Кількість } x_i \neq y_i)/i \quad (5)$$

Нами для проведення кластерного аналізу використана Евклідова відстань, що розраховується за початковими, а не за стандартизованими даними. *Евклідова відстань* характеризується геометричною відстанню в багатовимірному просторі. Ця відстань піддається обрахунку незалежно від типу запису даних, що характеризує об'єкти. Розраховується за формулою:

$$(x,y) = \{ \sum_i (x_i - y_i)^2 \}^{1/2} \quad (6)$$

Також, досліджувану сукупність сімей можна розбивати на класи (згустки) за допомогою - кластерного аналізу типу «Форель» [6, с. 69-74].

Кластерний аналіз проводить розбиття на основі попередньо визначених відстаней, має свою структуру та послідовність проведення.

Послідовність вирішення задач кластерного аналізу така:

1.Формування вибірки для аналізу.

Обсяг вибіркової сукупності, сформований при випадковому відборі для проведення маркетингового дослідження нами визначався за формулою [7, с. 252-259] і становив 654 особи-споживача освітніх послуг, з них за напрямками підготовки: гуманітарні - 176 осіб; економічні - 169 осіб; правові -149 осіб; інженерія -166 осіб. Дослідження (анкетування) проводилося в п'яти вищих навчальних закладах III- IV рівнів акредитації Волинської області.

2. Вибір ознак, що характеризують об'єкт.

Нами для проведення ідентифікації споживача освітньої послуги було проведено анкетне опитування студентів п'яти ВНЗ м. Луцька.

В анкеті були подані наступні питання: 1) Вкажіть Ваш вік; 2) Вкажіть Вашу стать; 3) Де Ви закінчили школу?; 4. В якій області Ви закінчили школу?; 5) Середній бал атестату про завершення ЗОШ (за 12 бальною шкалою); 6) Середній бал навчання у ВУЗі (за 5 бальною шкалою); 7) Кількість членів сім'ї батьків на час вступу до ВУЗу; 8) Курс навчання; 9) Ви навчаєтесь на комерційній основі чи за кошти держзамовлення; 10) Хто здійснює оплату за навчання?; 11) Плата за навчання в рік; 12) Ваші батьки мають диплом про вищу освіту батько; 13) Ваші

батьки мають диплом про вищу освіту мати; 14) Вкажіть напрям освітньої підготовки батьків; 15) Соціальний статус батька; 16) Соціальний статус матері; 17) Середньомісячний дохід Ваших батьків; 18) Напрямок підготовки за яким ви навчаєтесь у ВУЗі; 19) Потреба у гуртожитку. (19 ознак охарактеризовані в табл. 1, 2).

3. Вибір та розрахунок міри подібності (відстані) між об'єктами.

Табл. 1.2 , рис.1 сформовані на основі проведення реальних маркетингових дослідження споживачів освітніх послуг 5 ВНЗ III-IV р.а. міста Луцька.

(отримані результати опрацьовувалися за допомогою методів кластерного аналізу з використанням програми Statistica 6.0 [5]. (табл.1)).

Таблиця 1

Матриця відстаней між оцінками каналів розподілу

№	Оцінка каналів розподілу																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0,0	31	32	39	83	24,2	26,1	13,6	36	40	36	35	33,5	41,1	39,0	30,6	29	31	33
2	31	0	14	11	104	35,0	31,2	25,5	10	17	10	9	26,5	36,4	33,6	20,9	7	8	7
3	32	14	0	18	101	33,6	27,8	27,2	17	21	16	16	27,6	36,5	34,2	24,1	12	17	14
4	38	11	18	0	113	44,0	38,8	32,2	7	16	9	9	31,0	40,1	38,9	25,9	13	11	11
5	83	10	101	113	0	70,5	79,0	88,9	110	111	108	108	97,2	99,7	97,0	96,7	100	105	105
6	24	35	34	44	71	0,0	19,6	26,4	41	44	39	39	35,0	43,1	38,3	32,0	32	36	37
7	26	31	28	39	79	19,6	0,0	27,5	37	38	34	34	34,2	40,9	39,2	30,8	27	33	31
8	14	25	27	32	89	26,4	27,5	0,0	29	35	30	29	29,9	38,0	36,2	26,0	23	25	27
9	36,0	10	17	7	110	41,1	36,8	29,3	0	20	10	10	29,4	38,5	37,8	24,4	11	10	11
10	40,1	17	21	16	11	44,2	38,1	34,9	20	0	16	16	32,7	41,7	38,2	27,4	18	18	16
11	35,6	10	16	9	11	39,4	34,4	29,6	10	16	0	8	30,3	38,6	36,8	24,1	10	10	9
12	34,5	9	16	9	108	38,9	34,0	28,8	10	16	8	0	29,2	38,1	36,2	23,8	10	10	9
13	33,5	26	28	31	97	35,0	34,2	29,9	29	33	30	29	0,0	39,4	39,2	29,3	25	27	27
14	41,1	36	37	40	100	43,1	40,9	38,0	38	42	39	38	39,4	0,0	45,3	38,2	35	37	37
15	39,0	34	34	39	97	38,3	39,2	36,2	38	38	37	36	39,2	45,3	0,0	34,7	33	35	35
16	30,6	21	24	26	97	32,0	30,8	26,0	24	27	24	24	29,3	38,2	34,7	0,0	19	21	22
17	28,6	7	12	13	100	31,8	27,5	22,9	11	18	10	10	24,8	35,5	33,0	18,9	0	8	8
18	31,4	8	17	11	105	36,4	32,5	25,4	10	18	10	10	27,0	37,4	35,0	20,6	8	0	9
19	33,3	7	14	11	105	36,6	31,4	27,4	11	16	9	9	27,0	37,1	34,8	21,6	8	9	0

Визначення числа кластерів, за допомогою програми Statistica 6.0, здійснюється за допомогою К-середніх за алгоритмом (алгоритм є повністю визначеним кінцевим набором кроку, операцій чи процедур, які приводять до кінцевого результату).

У загальному випадку метод К- середніх буде задану кількість різних кластерів, розташованих на великих відстанях один від одного. Далі обчислення

починаються з випадково вибраних спостережень, які стають центрами груп, після чого об'єктний склад кластерів змінюється з метою мінімізації мінливості усередині кластерів і максимізації мінливості між кластерами. Кожне наступне спостереження (K+1) відноситься до тієї групи, міра схожості з центром тяжіння якої мінімальна. Після одержання результатів обраховується середнє значення показників по кожному кластеру, з метою оцінки відмінності між ними.

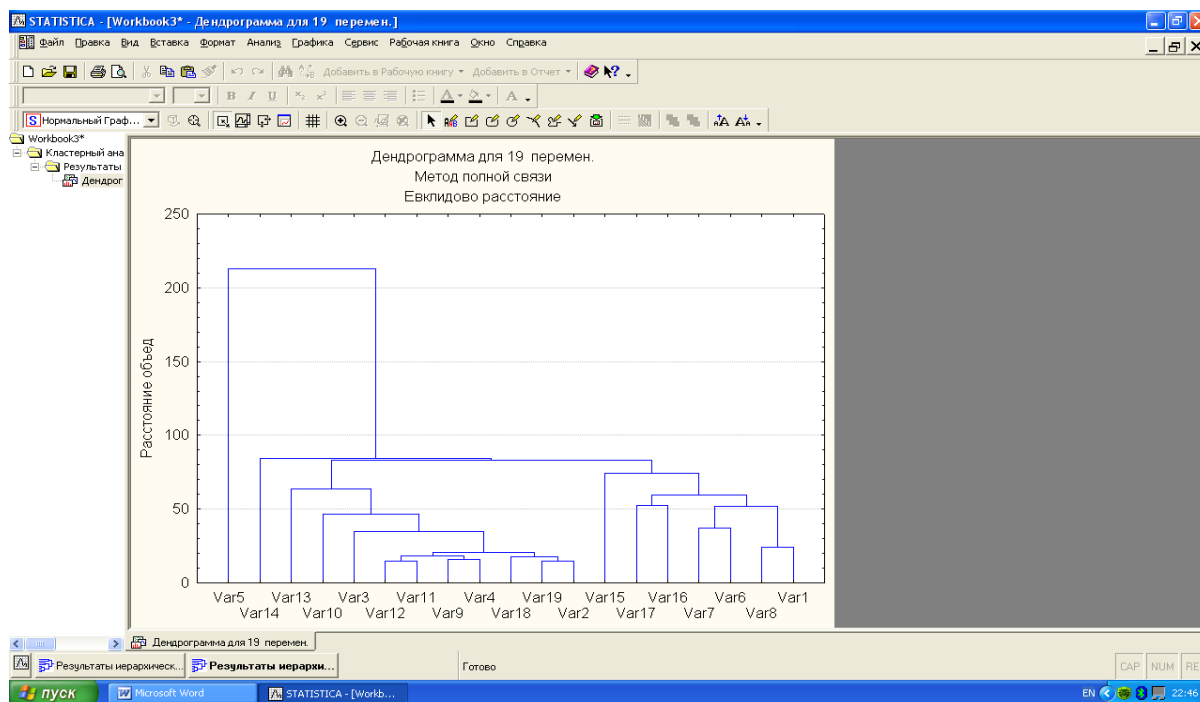


Рис. 1 Графічна інтерпретація розбиття сукупності на кластери із використанням Евклідової відстані

4. Формування кластерів здійснювалось за допомогою програми Statistica 6.0 на основі первинних даних, що одержані в результаті проведення маркетингового дослідження. Сформовано 8 кластерів (табл.2).

Таблиця 2

Розподіл респондентів за кластерами [5]

Кластер (респонденти)	Оцінка каналів розподілу																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 (60 осіб)	6	2	3	2	10	5	6	5	2	4	2	2	5	6	6	5	5	2	2
2 (41 осіб)	5	2	3	2	11	5	5	4	2	4	2	2	5	6	5	6	5	2	2
3 (52 осіб)	5	2	3	2	11	5	6	4	2	2	2	2	5	6	6	6	4	2	2
4 (81 осіб)	7	2	3	2	10	4	5	5	2	5	2	2	5	8	7	4	5	2	2
5 (47 осіб)	5	2	3	2	10	5	5	4	2	2	2	2	5	8	8	5	5	2	2
6 (40 осіб)	6	2	3	2	11	5	5	5	2	4	2	2	2	8	5	4	5	2	2
7 (290 осіб)	4	2	1	1	9	3	2	5	2	1	1	2	5	1	2	2	2	2	2
8 (42 осіб)	5	2	3	2	9	5	4	5	2	4	2	2	5	6	6	6	5	2	2



5. Аналіз одержаної інформації. Саме за результатами кластерного аналізу можна розрахувати середні для кожного кластера за кожним видом вимірювання, щоб оцінити, наскільки кластери різняться один від одного. Значення F-статистики, одержані для кожного вимірювання, є іншим індикатором того, наскільки добре відповідне вимірювання дискримінує кластери.

*Висновки.* Проведений літературний пошук та практична реалізація кластерного аналізу з метою ідентифікації образів споживачів освітніх послуг дають можливість стверджувати наступне:

1. Для проведення кластерного аналізу доцільно використати Евклідову відстань, що є геометричною відстанню у багатовимірному просторі ознак і піддається обрахунку незалежно від типу запису даних, що характеризує об'єкти.

2. Етапи проведення кластерного аналізу наступні: формування вибірки для аналізу; вибір ознак, що характеризують об'єкт; вибір та розрахунок міри подібності (відстані) для об'єктів; формування кластерів; аналіз одержаної інформації.

3. Одержанні у дослідженні результати ідентифікації образів споживачів освітніх послуг опрацьовувалися за допомогою методу кластерного аналізу із використанням програми Statistica 6. Опитано 653 респонденти та сформовано 8 кластерів (1 кластер - 60 осіб; 2 кластер - 41 особа; 3 кластер - 52 особи; 4 кластер - 81 особа; 5 кластер - 47 осіб; 6 кластер - 40 осіб; 7 кластер - 290 осіб; 8 кластер - 42 особи).

#### Література

1. Петруня Ю.Є. Маркетинг на ринку освітніх послуг / Ю.Є Петруня. //Маркетинг в Україні.– №1.– С.48– 51

2. Райс Э. Позicionирование: битва за узнаваемость / Э. Райс, Дж. Траут – СПб: Питер, 2001.– 256 с.

3. Лапач С. Н. Статистика в науке и бизнесе / С. Н. Лапач., А. В.Чубенко П. Н. Бабич – К.: МОРИОН, 2002. – с. 213– 218.

4. Редченко К.І. Техніка кластерного аналізу: групування з метою оцінки стратегічних перспектив. Стратегічні пріоритети розвитку регіону у системі

економічної політики в Україні/ К.І.Редченко, І.В.Островський //Науковий вісник.- Випуск II.– Чернівці ЧТЕІ, 2001.– С. 82– 84.

5. Отримані результати опрацьовувалися за допомогою методів кластерного аналізу з використанням програми Statistica 6.0.

6. Ліпич Л.Г. Методи визначення потреб індивідуальних споживачів. Монографія/ Л.Г. Ліпич, С.Я. Войтович Луцьк: Надстир'я.– 1999.– с. 205

7. Вашків П.Г. та інші. Статистика підприємництва: Навч. Посібник/ П.Г.Вашків, П.І. Пастер, В.П. Сторожук, Є.І. Ткач; за ред.. П.Г. Вашків, В.П. Сторожка.– К.: «Слобожанщина» 1999.– 600 с.