

*С.Я. Кравців, викладач-методист, НУЦЗУ,
О.М. Соболев, д.т.н., с.н.с., професор каф., НУЦЗУ*

ГРУПУВАННЯ АДМІНІСТРАТИВНО-ТЕРИТОРІАЛЬНИХ ОДИНИЦЬ УКРАЇНИ ПО РІВНЮ ІНТЕГРАЛЬНОГО ПОЖЕЖНОГО РИЗИКУ ЗА ДОПОМОГОЮ КЛАСТЕРНОГО АНАЛІЗУ

У даній роботі проведено групування адміністративно-територіальних одиниць України по рівню інтегрального пожежного ризику за допомогою кластерного аналізу двома методами, а саме, об'єднання (деревоподібна кластеризація) та кластеризації методом k-середніх з метою визначення областей, для яких мають бути застосовані відповідні типи моделей управління інтегральним ризиком.

Ключові слова: кластерний аналіз, інтегральний пожежний ризик, міра подібності, дендограма.

Постановка проблеми. На сьогодні велика увага приділяється реформуванню Державної служби України з питань надзвичайних ситуацій (ДСНС України), яка повинна забезпечувати належний рівень життєдіяльності населення, його захисту від надзвичайних ситуацій, пожеж та інших небезпечних подій. Значна увага у Стратегії реформування системи ДСНС України [1] приділяється застосуванню ризик-орієнтованого підходу для підвищення рівня пожежної безпеки на території нашої держави, оскільки ризик є об'єктивним показником, який характеризує величину потенційної небезпеки для людей, матеріальних цінностей та навколишнього середовища. Таким чином, виникає науково-прикладна проблема управління інтегральним пожежним ризиком на регіональному рівні. Розв'язанню даної проблеми сприятиме побудова моделей та методів управління зазначеним ризиком для кожного регіону України. У зв'язку з цим, проведення кластерного аналізу адміністративно-територіальних одиниць України дозволить визначити групи регіонів, для котрих мають бути застосовані відповідні типи моделей управління інтегральним ризиком, що призведе до зменшення обчислювальних ресурсів, необхідних для вирішення вказаної проблеми.

Аналіз останніх досліджень і публікацій. Огляд кластерного аналізу детально описано в книзі Дж.-О Кіма [2], яка представляє собою збірник робіт американських вчених у даному напрямі. В ній розглянуто апарат факторного, дискримінантного та кластерного аналізів. В статті [3] було використано факторний аналіз для оцінки параметрів впливу на інтегральний пожежний ризик. В роботі [4] наведена математична модель управління інтегральним пожежним ризиком для Харківської обла-

сті. Таким чином, проведення кластерного аналізу для адміністративно-територіальних одиниць України дозволить застосувати побудовану модель і для інших областей, які будуть входити в один кластер з Харківською областю.

Постановка завдання та його вирішення. Метою роботи є групування областей України та м. Києва в кластери з подальшим знаходженням подібності і можливості використання моделі управління інтегральним пожежним ризиком [4] для інших областей.

Головне призначення кластерного аналізу [5] – це розбиття множини досліджуваних об'єктів і ознак на однорідні у відповідному розумінні групи або кластери. Це означає, що вирішується завдання класифікації даних і виявлення відповідної структури в ній. Методи кластерного аналізу можна застосовувати в самих різних випадках, навіть в тих випадках, коли мова йде про просте угруповання, в якому все зводиться до утворення груп по кількісній подібності.

Рішенням задачі кластерного аналізу є розбиття, що задовольняють деякому критерію оптимальності. Цей критерій може являти собою деякий функціонал, що виражає рівні бажаності різних угруповань, який називають цільовою функцією.

Однією з проблем кластерного аналізу є неоднорідність одиниць виміру ознак, що ускладнює коректність розрахунків відстаней між точками. Ця проблема вирішується за допомогою попередньої стандартизації змінних. Стандартизація або нормування приводить значення всіх перетворених змінних до єдиного діапазону значень шляхом відношення цих значень до деякої величини, що відбиває певні властивості конкретної ознаки. Оберемо такий спосіб нормування вихідних даних

$$z = \frac{x - \bar{x}}{\sigma}, \quad (1)$$

де z – стандартизоване значення змінної; x – абсолютне значення змінної; \bar{x} , σ – відповідно середнє значення і середньоквадратичне відхилення змінної x .

Враховуючи статистичні дані щодо загибелі людей внаслідок пожеж за останній 7 років [6] розрахуємо інтегральний пожежний ризик R_3 [7] та опрацюємо результати за допомогою програмного забезпечення STATISTICA 10 (рис. 1). Слід відзначити, що у розрахунках не враховано Донецьку та Луганську області через некоректність інформації за 2014-2016 рр.

Data: Ризик по всіх обл України (7v by 23c)							
	1 2010	2 2011	3 2012	4 2013	5 2014	6 2015	7 2016
Вінницька	0,00006	0,00006	0,00006	0,00005	0,00005	0,00006	0,00005
Волинська	0,00005	0,00005	0,00005	0,00003	0,00004	0,00004	0,00003
Дніпропетровська	0,00008	0,00007	0,00006	0,00006	0,00006	0,00005	0,00005
Житомирська	0,00008	0,00008	0,00009	0,00008	0,00008	0,00008	0,00008
Закарпатська	0,00003	0,00004	0,00004	0,00004	0,00002	0,00003	0,00002
Запорізька	0,00008	0,00008	0,00006	0,00006	0,00006	0,00006	0,00005
Івано-Франківська	0,00004	0,00005	0,00004	0,00004	0,00004	0,00004	0,00003
м. Київ	0,00003	0,00002	0,00002	0,00002	0,00002	0,00002	0,00002
Київська	0,00008	0,00008	0,00008	0,00007	0,00006	0,00005	0,00008
Кіровоградська	0,00008	0,00008	0,00008	0,00008	0,00008	0,00008	0,00007
Львівська	0,00004	0,00003	0,00004	0,00003	0,00003	0,00003	0,00003
Миколаївська	0,00006	0,00007	0,00007	0,00007	0,00006	0,00007	0,00005
Одеська	0,00006	0,00006	0,00006	0,00005	0,00005	0,00005	0,00005
Полтавська	0,00008	0,00007	0,00006	0,00006	0,00006	0,00005	0,00005
Рівненська	0,00005	0,00007	0,00005	0,00004	0,00004	0,00004	0,00004
Сумська	0,00007	0,00006	0,00007	0,00008	0,00008	0,00007	0,00006
Тернопільська	0,00005	0,00005	0,00006	0,00005	0,00006	0,00004	0,00004
Харківська	0,00005	0,00006	0,00007	0,00006	0,00005	0,00005	0,00005
Херсонська	0,00007	0,00007	0,00007	0,00006	0,00006	0,00005	0,00005
Хмельницька	0,00004	0,00004	0,00004	0,00004	0,00004	0,00004	0,00004
Черкаська	0,00007	0,00008	0,00006	0,00005	0,00005	0,00005	0,00005
Чернігівська	0,00006	0,00006	0,00005	0,00006	0,00004	0,00004	0,00004
Чернівецька	0,00008	0,00005	0,00005	0,00005	0,00006	0,00006	0,00005

Рис. 1. Динаміка інтегрального пожежного ризику для областей України та м. Київ протягом 2010-2016 рр.

За допомогою ієрархічного методу кластерного аналізу розіб'ємо області України на групи (рис. 2), що будуть характеризуватися схожою ситуацією стосовно загибелі людей під час пожеж за одиницю часу. Під час проведення кластеризації для заходів подібності було використано метод Варда (Уарда), який полягає в тому, що спочатку в обох кластерах для всіх наявних спостережень проводиться розрахунок середніх значень окремих змінних. Потім обчислюються квадрати евклідових відстаней від окремих спостережень кожного кластера до визначеного середнього значення, причому зазначені відстані підсумовуються. Далі в один новий кластер поєднуються ті кластери, при об'єднанні яких відбувається найменший приріст загальної суми відстаней. Як міру відстані було обрано манхетенівську відстань (відстань міських кварталів), яка розраховується за формулою

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|, \quad (2)$$

де d_{ij} – відстань між об'єктами i та j ; x_{ik} (x_{jk}) – значення k -ї змінної для i -го (j -го) об'єкта ($k=1,2,3\dots p$).

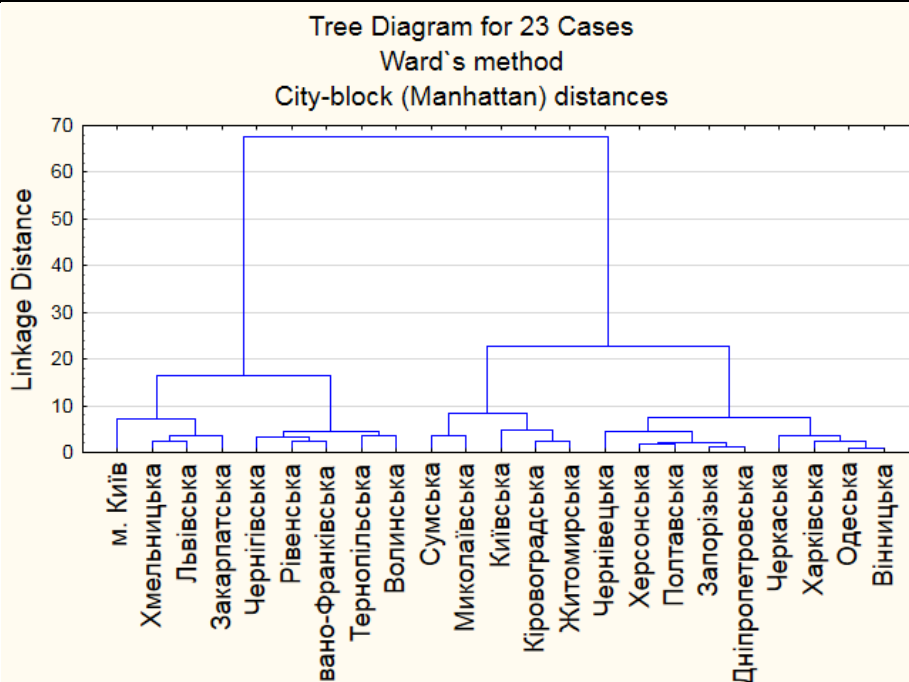


Рис. 2. Вертикальна деревоподібна дендограма за методом Варда, яка показує відстань об'єднання досліджуваних областей

На дендограмі, що наведена на рис. 2 горизонтальна ось являє собою область спостереження, вертикальна – відстань об'єднання. Таким чином, на першому етапі були об'єднані Одеська і Вінницька області з мінімальними відстанями, що становить 1,000364, а на останньому рівні було об'єднано в єдиний кластер усі області з максимальним значенням відстані 67,59186. Аналіз дендограми дозволив зробити висновок, на якому етапі кластеризації слід прийняти отриману класифікацію як остаточну. При кількості кластерів, яка дорівнює двом ($K=2$), отримаємо дві групи областей. До першої входять м. Київ, Хмельницька, Львівська, Закарпатська, Чернівецька, Рівненська, Івано-Франківська, Тернопільська та Волинська області, а до другої – всі інші. Більш доцільніше розглядати систему при $K=4$, оскільки тоді можна одержати чотири кластери із значеннями відстаней, які наведені в табл. 1.

Табл. 1. Відомості про угруповання при кількості кластерів, яка дорівнює 4 ($K=4$)

Групи кластерів при $K=4$	Області	Значення відстані кластеру
Кластер 1	Сумська, Миколаївська, Київська, Кіровоградська та Житомирська області	8,352732
Кластер 2	Чернівецька, Херсонська, Полтавська, Запорізька, Дніпропетровська, Черкаська, Харківська, Одеська та Вінницька області	7,410969
Кластер 3	м. Київ, Хмельницька, Львівська, Закарпатська області	7,20112
Кластер 4	Чернівецька, Рівненська, Івано-Франківська, Тернопільська та Волинська області	4,466044

Остаточно приймаємо кількість кластерів рівною 4, оскільки при подальшому збільшенні їх кількості втрачається наочність класифікації. Результати деревовидної кластеризації в графічному вигляді наведені на рис. 3.

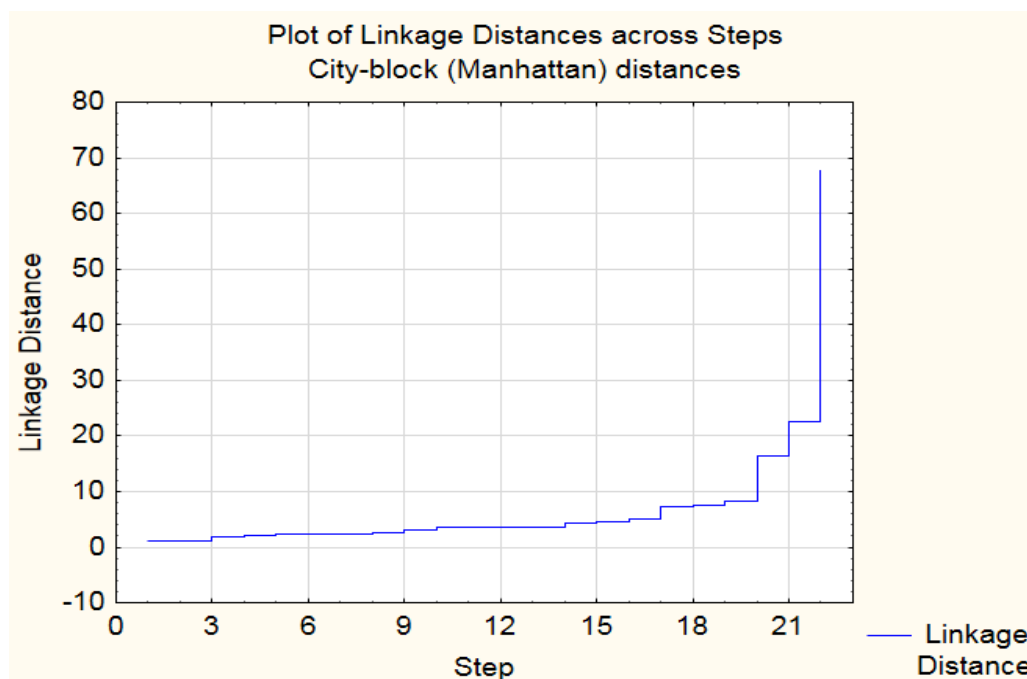


Рис. 3. Графік покрокового об'єднання основних змінних, що досліджуються

Для обґрунтування прийнятності класифікації за методом k-середніх в роботі проведено порівняння мінливості всередині кластерів з мінливістю поза кластерами. Оцінка порівняння мінливості всередині і поза кластерами проведена з використанням F-критерію Фішера [2]. В табл. 2 представлені результати дисперсного аналізу методом k-середніх. Визначаємо, що критичне (табличне) значення критерію Фішера буде дорівнювати 2,447, тобто емпіричне значення критерію Фішера перевищує табличне. У зв'язку з цим, нульова гіпотеза про рівність генеральних дисперсій на рівні значущості 0,05 приймається.

Табл. 2. Результати дисперсійного аналізу k-середніх значень основних змінних

Змінні	Мінливість між кластерами	Ступінь свободи	Мінливість всередині кластерів	Ступінь свободи	F - критерій	
					F	α
2010	18,00713	3	3,992870	19	28,56220	0,000000
2011	15,60396	3	6,396039	19	15,45098	0,000025
2012	17,87237	3	4,127626	19	27,42295	0,000000
2013	17,56214	3	4,437860	19	25,06318	0,000001
2014	18,22331	3	3,776688	19	30,55967	0,000000
2015	18,18678	3	3,813215	19	30,20626	0,000000
2016	19,87199	3	2,128006	19	59,14267	0,000000

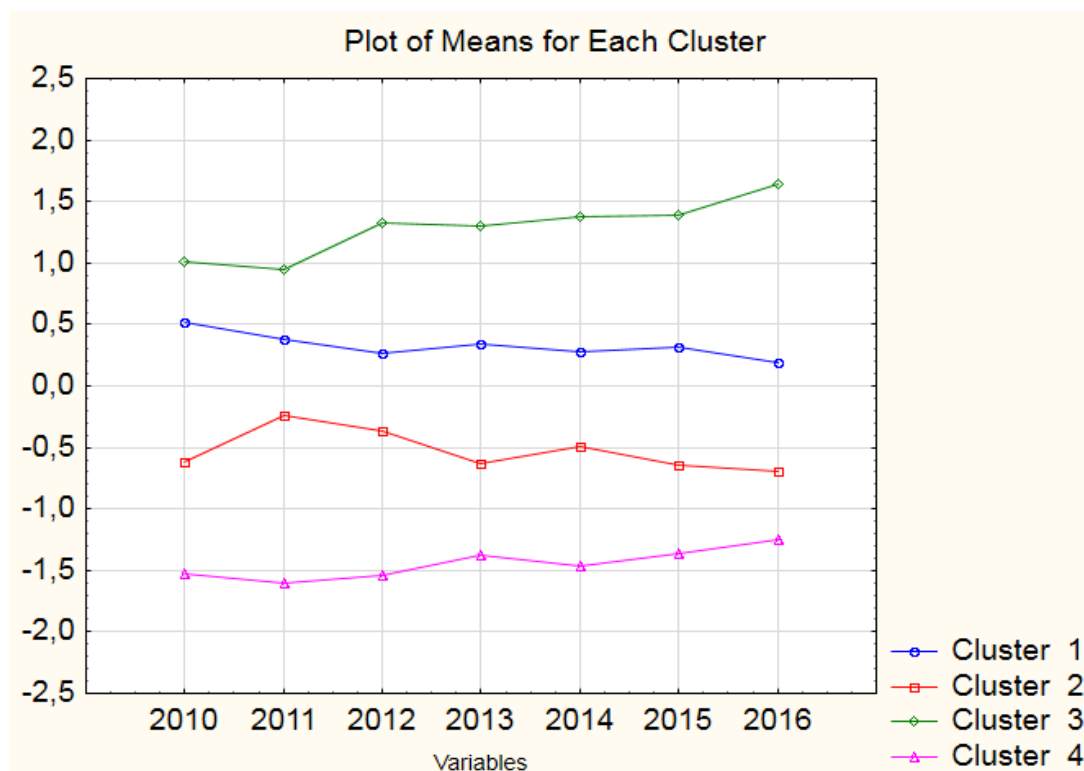


Рис. 4. Графік математичних очікувань для кожного кластера стандартизованих значень основних змінних

Табл. 3. Математичне відхилення (M_x), стандартне відхилення (σ_x) і дисперсія (σ_x^2) стандартизованих значень основних змінних (X)

Змінні (X)	I кластер			II кластер		
	M_x	σ_x	σ_x^2	M_x	σ_x	σ_x^2
2010	0,514287	0,544098	0,296043	-0,612282	0,344829	0,118907
2011	0,380769	0,550637	0,303201	-0,244578	0,544714	0,296714
2012	0,264614	0,406409	0,165168	-0,360602	0,372876	0,139037
2013	0,343631	0,429718	0,184657	-0,628507	0,633780	0,401677
2014	0,283000	0,248750	0,061876	-0,493584	0,596655	0,355997
2015	0,313017	0,373096	0,139201	-0,644811	0,177532	0,031518
2016	0,190168	0,206482	0,042635	-0,691584	0,276410	0,076403
Змінні (X)	III кластер			IV кластер		
	M_x	σ_x	σ_x^2	M_x	σ_x	σ_x^2
2010	1,011933	0,188861	0,035668	-1,53230	0,498616	0,248618
2011	0,952628	0,741283	0,549500	-1,59883	0,526585	0,277292
2012	1,327837	0,539238	0,290778	-1,53862	0,635776	0,404211
2013	1,304706	0,376612	0,141837	-1,37815	0,497905	0,247909
2014	1,377235	0,524213	0,274800	-1,46775	0,569038	0,323804
2015	1,386825	0,781341	0,610494	-1,36335	0,448277	0,200952
2016	1,644365	0,538173	0,289631	-1,25531	0,435809	0,189930

На рис. 4 графічно відображено значення, що наведені в табл. 3. Дані значення розраховані методом k-середніх та містять інформацію про

математичне відхилення (M_x), стандартне відхилення (σ_x) і дисперсію (σ_x^2) стандартизованих значень основних змінних (X) всіх 4-х кластерів.

Висновки. У даній роботі проведено групування адміністративно-територіальних одиниць України по рівню інтегрального пожежного ризику за допомогою кластерного аналізу. В результаті аналізу всі регіони України були розбиті на 4 групи. Це дозволить застосувати для кожної групи відповідну модель управління інтегральним пожежним ризиком. Подальші дослідження будуть направлені на розробку методу управління інтегральним пожежним ризиком, який передбачається застосувати для кожної групи регіонів.

ЛІТЕРАТУРА

1. Розпорядження Кабінету Міністрів України від 25.01.2017 р. № 61-р «Про схвалення Стратегії реформування системи Державної служби України з надзвичайних ситуацій» [Електронний ресурс] // Режим доступу: <http://zakon3.rada.gov.ua/laws/show/61-2017-p>.
2. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.О. Ким, Ч.У. Мюллер, У.Р. Клекка и др.; Под ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
3. Кравців С.Я. Оцінювання параметрів впливу на інтегральний пожежний ризик за допомогою факторного аналізу / С.Я. Кравців, О.М. Соболь, В.В. Тютюнник // Пожежна безпека: збірник наукових праць. – Львів: ЛДУБЖД, 2017. – Вип. 30. – С. 99-104.
4. Соболь О.М. Математична модель управління інтегральним пожежним ризиком та її особливості / О.М. Соболь, С.Я. Кравців // Вісник ХНТУ: збірник наукових праць. – Херсон: ХНТУ, 2017. – Вип. 3 (62) Т. 2. – С. 317-321.
5. Буреева Н.Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007. – 112 с
6. Аналіз масиву карток обліку пожеж – ДСНС Україна [Електронний ресурс] // Режим доступу: <http://undicz.dsns.gov.ua/ua/Analiz-masivu-kartok-obliku-pozhezh.html>.
7. Kravtsiv S.Ya. The analysis of integral risks of the territory of Ukraine / S.Ya. Kravtsiv, O.M. Sobol, A.V. Maksimov // Проблеми надзвичайних ситуацій: збірник наукових праць. – Харків: НУЦЗУ, 2016. – Вип. 23. – С. 53-60. – Режим доступу: <http://nuczu.edu.ua/sciencearchive/ProblemsOfEmergencies/vol23/Kravtsiv.pdf>.

Отримано редколегією 25.09.2017

С.Я. Кравцив, А.Н. Соболев

Кластерный анализ административно-территориальных единиц Украины по уровню интегрального пожарного риска

В данной работе проведено группировка административно-территориальных единиц Украины по уровню интегрального пожарного риска с помощью кластерного анализа двумя методами, а именно, объединения (древовидная кластеризация) и кластеризации методом k-средних с целью определения областей, для которых должны быть применены соответствующие типы моделей управления интегральным риском.

Ключевые слова: кластерный анализ, интегральный пожарный риск, степень сходства, дендограмма.

S.Ya. Kravtsiv, O.M. Sobol

Cluster analysis of administrative-territorial areas of Ukraine on the level of integral fire risk

In this paper, the grouping of the administrative-territorial units of Ukraine in the level of integral fire risk by means of cluster analysis by two methods, namely, association (tree-like clustering) and k-medium clustering with the purpose of identifying the areas for which the appropriate types Integral risk management models.

Key words: cluster analysis, integral fire risk, degree of similarity, dendrogram.