

## **Тема 8\_2. Приклад застосування кластерного аналізу в соціо-еколого-економічних дослідженнях.**

Розглянемо застосування кластерного аналізу для дослідження структури європейського союзу (в розрізі країн – членів) та місця в ній України за критеріями податкової політики.

Нагадаємо основні етапи кластерного аналізу:

**Етап 1 - *Вибір змінних – критеріїв для кластеризації.***

**Етап 2 - *Вибір засобу виміру відстані між кластерами.***

**Етап 3 - *Стандартизація (нормування) спостережень.***

**Етап 4 - *Формування кластерів.***

**Етап 5- *Інтерпретація результатів***

Нагадаємо їх зміст.

### **Етап 1 - Вибір змінних – критеріїв для кластеризації.**

Вибір об'єктів і показників для кластерного аналізу потребує одержання репрезентативної вибірки, що дозволяє переносити результати, отримані на обмеженій кількості об'єктів, на всю генеральну сукупність. Якщо ставиться завдання типології об'єктів, необхідно виділити систему інформативних ознак, які найбільше повно характеризують досліджувані об'єкти. Але, не обов'язково треба включати всі змінні до кластерного аналізу.

### **Етап 2 - Вибір засобу виміру відстані між кластерами.**

Найпоширенішою мірою для визначення відстані між двома точками на площині, утвореній координатними осями  $x$  і  $y$ , є евклідова міра. Часто за замовченням використовується квадрат евклідової відстані. Але поряд з евклідовою мірою відстані, можна обрати й інші дистанційні міри. Кластерний аналіз можна проводити не тільки зі змінними інтервальної шкали, але і з категорійними змінними, та у таких випадках застосовуються вже інші дистанційні міри.

### **Етап 3 - Стандартизація (нормування) спостережень.**

Рівні значень змінних часто дуже сильно відрізняються один від одного. Відповідно до формули евклідової міри, змінна, що має велике значення, практично цілком домінує над змінною з малими значеннями. Рішенням цієї проблеми є нормування (стандартизація) значень змінних за формулами (4.2) або (4.3). Стандартизація приводить значення всіх перетворених змінних до єдиного діапазону значень, і якщо вона виконується за формулою (4.2), то середнє кожної приводиться до 0, а середнє відхилення – до 1. Тоді всі спостереження змінюються приблизно у діапазоні від  $-3$  до  $+3$ .

### **Етап 4 - Формування кластерів.**

Існує дві основні групи методів формування кластерів : ієрархічні та неієрархічні. Перші поділяються на метод злиття та

метод подрібнення. У першому випадку існуючі кластери розширюються шляхом об'єднання, доки не буде сформований один – єдиний кластер, що об'єднує всі спостереження. Метод подрібнення заснований на зворотній операції: спочатку всі спостереження об'єднуються у єдиний кластер, а потім починається процес розділення його на частини. Частіше використовують метод злиття .

З неієрархічних методів найбільш поширений метод  $k$  – середніх. В прикладах будемо використовувати обидві групи методів. Причому, ієрархічні методи допоможуть нам визначити кількість кластерів.

#### **Етап 5- *Інтерпретація результатів* .**

Цей етап є достатньо складним і залежить від мети дослідника. На жаль, виразна картина відносин між змінними зустрічається не дуже часто. По-перше, структури кластерів, якщо отримуються, не так чітко розділені, особливо при наявності великої кількості спостережень. Скоріше навпаки: кластери розмиті і навіть проникають один в один. По-друге, як правило, кластерний аналіз проводиться з великою кількістю змінних, що ускладнює аналіз.

## Постановка задачі

Розглянемо кластерний аналіз країн по ступеню їх подібності на основі певних показників, що характеризують критерії податкової політики держави. Результатом аналізу стане формування регіональних кластерів, що дозволить враховувати особливості розвитку країн-членів ЄС при здійсненні податкової реформи в Україні.

Стратегія формування ефективної податкової політики є неоднозначною в різних країнах, що обумовлене істотними відмінностями регіональних податкових систем за рівнем оподаткування, їх складом і структурою, досягнутим рівнем розвитку економіки, умовами входження в ринкове середовище, конкурентними перевагами, а також визначальними факторами гармонізації оподаткування в умовах глобалізації економічних систем. Однак податковий потенціал держави характеризується певною системою показників податкової політики, які можна об'єднати в кілька груп, виходячи з науково обґрунтованих критеріїв її реалізації. В окремі групи можна об'єднати показники, які характеризують фіскальну достатність, економічну ефективність і соціальну справедливість податкової політики окремої країни. Також аналізуються показники, які враховують тенденції соціально-економічного розвитку країни.

Разом з тим є очевидним, що всі ці показники, які характеризують політику в сфері оподаткування, тісно зв'язані між собою й мають деякий ступінь подібності закономірностей їх розвитку.

Метою роботи є об'єднання європейських країн у більші групи по ступеню їх подібності на основі певних показників, що характеризують фіскальну достатність, економічну ефективність, соціальну справедливість і гнучкість національної податкової політики. Кластерний аналіз європейських країн за основними критеріями податкової політики включає кілька основних етапів.

Допустимо, що є набір об'єктів  $O_i$  (країн),  $i=1,2,\dots,28$  – податкової політики країн-членів ЄС і України, кожна з яких описується сукупністю ознак  $P_j$ ,  $j=1,2,\dots,m$  – показниками критеріїв податкової політики. Остаточо для проведення кластерного аналізу були відібрані такі показники:

- загальне податкове навантаження, %
- ефективна ставка податків на споживання, %

- ефективна ставка податків з капіталу, %
- ефективна ставка податків на працю, %
- темп росту реального ВВП, % до попереднього періоду.

Таблиця 4.1 - Дані для кластерного аналізу країн за податками

Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП
▶ UA	37,3	22,8	24,4	63,2	28,3	102,3
BG	33,3	26,4	16,9	71,7	42,6	106
EE	32,2	20,9	10,7	71,5	33,7	96,4
LT	30,3	17,5	12,4	78,5	33	102,8
LV	28,9	17,5	16,3	48,4	28,2	102,3
RO	28	17,7	19,6	26,2	29,5	107,3
SK	29,1	18,4	16,7	84,6	33,5	106,2
AT	42,8	22,1	27,3	150,2	41,3	102,2
BE	44,3	21,2	32,7	97,1	42,6	101
DE	39,3	19,8	23,1	193,8	39,2	101
ES	33,1	14,1	32,8	114,6	30,5	100,9
FI	43,1	26	28,1	114,5	41,3	100,9
FR	42,8	19,1	38,8	160,7	41,4	100,2
IT	42,8	16,4	35,3	187,4	42,8	98,7
NL	39,1	26,7	17,2	189,8	35,4	101,9
PT	36,7	19,1	19,6	143,4	29,6	100,2
SE	47,1	28,4	27,9	190,1	42,1	99,6
CZ	36,1	21,1	21,5	127,1	39,5	102,5
EL	32,6	15,1	15,8	102	37	102
HU	40,4	26,9	19,2	98	42,4	100,6
PL	34,3	21	22,5	108	32,8	105
SI	37,3	23,9	21,6	121,7	35,7	103,5
CY	39,2	20,6	36,4	110	24,5	103,6
DK	48,2	32,4	43,1	267,8	36,4	99,1
IE	29,3	22,9	15,7	153,1	24,6	97
LU	35,6	27,1	39,6	173,3	31,5	100,1
MT	34,5	20	42,1	197	20,2	101,7
UK	37,3	17,6	45,9	180,2	26,1	99,9

Наступним етапом кластерного аналізу є обчислення деякого ступеня подібності або відмінності між вибраними для аналізу об'єктами. Це означає перехід від таблиці «об'єкт – ознака» до таблиці «об'єкт – об'єкт», де  $d_{ij}$  – вимір подібності або відмінності між об'єктами  $O_i$  і  $O_j$ . Як вимір подібності був використаний найпоширеніший показник – евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

де  $x_{ik}$  - значення k-го показника в i-ом об'єкті,  $i = 1, 2, \dots, m$ .

Матриця евклідових відстаней відображає подібність і відмінність у податкових політиках різних країн. Чим менше значення, тим вище ступінь подібності двох країн і комбінацій у кластері. І навпаки, чим більше відповідне значення, тем більша відмінність між країнами.

У результаті цього етапу можна побудувати за допомогою пакетів STATISTICA 8 або MATLAB 7 деревоподібну діаграму, яка дає перше уявлення про кількість можливих кластерів (рис. 4.4).

У результаті побудови дендрограми можна сформуванати гіпотезу про наявність як мінімум чотирьох кластерів.

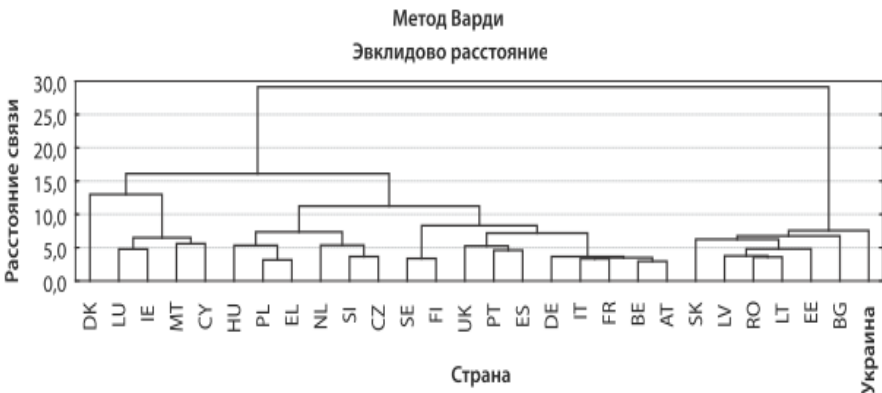


Рисунок 4.4 - Дендрограма відстаней між країнами за критеріями податкової політики

Наявність різкого стрибка середніх відхилень показників можна інтерпретувати як характеристику числа кластерів, які об'єктивно існують у досліджуваній сукупності. Тобто на кроці, де значення коефіцієнта збільшується стрибкоподібно, процес об'єднання в нові кластери необхідно зупинити, оскільки інакше були б об'єднані кластери, які знаходяться на відносно великій відстані один від іншого. Отже, гіпотеза про кількість кластерів ( а саме, чотири) країн, які формуються на основі відповідних критеріїв податкової політики, підтверджується і є остаточною.

## Кластерний аналіз у програмі Deductor

Спочатку необхідно здійснити імпорт розглянутих даних з файлу – Податки.txt. Результати імпорту показані в табл. 4.1.

Після цього вибираємо й запускаємо Майстер обробки "Кластеризація". При запуску Майстра необхідно настроїти призначення стовпців, тобто вибрати властивості, по яких буде відбуватися угруповання об'єктів. Укажемо стовпцю "Країна" призначення "Інформаційне", а іншим стовпцям – "Вхідне". (Рис. 4.5).

На наступному кроці Майстра необхідно настроїти спосіб поділу вхідної множини даних на тестове й навчальне, а також кількість прикладів у тій і іншій множині. Укажемо, що дані обох множин беруться випадковим чином, і визначимо всю множину як навчальну (100%).

Наступний крок пропонує настроїти параметри кластеризації, визначити на яку кількість кластерів буде розподілятися вхідна множина. Оберемо фіксовану кількість кластерів – чотири.

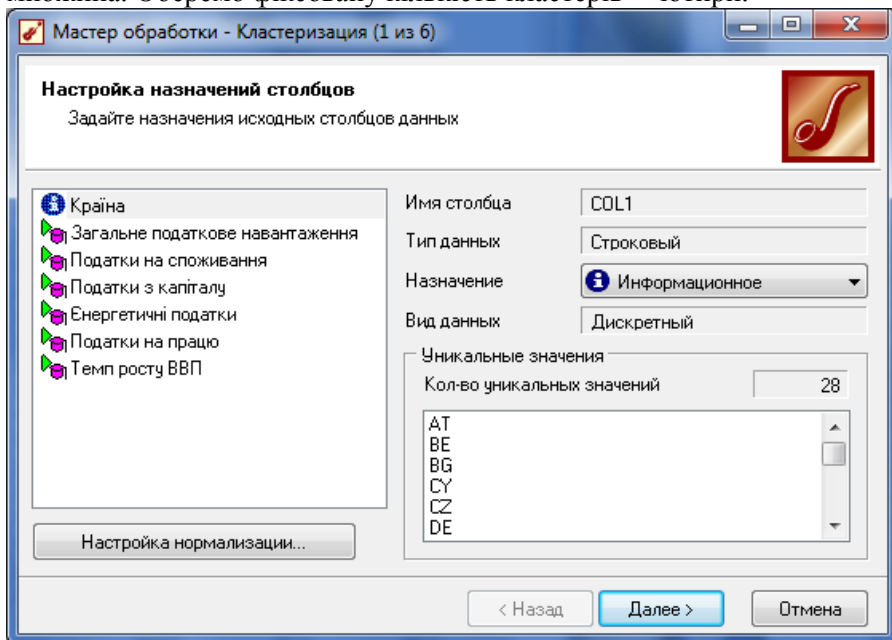


Рисунок 4.5 – Настроювання призначень стовпців

Для відображення отриманих груп кластерів виберемо в оброблювачі «Кластеризація» зі списку візуалізаторів такі способи відображення даних: 1) «Профілі кластерів» для визначення структури формування групи кластерів і 2) «Куб» для наочного перегляду отриманих результатів (рис. 4.6).

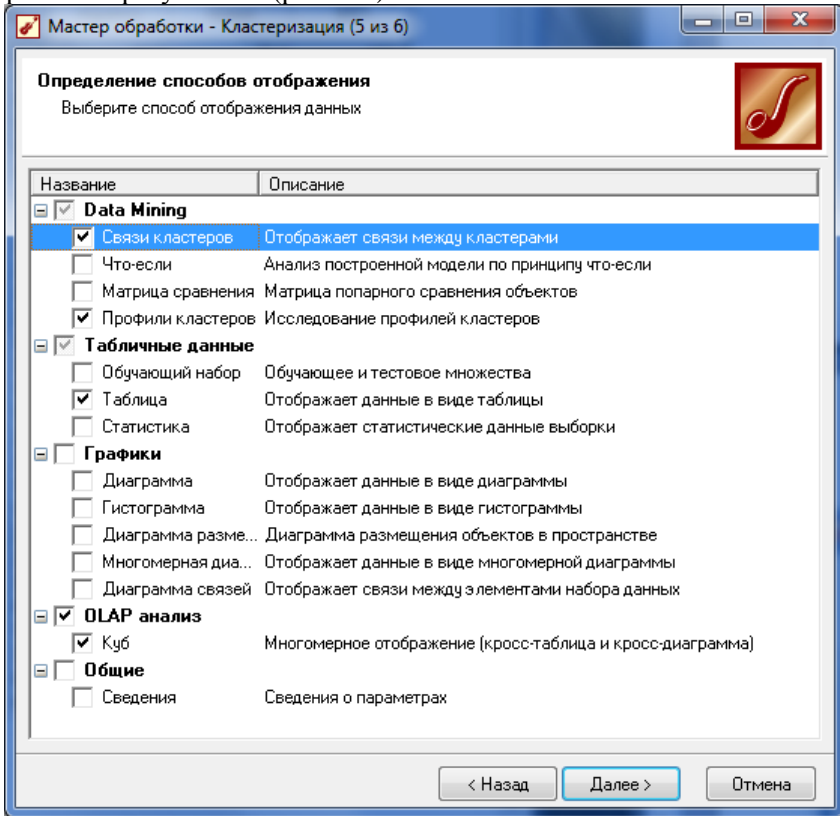


Рисунок 4.6 – Настроювання візуалізаторів

Для настроювання візуалізатора «Куб» необхідно вибрати розглянуті властивості як факти, а номер кластера як вимір (рис. 4.7).

Далі необхідно визначити як в таблиці розташовувати виміри і факти (рис. 4.8).

Найбільш правильно в подальших настроюваннях задати відображення фактів як середнє по розглянутій групі (рис. 4.9).



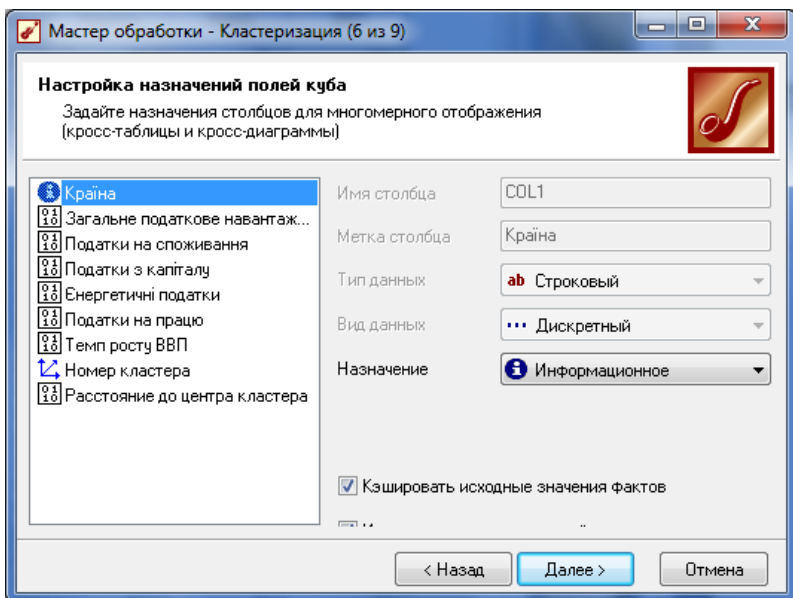


Рисунок 4.7 – Настроювання полів куба

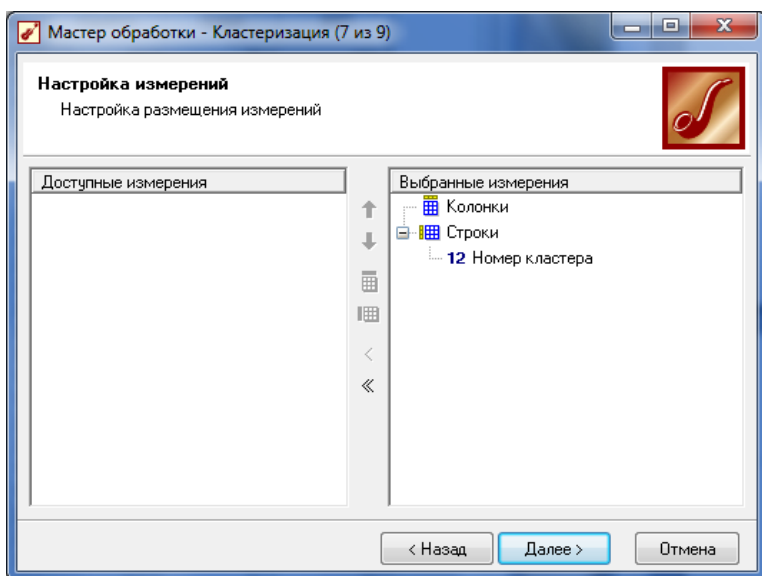


Рисунок 4.8 – Настроювання вимірів

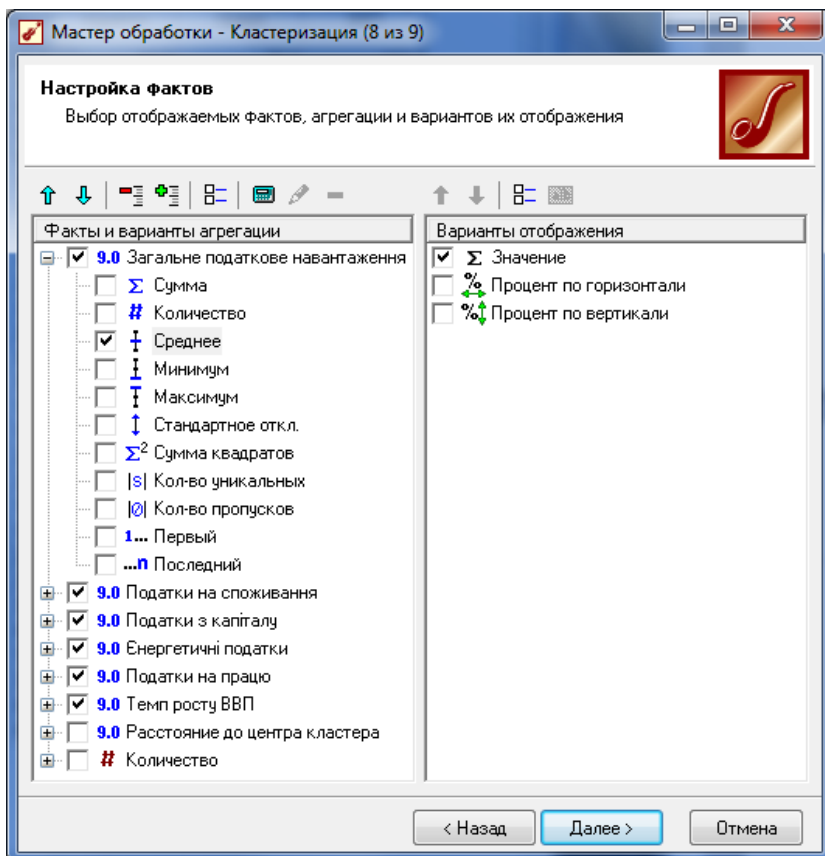


Рисунок 4.9 – Вибір фактів та їхньої агрегації

Загальну структуру сформованих алгоритмом кластерів можна переглянути у візуалізаторі «Профілі кластерів». У ньому представлені всі розглянуті властивості разом з характером впливу їх на склад кластера (рис. 4.10).

Основним фактором, що визначає склад кластера, є значимість властивостей, виражена у відсотках. Загальна значимість розглянутого поля визначається варіабельністю її розглянутих параметрів.

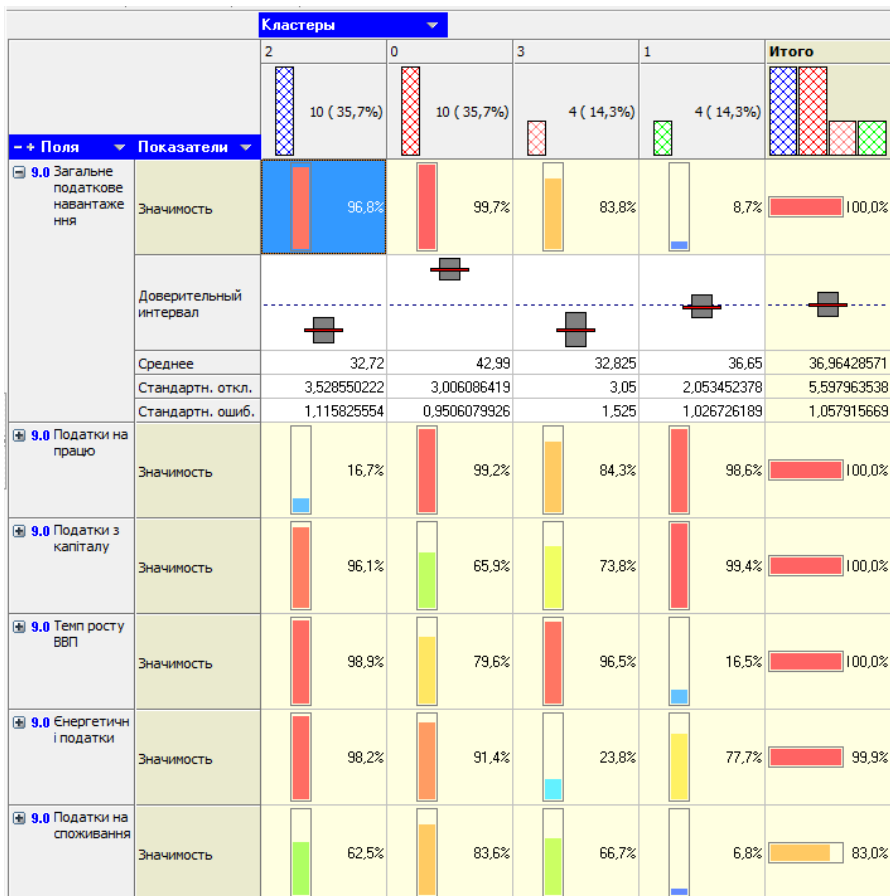


Рисунок 4.10 – Профілі кластерів

Значимість для безперервних і дискретних полів визначається по-різному. Значимість для безперервних полів встановлюється залежно від відхилення середнього значення розглянутої групи кластерів від загального середнього всієї вибірки, чим більше виражене дане відхилення, тим більше його значимість. Значимість для дискретних полів визначається наявністю індивідуальних відмінностей, між розглянутими групами, чим більше виражені відмінності, тим більше значимість. Для кожної розглянутої властивості в кластері обчислюється: довірчий інтервал, середнє, стандартне відхилення й стандартна помилка (рис. 4.11).

Показатель	Пример	Описание
Значимость		1 минус вероятность нулевой гипотезы. Значимость выражается в процентах. Для непрерывных полей используется <i>t</i> -критерий Стьюдента, а для дискретных полей – критерий хи-квадрат. Общая значимость поля определяется по <u>F-критерию Фишера</u> .
Доверительный интервал		Графическое изображение 95% доверительного интервала для среднего значения кластера (темно-серая область). Кроме этого, показываются: <ul style="list-style-type: none"> <li>▪ среднее значение по кластеру – красной линией;</li> <li>▪ среднее значение по всей выборке – синей штрихпунктирной линией.</li> </ul>
Среднее	–	Среднее значение по полю, рассчитанное для объектов, попавших в кластер.
Стандартное отклонение	–	Стандартное отклонение по полю, рассчитанное для объектов, попавших в кластер.
Стандартная ошибка	–	Стандартная ошибка по полю, рассчитанная для объектов, попавших в кластер.

Рисунок 4.11 – Визначення показників кластерів

Завершальним етапом кластерного аналізу є інтерпретація отриманих результатів. Результати аналізу свідчать, що навіть країни, які близько знаходяться на географічній карті, настільки сильно відрізняються за критеріями податкової політики, що не можуть бути віднесені до одного регіонального кластера. Усім кластерам треба надати умовні назви, які будуть відобразити їх характеристики на основі 6 основних показників.

Таблиця 4.1 – Середні показники по кожному кластеру

Номер кластера	Загальне пс	Податки на сп	Податки з капіт	Енергетичні под	Податки на пра	Темп росту ВВП
0	42,99	23,90	29,27	164,94	40,49	100,52
1	36,65	21,33	41,00	165,13	25,58	101,32
2	32,72	20,14	18,77	83,14	34,01	103,99
3	32,83	19,25	19,70	120,65	29,60	98,63
<b>Итого:</b>	36,96	21,53	25,83	129,43	34,49	101,60

Таблиця 4.2 – Країни, що віднесені до кластеру 0

Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶ AT	42,8	22,1	27,3	150,2	41,3	102,2	0	0,20
FI	43,1	26	28,1	114,5	41,3	100,9	0	0,25
BE	44,3	21,2	32,7	97,1	42,6	101	0	0,35
SE	47,1	28,4	27,9	190,1	42,1	99,6	0	0,36
DE	39,3	19,8	23,1	193,8	39,2	101	0	0,37
FR	42,8	19,1	38,8	160,7	41,4	100,2	0	0,38
HU	40,4	26,9	19,2	98	42,4	100,6	0	0,46
IT	42,8	16,4	35,3	187,4	42,8	98,7	0	0,49
NL	39,1	26,7	17,2	189,8	35,4	101,9	0	0,51
DK	48,2	32,4	43,1	267,8	36,4	99,1	0	0,82

Таблиця 4.3 – Країни, що віднесені до кластеру 1

Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶ UK	37,3	17,6	45,9	180,2	26,1	99,9	1	0,29
MT	34,5	20	42,1	197	20,2	101,7	1	0,30
CY	39,2	20,6	36,4	110	24,5	103,6	1	0,36
LU	35,6	27,1	39,6	173,3	31,5	100,1	1	0,43

Таблиця 4.4 – Країни, що віднесені до кластеру 2

Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶ PL	34,3	21	22,5	108	32,8	105	2	0,20
LT	30,3	17,5	12,4	78,5	33	102,8	2	0,29
SK	29,1	18,4	16,7	84,6	33,5	106,2	2	0,29
SI	37,3	23,9	21,6	121,7	35,7	103,5	2	0,36
EL	32,6	15,1	15,8	102	37	102	2	0,37
CZ	36,1	21,1	21,5	127,1	39,5	102,5	2	0,38
LV	28,9	17,5	16,3	48,4	28,2	102,3	2	0,42
UA	37,3	22,8	24,4	63,2	28,3	102,3	2	0,44
RO	28	17,7	19,6	26,2	29,5	107,3	2	0,51
BG	33,3	26,4	16,9	71,7	42,6	106	2	0,55

Таблиця 4.5 – Країни, що віднесені до кластеру 3

Країна	Загальне податкове навантаження	Податки на споживання	Податки з капіталу	Енергетичні податки	Податки на працю	Темп росту ВВП	Номер кластера	Расстояние до центра кластера
▶ PT	36,7	19,1	19,6	143,4	29,6	100,2	3	0,26
IE	29,3	22,9	15,7	153,1	24,6	97	3	0,42
EE	32,2	20,9	10,7	71,5	33,7	96,4	3	0,44
ES	33,1	14,1	32,8	114,6	30,5	100,9	3	0,51

Результати по сформованих кластерах найбільш зручно розглядаються за допомогою візуалізатора "Куб", у який вбудована

крос-діаграма, що зображує отримані кластери в графічному виді, що суттєво спрощує аналіз (рис. 4.12).

При побудові крос-діаграми на панелі інструментів вікна крос-діаграми натисніть кнопки «Нормалізація, приведення графіків до єдиного масштабу».

Додайте в крос-діаграму всі параметри, по яких проводилася кластеризація, і легенду, яка вкаже яким кольором який параметр відображається.

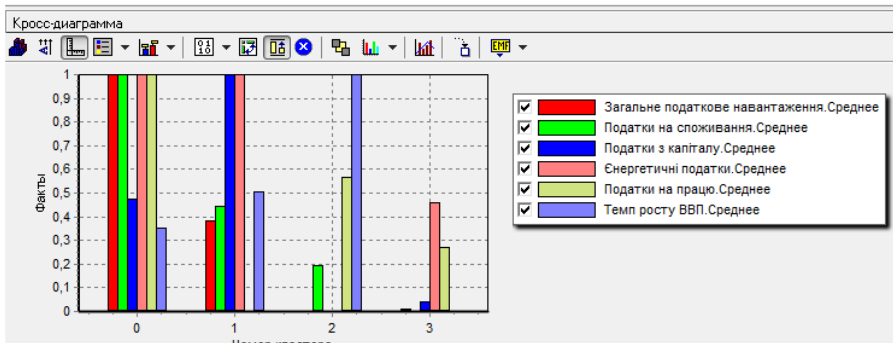


Рисунок 4.12 – Крос-діаграма – графіки приведені до єдиного масштабу

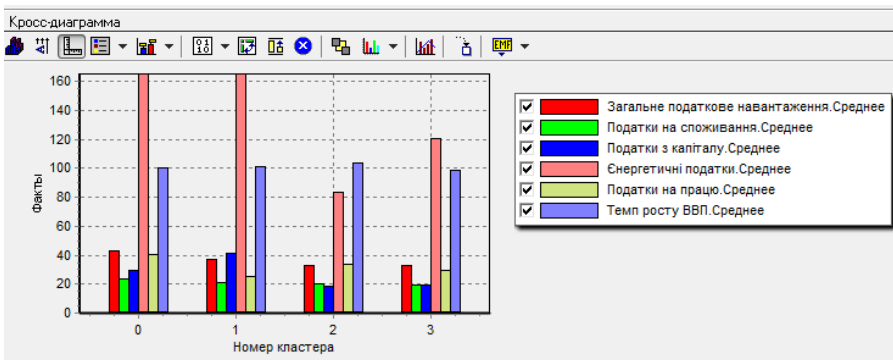


Рисунок 4.13 – Крос-діаграма – графіки не приведені до єдиного масштабу

Тепер починається головна робота менеджера-аналітика. Треба охарактеризувати кожен кластер та надати йому змістовне ім'я.

**Кластер 0** - “Політика високого оподаткування” утворюють найбільша кількість країн, і повністю представлений “старими” країнами-членами ЄС (Австрія, Бельгія, Німеччина, Іспанія, Фінляндія, Франція, Італія, Нідерланди, Португалія й Швеція). Рівень оподаткування в середньому по кластеру є найбільшим серед європейських країн. Забезпечуючи середній рівень оподаткування споживання, більшість країн даної групи мають досить високі бюджетні надходження екологічних податків і податків з капіталу, а відповідно високі ставки цих податків. До того ж, середня ефективність оподаткування праці в кластері є найбільшою серед країн-членів ЄС, що обумовлене прогресивністю шкали оподаткування трудових доходів і високими соціальними гарантіями громадянам у європейських країнах даного кластера. Гнучкість податкової політики країн, які поєднані у кластер, забезпечується стабільними показниками соціально-економічного розвитку.

**Кластер 1** - “Політика помірною оподаткування – стримуюча” поєднує 4 найбільш динамічні як в оподаткуванні, так і у власному економічному розвитку європейські країни (Кіпр, Люксембург, Мальта й Великобританія). Незважаючи на середній рівень оподаткування, більшість країн даної групи обмежують ефективність накопичення капіталу, обкладаючи його досить високими податками з капіталу. Вилучення екологічних податків і податків на споживання товарів і послуг у більшості країн цього кластера є найбільшими в Європі, тоді як ефективність оподаткування праці – найменшою. Останнє може бути пояснене відносно низькими ставками індивідуальних прибуткових податків із громадян і внесків на соціальне страхування, які компенсуються високими ставками корпоративних податків на прибуток і податків на споживання. У цілому ж, податкові реформи більшості країн даного кластера, а особливо Кіпру й Мальти, спрямовані на задоволення вимог, які пред'являються ЄС до країн-членів даного об'єднання, а також на збереження привабливості податкового режиму щодо міжнародного бізнесу і є логічними кроками на шляху розвитку країн як надійних європейських фінансових центрів.

У **Кластер 2** - “Політика низького оподаткування” увійшли 10 країн (Україна, Болгарія, Естонія, Литва, Латвія, Румунія й

Словаччина). Загальне податкове навантаження в більшості країн є значно меншим в порівнянні із середньоєвропейським. Незважаючи на те, що Україна й Болгарія мають вищий рівень оподаткування своїх економік у межах даного кластера, ефективність вилучення податків з капіталу, використання інструментів екологічного й споживчого оподаткування в порівнянні із середнім значенням по ЄС є низьким, що характерно для всіх європейських країн даного кластера. У той же час, рівень вилучення податків на працю відповідає майже середньому значенню по ЄС, при цьому більшість податків платяться роботодавцями. У цілому ж, економічні системи даного кластера схожі за рівнем соціально-економічного розвитку, динамічно ростуть по основних макроекономічних показниках і активно здійснюють реформування власних національних податкових систем, вектори якого спрямовані на підвищення економічного росту в країні.

**Кластер 3.** Чехія, Греція, Угорщина, Польща й Словенія належать до кластера 3 “Політика помірною оподаткування – стимулююча”. Податковий тиск на економіки цих країн відповідає середньому рівню по ЄС, але в порівнянні з іншими “новими” країнами-членами ЄС, держави даного кластера мають високу ефективність оподаткування споживання, капіталу й праці з мінімальною кількістю цих податків. Однак податки з капіталу й екологічні податки не обмежують активність економічних суб'єктів, а стимулюють їх до ефективного використання капіталу й енергозберігаючих технологій за рахунок низьких ставок корпоративних податків на прибуток і податків на енергетичну продукцію. Особливістю даного кластера є те, що країни даного об'єднання мають досить високий рівень оподаткування праці з відносно низькими ставками індивідуальних податків на трудові доходи й соціальних внесків з роботодавців і працівників у порівнянні із середньоєвропейськими показниками. У цілому ж, економічні системи країн близькі по своєму розвитку по більшості соціально-економічним показникам і мають більш інтенсивну економічну політику.

**Висновки.** За допомогою кластерного аналізу європейських країн і України за ступенем подібності показників, які характеризують критерії ефективності податкової політики, виділено чотири регіональні кластери. Отримані результати можуть бути використані при розробці основних напрямків реформування податкової політики



й підвищення рівня соціально-економічного розвитку України, враховуючи регіональні особливості розвитку країн-членів ЄС, а разом з комплексом ефективних законодавчих рішень здатні обґрунтувати адекватність тієї або іншої стратегічної альтернативи в контексті інтеграції України до європейського економічного простору.