

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний аерокосмічний університет ім. М.Є. Жуковського
„Харківський авіаційний інститут”

Ю.Ю. Гусєва, О.В. Гребенікова

СТАТИСТИЧНІ МЕТОДИ АНАЛІЗУ ФІНАНСОВО-ЕКОНОМІЧНИХ
ПРОЦЕСІВ

Навчальний посібник

Харків „ХАІ” 2010

УДК 336 (075.8)

Гусєва Ю.Ю. Статистичні методи аналізу фінансово-економічних процесів: навч. посіб. / Ю.Ю. Гусєва, О.В. Гребенікова. – Х.: Нац. аерокосм. ун-т „Харк. авіац. ін-т”, 2010. – 36 с.

Розглянуто основні принципи й методи наукового дослідження, методологічні принципи багатовимірної класифікації.

Аналітичні можливості та межі застосування конкретних моделей проілюстровано на фінансово-економічних прикладах. Розрахунки виконано за технологіями статистичного аналізу й оброблення даних, реалізованими в системі Statistica.

Для студентів, аспірантів і викладачів, діяльність яких пов'язана з обробленням та аналізом статистичної інформації.

Іл. 17. Табл. 8. Бібліогр.: 12 назв

Рецензенти: д-р екон. наук, проф. П.Т. Бубенко,
проф. В.І. Успенко

© Національний аерокосмічний університет ім. М.Є. Жуковського
„Харківський авіаційний інститут”, 2010 р.

ЗМІСТ

ВСТУП	5
1 ЗАГАЛЬНА МЕТОДИКА НАУКОВОЇ ТВОРЧОСТІ	6
1.1 Наукове вивчення як основна форма наукової роботи	6
1.2 Методи емпіричного дослідження.....	7
1.3 Методи теоретичних досліджень	8
1.4 Методи, що застосовуються на емпіричному та теоретичному рівнях досліджень	10
1.5 Загальна схема наукового дослідження.....	11
2 ОСНОВИ РОБОТИ З СИСТЕМОЮ STATISTICA	13
2.1 Формування інформаційної бази моделі	13
2.2 Розвідувальний аналіз даних	16
3 МЕТОДИ БАГАТОВИМІРНОЇ КЛАСИФІКАЦІЇ	18
3.1 Однорідність і типологія.....	19
3.2 Кластерні процедури класифікації	23
3.3 Дискримінантний аналіз.....	29
ЗАПИТАННЯ ДЛЯ САМОПЕРЕВІРКИ	34
БІБЛІОГРАФІЧНИЙ СПИСОК.....	35

ВСТУП

У сучасних умовах розвитку науково-технічного прогресу, інтенсивного збільшення обсягу наукової та науково-технічної інформації, швидкої змінюваності та оновлення знань особливого значення набуває підготовка у вищій школі висококваліфікованих фахівців, здатних до самостійної, творчої роботи, до впровадження у виробничий процес новітніх технологій.

З цією метою до навчального плану включено дисципліну «Основи наукових досліджень», яка розкриває перед студентами зміст наукового дослідження, дає уявлення про методи й методики проведення наукового дослідження, формує потребу в отриманні нових знань та інтерес до науки. Водночас ця дисципліна розкриває великі перспективи творчих аспектів вибраної спеціальності.

Оволодіння багатим арсеналом методів статистичного оброблення даних з використанням комп'ютерних технологій є важливою складовою професійної підготовки економіста. Відповідно до програми курсу в навчальному посібнику розглядаються методи багатовимірної класифікації.

Логічна структура аналізу ілюструється на конкретних прикладах соціально-економічного змісту. Для кожного типу моделей розглядаються принципи формування інформаційної бази, вибору процедур аналізу, інтерпретації результатів.

Акцентуючи увагу студентів на параметрах моделей, таблиці з результатами аналізу і графіки наводяться у стандартному вигляді англійською мовою. Специфікація включених у модель ознак і змістова інтерпретація параметрів моделі розкриваються в коментарях до таблиць і графіків.

1 ЗАГАЛЬНА МЕТОДИКА НАУКОВОЇ ТВОРЧОСТІ

1.1 Наукове вивчення як основна форма наукової роботи

Будь-яке наукове дослідження – від творчого задуму до кінцевого оформлення наукової праці – здійснюється індивідуально. Проте можна визначити і деякі загальні методологічні підходи до його проведення, що прийнято називати вивченням у науковому розумінні.

Накопичення наукових фактів під час дослідження – завжди творчий процес, в основі якого лежить задум ученого, його *ідея*. У філософському розумінні ідея являє собою продукт людської думки, форму відображення дійсності. Ідея відрізняється від інших форм мислення й наукового знання тим, що в ній не лише відображено об'єкт вивчення, але й міститься усвідомлення мети, перспективи пізнання і практичного перетворення дійсності.

Ідеї народжуються із практики, спостереження навколишнього світу і потреб життя. У їхній основі лежать реальні факти і події. Життя висуває конкретні завдання, але не завжди відразу знаходяться продуктивні ідеї для їх вирішення. Тоді на допомогу приходить спроможність дослідника пропонувати новий, зовсім незвичний аспект розгляду завдання, котре довго не могли вирішити при звичайних підходах до справи.

Нова ідея – не просто зміна уявлення про об'єкт дослідження шляхом суворого обґрунтування, це якісний стрибок думки за межі сприйнятих почуттями даних і перевірених рішень. Нові ідеї можуть виникати під впливом парадоксальних ситуацій, коли виявляється незвичний, несподіваний результат, який надто розбіжний із загальноприйнятими положеннями науки – парадигмами. При цьому отримання нових знань відбувається за такою схемою: парадигма – парадокс – нова парадигма. Можна стверджувати, що розвиток науки – це зміна відмінних парадигм, методів, стереотипів мислення. Перехід від однієї парадигми до іншої не піддається логічному опису, бо кожна з них відкидає попередню і несе принципово новий результат дослідження, котрий не можна логічно вивести із відомих теорій. Особливу роль тут відіграють інтуїтивні механізми наукового пошуку, які не ґрунтуються на формальній логіці.

Розвиток ідеї до стадії вирішення завдання зазвичай здійснюється як плановий процес наукового дослідження. Хоча в науці й відомі випадкові відкриття, проте тільки планове, добре обладнане сучасними засобами наукове дослідження дає змогу розкрити і глибоко пізнати об'єктивні закономірності в природі. Згодом іде процес продовження цільового оброблення первинного задуму, уточнення, зміни,

доповнення і розвитку накресленої схеми дослідження з використанням різних методів пізнання.

Метод – це сукупність прийомів чи операцій практичного або теоретичного освоєння дійсності, підпорядкованих розв'язанню конкретної задачі. Фактично різниця між методом і теорією має функціональний характер: формуючись як теоретичний результат попереднього дослідження, метод виступає як вихідний пункт та умова майбутніх досліджень.

У кожному науковому дослідженні можна виділити два рівні: емпіричний, на якому відбувається процес накопичення фактів; теоретичний – досягнення синтезу знань (у формі наукової теорії).

Згідно з названими рівнями загальні методи пізнання можна поділити на три групи, грані між якими визначено приблизно:

- методи емпіричного дослідження;
- методи теоретичного дослідження;
- методи, що використовуються на емпіричному та теоретичному рівнях.

1.2 Методи емпіричного дослідження

Спостереження – це систематичне цілеспрямоване вивчення об'єкта. Аби бути плідним, спостереження мусить відповідати:

- завчасному задуму (спостереження провадиться для певного, чітко поставленого завдання);
- планомірності (виконується за планом, складеним відповідно до завдання спостереження);
- цілеспрямованості (спостерігаються лише певні сторони явища, котрі викликають інтерес при дослідженні);
- активності (спостерігач активно шукає потрібні об'єкти, риси явища);
- систематичності (спостереження ведеться безперервно або за певною системою).

Порівняння – це процес визначення подібності або відмінності предметів та явищ дійсності, а також знаходження спільного, притаманного двом або кільком об'єктам.

Застосування методу порівняння має результат, якщо:

- порівнюються тільки такі явища, між якими можлива деяка об'єктивна спільність;
- порівняння здійснюється за найбільш важливими, суттєвими (в плані конкретного завдання) рисами.

Різні об'єкти чи явища можуть порівнюватися безпосередньо або опосередковано через їх порівняння з будь-яким іншим об'єктом (еталоном). У першому випадку зазвичай отримують якісні результати (більше-менше, вище-нижче). Порівняння ж об'єктів з еталоном дає можливість отримати кількісні характеристики. Таке порівняння називають вимірюванням.

Вимірювання – це визначення числового значення певної величини за допомогою одиниці виміру. Вимірювання передбачає наявність таких основних елементів: об'єкт вимірювання, еталон, вимірювальні прилади, метод вимірювання.

Експеримент – це такий метод вивчення об'єкта, за яким дослідник активно і цілеспрямовано впливає на нього завдяки створенню штучних умов або використанню природних умов, необхідних для виявлення відповідної властивості. Переваги експериментального вивчення об'єкта порівняно зі спостереженням такі:

- під час експерименту можна вивчати явище «у чистому вигляді», звільнившись від побічних факторів, які затінюють основний процес;
- в експериментальних умовах можна дослідити властивості об'єктів;
- повторюваність експерименту: можна проводити дослідження стільки разів, скільки це потрібно.

У науковому дослідженні експеримент і теорія найтісніше взаємозв'язані. Усіляке ігнорування експерименту неодмінно призводить до помилок, тому всебічне розгортання експериментальних досліджень являє собою один із найважливіших шляхів розвитку сучасної науки.

1.3 Методи теоретичних досліджень

Ідеалізація – це конструювання подумки об'єктів, які не існують насправді або майже нездійсненні (наприклад, абсолютно тверде тіло, абсолютно чорне тіло, лінія, площина).

Формалізація – метод вивчення різноманітних об'єктів шляхом відображення їхньої структури у знаковій формі за допомогою штучних мов, наприклад мовою математики. Переваги формалізації:

- вона забезпечує узагальненість підходу до вирішення проблем;
- символіка надає стислості й чіткості фіксації значень;
- однозначність символіки (немає багатозначності звичайної мови);

- дає змогу формувати знакові моделі об'єктів та замінювати вивчення реальних речей і процесів вивченням цих моделей.

Аксиоматичний метод – метод побудови наукової теорії, при якому деякі твердження приймаються без доведень, а всі інші знання виводяться з них відповідно до певних логічних правил.

Гіпотеза й припущення. У становленні теорій як системи наукового знання найважливішу роль відіграє гіпотеза. Гіпотеза є формою осмислення фактичного матеріалу, переходу від фактів до законів.

Розвиток гіпотези відбувається за трьома стадіями:

- накопичення фактичного матеріалу і висловлення на його основі припущень;
- формування гіпотези, тобто виведення висновків із зробленого припущення, розгортання на його основі прийнятної теорії;
- перевірка отриманих результатів на практиці і на її основі уточнення гіпотези; якщо під час перевірки висновок відповідає дійсності, то гіпотеза перетворюється на наукову теорію.

Гіпотези (як і ідеї) мають імовірнісний характер. На їхній основі відбувається систематизація накопичених раніше знань і здійснюється пошук нових наукових результатів – у цьому суть і призначення гіпотези як форми розвитку науки.

Історичний метод дає змогу дослідити виникнення, формування й розвиток процесів і подій у хронологічній послідовності з метою виявлення внутрішніх і зовнішніх зв'язків, закономірностей й суперечностей. Цей метод дослідження використовується переважно у суспільних науках. У прикладних – він застосовується, приміром, при вивченні розвитку і формування тих чи інших галузей науки і техніки.

Системний підхід полягає у комплексному дослідженні великих і складних об'єктів (систем), дослідженні їх як єдиного цілого із узгодженим функціонуванням усіх елементів і частин. З огляду на цей принцип треба вивчити кожен елемент системи у його зв'язку та взаємодії з іншими елементами, виявити вплив властивостей окремих частин системи на її поведінку в цілому, визначити емерджентні властивості системи і оптимальний режим її функціонування.

Ускладнення задач та об'єктів дослідження спричиняє необхідність розподілення (декомпозиції) системи на системи нижчого рівня (підсистеми), які досліджуються автономно, причому з обов'язковим урахуванням подальшого узгодження цілей кожної підсистеми із загальною метою системи. Таким чином, декомпозиція наперед визначає створення ієрархії системи. Застосування декомпозиції зумовлене не тільки неможливістю охопити неосяжне, але й різномірністю елементів

складної системи і, як наслідок, необхідністю залучення фахівців різного профілю.

1.4 Методи, що застосовуються на емпіричному та теоретичному рівнях досліджень

Абстрагування – відхід у думці від несуттєвих властивостей, зв'язків, відношень предметів і виділення декількох рис, котрі цікавлять дослідника.

Аналіз і синтез. Аналіз – метод пізнання, який дає змогу поділяти предмети дослідження на складові частини (природні елементи об'єкта або його властивості й відношення). Синтез, навпаки, припускає з'єднання окремих частин або рис предмета в єдине ціле. Аналіз і синтез взаємозв'язані, вони являють собою єдність протилежностей.

Аналіз і синтез можуть бути:

- прямими, або емпіричними (використовуються для виокремлення частин об'єкта, виявлення його властивостей, найпростіших вимірювань і т. ін.);
- зворотними, або елементарно-теоретичними (базуються на деяких теоретичних міркуваннях стосовно причинно-наслідкового зв'язку різних явищ або дії будь-якої закономірності; при цьому виділяються та з'єднуються явища, які здаються суттєвими, а другорядні – ігноруються);
- структурно-генетичними (потребують виокремлення у складному явищі таких елементів, які мають вирішальний вплив на всі інші сторони об'єкта).

Індукція та дедукція. Дедуктивною називають таку розумову конструкцію, у якій висновок щодо якогось елемента множини робиться на підставі знання загальних властивостей усієї множини. Змістом дедукції як методу пізнання є використання загальних наукових положень при дослідженні конкретних явищ.

Під індукцією розуміють перехід від часткового до загального, коли на підставі знання про частину предметів класу робиться висновок стосовно класу в цілому. Дедукція та індукція – взаємопротилежні методи пізнання.

Моделювання – метод, який ґрунтується на використанні моделі як засобу дослідження явищ і процесів природи. Під моделями розуміють системи, що замінюють об'єкт пізнання і є джерелом інформації стосовно нього. Моделі – це такі аналоги, подібність яких до оригіналу є суттєвою, а розбіжність – несуттєвою. Моделі поділяють на два види: матеріальні й ідеальні. Матеріальні моделі втілюються у певному матеріалі – дереві, металі, склі й т. ін. Ідеальні моделі фіксуються в

таких наочних елементах, як креслення, рисунок, схема, комп'ютерна програма тощо.

Метод моделювання має таку структуру: а) постановка завдання; б) створення або вибір моделі; в) дослідження моделі; г) перенесення знань із моделі на оригінал.

1.5 Загальна схема наукового дослідження

Увесь процес наукового дослідження можна приблизно зобразити у вигляді такої логічної схеми:

- обґрунтування актуальності вибраної теми;
- постановка мети й конкретних завдань дослідження;
- визначення об'єкта й предмета дослідження;
- вибір методів (методики) проведення дослідження;
- опис процесу дослідження;
- обговорення результатів дослідження;
- формулювання висновків і оцінювання одержаних результатів.

Обґрунтування актуальності вибраної теми – початковий етап будь-якого дослідження.

Висвітлення актуальності не має бути багатослівним. Досить кількома реченнями висловити головне – суть проблеми, з чого й впливе актуальність теми. Проблема завжди виникає тоді, коли колишнє знання вже виявило свою неспроможність, а нове ще не набуло розвиненої форми. Таким чином, проблема в науці – це суперечлива ситуація, яка потребує свого вирішення. Така ситуація найчастіше виникає внаслідок відкриття нових фактів, які явно не вкладаються у рамки колишніх теоретичних уявлень, тобто коли жодна з теорій не може пояснити щойно виявлені факти.

Правильна постановка та ясне формулювання нових проблем часом мають не менше значення, ніж вирішення цих проблем. По суті, саме вибір проблеми якщо не цілком, то дуже великою мірою визначає як стратегію дослідження взагалі, так і напрямок наукового пошуку зокрема. Невипадково вважається, що сформулювати наукову проблему – означає показати вміння відокремити головне від другорядного, виявити відоме й невідоме науці про предмет дослідження.

Від доведення актуальності вибраної теми слід перейти до *формулювання мети дослідження*, а також указати конкретні завдання, які мають бути вирішені відповідно до цієї мети. Це зазвичай робиться у формі перелічення (вивчити..., описати .., встановити..., з'ясувати..., вивести формулу... і т. ін.).

Мета дослідження – це запланований результат, який має бути конструктивним, тобто спрямованим на вироблення суспільно корисного продукту з ліпшими, ніж було раніше, показниками якості або процесу її досягнення.

Надалі формулюються *об'єкт і предмет дослідження*. Об'єкт – це процес або явище, що породжують проблемну ситуацію і вибрані для вивчення. Предмет – це те, що міститься в межах об'єкта.

Об'єкт і предмет дослідження як категорії наукового процесу співвідносяться як загальне і часткове. В об'єкті виділяється та його частина, яка є предметом дослідження. Саме на нього і спрямована основна увага дослідника, саме предмет дослідження визначає тему наукової роботи, що виноситься на титульний аркуш як заголовок.

Дуже важливим етапом наукової праці є *вибір методів дослідження* – інструменту отримання фактичного матеріалу і необхідної умови досягнення поставленої в роботі мети.

Опис процесу дослідження – основна частина наукової праці, де висвітлюються методика і техніка дослідження з використанням логічних законів і правил.

Дуже важливий етап процесу наукового дослідження – *обговорення його результатів* (на конференціях, засіданнях профілюючих кафедр, учених і науково-технічних радах, наукових семінарах).

Заключним етапом процесу наукового дослідження є *висновки*, які містять те нове і суттєве, що становить наукові та практичні результати виконаної наукової роботи.

Наукові результати можна поділити на два види: теоретико-методологічні (для теоретичних досліджень), зокрема концепція, гіпотеза, класифікація, закон, метод; інструментальні (для прикладних і емпіричних досліджень), зокрема спосіб, технологія, методика, алгоритм, речовина.

Найбільш значущі результати наукового дослідження зазвичай прийнято відображати в наукових статтях.

Наукова стаття є одним із видів публікацій, в якій подаються проміжні або кінцеві результати, висвітлюються конкретні окремі питання за темою дослідження, фіксується науковий пріоритет автора. Стаття – це самостійний науковий твір, що являє собою висловлення своїх думок з актуальної наукової проблеми. У статті зі змістовного боку можуть розкриватися конкретні питання теоретичної і прикладної роботи дослідника.

Основні етапи роботи над статтею:

- визначення теми, її аналіз, складання плану (тез);
- робота над першим варіантом статті на основі плану (тез);
- роздум, обдумування, вирішення проблеми.

- послідовне викладення думки, доказове міркування, точний вибір слів;

- критичне оцінювання результатів роботи, пошук слабких місць, помилок і їх виправлення.

Відповідно до Постанови президії ВАК (Вищої атестаційної комісії) України від 15.01.2003 р. № 7-05/1 “Про підвищення вимог до фахових видань, внесених до переліків ВАК України” (Бюлетень ВАК України. – 2003. – № 1) наукова стаття має містити такі обов’язкові елементи:

- постановка проблеми у загальному вигляді та її зв’язок з важливими науковими або практичними завданнями;

- аналіз останніх досліджень і публікацій, в яких започатковано вирішення проблеми і на які спирається автор;

- виділення не вирішених раніше частин загальної проблеми, яким присвячується певна стаття;

- формулювання цілей статті (постановка завдання);

- викладення основного матеріалу дослідження з повним обґрунтуванням отриманих наукових результатів;

- висновки щодо результатів і перспективи подальших досліджень у даному напрямку.

Наукова стаття подається до редакції завершеною, відповідно до вимог, які публікуються в окремих номерах журналів або збірниках у вигляді пам’ятки автору. Обсяг наукової статті 0,5 - 0,7 авт.арк. вважається оптимальним. Рукопис статті повинен мати повну назву роботи, прізвище та ініціали автора, анотацію, список використаної літератури.

2 ОСНОВИ РОБОТИ З СИСТЕМОЮ STATISTICA

2.1 Формування інформаційної бази моделі

Масив первинних даних у системі *Statistica* організується та зберігається у вигляді електронної таблиці *Spreadsheet* з рядками *Cases* і стовпцями *Variables*, тобто по рядках електронної таблиці розміщуються елементи статистичної сукупності ($j = 1, 2, \dots, n$), по стовпцях – ознаки ($i = 1, 2, \dots, m$). Клітинки таблиці призначено для введення числової або текстової інформації. В системі реалізовано так званий механізм «подвійного запису», за яким встановлюється еквівалент число = текстове значення. Скажімо, для ознаки «стать»: 1 = чол., 2 = жін. Під час опрацювання інформації можна переключатися з од-

ного типу даних на інший.

Уведення первинних даних у таблицю здійснюється різними способами, а саме:

- безпосередньо з клавіатури;
- перетворенням існуючого масиву даних за допомогою певних операцій (ранжування, стандартизації), математичних або статистичних функцій (ln, sin тощо);
- імпортуванням даних з інших додатків Windows, наприклад з Excel.

При формуванні нового файлу через команду *New* автоматично створюється таблиця розміром 10x10: по стовпцях – 10 ознак (VAR1, VAR2 ...) і по рядках – 10 спостережень. За допомогою кнопок *Vars* і *Cases* (або пункту меню *Insert*) можна додати або видалити стовпці й строки (команди *Delete/Add Variables* і *Delete/Add Cases*).

Необхідним етапом формування інформаційної бази є специфікація ознак, тобто визначення основних параметрів кожної з них: імені, формату, коду для пропущених даних, формули.

Найпростіший спосіб надання специфікації ознакам – через команду *Specs* – *Специфікації* (кнопка VARS). Наведемо основні з них:

Name – ім'я, не більше восьми символів. У протилежному разі необхідно використати мітку *Label* у полі *Long Name*, скажімо, *Name: ВВП; Label: Валовий внутрішній продукт 1999 р. у поточних цінах, млрд грн*;

MD code – код, який приписується пропущеним даним. Можна вилучити дані з розрахунків, замінити їх середніми значеннями або інтерполювати. За замовчуванням *MD code* становить 9999.

Decimals – кількість розрядів після коми.

Display Format – спосіб подання ознак різних типів: числа, дати, час, науковий формат, грошовий формат, проценти.

Якщо значення певної ознаки необхідно розрахувати, то в текстовому полі *Long Name* задається формула, за якою ведеться розрахунок. Формула починається із знака «=». Якщо потрібно, у формулі використовуються математичні або статистичні функції – *Functions*.

Інші параметри електронної таблиці встановлюються за допомогою функціональних кнопок на панелі інструментів (імена спостережень (рядків) – за допомогою кнопки *Cases*). Після того, як усі параметри таблиці задано, можна вводити дані.

Процедуру створення файлу первинних даних розглянемо на прикладі 25 комерційних банків України, які характеризуються сьома ознаками. Специфікацію ознак показано на рисунку 2.1, а отриману в результаті таблицю даних – на рисунку 2.2.

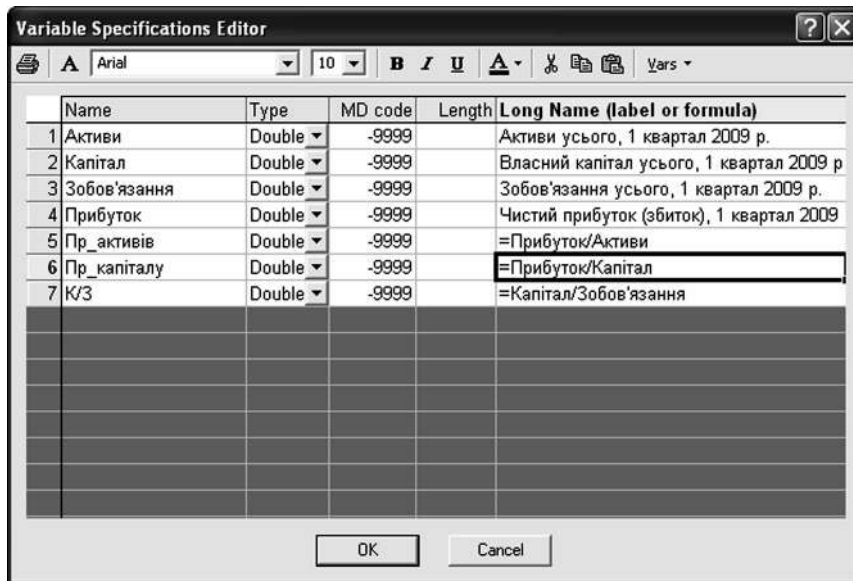


Рисунок 2.1 – Специфікація ознак

	1	2	3	4	5	6	7
	Активи	Капітал	Зобов'язання	Прибуток	Пр_активів	Пр_капіталу	К/З
1	79 308 914	8 572 810	70 736 104	422 272	0,51%	4,72%	175,33321
2	37 222 685	6 454 542	30 768 061	-1 664 896	-1,50%	-18,81%	-46,7914434
3	56 730 780	8 106 254	48 624 526	12 205	0,02%	0,15%	3683,9841
4	53 242 332	16 462 963	37 130 267	733 679	1,41%	4,59%	49,2524836
5	99 971 396	5 277 259	45 694 637	-289 665	-1,63%	-16,74%	-65,0307296
6	44 975 616	5 242 440	39 733 172	61 637	0,14%	1,18%	644,422726
7	32 382 326	3 141 326	29 241 001	-192 818	-0,60%	-6,14%	-151,851822
8	29 732 368	1 923 538	27 808 770	-60 324	-1,50%	-29,25%	-49,1756222
9	27 866 166	3 268 295	24 597 870	-371 211	-1,36%	-11,48%	-66,557433
10	27 283 472	2 461 519	24 821 953	1 743	0,01%	0,07%	142,491119
11	36 392 601	797 374	35 595 227	-1 640 885	-3,38%	-130,51%	-24,4833716
12	19 131 633	2 246 906	16 884 727	-283 290	-1,48%	-13,61%	-69,6022696
13	17 756 640	2 871 260	14 885 382	-694 127	-3,40%	-21,04%	-24,6394914
14	17 389 472	2 289 080	15 100 392	-187 364	-1,07%	-8,19%	-81,6623808
15	14 309 214	2 322 960	11 986 254	9 939	0,07%	0,43%	128,64863
16	13 949 944	1 847 752	11 993 193	-478 242	-3,46%	-26,92%	-25,0410133
17	12 139 814	1 372 698	10 767 115	-76 469	-0,63%	-5,57%	-140,395076
18	9 401 491	-2 677 407	12 147 602	-4 653 184	-46,51%	171,95%	-2,6390229
19	13 154 323	1 401 538	11 752 784	-166 471	-1,27%	-11,71%	-70,4864683
20	10 480 382	1 444 733	9 035 646	41 270	0,39%	2,86%	216,944657
21	10 397 931	1 022 960	9 174 969	-340 148	-3,27%	-33,55%	-36,7396801
22	9 012 392	771 197	8 241 196	-20 211	-0,29%	-6,34%	-154,965444
23	8 730 079	-211 802	8 941 881	-1 510 894	-17,31%	713,35%	-5,91827157
24	8 239 182	1 349 313	6 910 870	92 656	1,12%	6,87%	74,5627177
25	7 724 003	767 234	6 956 769	-49 079	-1,29%	-13,02%	-69,7030256

Рисунок 2.2 – Таблиця даних у системі Statistica

Для роботи з даними в системі Statistica реалізовано всі стандартні операції методу *Drag-and-Drop*, зокрема:

- копіювання, переміщення, вставка;
- автозаповнення блоків з регулярною структурою;
- стандартизація даних;
- транспонування;
- ранжування даних;
- перекодування значень ознаки, перехід від однієї шкали вимірювання до іншої.

2.2 Розвідувальний аналіз даних

Статистичне оброблення даних у будь-якому модулі генерує велику кількість вихідної інформації у вигляді електронних таблиць *Scrollsheet* і графіків. Таблиці *Scrollsheet* підтримують усі стандартні операції з виділеними блоками значень (копіювання, переміщення, вставка, екстраполяція, стандартизація даних тощо). Таблицю *Scrollsheet* можна відредагувати, зберегти як файл результатів (з розширенням *scr*) або конвертувати в первинні дані (файл з розширенням *sta*), можна експортувати в інші додатки *Windows*.

Діалогове вікно процедури (рисунок 2.3) пропонує вибрати тип аналізу:

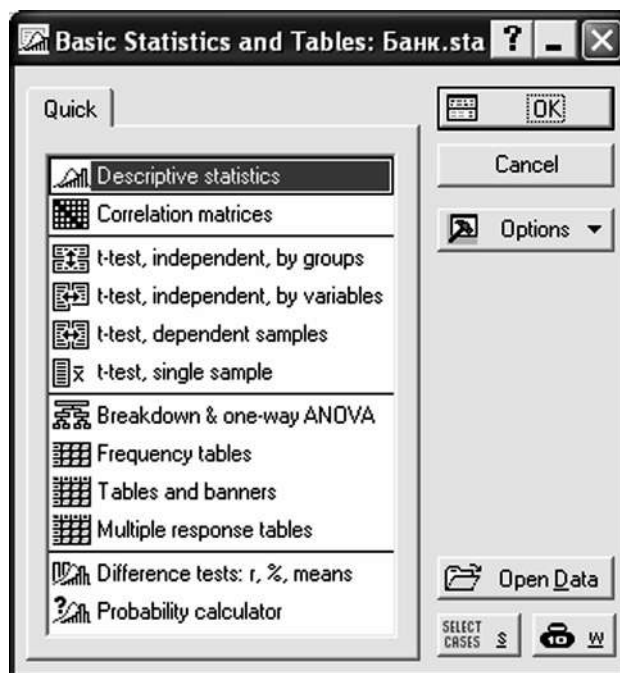


Рисунок 2.3 – Діалогове вікно *Basic Statistics/Tables*

1 Описові статистики і графіки. Програма обчислює описові статистики загального характеру: медіану, моду, квартилі, задані користувачем процентілі, середнє значення і стандартне відхилення, квартильний розмах, довірчі інтервали для середнього, асиметрію і ексцес (і їхні стандартні помилки), гармонійне і геометричне середнє.

У таблиці 2.1 наведено результати обчислень для 25 комерційних банків (*Statistics – Basic Statistics/Tables – Descriptive Statistics – Advanced*): *Mean* – середня величина, *Minimum*, *Maximum* – мінімальне та максимальне значення, *Standard Deviation* – стандартне відхилення, *Skewness* – коефіцієнт асиметрії та *Kurtosis* – коефіцієнт ексцесу.

Таблиця 2.1 – Описові статистики

Показники	Mean	Minimum	Maximum	Std.Dev.	Skewness	Kurtosis
Активи	26332613	7729003	79508914	19799383	1,186	0,609
Капітал	3140563	-2677417	16422065	3735416	2,059	5,983
Зобов'язання	23192050	6910870	70936104	16968193	1,260	1,086
Прибуток	-444748	-4603184	753673	991543	-3,272	13,398

2 Кореляції (Statistics – Basic Statistics/Tables – Descriptive Statistics – Correlation matrice). У системі є великий набір методів для дослідження кореляцій між змінними. У вікні Product-Moment and Partial Correlation вибираємо тип матриці – *One variable list* – *Один список ознак*, а у вікні вибору ознак – ознаки для аналізу. За командою на виконання процедури створюється нова таблиця, елементами якої є парні коефіцієнти кореляції (таблиця 2.2).

Таблиця 2.2 – Кореляції

Показники	Активи	Капітал	Зобов'язання	Прибуток
Активи	1,00	0,80	0,99	0,24
Капітал	0,80	1,00	0,71	0,50
Зобов'язання	0,99	0,71	1,00	0,17
Прибуток	0,24	0,50	0,17	1,00

3 Т-критерії (та інші критерії групових відмінностей). У цьому вікні можуть бути обчислені Т-критерії для залежних і незалежних виборок.

4 Калькулятор імовірнісних розподілів. Він підтримує стандартні розподіли: бета, Коші, χ -квадрат, експоненціальне, екстремальне (Гумбеля), F, гамма, Лапласа, лог-нормальне, логістичне, Парето, Релея, t (Стюдента), Вейбулла і Z (нормальне).

Статистичне оброблення даних у будь-якому модулі генерує велику кількість вихідної інформації у вигляді електронних таблиць *Scrollsheet* і графіків. Таблиці *Scrollsheet* підтримують усі стандартні операції з виділеними блоками значень (копіювання, переміщення, вставка, екстраполяція, стандартизація даних тощо). Відредаговану числову, текстову та графічну інформацію можна експортувати в інші додатки *Windows*, на принтер, у текстовий файл.

3 МЕТОДИ БАГАТОВИМІРНОЇ КЛАСИФІКАЦІЇ

Соціальні та економічні об'єкти, як правило, характеризуються чималою кількістю параметрів, утворюючих багатовимірні вектори, і особливого значення в економічних і соціальних дослідженнях набувають задачі вивчення взаємозв'язків між компонентами цих векторів, причому ці взаємозв'язки необхідно виявляти на підставі обмеженої кількості багатовимірних спостережень.

Поділ сукупності на однорідні класи (групи, кластери) називають *класифікацією*. Ідея класифікації ґрунтується на поняттях подібності й відмінності. Методологічний принцип класифікації містить два фундаментальних положення:

- в один клас об'єднуються подібні, схожі між собою одиниці сукупності;
- ступінь подібності, схожості одиниць, які належать до одного класу, вища, ніж ступінь подібності одиниць, віднесених до різних класів.

Оцінювання подібності здійснюється на основі однієї або декількох ознак, які, на думку експертів, формують «образ класу». В традиційній схемі класифікації ці ознаки ієрархічно впорядковуються за своєю вагомістю, наприклад, класифікація шахт – за гірничо-геологічними умовами виробництва: потужністю пласта, нахилом його залягання, глибиною розробки лав, загазованістю лав і т. д. Саме так будується більшість комбінаційних групувань. На кожному кроці поділу сукупності до уваги береться лише одна ознака, тобто відбувається послідовне формування, покрокове уточнення, детальніше описування класів. У невеликих за обсягом сукупностях можливості використання такої схеми класифікації обмежені.

У другій схемі класифікації використовується множина класифікаційних ознак одночасно. Будь-яка одиниця сукупності, описана множиною ознак, геометрично інтерпретується як точка у багатовимірному просторі, а близькість двох точок розглядається як подібність їх, однорідність. Існують різні варіанти реалізації багатовимірної схеми класифікації. Їх можна об'єднати в два блоки:

- конструювання багатовимірних інтегральних оцінок (індексів, рейтингів), на основі яких проводиться класифікація за традиційною схемою;
- автоматична багатовимірна класифікація методами кластерного аналізу, коли поняття однорідності задається певними метриками.

Слід зазначити, що класифікація за будь-якою схемою є певною

мірою суб'єктивною, оскільки результати її визначаються передусім множиною класифікаційних ознак та їхніми розмежувальними властивостями.

Багатовимірним статистичним аналізом називають розділ математичної статистики, що вивчає методи збору і оброблення багатовимірних статистичних даних, їх систематизації з метою виявлення характеру й структури взаємозв'язків між компонентами досліджуваної багатовимірної ознаки, отримання практичних висновків.

Методи зменшення розмірності багатовимірного простору дають можливість без істотної втрати інформації перейти від початкової системи великої кількості досліджуваних взаємозв'язаних чинників до системи істотно меншої кількості прихованих чинників, що визначають варіацію первинних ознак.

Методи багатовимірної класифікації призначено для поділу сукупностей об'єктів (що характеризуються великою кількістю ознак) на класи, до кожного з яких мають входити об'єкти в певному розумінні однорідні або близькі. Таку класифікацію можна провести методами кластерного і дискримінантного аналізу.

3.1 Однорідність і типологія

Однією з умов статистичного моделювання є однорідність сукупності. Лише в однорідній сукупності виявлені закономірності є сталими і їх можна застосувати до всіх одиниць сукупності.

Поняття *однорідності* пов'язують з наявністю в усіх одиницях сукупності таких спільних властивостей і рис, які визначають їх належність до одного й того ж типу. Оцінювання ступеня однорідності здійснюється за допомогою критеріїв математичної статистики, більшість з яких орієнтовано на аналіз форми одновершинних розподілів.

Однорідними вважаються сукупності, яким властивий симетричний, нормальний розподіл. Звісно, в соціально-економічних явищах нормальний розподіл у чистому вигляді не існує. Але він близький до інших одновершинних розподілів, його часто використовують як перше наближення при моделюванні. Деякі одновершинні розподіли зводяться до нормального виду перетворенням значень ознак, скажімо, заміною їх логарифмами. Логонормальною кривою можна описати низку асиметричних розподілів, передусім з правосторонньою асиметрією.

Основні властивості нормального розподілу:

- крива розподілу симетрична відносно максимальної ординати, яка відповідає середньому арифметичному значенню \bar{x} ;

- у межах $\bar{x} \pm \sigma$ міститься 68,3 % усіх частот ряду розподілу, в межах $\bar{x} \pm 2\sigma$ – 95,4 % частот, у межах $\bar{x} \pm 3\sigma$ – 99,7 % частот;

- співвідношення стандартного відхилення σ і середнього модуля відхилень \bar{l} становить $\frac{\bar{l}}{\sigma} = \sqrt{\frac{2}{\pi}} = 0,8$ або $\frac{\sigma}{\bar{l}} = 1,25$. Значення його

залежить від наявності в сукупності нетипових, аномальних спостережень і може бути індикатором її «засміченості»;

- коефіцієнт асиметрії розподілу дорівнює нулю, коефіцієнт ексцесу – трьом.

Завдяки цим властивостям нормальна крива застосовується як *стандарт* і відіграє значну роль при використанні методів вибіркового, регресійного, факторного аналізу. Оцінювання ступеня наближення до цього стандарту ґрунтується на порівнянні емпіричних f_i і теоретичних \tilde{f}_i частот розподілу, де i – номер інтервалу. Теоретичні частоти визначають за формулою

$$\tilde{f}_i = n |F_{x_i} - F_{x_{i-1}}|,$$

де $F_x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$ – інтегральна функція нормального розподілу;

$z = \frac{x - \bar{x}}{\sigma}$ – стандартизоване значення ознаки x ;

$n = \sum f_i$ – обсяг сукупності.

3-поміж критеріїв, що застосовуються для оцінювання узгодженості емпіричного розподілу з нормальним, найпоширенішими є критерії Пірсона χ^2 та Колмогорова – Смирнова d .

Фактичне значення критерію χ^2 порівнюється з критичним значенням $\chi^2_{1-\alpha}(k)$ (для ймовірності $1 - \alpha$ і відповідної кількості ступенів свободи $k = m - q - 1$, де m – кількість інтервалів групування, q – кількість параметрів функції (для нормального розподілу $q = 2$)). Якщо $\chi^2 < \chi^2_{1-\alpha}(k)$, відхилення емпіричного розподілу від нормального є неістотним. У разі, коли фактичне значення критерію перевищує критичне, гіпотеза про узгодженість розподілів відхиляється. Критичні значення критерію χ^2 наведено в таблиці 3.1.

Висновки за критерієм χ^2 значною мірою залежать від кількості груп, частота яких має бути не менше п'яти. Окрім того, χ^2 не враховує послідовність знаків відхилень частот (часток), за наявності серій

знаків надійність висновку зменшується. Цих вад позбавлений критерій Колмогорова – Смирнова d . Критичні значення d наведено в таблиці 3.2.

Таблиця 3.1 – Критичні значення $\chi_{1-\alpha}(k)$

k	1	2	3	4	5	6	7	8	9	10	11
$\alpha = 0,10$	2,71	4,61	6,25	7,78	9,24	10,64	12,02	13,36	14,68	15,99	17,28
$\alpha = 0,05$	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31	19,67

Таблиця 3.2 – Критичні значення критерію Колмогорова – Смирнова

Обсяг вибірки n	$\alpha = 0,10$	$\alpha = 0,05$	Обсяг вибірки n	$\alpha = 0,10$	$\alpha = 0,05$
5	0,510	0,563	15	0,304	0,338
6	470	521	16	295	328
7	438	486	17	286	318
8	411	457	18	278	309
9	388	432	19	272	301
10	368	409	20	264	294
11	352	391	25	240	264
12	338	375	30	220	242
13	325	361	$n > 30$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$

У системі *Statistica* закономірність одновимірного розподілу можна аналізувати за допомогою процедур *Distribution* стартової панелі *Descriptive statistics* (модуль *Basic Statistics and Tables*). Розподіл сукупності за варіаційною ознакою подається у вигляді таблиці *Frequency Table*, де вказуються інтервали групувань, частоти і частки розподілу за інтервалами, а також кумулятивні частоти і частки. Опція *Normal expected frequencies* додає в таблицю теоретичні частоти і частки (групові й кумулятивні). Перевірка на нормальність розподілу здійснюється за опцією *K-S and Lilliefors test for normality*.

У таблиці 3.3 наведено розподіл 25 комерційних банків за розміром власного капіталу.

У верхньому лівому куті таблиці подано значення критерію $d = 0,24636$. Значення d менше за критичне (0,264) при $\alpha = 0,05$, отже, гіпотеза про нормальний розподіл комерційних банків за рівнем власного капіталу не відхиляється.

Таблиця 3.3 – Характеристики розподілу (власний капітал)

Frequency table: Капітал: Власний капітал усього, 1-й квартал 2009 р. (Банк.sta) K-S d=,24636, p<,10 ; Lilliefors p<,01						
Category	Count	Cumulative	Percent	Cumul %	% of all	Cumulative %
-5E6 <x<=0,000000	2	2	8,00	8,00	8,00	8,00
0,000000<x<=5000000,	17	19	68,00	76,00	68,00	76,00
5000000,<x<=1E7	5	24	20,00	96,00	20,00	96,00
1E7 <x<=15E6	0	24	0,00	96,00	0,00	96,00
15E6 <x<=2E7	1	25	4,00	100,00	4,00	100,00
Missing	0	25	0,00		0,00	100,00

Узгодженість емпіричного розподілу комерційних банків за рівнем власного капіталу з нормальним розподілом показано на рисунку 3.1.

Якщо необхідно перевірити гіпотезу про узгодженість даних з іншими розподілами (лог-нормальним, експоненціальним тощо), використовується модуль *Nonparametrics / Distribution – Непараметричні статистики / Розподіли*.

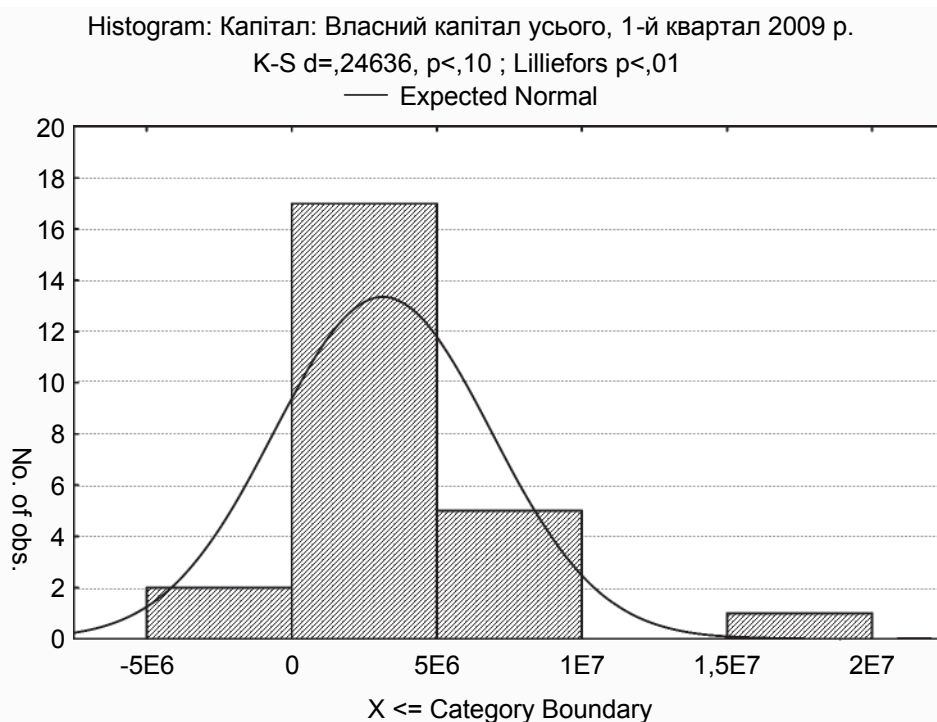


Рисунок 3.1 – Емпіричний і теоретичний розподіли комерційних банків за рівнем власного капіталу

Стандартизація. Безпосереднє використання абсолютних значень ознак при аналізі може призвести до того, що класифікацію визначатимуть ознаки, що мають найбільший розкид значень. Тому застосовуються такі види стандартизації:

1 Z-шкали (Z-Scores). Із значень змінних віднімається їх середнє значення, і ці значення діляться на стандартне відхилення.

2 Розкид від -1 до 1. Лінійним перетворенням змінних добиваються розкиду значень від -1 до 1.

3 Розкид від 0 до 1. Лінійним перетворенням змінних добиваються розкиду значень від 0 до 1.

4 Максимум 1. Значення змінних діляться на їхній максимум.

5 Середнє 1. Значення змінних діляться на їхнє середнє значення.

6 Стандартне відхилення 1. Значення змінних діляться на стандартне відхилення.

3.2 Кластерні процедури класифікації

Кластер – це група, клас однорідних одиниць сукупності. Основне завдання кластерного аналізу – формування таких груп у багатовимірному просторі. Однорідність сукупності задається правилом обчислення певної метрики, що характеризує ступінь подібності (схожості) j -ї та k -ї одиниць сукупності. Такою метрикою може бути відстань між ними s_{jk} або коефіцієнт подібності r_{jk} . Близькі (схожі) за вибраними метриками одиниці вважаються належними до одного типу, однорідними. Вибір метрики є важливим моментом кластерного аналізу, від якого залежить кінцевий варіант поділу сукупності на класи.

Приклад 3.1 Двадцять банків, акції яких котируються на ринку, надали інформацію, наведену в таблиці 3.3, де X – витрати за минулий період, Y – прибуток за минулий період.

Таблиця 3.3 – Вихідні дані (витрати й прибуток комерційних банків)

Номер банку	X	Y	Номер банку	X	Y
1	4	2	11	5	4
2	6	10	12	3	8
3	5	7	13	13	5
4	12	3	14	15	3
5	17	4	15	5	9
6	3	10	16	8	3
7	6	1	17	6	4
8	6	3	18	14	3
9	15	1	19	4	8
10	15	4	20	4	3

Необхідно з'ясувати, акції яких банків має сенс придбати (*Buy*), яких – притримати (*Hold*), а від яких – позбавитися (*Sell*). Графічне подання вихідних даних показано на рисунку 3.2 (графік побудовано за допомогою команди *Scatterplots* меню *Graphs*).

Очевидно, що існують три відокремлені групи банків. Підтвердимо цей висновок за допомогою кластерного аналізу.

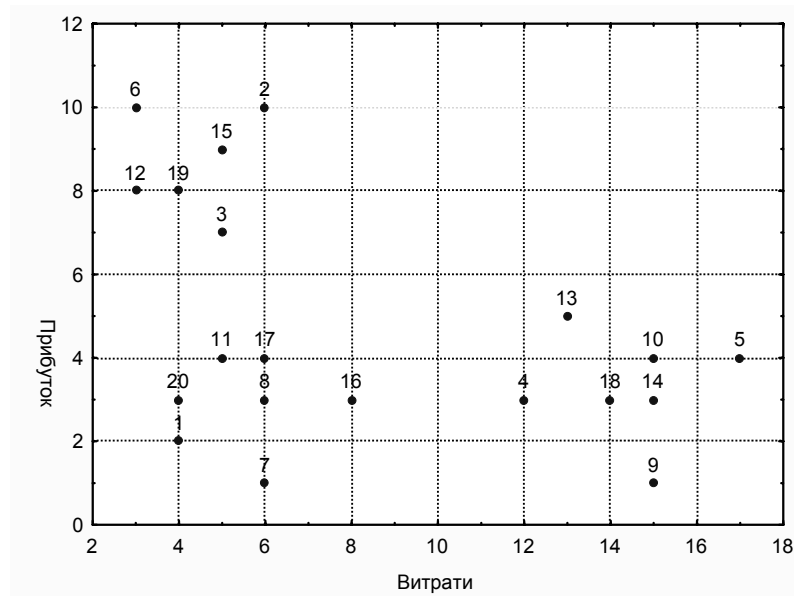


Рисунок 3.2 – Графічне подання вихідних даних

Кластерний аналіз у системі *Statistica* здійснюється за допомогою пункту меню *Статистика – Методи багатовимірного аналізу – Кластерний аналіз* (рисунок 3.3).

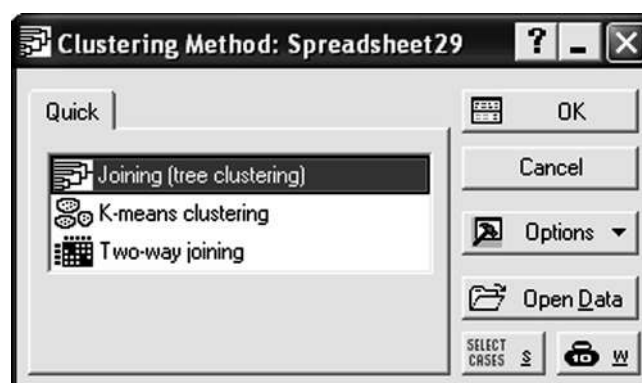


Рисунок 3.3 – Вибір типу кластерного аналізу

Побудова деревоподібної діаграми (*Joining (tree clustering)*).
У системі *Statistica* ієрархічну процедуру класифікації – *Joining (Tree*

clustering) – реалізовано в модулі *Cluster Analysis*. Діалогове вікно процедури (рисунок 3.4) пропонує вибрати:

- ознакову множину;
- тип вихідних даних: *Raw Data* – дані типу «об’єкт – ознака» або *Distance Matrix* – матриця відстаней;
- варіант класифікації: по стовпцях (*columns*) – класифікацію ознак або по рядках (*rows*) – класифікацію об’єктів;
- алгоритм об’єднання – *Amalgamation (linkage) Rules*; за замовчуванням – алгоритм одиничного зв’язку – *Single linkage (nearest neighbor)*;
- метрику відстаней – *Distance measure: Euclidean distances, City-block (Manhattan) distance*, тощо.

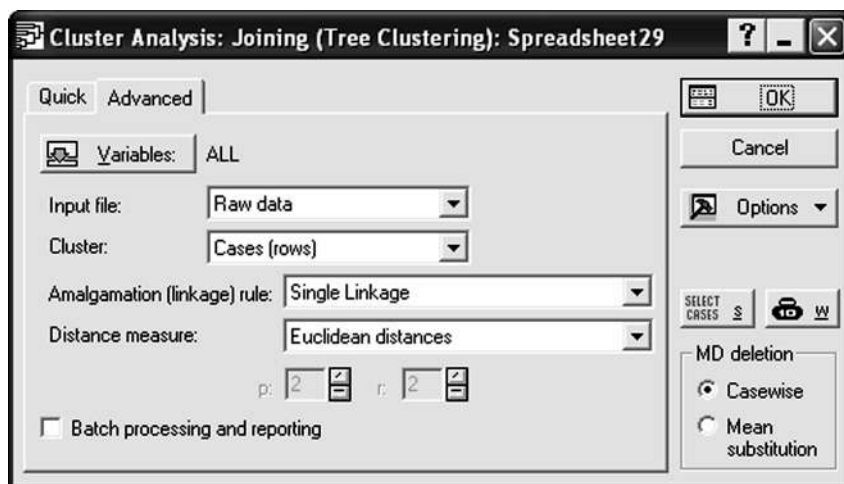


Рисунок 3.4 – Діалогове вікно процедури Joining (tree clustering)

За командою на виконання система видає *Joining Results* з опціями вибору типу дендрограми – горизонтального або вертикального (рисунки 3.5, 3.6).

На рисунку 3.6 показано результати кластерного аналізу. Відокремлено три кластери (до першого ввійшли банки 1, 20, 8, 17, 11, 7, 16; до другого – банки 2, 15, 12, 19, 3, 6; до третього – банки 4, 10, 14, 18, 5, 9, 13).

Узагалі ієрархічна кластер-процедура досить проста й прийнята для інтерпретації. Проте для численної сукупності вона виявляється громіздкою. У таких випадках перевагу віддають ітераційним процедурам.

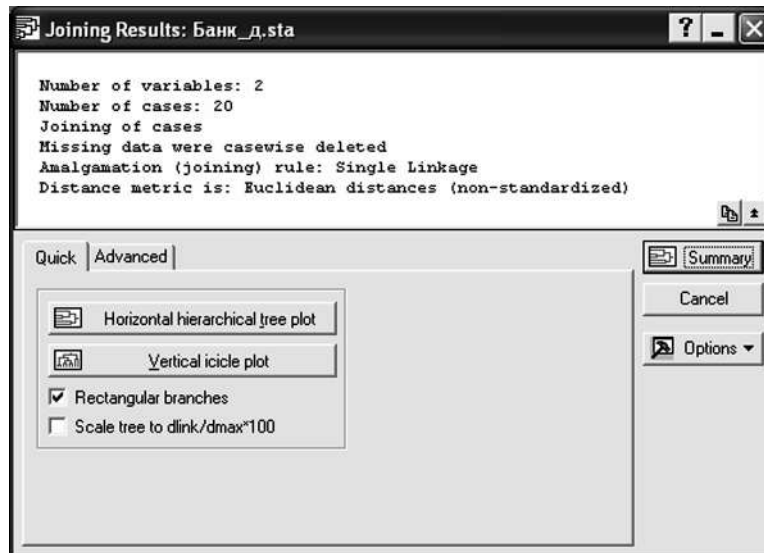


Рисунок 3.5 – Діалогове вікно Joining Results

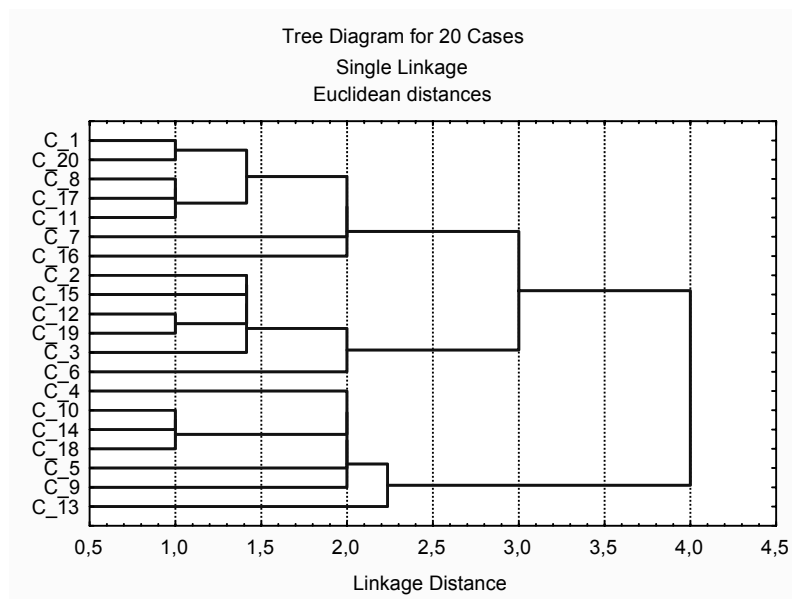


Рисунок 3.6 – Результати Joining (tree clustering)

Метод k -середніх (k -means clustering). На відміну від ієрархічної процедури, яка потребує розрахунку і збереження матриці подібності, ітераційна процедура оперує безпосередньо первинними даними; формуються кластери одного рангу, ієрархічно не підпорядковані. Основні риси ітераційних кластер-процедур розглянемо на прикладі алгоритму k -середніх (рисунок 3.7), який реалізує ідею утворення груп за принципом «найближчого центра».

На першому кроці ітераційного процесу здійснюється орієнтовний поділ сукупності на класи і визначаються центри тяжіння (багатовимірні середні) цих класів.

На другому кроці визначаються евклідові відстані одиниць сукупності до центрів тяжіння виділених кластерів і кожна з них відноситься до того кластера, центр тяжіння якого найближчий.

На третьому кроці розраховуються нові центри тяжіння кластерів.

Кроки 2 і 3 повторюються доти, доки склад кластерів не стабілізується. Ітерації за принципом *k*-середніх у явному вигляді не використовують критеріїв якості класифікації, проте неявно вони мінімізують внутрішньогрупові дисперсії, забезпечуючи тим самим однорідність сформованих кластерів.

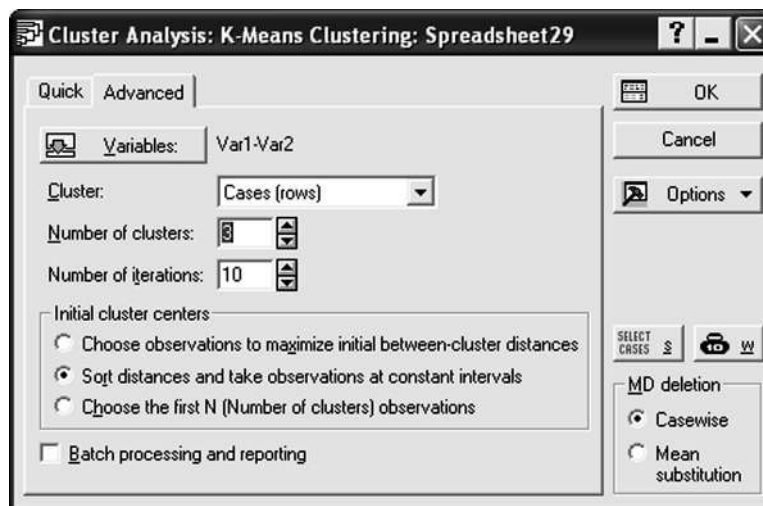


Рисунок 3.7 – Діалогове вікно процедури k-means clustering

Діалогове вікно процедури (див. рисунок 3.7) пропонує вибрати:

- ознакову множину;
- варіант класифікації: по стовпцях (*columns*) – класифікацію ознак або по рядках (*rows*) – класифікацію об'єктів;
- кількість кластерів;
- кількість ітерацій;
- метод визначення центра кластерів та ін.

За командою на виконання система видає вікно *k-means clustering Results* (рисунок 3.8).

Вкладка *Members of each cluster & distances* демонструє інформацію щодо складу кожного з кластерів і відстаней членів кластера від кластерного центра (рисунок 3.9).

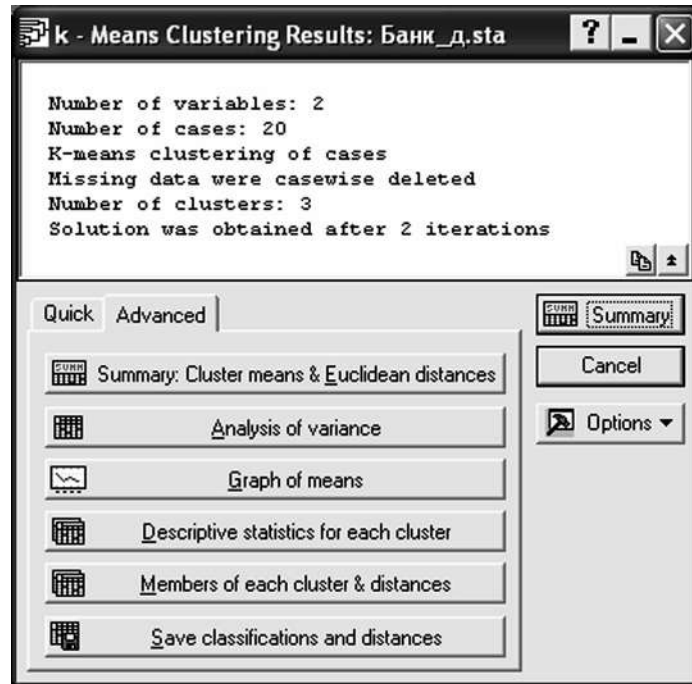


Рисунок 3.8 – Результати k-means clustering

За результатами аналізу отримано три кластери: перший (банки 4, 5, 9, 10, 13, 14, 15); другий (банки 1, 7, 8, 11, 16, 17, 20) і третій (банки 2, 3, 6, 12, 15, 19).

Members of Cluster Number 1 (Банк_д.sta) and Distances from Respective Cluster Center Cluster contains 7 cases							
Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.	Case No.
C 4	C 5	C 9	C 10	C 13	C 14	C 18	
Distance	1,729103	1,887121	1,665986	0,646813	1,577909	0,451754	0,364216

Рисунок 3.9 – Склад кластерів

За допомогою процедури класифікації отримано три групи банків: для першої характерно отримання високого прибутку при малих витратах (акції цих банків доцільно купувати (*Buy*)), для другої – середні величини прибутку та витрат (рекомендація *Hold*), для третьої – високі витрати та невеликий прибуток (рекомендація *Sell*). Рекомендації наведено в таблиці 3.4.

Таблиця 3.4 – Рекомендації

Номер банку	Рекомендація	Номер банку	Рекомендація	Номер банку	Рекомендація	Номер банку	Рекомендація
1	Hold	6	Buy	11	Hold	16	Hold
2	Buy	7	Hold	12	Buy	17	Hold
3	Buy	8	Hold	13	Sell	18	Sell
4	Sell	9	Sell	14	Sell	19	Buy
5	Sell	10	Sell	15	Sell	20	Hold

Слід пам'ятати, що у кластерному аналізі розбиття на кластери істотно залежить від абсолютних значень вихідних даних. Цю проблему вирішують за допомогою нормування (стандартизації).

3.3 Дискримінантний аналіз

З-поміж методів розпізнавання образів особливе місце посідає дискримінантний аналіз. На відміну від кластерного аналізу дискримінантний не утворює нових класів, а допомагає виявити різницю між існуючими класами і віднести новий (нерозпізнаний) об'єкт до одного з них за принципом максимальної схожості. Наприклад, банк, спираючись на певну систему характеристик фінансового стану клієнтів, які звертаються за позиками, поділяє їх на дві категорії: надійні та ненадійні. Дискримінантний аналіз використовується в медичній діагностиці, при визначенні ризику відмови приладів у технічних системах тощо. Основна проблема — звести помилку класифікації до мінімуму.

Дискримінантна функція — це лінійна комбінація певної множини ознак, які називаються класифікаційними і на основі яких ідентифікуються класи. Особливість дискримінантної функції полягає в тому, що класи представляються шкалою найменувань, а класифікаційні ознаки x_i , де $i = 1, 2, \dots, m$, вимірюються метричною шкалою. Кількість останніх не може перевищувати $(n - 2)$, де n — обсяг сукупності. Функціонально зв'язані й висококорельовані ознаки до ознакового простору моделі не включаються.

Дискримінантна функція f_j визначається для кожного j -го класу ($j = 1, 2, \dots, p$):

$$f_j = a_{0j} + a_{1j}\bar{x}_{1j} + a_{2j}\bar{x}_{2j} + \dots + a_{mj}\bar{x}_{mj},$$

де a_{ij} — коефіцієнт функції (змістовної інтерпретації не має);

\bar{x}_{ij} — середнє значення i -ї ознаки в j -му класі.

Для оцінювання спроможності дискримінантної функції розпізнавати класи у багатовимірному ознаковому просторі використовують також λ -статистику Вілкса (*Wilks lambda*):

$$\lambda = \prod_{j=1}^p \frac{1}{1 + \lambda_j},$$

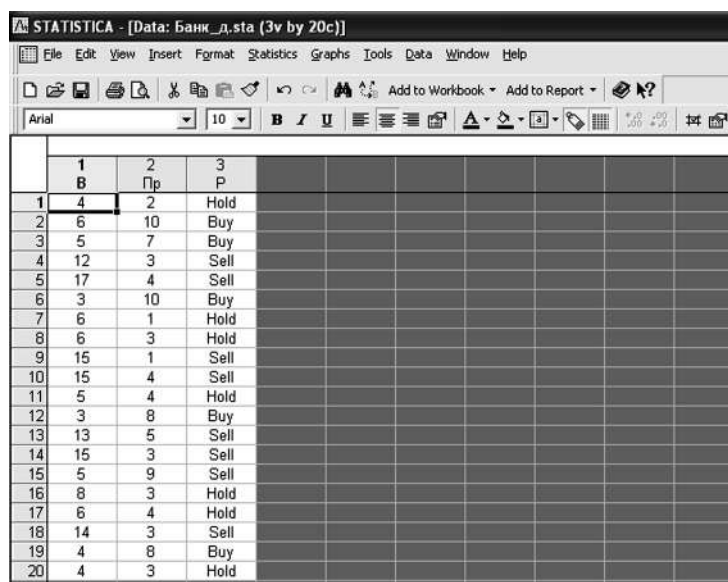
де λ_j – властиві значення матриці коваріацій.

λ -статистика враховує як відмінності між класами, так і однорідність кожного класу. Оскільки λ розраховується як обернена величина, то чим більше різняться центроїди, тим менше її значення, і навпаки, якщо центроїди збігаються, то λ прямує до 1. Отже, близькі до 0 значення λ свідчать про високу розпізнавальну спроможність дискримінантної функції. Істотність різниці значень центроїдів перевіряється також за допомогою критерію χ^2 або дисперсійного F -критерію, які функціонально зв'язані з λ -статистикою.

У системі *Statistica* процедури дискримінантного аналізу об'єднані в модулі *Discriminant Analysis – Дискримінантний аналіз*. Порядок використання модуля розглянемо на прикладі.

Приклад 3.2 За даними прикладу 3.1 сформулювати процедуру класифікації акції комерційного банку (віднесення її до класу *Sell*, *Buy* або *Hold*).

Вихідні дані для аналізу наведено на рисунку 3.10.



	1 B	2 Пп	3 P
1	4	2	Hold
2	6	10	Buy
3	5	7	Buy
4	12	3	Sell
5	17	4	Sell
6	3	10	Buy
7	6	1	Hold
8	6	3	Hold
9	15	1	Sell
10	15	4	Sell
11	5	4	Hold
12	3	8	Buy
13	13	5	Sell
14	15	3	Sell
15	5	9	Sell
16	8	3	Hold
17	6	4	Hold
18	14	3	Sell
19	4	8	Buy
20	4	3	Hold

Рисунок 3.10 – Дані для проведення дискримінантного аналізу

Для запуску модуля *Дискримінантний аналіз* обираємо пункт меню *Statistics – Multivariate Exploratory Techniques – Discriminant*

Analysis.

Діалогове вікно процедури (рисунок 3.11) пропонує вибрати ознаки (*Variables*) для аналізу. В цьому випадку незалежними ознаками є витрати (*B*) і прибуток (*Пр*). Ознакою, за якою здійснюється групування, є рекомендація (*P*).

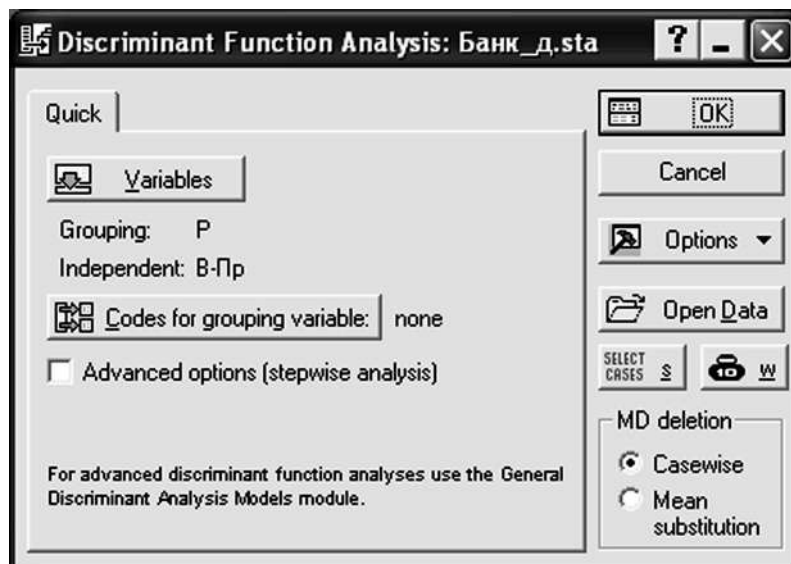


Рисунок 3.11 – Діалогове вікно процедури

За командою на виконання система видає вікно *Discriminant Function Analysis Results* (рисунок 3.12).

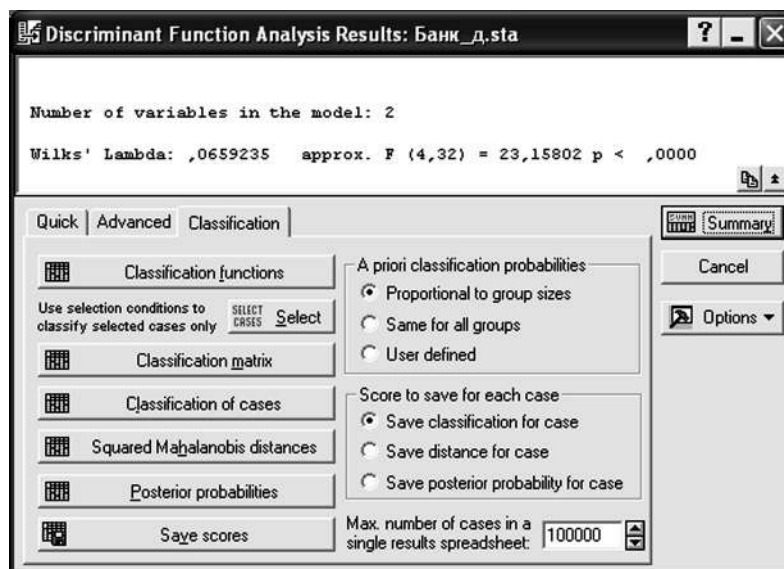


Рисунок 3.12 – Результати дискримінантного аналізу – Discriminant Function Analysis Results

Інформаційна панель містить дані про кількість змінних – 2, про значення лямбди Уїлкса, значення F -статистики, що пов'язане з лямбдою Уїлкса. Значення лямбди Уїлкса 0,066 свідчить про достатньо високу точність аналізу.

Вкладка *Classification functions* дає можливість отримати функції класифікації для кожного класу (рисунок 3.13).

Отже, отримано такі функції класифікації:

$$P_{hold} = -10,508 + 1,993B + 2,734Pr;$$

$$P_{buy} = -30,977 + 2,927B + 5,452Pr;$$

$$P_{sell} = -38,295 + 4,123B + 5,031Pr.$$

Для класифікації нового об'єкта (акції) в формули підставляються значення ознак (витрати, прибуток). Новий об'єкт буде віднесено до того класу, класифікаційне значення для якого буде найбільшим.

Variable	Hold	Buy	Sell
	p=,35000	p=,25000	p=,40000
B	1,993	2,927	4,123
Pr	2,734	5,452	5,031
Constant	-10,508	-30,977	-38,295

Рисунок 3.13 – Функції класифікації – Classification functions

Припустимо, що необхідно класифікувати акцію банку з такими характеристиками: витрати дорівнюють 6 д.о., прибуток – 8 д.о.

Підставляємо ці значення в класифікаційні функції і отримуємо

$$P_{hold} = 23,322,$$

$$P_{buy} = 30,20,$$

$$P_{sell} = 26,691.$$

Отже, акцію треба віднести до групи *Buy*.

Вкладка *Апостеріорні ймовірності* (Posterior Probabilities) (рисунок 3.14) демонструє ймовірності, за якими той чи інший об'єкт належить до певного класу, а також показує неправильно класифіковані об'єкти (позначено знаком «*»).

Posterior Probabilities (Банк д. ста)				
Incorrect classifications are marked with *				
Case	Observed Classif.	Hold p=.35000	Buy p=.25000	Sell p=.40000
1	Hold	1,000	0,000	0,000
2	Buy	0,000	0,987	0,013
3	Buy	0,038	0,949	0,013
4	Sell	0,009	0,003	0,988
5	Sell	0,000	0,000	1,000
6	Buy	0,000	1,000	0,000
7	Hold	1,000	0,000	0,000
8	Hold	0,998	0,001	0,000
9	Sell	0,002	0,000	0,998
10	Sell	0,000	0,000	1,000
11	Hold	0,992	0,007	0,000
12	Buy	0,017	0,983	0,001
13	Sell	0,000	0,002	0,998
14	Sell	0,000	0,000	1,000
* 15	Sell	0,000	0,994	0,006
16	Hold	0,972	0,008	0,021
17	Hold	0,979	0,018	0,003
18	Sell	0,000	0,000	1,000
19	Buy	0,007	0,991	0,003
20	Hold	1,000	0,000	0,000

Рисунок 3.11 – Апостеріорні ймовірності (Posterior Probabilities)

Дані таблиці 3.5 (вкладка *Classification matrix*) показують точність класифікації. Загальна точність моделі – 95 % (помилка – в 5 % випадків). При цьому всі акції типу *Hold* і *Buy* були класифіковані правильно (точність 100 %). При класифікації акцій типу *Sell* допущено одну помилку (одну акцію було помилково віднесено до типу *Buy*, точність – 87,5 %).

Таблиця 3.5 – Матриця класифікації

Матриця класифікації	Percent	Hold	Buy	Sell
Hold	100,0	7	0	0
Buy	100,0	0	5	0
Sell	87,5	0	1	7
Total	95,0	7	6	7

Слід зазначити, що дискримінантний аналіз проводиться за таких основних припущень:

- множину об'єктів M для аналізу розбито на дві або більше підмножини;
- у кожній підмножині знаходяться, принаймні, два об'єкти (причому всі об'єкти множини M повинні належати до однієї з підмножин);
- кількість об'єктів N перевищує кількість дискримінантних змінних p на дві одиниці ($0 < p < N - 2$);

- лінійна незалежність ознак, тобто жодна з ознак не є лінійною комбінацією інших ознак (інакше, вона не несе нової інформації);
- закон розподілу дискримінантних змінних (ознак об'єкта) нормальний.

ЗАПИТАННЯ ДЛЯ САМОПЕРЕВІРКИ

- 1 Назвіть специфічні ознаки науки як системи наукових знань.
- 2 Сформулюйте й охарактеризуйте основні етапи наукового дослідження.
- 3 У чому суть кластерного аналізу?
- 4 Які обмеження має дискримінантний аналіз?
- 5 Що є результатом дискримінантного аналізу?
- 6 Що є метою методів класифікації?
- 7 Для аналізу фінансового стану банків самостійно визначте сукупність банків (не менше 25) і по кожному з них випишіть інформацію щодо розміру активів (k_1), капіталу (k_2), зобов'язань (k_3) і прибутку (k_4) (за даними офіційного сайту Національного банку України www.bank.gov.ua). По кожному банку визначте: прибутковість активів (k_5) і капіталу (k_6), співвідношення капіталу і зобов'язань (k_7). У системі *Statistica* сформулюйте файл первинних даних, здійсніть специфікацію показників.
- 8 Використовуючи процедури модуля *Basic Statistics and Tables*, здійсніть розвідувальний аналіз даних завдання 7: а) для показників k_5 і k_6 визначте характеристики розподілу: *Mean* – середню величину, *Median* – медіану, *Standard Deviation* – стандартне відхилення, *Skewness* – коефіцієнт асиметрії та *Kurtosis* – коефіцієнт ексцесу; б) складіть ряд розподілу банків за рівнем прибутковості активів (*Frequency tables*), здійсніть частотний аналіз розподілу; в) для візуалізації розподілу банків за всіма показниками скористайтеся графіками.
- 9 Використовуючи процедуру *Correlation matrices*, оцініть взаємозв'язки між показниками прибутковості (k_5 , k_6) та коефіцієнтом співвідношення капіталу і зобов'язань (k_7), подайте їх у матричному вигляді й графічно. Зробіть висновки.
- 10 Проведіть кластерний аналіз і здійсніть класифікацію комерційних банків:
 - за допомогою деревоподібної діаграми;
 - за допомогою методу *k*-середніх (*k-means*); кількість кластерів – 3 – 4.
 Проаналізуйте отримані результати.

БІБЛІОГРАФІЧНИЙ СПИСОК

Андреев Г.И. В помощь написания диссертации и рефератов: основы научной работы и оформление результатов научной деятельности: учеб. пособие / Г.И. Андреев, С.А. Смирнов, В.А. Тихомиров. – М.: Финансы и статистика, 2004. – 272 с.

Боровиков В.П. Популярное введение в программу STATISTICA / В.П. Боровиков. – М.: КомпьютерПресс, 1998. – 267 с.

Боровиков В. П. Statistica® – статистический анализ и обработка данных в среде Windows® / В. П. Боровиков, И. П. Боровиков. – М.: Информ.-изд. дом «Филинь», 1998. – 608 с.

Єріна А. М. Статистичне моделювання та прогнозування: навч. посіб. / А. М. Єріна. – К.: КНЕУ, 2001. – 170 с.

Замков О.О. Математические методы в экономике: учебник / О.О. Замков, А.В. Толстопятенко, Ю.Н. Черемных. – 2-е изд. – М.: Изд-во «Дело и Сервис», 1999. – 368 с.

Ковалев А. И. Прологомены к методам научных исследований: учеб. пособие / А. И. Ковалев. – Х.: Инжэк, 2005. – 312 с.

Основы научной речи: учеб. пособие для студ. нефилол. высш. учеб. заведений / Н.А. Буре, М.В. Быстрых, С.А. Вишнякова и др.; под ред. В.В. Химика, Л.Б. Волковой. – СПб.: Филологический факультет СПбГУ; М.: Изд. центр «Академия», 2003. – 272 с.

Основы научных исследований: учеб. пособие / под ред. А.А. Лудченко. – 2-е изд. – К.: Знання, 2001. – 113 с.

Плюта В. Сравнительный многомерный анализ в эконометрическом моделировании: пер. с польск. / В. Плюта. – М.: Финансы и статистика, 1989. – 175 с.

Факторный, дискриминантный и кластерный анализ: пер. с англ. / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др.; под ред. И. С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.

Цехмістрова Г.С. Основи наукових досліджень: навч. посіб. / Г.С. Цехмістрова. – К.: Вид. дім «Слово», 2003. – 240 с.

Як підготувати і захистити дисертацію на здобуття наукового ступеня: метод. поради / упоряд. Л.А. Пономаренко. – К.: Толока, 2001. – 80 с.

Гусєва Юлія Юрїївна
Гребенїкова Олена Володимирївна

СТАТИСТИЧНІ МЕТОДИ АНАЛІЗУ ФІНАНСОВО-ЕКОНОМІЧНИХ
ПРОЦЕСІВ

Редактор А.М. Ємленїнова

Зв. план, 2010

Підписано до друку 16.06.2010

Формат 60x84 1/16. Папір офс. № 2. Офс. друк

Ум. друк. арк. 2. Обл.-вид. арк. 2,25. Наклад 150 прим. Замовлення 192.

Цїна вільна

Національний аерокосмічний університет ім. М.Є. Жуковського

«Харківський авіаційний інститут»

61070, Харків-70, вул. Чкалова, 17

<http://www.khai.edu>

Видавничий центр «ХАІ»

61070, Харків-70, вул. Чкалова, 17

izdat@khai.edu