

## РОЗДІЛ 10

# МЕТОДИ АНАЛІЗУ РОЗМІРІВ ТА ФОРМИ ТІЛА

Розглядаючи методологію екологічних досліджень, необхідно зупинитися також на такому важливому популяційному показнику як загальні розміри та форма організму. Саме розміри і форма особин даної популяції виявляють собою миттєвий продукт природного відбору, інтенсивність і спрямованість якого безпосередньо визначаються умовами існування популяції. Загальна морфологія являє собою, з одного боку, результат генетичної передісторії виду (тобто, є фенотиповим відображенням історично сформованого на даній території популяційного генофонду), а з другого боку — формується в характерних для даного часового моменту умовах середовища, що як скульптор “відтинає” усі нежиттєздатні генні сполучення. Як не існує в популяціях організмів зі статевим розмноженням двох особин із однаковим генотипом, так і не існує двох тварин чи рослин зовсім ідентичних у відношенні загальних розмірів і форми тіла. Причина цього полягає в тому, що успадковуються не якісь стандартні (середньопопуляційні) розміри тих або інших органів (або організму в цілому), а їхній рівень мінливості (тобто, норма реакції), при цьому, генний набір визначає не стільки загальні розміри тварини або рослини, а більшою мірою — *основні пропорції тіла*, тобто його форму.

У багатьох випадках саме форма тіла є основним результатом екологічного преферендуму виду. На конвергентній подібності форми тіла базується виділення життєвих форм (екотипів), що характеризують екологічну близькість видів (навіть з віддалених таксономічних груп) у відповідності з подібністю умов їхнього місцеперебування. На закономірній зміні загальних розмірів і форми організмів (або їхніх окремих частин) залежно від градієнта кліматичних умов ґрунтуються відомі екологічні закономірності — правила Бергмана й Аллена.

На відміну від класичної морфології, що більшою мірою служить інтересам систематики і класифікації, *екологічна морфологія* являє собою інструмент для вивчення *селектогенезу*, тобто вивчає закономірності зміни загальних розмірів і форми організмів залежно від їхніх умов існування. Методи і підходи екоморфології частково вже розглядалися в розділі 4 (щодо якісних ознак) та в розділі 7 (щодо ростових процесів). Тут же основну увагу буде приділено багатовимірним методам аналізу морфометричних ознак.

Оскільки кількісні (метричні) ознаки успадковуються, з одного боку, полігенно (на відміну від якісних, котрі успадковуються моно- або



олігогенно), а з другого боку — плейотропно, то для їх аналізу необхідні й особливі методики, які призначені для дослідження їх *спільної і сполученої мінливості*.

Екоморфологічні дослідження, також як і аналіз якісних ознак, можуть проводитися як у часі, так і в просторі. У першому випадку ми маємо справу із популяційним генофондом, зміни якого із покоління в покоління відбуваються насамперед за рахунок нестабільності мікрокліматичних (погодних, сезонних, циркадних, катастрофічних тощо) умов середовища його носіїв (тобто, особин популяції), що може і повинне відбиватися на загальній морфології останніх. Часовий аспект екоморфологічного напрямку є фундаментом для *мікроеволюційних* змін, особливо в тих випадках, коли мається постійно діючий (нехай навіть слабкий) вектор природного відбору.

В другому випадку аналіз загальних розмірів і форми проводиться для групи популяцій одного виду, що ізольовані як просторово, так і екологічно. Основною задачею таких досліджень є виявлення факторів і чинників внутрішньовидового морфометричного розмаїття, тобто пошук сполученої мінливості розмірів та форми об'єкту на фоні прояви генотипової і паратипової складових, і аналіз механізмів таких змін у термінах теорії оптимальності.

Теорія оптимальності в застосуванні до біологічних об'єктів затверджує, що будь-які органи й організми при формуванні і для стабільного їхнього існування в часі повинні відбивати **принцип максимальної простоти й оптимальності конструкції** (щодо необхідних витрат речовини й енергії для їхнього формування), а також **принцип адекватної конструкції організму** умовам середовища, що змінюються. При цьому останній допускає, мабуть, не єдине рішення проблеми форми і функції — часто зустрічаються організми й органи, що виконують в основному однаковій функції, проте різні за своєю формою. Бажано (хоча і не завжди він виявляється) виконання **принципу поліфункціональності**, коли той самий ознак (або орган) може виконувати різні біологічні функції (наприклад, одні і теж протоки у ссавців виводять і сечу, і статеві продукти).

Генетичні механізми, що лежать в основі такого морфометричного розмаїття найчастіше базуються не на глобальній перебудові онтогенетичних процесів, а діють більш локально. У таких випадках частіше відбувається зміна швидкості ростового процесу того або іншого органа, тієї або іншої системи, а також змінюються терміни вмикання/вимикання цього процесу (див. розділ 7).

Головною проблемою екоморфології є проблема співвідношення розміру і форми тіла. Більш великі особини одного виду з однієї популяції являють собою не просто збільшені копії своїх більш дрібних побратимів.



При збільшенні розміру обсяг (біомаса) організму збільшується пропорційно кубу, а площа його поверхні — квадрату лінійних розмірів. Отже виникають фізіологічні обмеження, що навіть при незначному підвищенні розмірів тіла повинні враховувати зміни таких життєво важливих параметрів, як потреба організму в їжі, водо- та газообмін, форма та міцність кістякових елементів. Ці обмеження і приводять до формування нелінійних зв'язків між розмірами і екофізіологічними потребам організму, що найчастіше адекватно описуються аллометричними рівняннями (див. розділ 7). Аллометричні рівняння є окремим випадком багатовимірних методик, коли аналізується зв'язок між будь-якою парою ознак. Однак набагато цікавіше простежувати зв'язок між цілим набором ознак, що характеризують загальну морфологію організму з різних позицій (використовуються розмірні показники на різних стадіях онтогенезу, вагові показники, показники площі поверхні тощо). Крім того, використання багатовимірних методів дозволяє розрізняти зміни загальної морфології особин, що лежать в основі таксономічної роз'єднаності популяцій (видів) від внутрішньовидової мінливості, що формується залежно від умов існування.

Класична морфологія оперує такими поняттями, як середні розміри (а також розмах їхньої мінливості) і форма (або пропорції) тіла, і на основі рівня їхньої відмінності проводить таксономічний поділ групи об'єктів (наприклад видів одного роду), при цьому зовсім не враховується наявність зв'язків між розмірами і формою організму і, особливо, нелінійність цих зв'язків. Екоморфологія (як і екофізіологія), розглядаючи весь узагальнений морфотип дозволяє кількісно охарактеризувати аспекти різних стадій (під)видоутворення. Екоморфологія дає відповідь на питання: чи має морфологічне розмаїття серед досліджуваного набору родинних популяцій більшою мірою екологічну складову, або ж характер мінливості в різних популяціях відрізняється значно і, отже, останні являють собою таксономічно відособлені угруповання визначеного статусу (підвиди, напіввиди, надвиди, види тощо).

Аналіз взаємозв'язків між формою і розмірами організму проводиться з використанням методу онтогенетичних траєкторій на підставі результатів, отриманих методом головних компонентів. Визначення ступеня і спрямованості зв'язків між комплексом ознак загальної морфології і факторами середовища здійснюється, з використанням канонічних кореляцій. Нарешті, класифікація та ординація об'єктів (популяцій) у багатовимірному просторі їхніх ознак здійснюється на підставі або параметричного дискримінантного аналізу, або непараметричного методу багатовимірного шкалювання. У даному розділі розглядаються основні алгоритми цих процедур, області застосування, задачі та приклади їхнього використання.



## 10.1. Вибірки й ознаки

Для проведення екоморфологічного аналізу використовуються тільки статевозрілі організми для того, щоб нівелювати вікову мінливість. Багато організмів мають детермінований ріст і чіткі критерії настання статевої зрілості, яке викликає припинення збільшення розмірів. Аналіз гетерогенного у віковому відношенні матеріалу проводиться частіше навмисно — при порівнянні особливостей ростових процесів (див. розділ 7).

**Вибірки** являють собою миттєвий зріз через популяцію — серію особин з розмірами, що варіюють. Вони являють собою *середні зразки росту в межах набору аналізованих популяцій*. Мінімальний обсяг вибірки при використанні багатовимірних підходів складає 30 - 40 особин.

Збір матеріалу повинен вироблятися із серії популяцій одного виду, що існують у різних районах досліджуваної території, бажано, у контрастних умовах середовища.

Вибір ознак для оцінки загальної морфології повинен враховувати кілька умов:

— ознаки повинні як можна більш різнобічно характеризувати об'єкт дослідження;

— проміри кожної ознаки повинні мати чіткі границі, однозначно обумовлені на кожній особині;

— проміри повинні вироблятися з визначеним ступенем точності.

---

Примітка 10.1. Іноді використовуються рангові критерії для характеристики метричних ознак. Використання їх, поряд з безпосередньо вимірюваними коректно, якщо характер розподілу цих рангових характеристик не істотно відрізняється від нормального. В протилежному випадку для аналізу рангових ознак використовуються особливі методики (у тому числі й багатовимірні), наприклад медіана, тест Краскела-Уолліса тощо. Опис багатьох напараметричних методик дано в книзі Е.В.Гублера та А.А.Генкіна (1973).

---

у багатовимірних аналізах необхідно використовувати *тільки результати безпосередніх вимірів*; використання яких-небудь похідних (наприклад індексів) повинне бути виключено.

У результаті вимірів формується первинна матриця, розміром  $i \times j$ , де  $i$  — число використовуваних об'єктів;  $j$  — число аналізованих ознак для кожного об'єкта. Кількість останніх, як правило, повинне бути досить велике (наприклад, 20-30).



Первинна матриця вихідних значень характеризується наступними показниками:

- вектор-строкою середніх значень для досліджуваної сукупності за кожною аналізованою ознакою;
- вектором-рядком дисперсій цих ознак;
- матрицею кореляцій-ковариаций розміром  $j \times j$  (тобто, між кожною парою ознак).

---

Примітка 10.2. Основи матричної алгебри, у тому числі і принцип розрахунку основних чисел матриця в досить доступній формі викладено в книзі С.Гроссмана і Дж.Тернера [6].

---

Перед початком аналізу необхідно трансформувати вихідні дані для того, щоб уникнути впливу шкали (наприклад, деякі ознаки можуть бути обмірювані в десятках міліметрів, а інші — у частках міліметрів), а також нівелювати розходження при використанні різних одиниць (наприклад, міліметрів, градусів, рангів тощо). Найбільш прийнятним методом трансформації вихідних ознак є їхня стандартизація:

$$z_{ij} = \frac{x_{ij} - M_j}{s_j}, \quad (10.1)$$

де  $z_{ij}$  — стандартизоване значення (т.зв.  $z$ -мітки, або  $z$ -scores);  $x_{ij}$  — вихідне значення для  $i$ -тої особи по  $j$ -тій ознаці;  $M_j$  та  $s_j$  — середнє арифметичне значення та стандартне відхилення для  $j$ -тої ознаки для всіх аналізованих особин матриці.

Після стандартизації середнє арифметичне для кожної ознаки дорівнює 0, а значення його стандартного відхилення (і, отже, варіанси) — 1. У сучасних пакетах статистичних програм процедура стандартизації вихідної матриці даних входить до складу опцій. Характерною рисою стандартизованої матриці є те, що її сумарна дисперсія дорівнює  $j$ , тобто числу використовуваних ознак.

Екоморфологічний аналіз має на увазі також наявності ознак, що характеризують середовище існування кожної досліджуваної популяції. Це можуть різні показники — едафічні, орографічні, метеорологічні, фізико-хімічні, фітоценотичні тощо. Розмаїття спектра таких параметрів дозволяє провести більш заглиблений екоморфологічний і екофізіологічний аналіз. Звичайно, не завжди характер розподілу таких паратипових (тобто, середовищних) параметрів буде відповідати нормальному (обов'язкова умова при використанні багатовимірних методів аналізу), але необхідно прагнути до цього, наприклад, використовуючи різні трансформації, що дозволяють уникнути ненормальності. Для ознак, що мають тільки



позитивні значення й лівосторонню асиметрію, найбільш частіше використовують логарифмування набору вихідних даних чи взяття квадратного кореня від них.

Мається два типи аналізу: у першому випадку *аналізується сполучена мінливість для окремих особин* (з однієї або декількох популяцій) — **R-аналіз**; у другому випадку *аналізується сполучена мінливість матриці, рядки якої являють собою середні арифметичні значення по кожній ознаці для групи досліджуваних вибірок* — **Q-аналіз**.

## 10.2. Аналіз головних компонентів (АГК)

Аналіз головних компонентів є окремим випадком факторного аналізу. Зміст даної процедури полягає в наступному. Якщо уявити собі багатовимірний простір (його вимірність визначається числом використовуваних ознак), то значення за кожною ознакою для кожної особини (або популяції при Q-аналізі) являють собою координати цих об'єктів у даному просторі. Формується *еліпсоїд розсіювання* (аналогічно двовірному випадку, коли формується еліпс розсіювання значень уздовж кожної з двох осей). Процедура АГК полягає у визначенні такого напрямку, уздовж якого еліпсоїд розсіювання має максимальну довжину — тобто описує максимальну частку загальної мінливості вихідної матриці. Цей напрямок відповідає першій головній компоненті (personal component 1, PC1). В алгебраїчному вираженні рівняння цієї осі є лінійною комбінацією вихідних ознак:

$$PC1 = a_1x_1 + a_2x_2 + \dots + a_jx_j, \quad (10.2)$$

де  $x_1, \dots, x_j$  — значення вихідних ознак,  $a_1, \dots, a_j$  — якісь коефіцієнти (власні вектори, eigenvector).

Аналогічно знаходяться коефіцієнти для другої головної компоненти (PC2). Однак при цьому виконується дві необхідні умови:

— друга головна компонента повинна описувати *максимальну частку залишкової мінливості*;

— крім цього, вона повинна бути *ортогональною* (тобто перпендикулярною) стосовно першої головної компоненти.

У такий спосіб для еліпсоїда розсіювання вихідних ознак визначається така вісь, що була б перпендикулярною до першої, а її напрямок також відповідав би такому радіусу еліпсоїда, що має наступне за першою головною компонентою за величиною значення.

На рис.10.1 наведено двовимірний випадок для більш наглядного представлення процедури визначення головних компонентів.

Аналогічно знаходиться третя головна компонента і всі наступні. Кожна з них описує максимальну частку *залишкової дисперсії* й



ортогональна до попередніх. У такий спосіб уся дисперсія вихідної матриці (її значення дорівнює  $j$ ; див. вище) зберігає своє значення, однак вона перерозподіляється між головними компонентами. Перша головна компонента тоді буде описувати більшу частку мінливості вихідної матриці (як правило, понад 50-70%). Частка дисперсії матриці, що описується кожною головною компонентою називається її власним числом (eigenvalue). Також, приділяється увага аналізу мінливості по першим двом-трьом головним компонентам, що описують більшу частку вихідної інформації. Інші компоненти є малоінформативними.

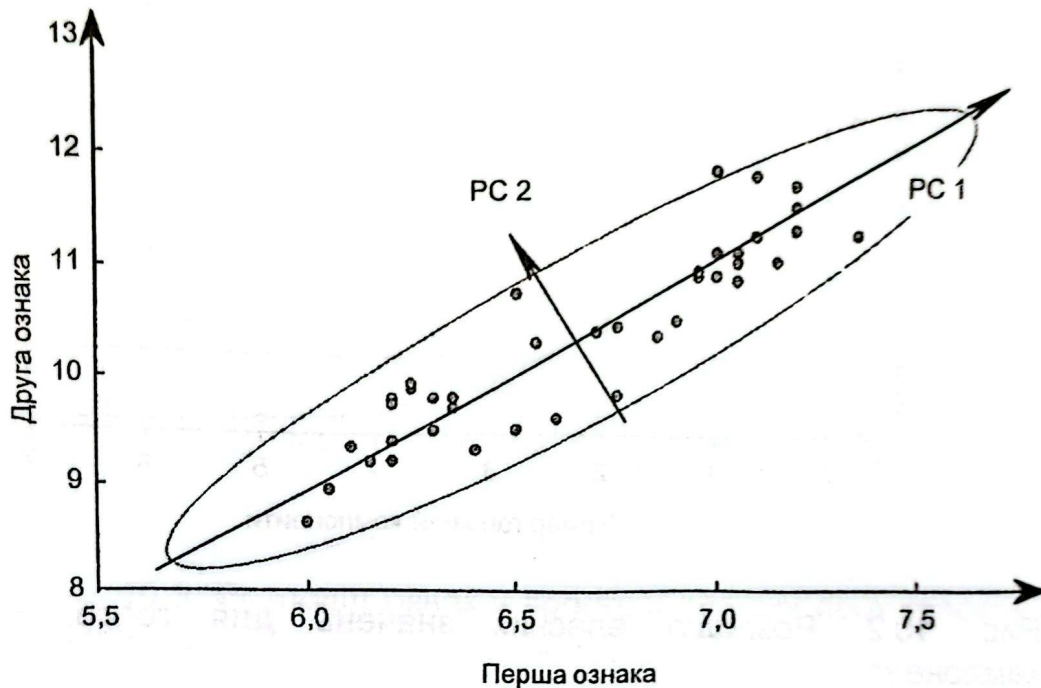


Рис. 10.1. Методика визначення головних компонент у випадку використання двох ознак

Використання АГК дозволяє значно знизити мірність вихідного набору даних (аналіз переходить від  $j$ -вимірного випадку до двох- або тривимірного), при цьому втрачається лише незначна частка вихідної інформації. Таке зниження розмірності дозволяє також візуалізувати вихідну мінливість, тобто, на двовимірному графіку розглянути ступінь близькості між різними об'єктами. При цьому об'єкти (особи чи популяції), що слабо розрізняються між собою за загальними розмірами або формою тіла, в просторі перших головних компонентів виявляються поруч, а ті, що сильно відрізняються — виявляються віддаленими. При цьому можна також оцінити, за якою ознакою вони найбільш розрізняються — уздовж якої з головних компонентів відстань між ними максимальна.



Вибір значущих головних компонентів виробляється на підставі частки описуваної ними долі вихідної мінливості. Для включення в число аналізованих головна компонента повинна описувати не менше  $1/j$  сумарної мінливості.

Як правило, розподіл за власними значеннями для головних компонентів (від першої до останньої) являє собою характерну ламану, що нагадує осип на схилі (рис.10.2). В аналіз включаються лише ті головні компоненти, чиї власні значення перевищують одиницю.

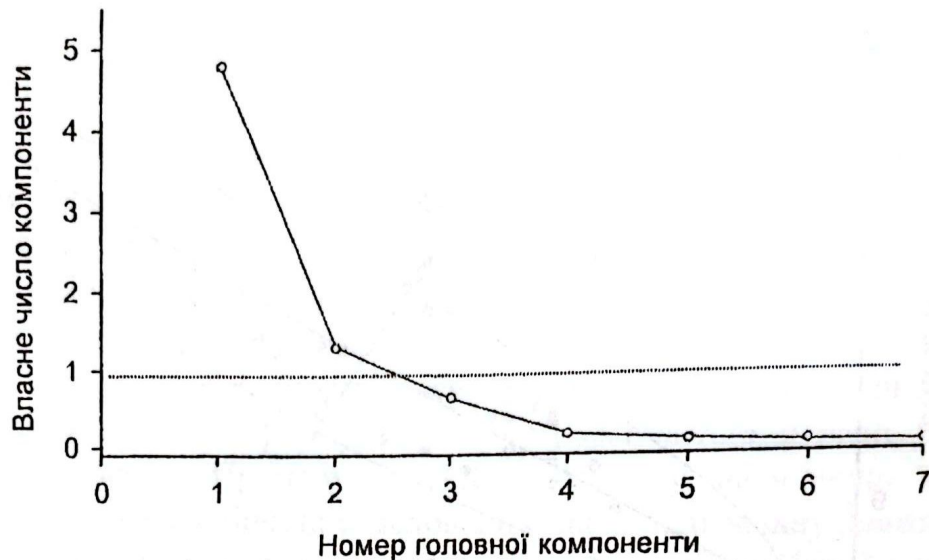


Рис. 10.2. Розподіл власних значень для головних компонент

Основною проблемою є інтерпретація головних компонентів, тобто додання їм біологічного змісту. Інтерпретація головних компонентів виробляється на підставі ступеня корельованості їх з набором вихідних ознак. Вихідні ознаки, що виявляються найбільш корельованими з розглянутою головною компонентою, вносять максимальний вклад у її інтерпретацію (з урахуванням знака коефіцієнта кореляції).

На скаттері розподілу вихідних ознак у просторі перших двох головних компонентів (координати ознак відповідають значенням коефіцієнтів кореляції кожного вихідних з ознак з відповідним головним компонентом — факторні навантаження, factor loading) добре видно, що з першою головною компонентою (PC1) сильно і позитивно корельовані ознаки №№ 1, 3-6, а з другою (PC2) — сильно, але негативно ознака № 2 (рис.10.3).

Біологічний зміст АГК полягає в тому, що набір вихідних ознак має цілий ряд таких, котрі сильно корельовані один з одним (і, отже, більшість з них малоінформативні), але з другого боку — у набір ознак могли і не



попасти такі, котрі значно впливають на розглянуті закономірності варіювання розмірів та форми тіла. Перші головні компоненти і вважаються такими ознаками, що не потрапили в число обмірюваних, але впливають на характер мінливості досліджуваної сукупності даних.

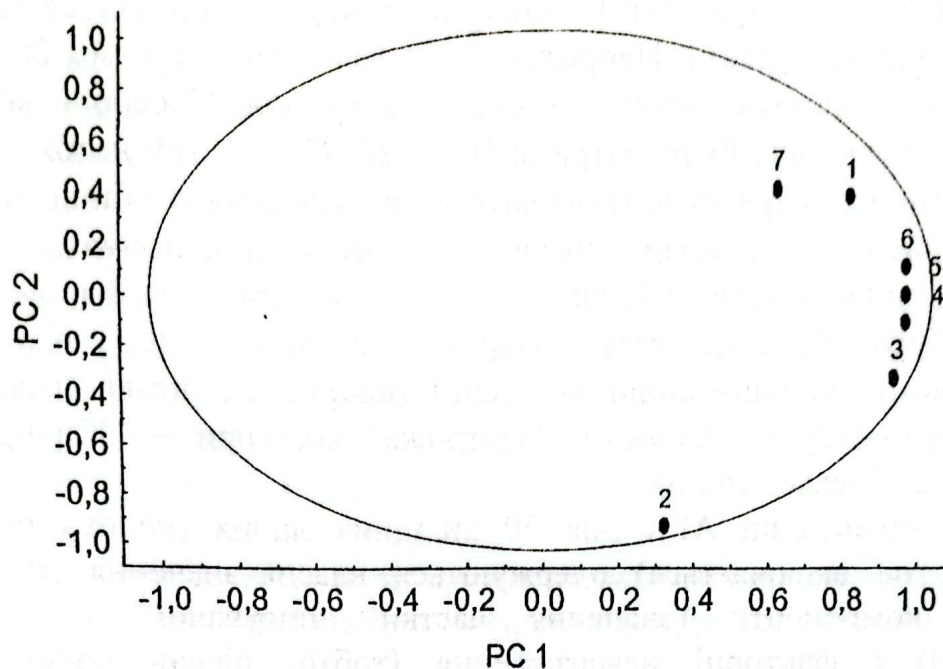


Рис. 10.3. Розподіл ознак у просторі перших двох головних компонент

На відміну від вихідних ознак, що являють собою статичні характеристики, головні компоненти часто характеризують динамічні процеси (наприклад формоутворення). Найчастіше перша головна компонента є розмірної, тобто описує *мінливість щодо загальних розмірів тіла* організмів, тоді як друга головна компонента описує *мінливість за формою* тіла. При цьому найбільш важливе те, що друга компонента описує мінливість форми без тієї частки мінливості, що описується першою головною компонентою. Перші дві головні компоненти можна розглядати як дві *незалежні* характеристики загального морфотипу особини — одна описує мінливість за розмірами, а друга — за формою (або за швидкістю ростових процесів).

Для кожної точки (тобто для кожної особини) розраховуються їхні нові координати (**факторні мітки**, factor scores) у просторі головних компонентів. Ці значення далі можуть бути використовувані для побудови **онтогенетичних траєкторій**. Але ці координати також можна розглядати у просторі перших двох-трьох компонент. Цей підхід дозволяє провести ординацію й класифікацію об'єктів чи їх груп (проблеми ординації та



класифікації якісних ознак та біоценотичних розглянуто в розділі 4 та розділі 9).

### 10.3. Аналіз онтогенетичних траєкторій

Для аналізу онтогенетичних траєкторій використовуються результати R-АГК, при якому вихідна матриця формується для ряду досліджуваних популяцій. Наприклад, якщо використовується 20 ознак і 3 популяції (що містять набори вихідних даних для 35 особин кожна), то первинною матрицею буде матриця 105 x 20. При цьому необхідна також ще одна змінна, котра буде позначати приналежність особини до тієї або іншої популяції (наприклад, для всіх особин першої популяції значення цієї ознаки приймається за 1, другої — за 2, третьої — за 3 тощо). Або це можуть бути й які-небудь літерні позначення, що відповідають географічному розташуванню популяції (наприклад, назва найближчого населеного пункту — “Іванівка”, “Петрівка”, або гори — “Кара-Даг”, або моря — “Балтійське” тощо).

При проведенні АГК для 20 вихідних даних (кодова перемінна, природно, не включається) одержуються власні значення для кожної головної компоненти (значення частки описуваної нею вихідної мінливості) і факторні навантаження (тобто, рівень корельованості головних компонент із ознаками вихідних даних). Крім того, для кожної особини розраховуються нові значення (факторні мітки) координат у просторі перших двох-трьох головних компонентів (інші компоненти частіше за все малоінформативні і з аналізу виключаються). У такий спосіб вихідна матриця, наприклад 105 x 20, трансформується в нову матрицю 105 x 2 (або 3). Тепер можна вже на двомірному графіку представити розкид точок у просторі пари перших головних компонент і використовувати для подальшого аналізу сполученої мінливості за розмірами і формою тіла двовимірні процедури (кореляційний і регресійний аналізи).

Крім того, можна провести АГК окремо для кожної з досліджуваних популяцій (природно, за тими самими ознаками), одержати факторні навантаження для кожного з часткових головних компонент за кожною ознакою і порівняти їхній як між собою, так і з факторними навантаженнями загального АГК. Порівняння по кожній компоненті має на увазі обчислення кута ( $\alpha$ ) між ними за формулою:

$$\alpha = \arccos \left( \frac{\sum (a_{jk} b_{jk})}{\sqrt{\sum a_{jk}^2 \sum b_{jk}^2}} \right), \quad (10.3)$$

де  $a_{jk}$  і  $b_{jk}$  — величини факторних навантажень за  $j$ -тою вихідною ознакою для  $k$ -тої головної компоненти (для порівняння, див. розділ 4).



Чим ближче величина  $\alpha$  до  $0^\circ$ , тим більше подібною виявляється структура мінливості в двох порівнюваних наборах вихідних даних для даного набору ознак. І, навпаки, чим ближче це значення до  $90^\circ$ , тим сильніше вона розрізняється.

На скаттері в просторі перших двох головних компонент ступінь і характер зв'язку для них аналізуються методом *лінійної регресії* на підставі значень факторних міток для кожної з досліджуваних популяцій (або інших угруповань) окремо. (Для узагальненого набору даних кореляція між значеннями об'єктів по PC1 і PC2, природно, відсутня через ортогональність головних компонент). Для цього розраховуються значення коефіцієнтів "a" і "b" рівняння:  $PC2 = a + b \cdot PC1$  і перевіряється вірогідність їх відмінності від 0.

Через те, що перша головна компонента найчастіше визначає загальні розміри, а друга — незалежну від розмірів форму тіла, то оцінка коефіцієнту "a" (якщо він, звичайно, вірогідно відрізняється від 0) відповідає *значенню форми тіла при середніх розмірах тіла* (оскільки вихідні дані спочатку були стандартизовані, то їх середні арифметичні дорівнюють нулю; див. вище).

*Приклад 10.1.* Була проаналізована мінливість за 19 ознаками черепашки для двох груп особин моллюска *Chondrula tridens* (відповідно, 25 і 41 особина).

Спочатку був проведений АГК і факторні навантаження для перших трьох головних компонент (PC1-PC3) (табл. 10.1) для узагальненої матриці вихідних даних і для кожної з груп окремо (стандартизація вихідних даних проводилася на узагальненій матриці). Жирним курсивом виділено вірогідні значення коефіцієнтів кореляції для ознак, що роблять найбільший внесок в інтерпретацію головних компонент.

У двох останніх рядках матриці наведено власні значення для кожної компоненти і частка мінливості, яку вона описує. Частка мінливості розраховується як власне значення, віднесене до числа використовуваних ознак; наприклад, для першої головної компоненти узагальненої матриці частка описуваної нею мінливістю складає:

$$(13,11/19) \cdot 100 = 69,0 \%$$

Проаналізуємо тепер отримані результати. Перша головна компонента узагальненої матриці описує 69,0 % сумарної мінливості, друга — 15,7 % і третя — 5,6 %. Сумарно три перші компоненти описують більше 90 % вихідної мінливості. Отже, використовуючи перші три компоненти для аналізу характеру мінливості матриці вихідних даних, втрачається лише близько 10 % інформації, зате різко знижується вимірність — від 19-вимірного простору до 3-вимірного.



Перші три головні компоненти для кожної з груп описують сумарно по 74 - 75 %.

Таблиця 10.1

Ознаки черепашки	Для узагальнених даних			Для 1-ої групи			Для 2-ої групи		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Ширина апексу	0,89	0,18	0,02	0,73	0,00	0,38	0,60	0,52	-0,02
Ширина 3-х обертів	0,97	0,16	0,06	0,90	0,16	0,22	0,74	0,44	0,20
Ширина 4-х обертів	0,90	0,10	0,02	0,51	-0,03	-0,31	0,69	0,38	0,46
Ширина 5-ти обертів	0,96	-0,02	0,09	0,79	-0,21	0,15	0,40	0,67	0,54
Ширина 6-ти обертів	0,95	-0,17	0,05	0,74	-0,36	0,03	-0,19	0,67	0,6
Висота апексу	0,59	0,55	-0,42	0,65	0,53	-0,32	0,57	0,32	-0,39
Висота 3-го оберту	0,70	0,59	-0,28	0,67	0,48	-0,39	0,77	0,30	-0,38
Висота 4-го оберту	0,70	0,60	-0,19	0,88	0,30	-0,10	0,81	0,11	-0,39
Висота 5-го оберту	0,85	0,48	-0,06	0,89	0,37	0,00	0,87	0,28	-0,27
Висота 6-го оберту	0,91	0,31	0,03	0,81	0,44	0,10	0,80	0,38	0,01
Кількість обертів	0,18	-0,85	-0,41	-0,15	-0,85	-0,30	-0,88	0,19	-0,17
Висота черепашки	0,89	-0,43	-0,13	0,30	-0,88	-0,16	-0,75	0,51	-0,20
Ширина черепашки	0,95	-0,26	0,04	0,43	-0,76	-0,03	-0,49	0,74	0,15
Висота завитку	0,72	-0,58	-0,29	0,16	-0,81	-0,27	-0,75	0,39	-0,27
Висота устя	0,94	-0,29	0,00	0,41	-0,77	-0,05	-0,60	0,66	-0,05
Ширина устя	0,93	-0,28	-0,02	0,58	-0,58	-0,01	-0,52	0,62	0,04
a	0,90	-0,24	0,19	0,21	-0,18	0,76	-0,41	0,52	-0,37
b	0,91	-0,16	0,28	0,09	-0,24	0,91	-0,05	0,29	-0,58
c	0,53	0,05	0,60	-0,20	0,01	0,70	0,22	0,32	-0,55
Власне значення	13,1	2,99	1,06	6,78	4,91	2,68	7,55	4,22	2,37
Частка мінливості, %	69,0	15,7	5,6	35,6	25,8	14,1	39,7	22,2	12,4

Для першої головної компоненти узагальненої вибірки факторні навантаження практично для усіх використаних ознак досить великі і приблизно рівні. Це свідчить про те, що PC 1 узагальненої вибірки описує мінливість за загальними розмірами тіла, у даному випадку - за загальними розмірами черепашки.

Друга головна компонента узагальненої вибірки має відносно високі позитивні навантаження для ознак, що описують висоти початкових обертів черепашки і високе негативне навантаження для кількості обертів (а також для максимальної висоти черепашки і висоти завитка черепашки). З огляду на це, PC 2 може бути інтерпретована як швидкість наростання початкових обертів черепашки.



Третя головна компонента має максимальний рівень кореляції з показником "с" (відстань між виростами краю устя).

На рис.10.4 ясно видно, що об'єкти, що належать до групи 1 і групи 2, формують дві чітких хмари. Основним фактором, що розділяє ці групи, є перша головна компонента, тобто загальні розміри тіла. Тоді як за другою компонентою розходження між двома групами практично відсутні.

Метод онтогенетичних траєкторій має на увазі проведення регресійного аналізу на підставі значень факторних міток у просторі перших двох головних компонент для кожної з груп окремо (рис.10.5).

Для обох груп мається висока достовірна кореляція між першою і другою головними компонентами (в обох випадках коефіцієнт кореляції був більше 0,40).

Значення коефіцієнтів "a" і "b" рівняння лінійної регресії склали для першої групи:

$$a = -1,37 \pm 0,57; b = 1,00 \pm 0,45;$$

для другої групи:

$$a = 1,29 \pm 0,42; b = 1,61 \pm 0,53.$$

В обох випадках коефіцієнти регресії вірогідно відрізнялися від нуля (для розрахунків див. розділ 2).

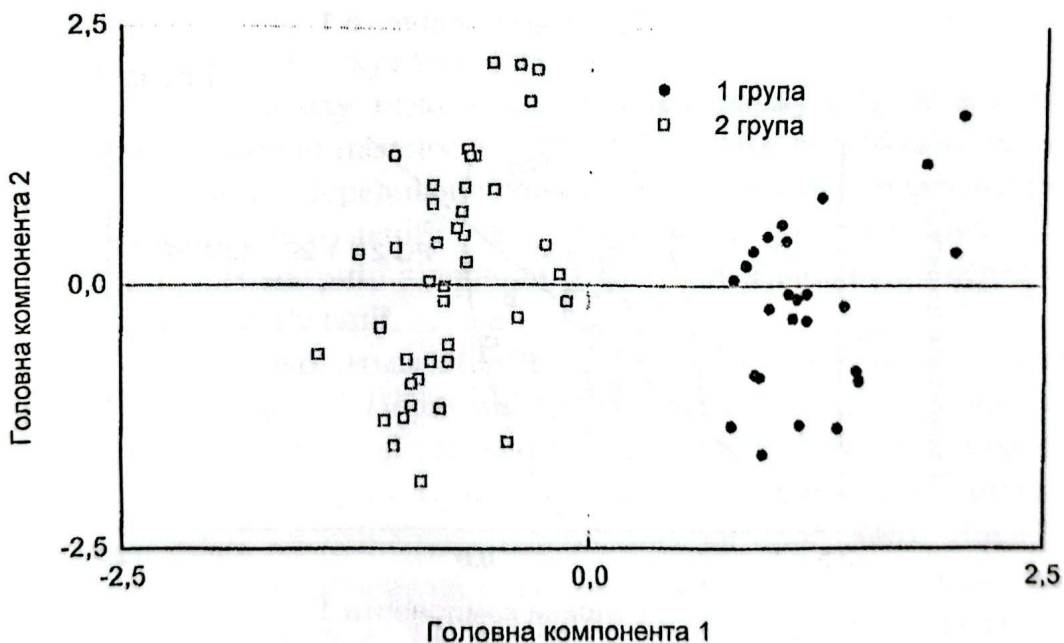


Рис. 10.4. Розподіл особин з прикладу 10.1 в просторі перших двох головних компонент

Крім того, коефіцієнти кута нахилу лінії регресії ("b") вірогідно не розрізнялися в двох вибірках, на відміну від коефіцієнтів інтерцепту (коефіцієнт "a" — точка перетинання лінії регресії з віссю OY).



Отже, при аналізі характеру мінливості даної групи організмів ми маємо результат гетерохронічних змін, у результаті яких сформувалося два морфотипи — великі особини з більш повільною швидкістю наростання початкових обертів (з першої групи) і дрібні, педоморфні особини (з другої групи).

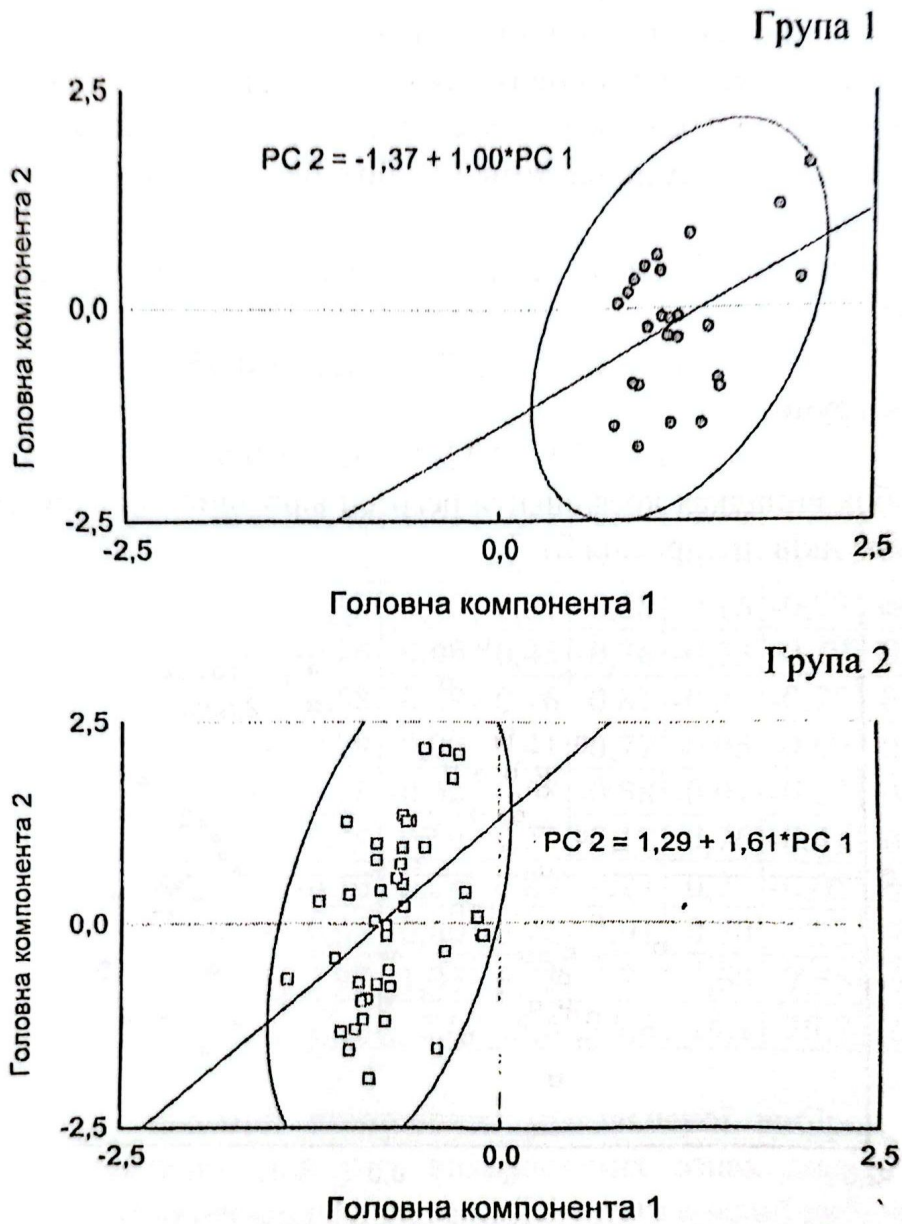


Рис. 10.5. Онтогенетичні траєкторії особин з першої та другої груп

Тепер проаналізуємо структуру мінливості для кожної з груп окремо. Перша і друга головні компоненти для особин з першої групи практично ідентичні відповідним головним компонентам для узагальненої вибірки. Тоді як для особин із другої групи характер мінливості відмінний — її перша головна компонента подібна за структурою другій компоненті



узагальненої вибірки, тобто, описує мінливість за швидкістю наростання початкових обертів черепашки. Це вказує на те, що загальні закономірності формоутворення черепашки зберігаються як у великих, так і у дрібних особин, але при цьому в останніх центр ваги загальної мінливості переноситься на утворення початкових обертів черепашки. Мінливість за загальними розмірами черепашки для матриці їхніх даних нівелюється, а набагато сильніше проявляється мінливість щодо форми (або, більш точно, процесу формоутворення). Це свідчить про те, що зі зміною розмірів тіла форма не може не змінюватися (принцип алометричного росту; див. розділ 7).

На підставі отриманих значень факторних навантажень розраховуються величини кутів між першими головними компонентами особин із двох груп, а також між їхніми значеннями і значеннями для узагальненої вибірки.

Кут між першими головними компонентами для особин відповідно із першої і другої груп складає  $60,3^\circ$ , що свідчить про те, що структура мінливості в першій та другій вибірках досить сильно відрізняється. Перші головні компоненти матриці даних першої групи та матриці узагальнених даних виявляються відносно паралельними — кут між ними складає  $\alpha = 29,5^\circ$ , а перші головні компоненти для матриці узагальнених даних і матриці даних особин другої групи — майже ортогональні (як і впливало очікувати) — кут між ними складає  $\alpha = 79,6^\circ$ .

Результати аналізу головних компонент можуть бути використані також і при визначенні наявності і рівня зв'язку між двома наборами даних — для матриці середньопопуляційних значень морфометричних показників серії популяцій, що існують в різноманітних умовах середовища та для матриці відповідних факторів зовнішнього середовища, в яких існують ці популяції.

Для цього проводиться стандартний Q-АГК (див. вище) окремо для кожного з наборів даних. Перші дві-три головні компоненти з обох наборів інтерпретуються. Потім визначається наявність зв'язку (як правило, з використанням коефіцієнта парної кореляції) між факторними мітками по першим головним компонентам відповідних наборів даних. Вірогідність отриманих коефіцієнтів кореляції вказує на наявність зв'язку між умовами зовнішнього середовища та морфологічними середньопопуляційними показниками.

*П р и к л а д 10.2.* Була вивчена мінливість п'яти показників черепашки молюска *Brephulopsis cylindrica* зі 100 популяцій у Криму. Для кожної популяції, крім того, було визначено шість гідрокліматичних показників. На підставі обох вихідних матриць було проведено Q-АГК і значення розрахованих факторних навантажень наведено в таблиці 10.2.



На підставі значень факторних навантажень перша компонента для морфометричних даних була інтерпретована, як “загальні розміри черепашки”, а друга — як “стрункість черепашки”.

Таблиця 10.2

Навантаження для морфометричних ознак		
Ознаки	PC1	PC2
Висота черепашки	0,46	0,26
Ширина черепашки	0,45	-0,44
Висота устя	0,46	-0,17
Ширина устя	0,47	-0,34
Кількість обертів	0,40	0,77
Навантаження для гідрокліматичних показників		
Показник	PC1	PC2
Число днів у році без морозу	0,40	0,42
Середня температура липня	0,45	0,01
Середня температура січня	0,35	0,67
Річна сума опадів	-0,44	0,36
Сума опадів днями з температурою вище 10° С	-0,36	0,46
Коефіцієнт зволоження	-0,44	0,18

Для гідрокліматичних показників перша головна компонента була інтерпретована як “рівень аридності”, а друга — як “м'якість зими”.

Між факторними мітками по отриманим компонентам були розраховані коефіцієнти парної кореляції Пірсона (табл.10.3).

Як видно, загальні розміри мають тенденцію зменшуватися з підвищенням рівня аридності місцеперебування. В умовах м'якої зими (що характерна для субтропічної зони Південного берегу Криму) розміри черепашки, навпроти, збільшуються. При цьому підвищується і їх “стрункість”, тобто формуються вузькі, високі черепашки, що мають велику кількість обертів.

Подібний аналіз може бути проведений із використанням канонічних кореляцій. На ньому ми зупинимося нижче.

#### 10.4. Метод канонічних кореляцій

Канонічний аналіз дозволяє одночасно досліджувати дві різні групи змінних, що оцінюють ті самі об'єкти. При проведенні екоморфологічних досліджень виникає необхідність сумісного аналізу двох матриць даних (як у попередньому випадку) — перша містить значення морфометричних



показників, а друга — кількісні (рідко, якісні) характеристики середовища, в якому існують дослідженні популяції.

Таблиця 10.3

Морфометричні ознаки	Гідрокліматичний індекс	
	Рівень аридності	М'якість зими
Загальні розміри	-0,25; $p = 0,013$	0,38; $p < 0,001$
“Стрункість”	-0,09; $p > 0,030$	0,32; $p = 0,001$

При канонічному аналізі, також як і при АГК, розраховуються нові лінійні комбінації вихідних змінних у межах обох наборів даних, що мають назву **канонічних факторів**. Розрахунок цих нових змінних проводиться так, щоб коефіцієнти кореляції між першою парою ( $i$ , відповідно, кожною наступною) канонічних факторів були *максимальними*. Якщо в одному з наборів даних є лише одна кількісна змінна, процедура канонічного аналізу зводиться до методу множинної лінійної регресії. Якщо ця змінна якісна — процедура зводиться до дискримінантного аналізу (див. нижче).

Важливою рисою, яка доставляє особливі труднощі, методу канонічної кореляції є проблема інтерпретації канонічних факторів. Математично це проводиться аналогічно тому, як інтерпретуються головні компоненти при АГК, тобто на підставі значущості коефіцієнтів кореляції між вихідними даними і розрахованими канонічними факторами. Таким чином, знову проводиться зниження вимірності вихідних матриць даних при незначній втраті вихідної інформації, що дозволяє візуалізувати розташування об'єктів у двох- або тривимірному просторі. Остання особливість методу канонічних кореляцій дозволяє використовувати його для графічного представлення результатів дискримінантного аналізу.

*Приклад 10.3.* Для 63 вибірок одного виду морських молюсків були отримані дані для 18 промірів черепашки (тобто, були розраховані середьопопуляційні значення) і три показники, що визначають температурний режим вод, у яких ці популяції існують.

Результати канонічного аналізу занесено в таблицю 10.4.

Інші 14 ознак черепашки мали дуже низькі навантаження і не вносили значимого внеску в інтерпретацію першого канонічного фактору.



Ознаки	Навантаження для кожної ознаки
Максимальна температура поверхні води за рік	0,946
Середня температура поверхні води за рік	0,929
Мінімальна температура поверхні води за рік	0,828
Довжина дорсальної плями	-0,449
Ширина дорсальної плями	-0,552
Кут черепашки	-0,839
Ширина устя	-0,485

На підставі розрахованих коефіцієнтів даний канонічний фактор для морфометричних показників може бути визначений як "форма черепашки". Цей фактор виявляється високо корельованим із температурним режимом — коефіцієнт канонічної кореляції між морфометричними показниками і температурою води дорівнює 0,788.

Розподіл вивчених популяцій у просторі перших канонічних факторів для обох наборів показників представлено на рис.10.6.

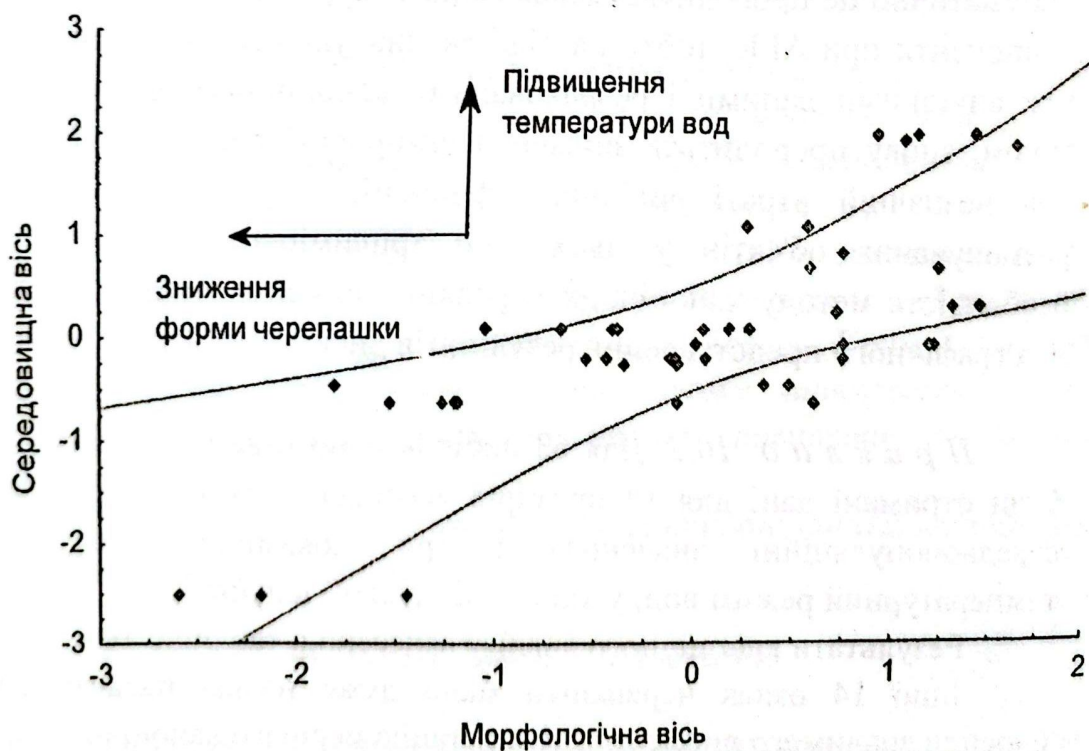


Рис. 10.6. Розподіл вибірок в просторі перших канонічних осей



## 10.5. Факторний дискримінантний аналіз (ФДА)

При розгляді групи популяцій особин одного виду, які існують в різних умовах середовища, може виникнути необхідність визначити ті показники, що виявляються найбільш мінливими в міжпопуляційному масштабі. В одновимірному випадку для цього застосовується *t*-критерій Ст'юдента. Однак, якщо використовується велика кількість вибірок і велика кількість ознак, то виникає ситуація, коли за одними ознаками дві окремо узяті популяції розрізняються, а за іншими виявляються східними. Крім цього, при такому порівнянні частіше аналізується *статична мінливість*. Однак набагато краще і цікавіше аналізувати *динамічну мінливість*, тобто процес формування розбігу за розмірам та формою об'єкту аналізу, для чого використовувати не окремі порівняння за кожною ознакою, а якийсь узагальнений показник, що вносить найбільшу дискримінацію при розташуванні об'єктів у багатовимірному просторі (ця вимірність визначається кількістю одночасно проаналізованих ознак). Таким узагальненим показником є **класифікаційні функції**, що розраховуються при проведенні факторного дискримінантного аналізу.

**Факторний дискримінантний аналіз (ФДА)**, також як і попередні методи багатовимірної аналізу, використовує лінійні комбінації вихідних ознак для формування нових показників, що найкраще розподіляють розглянуті популяції в просторі вихідних ознак. Виділення ознак, що вносять найбільшу дискримінацію, здійснюється на підставі рівня кореляції між вихідними ознаками та осями канонічних кореляцій (див. вище).

Розглянемо особливості застосування ФДА на прикладі класичних даних Р.Фішера (1936) за мінливістю різних показників квітки трьох видів ірисів.

*Приклад 10.4.* Було обміряно чотири ознаки, що характеризують довжину чашолистка, ширину чашолистка, довжину маточки і ширину маточки у 50 особин кожного з трьох видів ірисів. Вихідною є матриця 150 x 5 (чотири морфометричні ознаки та кодова змінна, що визначає приналежність кожної особини до одному з трьох видів).

Насамперед вихідні морфометричні дані були стандартизовані (див. вище). Потім було проведено стандартний ФДА.

Основним показником, що визначає рівень дискримінації є показник лямбда ( $\lambda$ ) Уїлкса (Wilks' Lambda). Значення даної статистики знаходяться в інтервалі від 0 до 1. При цьому, значення лямбди Уїлкса, що лежать близько до 0, свідчать про добру дискримінацію і, навпаки, якщо даний показник близький до 1 — дискримінація об'єктів незадовільна. У



розглянутому прикладі значення  $\lambda = 0,023$ , що свідчить про досить задовільний рівень дискримінації (тобто досліджувані угруповання досить гетерогенні і знайдено ознаки, що вносять максимальний внесок у поділ об'єктів на підставі відповідності їх кожному з видів).

Якість дискримінації перевіряється з використанням матриці класифікації (табл. 10.5).

Що ж означають ці цифри? У ході проведення аналізу для кожної групи об'єктів були розраховані класифікаційні функції, що є лінійними комбінаціями вихідних ознак. Потім для кожного об'єкта з кожної групи були розраховані класифікаційні значення (мітки) для об'єктів.

Таблиця 10.5

Група (фактична)	Частка вірно класифікованих об'єктів (%)	Група (передвіщена)			Сума
		setosa	versicol	virginic	
setosa	100,0	50	0	0	50
versicol	96,0	0	48	2	50
virginic	98,0	0	1	49	50
Сума	98,0	50	49	51	

Віднесення об'єктів до тієї або іншої групи здійснюється на основі величини їхнього класифікаційного значення — об'єкт відноситься до тієї групи, для якої значення його мітки **максимально**. Потім проводиться перевірка вірності віднесення кожного об'єкта до його групи. Так, для групи “setosa” усі 50 об'єктів виявилися віднесені в групу “setosa”, тобто в 100 % випадків дискримінація проведена вірно. Для групи “versicol” 48 об'єктів виявилися віднесеними до групи “versicol”, а два об'єкти — до групи “virginic”. Отже, правильність дискримінації складає:

$$(48:50) \cdot 100 = 96,0 \%$$

Аналогічно, для об'єктів із групи “virginic”.

Останній стовпчик таблиці вказує скільки було використано об'єктів з кожної групи на початку аналізу, а в останньому рядку таблиці наведено скільки об'єктів належить до кожної групи після проведення класифікації на основі розрахованих класифікаційних функцій, тобто наприкінці аналізу. Рівень точності класифікації розраховується окремо для кожної групи, а крім того — наводиться підсумкове значення рівня класифікації (як їх середнє арифметичне); у даному випадку воно складає **98,0%**.



Зрозуміло, необхідно прагнути до того, щоб це значення мало відхилятися від 100 %.

У таблиці 10.6 наведено класифікаційні функції і на їхньому прикладі покажемо, як відбувається теоретичне віднесення об'єктів до наявних груп.

Таблиця 10.6

Ознака	Група		
	setosa	versicol	virginic
Довжина чашолистка (ДЧ)	5,23	-1,27	-3,96
Ширина чашолистка (ШЧ)	5,29	-1,91	-3,38
Довжина маточки (ДП)	-29,92	8,29	21,63
Ширина маточки (ШП)	-15,83	2,33	13,50
Постійний член рівняння	-30,06	-3,03	-17,95

Для групи "setosa" класифікаційна функція має вигляд:

$$\text{setosa}^* = 5,23 \cdot \text{ДЧ} + 5,29 \cdot \text{ШЧ} - 29,92 \cdot \text{ДП} - 15,83 \cdot \text{ШП} - 30,06;$$

для групи "versicol":

$$\text{versicol}^* = -1,27 \cdot \text{ДЧ} - 1,91 \cdot \text{ШЧ} + 8,29 \cdot \text{ДП} + 2,33 \cdot \text{ШП} - 3,03;$$

для групи "virginic":

$$\text{virginic}^* = -3,96 \cdot \text{ДЧ} - 3,38 \cdot \text{ШЧ} + 21,63 \cdot \text{ДП} + 13,50 \cdot \text{ШП} - 17,95.$$

Наприклад, особина мала наступні значення промірів: ДЧ = 0,1; ШЧ = -0,5; ДП = -1,1; ШП = -1,2 (дані, природно, стандартизовані). Для віднесення цієї особини у відповідну групу ці значення підставляються в усі три рівняння:

$$\text{setosa}^* = 5,23 \cdot 0,1 + 5,29 \cdot (-0,5) - 29,92 \cdot (-1,1) - 15,83 \cdot (-1,2) - 30,06 = 19,73;$$

$$\text{versicol}^* = -1,27 \cdot 0,1 - 1,91 \cdot (-0,5) + 8,29 \cdot (-1,1) + 2,33 \cdot (-1,2) - 3,03 = -14,12;$$

$$\text{virginic}^* = -3,96 \cdot 0,1 - 3,38 \cdot (-0,5) + 21,63 \cdot (-1,1) + 13,50 \cdot (-1,2) - 17,95 = -56,65.$$

Як видно, максимальне значення отримано для групи "setosa\*" (+19,73), отже даний організм належить до групи "setosa".



Для наочності об'єкти можуть бути представлені в просторі перших двох канонічних осей (рис. 10.7).

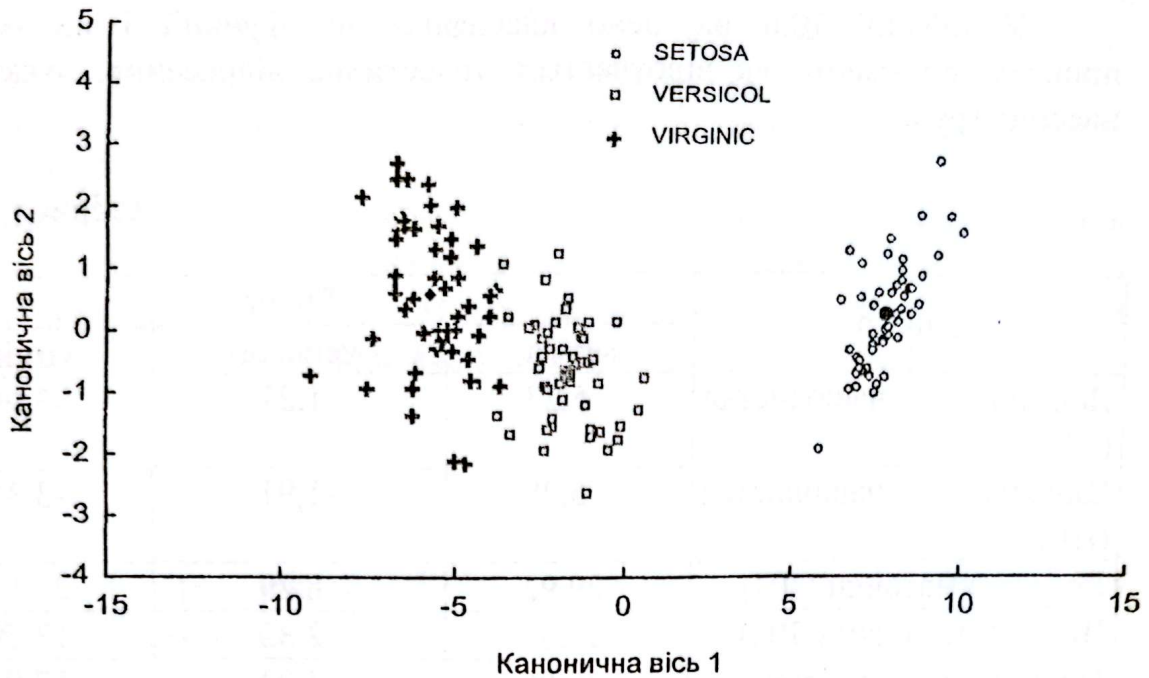


Рис. 10.7. Розподіл об'єктів з прикладу 10.4 у просторі перших двох канонічних осей

Для інтерпретації осей використовуються коефіцієнти кореляції між вихідними ознаками і канонічними осями. Значення, отримані в даному прикладі, наведено в таблиці 10.7 (вірогідні значення виділено).

Таблиця 10.7

Ознаки	Канонічна вісь 1	Канонічна вісь 2
ДЧ	-0,22	0,31
ШЧ	0,12	<b>0,86</b>
ДП	-0,71	0,17
ШП	-0,63	<b>0,73</b>

Отже, перша канонічна вісь може бути охарактеризована як “загальні розміри маточки”, а друга — як “ширини органів”. Перша канонічна вісь розділяє групи на підставі їхніх розмірів маточки: у групі “setosa” даний показник самий мінімальний із усіх порівнюваних груп (при великих значеннях міток мається негативна кореляція між розмірами маточки і першою канонічною віссю), а в групі “virginic”, навпаки, самий максимальний. За другою канонічною віссю незначно відрізняється група “versicol”, для особин якої характерні менші значення ширинів органів.



Середні значення міток по обох канонічних осях для всіх трьох груп являють собою координати центроїдів груп, і на рис.10.7 виділено чорним кольором.

У даному прикладі було використано тільки 4 ознаки, тому інтерпретація канонічних осей виявилася не скрутною. У випадках, коли використовуються 20 - 30 ознак, доцільно використовувати покрокові алгоритми ФДА — Послідовного включення (Forward stepwise) або Послідовного виключення (Backward stepwise) ознак. В останньому випадку для розрахунку класифікаційних функцій використовуються тільки ті ознаки, що вносять *максимальну дискримінацію* і, відповідно, кількість ознак значно знижується (як правило, зводиться до 2 - 5). При цьому необхідно стежити, щоб значення  $\lambda$  Уїлкса було якнайменше.

Для наведеного вище приклада застосування методу Послідовного виключення дає нове значення —  $\lambda = 0,025$ , а точність класифікації зберігається тією ж, при цьому класифікаційна функція і канонічні осі розраховуються на основі трьох останніх ознак. (Показник “Довжина чашолистка” виявляється дуже подібним у всіх трьох порівнюваних групах.) У такий спосіб можна значно знизити мірність простору вихідних ознак і виділити лише ті, котрі значною мірою визначають міжпопуляційну мінливість об'єктів.

Надалі розраховані мітки для кожної групи (для значущих канонічних кореляцій) можна використовувати для визначення наявності зв'язку між канонічними осями (що визначають ті або інші морфологічні характеристики об'єкту дослідження) і групою показників, що характеризують різноманітні умови існування популяцій, як це проводилося при використанні АГК (див. вище).