

Розділ 2. Перевірка статистичних гіпотез

Тема 5. Оцінка належності варіанти до вибірки

Варіанти які сильно відрізняються від інших. Методи визначення відноситься дана варіанта разом з іншими варіантами досліджуваної вибірки до однієї й тієї ж генеральної сукупності, тобто сформовано дане значення варіанти під дією тих же домінуючих та випадкових факторів, що і всі інші варіанти даної вибірки. Варіанти, що виходять за зазначені межі, повинні бути відкинуті для розрахунку більш точних оцінок генеральних параметрів. Нормоване відхилення є безрозмірною характеристикою відхилення окремої варіанти від середньої арифметичної. Критерій випадку (виключення). Нормоване відхилення – найпростіший статистичний критерій, який допомагає визначати варіанти що «вискакують» і вирішувати питання про можливість їх виключення з подальшої обробки. До оцінки чужорідності варіант не можна підходити формально; мета біометричного дослідження завжди полягає в тому, щоб зрозуміти специфіку явища. Зокрема, варіанти що «вискакують» може бути наслідком того, що ознака має інший, ненормальний розподіл. Нормоване відхилення як статистичний критерій. Критерій як метод перевірки статистичних гіпотез заснований на розподілі деяких безрозмірних випадкових величин тому не повинен мати одиниці виміру, щоб підходити до будь-якої біометричної задачі. Статистичний критерій повинен мати відомий закон розподілу, щоб давати ймовірні прогнози поведінки випадкових величин. t-статистика – безрозмірна випадкова величина, яка має відомий закон розподілу і може використовуватися в якості критерію для перевірки статистичних гіпотез.

Тема 6. Оцінка відмінностей двох вибірок

Оцінка статистичної достовірності різниці – визначення, чи можна дану відмінність вважати закономірним, характерним для всієї генеральної сукупності та розглядати його як результат дії особливих факторів, або ж воно випадково та є наслідком недостатньої кількості даних та у наступних дослідах може не виявитися. Виявлення достовірних відмінностей статистичних параметрів – перший крок до пізнання нових біологічних закономірностей, причому кількісно доведених. Параметричні статистичні критерії – використовують стандартні параметри розподілу. Статистичні критерії про достовірність або випадковості відмінностей. Порівняння середніх арифметичних та порівняння часток – порівняння вибірок за ступенем вираженості ознаки; достовірність (недостовірність) відмінностей середніх арифметичних та часток. Порівняння середніх арифметичних за критерієм критерії t Стьюдента. Порівняння показників мінливості – достовірність(недостовірність) відмінностей стандартних відхилень (дисперсій) та коефіцієнтів варіації. Порівняння стандартних відхилень за критерієм t Стьюдента. Порівняння дисперсій за F Фішера. Порівняння коефіцієнтів варіації за критерієм t Стьюдента. Порівняння двох вибірок за характером розподілу (достовірність відмінності частот). Загальні відмінності вибірок без вказівки певних параметрів (для ознак у напівкількісних одиницях). Порівняння вибірок за допомогою непараметричних критеріїв, які спрямовані на

вивчення співвідношень рангів вихідних значень варіант. Ранг – число натурального ряду, яким позначається порядковий номер кожного члена впорядкованої сукупності варіант. Порівняння двох вибірок у цілому (непараметричні критерії): критерій U Уїлкоксона– Манна– Уїтні, критерій T Уайта, критерій Q Розенбаума. Порівняння двох вибірок по силі кореляції двох ознак. Порівняння двох ліній регресії. Порівняння двох вибірок за характером розподілу: критерій χ^2 Пірсона, критерій λ Колмогорова-Смирнова. Відносини між статистиками t, T, F та χ^2 .

Тема 7. Оцінка впливу фактора

Факторіальна біометрія. Однофакторний дисперсійний аналіз кількісних ознак. Дисперсійний аналіз – оцінка ступеня та достовірності відмінності декількох вибірових середніх одночасно, тобто вивчення впливу одного контрольованого фактора на результативну ознаку шляхом оцінки його відносної ролі у загальній мінливості цієї ознаки, викликаній впливом усіх факторів. Логіко-теоретичні основи дисперсійного аналізу. Техніка розрахунку. Дисперсійний аналіз для якісних ознак. Парне порівняння вибірових середніх методом Шеффе. Непараметричний однофакторний дисперсійний аналіз. Порівняння декількох вибірок по мінливості ознаки. Порівняння декількох вибірок за величиною двох ознак (двофакторний дисперсійний аналіз). Двофакторний дисперсійний аналіз досліджує вплив на результативну ознаку двох факторів як порізно, так і спільно. Важливою перевагою двофакторного дисперсійного аналізу перед однофакторним є те, що з його допомогою вдається визначити варіювання по поєднанню градацій, що дозволяє отримати оцінку впливу поєднаної дії (взаємодії) факторів. Логіко-теоретичні основи техніки розрахунків. Техніка розрахунків. Дисперсійний аналіз у середовищах Excel та StatGraphics.

Тема 8. Оцінка залежності між ознаками

Вивчення залежності між варіацією двох або декількох ознак. Визначення змінюються дві змінні самостійно, незалежно один від одного, або варіювання однієї ознаки в якій-о ступеня пов'язано з мінливістю іншого. Дисперсійний аналіз як метод вирішення задачі порівняння декількох вибірок. Особливості техніки дисперсійного аналізу. Методи вивчення поєднаної мінливості – кореляційний та регресійний аналізи. Двовимірний розподіл – координація варіант на площині осей двох ознак. Еліпс розсіювання – сфера поширення варіант однієї сукупності. Ефекти двовимірного розподілу: синхронна зміна двох ознак та розмивання цієї синхронності, тобто дія факторів спряження ознак уздовж осі еліпса та дію випадкових факторів – поперек неї. Регресійний аналіз залежності двох ознак. Коефіцієнт регресії показує на яку величину в середньому змінюється одна ознака при зміні іншого на одиницю виміру. Логіко-теоретичні основи. Техніка розрахунку лінійної регресії. Криволінійна регресія. Регресійний аналіз у середовищах Excel та StatGraphics. Кореляційний аналіз. Кореляція як взаємний зв'язок (взаємна залежність) двох ознак при їх мінливості, тобто спряженість їх варіації. Коефіцієнт кореляції як чисельне вираження частки сполученої варіації двох ознак в загальній їх варіації. Логіко-теоретичні основи. Біологічна інтерпретація коефіцієнту кореляції.

Напря́м мінливості. Техніка розрахунку лінійного коефіцієнту кореляції. Хибна кореляція – коли величина коефіцієнта кореляції визначається в першу чергу способом підбору варіант у вибірку, а не реальною залежністю між досліджуваними ознаками. Множинна кореляція – дослідження множинних зв'язків між великим числом взаємодіючих змінних, які виступають у вигляді цілої системи взаємозалежних ознак організму, так і у формі спільного впливу сукупності факторів на досліджуване явище. Часткова кореляція оцінює зв'язок між першою та другою ознаками при постійних значеннях третьої. Кореляційне відношення та критерій лінійності. Рангові коефіцієнт кореляції Спірмена. Кореляція між якісними ознаками Коефіцієнтом контингенції Шарл'є вимірюється ступінь спряженості (сполучності) двох можливих станів двох якісних ознак. Коефіцієнт асоціації. Кореляційний аналіз у середовищах Excel та StatGraphics.

Тема 9 Класифікування об'єктів

Методи багатовимірної аналізу. Методи багатовимірної статистики дають можливість більш повного (багатопланового) кількісного опису біологічних об'єктів та навколишнього середовища й представити величезні масиви інформації у більш наочному, інтегрованому, узагальненому вигляді. Теоретична основа для методів багатовимірної статистики – поняття гіперпростору (багатовимірний простору). Головна характеристика об'єктів – відстань між ними у гіперпросторі. Головна особливість всієї вибірки – форма хмари розсіювання зі своїми пустотами та згущеннями об'єктів. Методи багатовимірної статистики вивчають інформацію, «закодовану» у порядку розташування об'єктів. Основи кластерного аналізу. Класифікація, кластеризація – методи, які дозволяють наочно уявити схожість або відмінність біологічних об'єктів, надати характеристику за багатьма параметрами. Суть кластерного аналізу. Сенс кластеризації полягає у послідовному об'єднанні об'єктів у кластери, у групи, всередині яких схожість між об'єктами вище, ніж з іншими об'єктами або кластерами. Рівень кожного об'єднання фіксується і потім відображається на графіку. Основи дискримінантного аналізу. Дискримінантний аналіз застосовується для розрізнення (диференціації) та діагностування (розпізнавання) біологічних об'єктів та явищ, відмінності між якими неочевидна. Тобто встановлення групової приналежності окремих об'єктів. Основи методу головних компонент. Компонентний аналіз дозволяє замість численних вихідних характеристик об'єктів дослідження розрахувати кілька нових ознак, лінійних індексів (головних компонентів). Головні компоненти як фактори. Факторні навантаження. Вимога максимуму дисперсії. Факторні навантаження. Розрахунок кореляційних компонент. Вимога ортогональності компонент. Компонентний аналіз. Інформативність і значимість компонент. Етапи компонентного аналізу: організація масиву даних з мітками об'єктів та іменами змінних; вивчення напрямків мінливості вихідних ознак; виконання розрахунків в середовищі спеціальних пакетів; вивчення факторних навантажень; вивчення ординації об'єктів в осях значущих головних компонентів; присвоєння назв значущим компонентам; висновок про основні напрямки (факторах) мінливості даних; відсів або відбір ознак та повторення розрахунків; ітерації дозволяють глибше зрозуміти структуру зв'язків між ознаками.

Варіанти представлення результатів. Компонентний аналіз у середовищах Excel та StatGraphics.

Тема 10. Імітаційне моделювання

Імітаційна модель – як комп'ютерна програма, яка служить для кількісного відображення поведінки реальних об'єктів в різних умовах. Сенс побудови імітаційних моделей: встановити (виразити рівнянням) кількісні закономірності перебігу явищ природи; оцінити модельні параметри (коефіцієнтів пропорційності між змінними рівнянь). Особливості побудови моделей на аркуші Excel – табличне програмування. Компоненти імітаційної системи: блок вихідних даних; блок розрахунку модельних даних, власне імітаційна модель; блок параметрів, що беруть участь в розрахунку модельних даних і змінюються в процесі настройки; блок розрахунку відмінностей реальних і розрахункових значень змінних; значення суми відмінностей між моделлю і реальністю; блок процедури налаштування (програма «Пошук рішення»); блок графічного представлення результатів; блок статистичної оцінки результатів. Завдання апроксимації даних (статичні моделі). Моделювання – це не тільки інструмент перетворення нагромадження фактів в ємну модельну формулу. За допомогою імітаційної моделі будь-яка складна крива залежності може бути представлена як сума більш простих кривих, параметри рівнянь яких набувають біологічний сенс. Динамічні моделі покликані відображати хід досліджуваних процесів. Головний момент імітаційного моделювання полягає в тому, щоб висловити відомі змінні через невідомі параметри. Сучасні пакети прикладних програм для статистичних обчислень: устрій пакету; робота з даними; проведення розрахунків; графічне представлення результатів аналізу; збереження та друкування результатів.