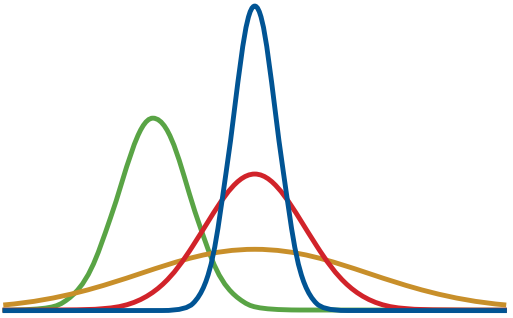
НАЦІОНАЛЬНА АКАДЕМІЯ АГРАРНИХ НАУК УКРАЇНИ

ІНСТИТУТ БІОЛОГІЇ ТВАРИН

І. Р. ПЕТРОВСЬКА
Ю. Т. САЛИГА
І. В. ВУДМАСКА

СТАТИСТИЧНІ МЕТОДИ в БІОлогічних дослідженнях

Навчально-методичний посібник



Київ
АГРАРНА НАУКА
2022

УДК 57.087.1:519.22
С 78

Рекомендовано до друку
вченою радою Інституту біології тварин НААН
22 вересня 2021 р. (*протокол № 11*)

Рецензенти:

А. М. Бабський –
доктор біологічних наук, професор,
завідувач кафедри біофізики та біоінформатики
(*Львівський національний університет імені Івана Франка*);

А. В. Фечан –
доктор технічних наук, професор кафедри програмного забезпечення
(*Національний університет «Львівська політехніка»*)

Петровська І. Р., Салига Ю. Т., Вудмаска І. В.

С 78 Статистичні методи в біологічних дослідженнях: навчально-методичний посібник. Київ: Аграрна наука, 2022. 172 с.

ISBN 978-966-540-551-1

Навчально-методичний посібник знайомить читача з низкою методів, які застосовують для статистичної обробки даних у біологічних дослідженнях. Описано методологію та процедурні особливості методів статистичного аналізу даних з урахуванням специфіки біологічних об'єктів, розкрито сутність основних категорій та методів математичної статистики і біометрії, основні вимоги й умови їх застосування. У виданні на прикладах реальних фізіологічних, біохімічних, генетичних токсикологічних, гематологічних тощо експериментів продемонстровано способи реалізації ключових статистичних методів за допомогою програмного пакета STATISTICA, розробленого компанією StatSoft для всебічного статистичного аналізу.

Розраховано як на досвідчених науковців і викладачів, так і на молодих вчених, аспірантів, студентів як навчально-методичний посібник, та всіх тих, хто цікавиться математично-статистичними методами обробки даних емпіричних досліджень у сфері біології, ветеринарії, експериментальної медицини і сільського господарства.

УДК 57.087.1:519.22

ISBN 978-966-540-551-1

У практиці біолога-експериментатора
окремі результати спостережень або аналізів,
як правило, мають інтерес тільки тоді, коли на їх основі
можна зробити обґрунтовані узагальнення.

Мирон Деркач

ЗМІСТ

ПЕРЕДМОВА

Біологія *– наука точна. Попри значну частку описовості, переважна більшість, якщо не усі закономірності, принципи та механізми функціонування живих систем, починаючи від молекулярного і закінчуючи біосферним рівнями їх організації, можуть бути охарактеризовані математично. Своєю чергою, відкриття, з'ясування, дослідження всіх біологічних процесів, накопичення, систематика та аналіз даних, гіпотез і теорій, формулювання на основі цього наукових парадигм неможливе без застосування математичних методів. Біологічна статистика є саме тим математичним інструментом, котрий, за словами професора Вашингтонського університету Патрика Хітерті, «перетворює дані у знання». Біостатистика – це розділ математичної статистики для обробки результатів наукових досліджень у біології, гуманній медицині і ветеринарії, сільському господарстві, екології, а також розробка нових інструментів для вивчення цих сфер. Як наука біологічна статистика почала формуватися з кінця XVIII ст., коли французько-бельгійський математик, астроном і соціолог Адольф Кетле (1796–1874) заклав її основи. Але перші – ще «несвідомі» приклади застосування елементів біологічної статистики, вочевидь, були наявні тисячоліттями раніше. Ще у первісні часи, коли наші далекі предки почали цілеспрямовано полювати, вони бачили, що частіше чи рідше успіх у цій справі їх супроводжував за певних умов, у певний час чи у певних місцях. Наприклад, ймовірність впіймати рибу на світанку була вищою, ніж у полудень, або шанси вислідкувати звіра зростали у місцях міграцій чи біля водопою. Отже, ще тоді, хоча й неусвідомлено, але люди вже проводили*

примітивний статистичний аналіз, результати якого врешті забезпечували їх кращою їжею, давали змогу менше зазнавати ризиків та економніше використовувати сили й енергію.

Але повернемося до творення біологічної статистики, чи біометрії, як колись її частіше називали, як наукової галузі. Отже, на закладений А. Кетле фундамент було надбудовано ефективний математичний апарат цієї науки завдяки зусиллям англійської школи біометрії XIX ст. Її засновниками були двоюрідний брат Чарльза Дарвіна антрополог, географ, соціолог і психолог Френсіс Гальтон (1822–1911), який, до речі, першим запропонував спосіб обчислення коефіцієнта кореляції і ввів термін «біометрія» у 1889 р. та його учень – математик, біолог та філософ Карл Пірсон (1857–1936), який, своєю чергою, розробив теорію кореляції, критерії узгодженості, алгоритм прийняття рішень з оцінювання параметрів. Він також був основним співзасновником (разом із Ф. Гальтоном і Р. Вельдоном) першого у світі періодичного наукового видання присвяченого застосовуванню статичних методів у біології – журналу «Biometrika», який успішно видає видавництво Оксфордського університету донині.

Ґрунтовний внесок у розвиток методів біологічної статистики зробив ще один англійський вчений – ботанік, генетик і еволюціоніст Рональд Фішер (1890–1962). Поштовхом до цього став неординарний дуже цікавий випадок на агробіологічній станції неподалік Лондона, де він працював у 1910–1914 рр. Отже, одного разу, коли працівники агробіостанції пили чай, у них виникла суперечка – який спосіб приготування чаю правильніший – з доливанням молока до рослинної заварки, чи, навпаки, – коли її додають до чашки з молоком. Учасниця цього історичного чаювання дослідниця водоростей Мюріель Брістоль сказала, що зможе на смак розрізнити напій, залежно від порядку додавання його складників і, як не дивно, бездоганно це зробила, коли їй дали продегустувати кілька приготованих двома способами в іншій кімнаті напоїв. Рональд Фішер вважав, що це сталося випадково і задумався, чи був би результат леді Брістоль таким самим у разі кількох повторів аналогічного експерименту. Ці сумніви Фішера згодом переросли у запропоновану ним методологію планування експерименту, яка була доклад-

но описана в однойменній книзі «Планування експериментів», що побачила світ у 1935 р. Щоправда, передували цьому ґрунтовному виданню низка публікацій вченого у науковій періодиці, зокрема класична праця «Статистичні методи для дослідників», видана у 1925 р. в Единбурзі.

Це далеко не повна історія розвитку і впровадження статистичних методів у біологічні дослідження, а швидше – заклик читача до ознайомлення з нею, що можна зробити завдяки багатьом літературним джерелам, зокрема і тим, які наведено наприкінці нашого посібника.

Розглядаючи внесок українських вчених у розвиток біометричних досліджень, варто, вочевидь, розпочати із книги «Курс статистики» (Київ,1865 р.) авторства Миколи Християновича Бунге – вченого-економіста, академіка, ректора (1859–1862 рр., 1871–1875 рр., 1878–1880 рр.) Київського Університету Св. Володимира. Але вагоме застосування статистичних методів саме у біологічній науці серед українських вчених розпочав видатний нейрофізіолог, гістолог, академік АН України з 1929 р. Олександр Васильович Леонтович (1869–1943). Упродовж 1909–1911 рр. у відомій київській типографії Стефана Кульженка було видрукувано три частини його книги «Елементарний посібник до застосування методів Гаусса та Пірсона при оцінці помилок у статистиці та біології». У 1922 р. Олександр Леонтович публікує книгу «Біологічна статистика у застосуванні до сільського господарства», а у 1935 р. стає співавтором ще одного видання – «Варіаційна статистика».

Статистичну методологію у медичну галузь активно впроваджував всесвітньо відомий український хірург і вчений Микола Іванович Пирогов. Він стверджував, що застосування статистики для визначення діагностичної важливості симптомів можна розглядати як значне надбання новітньої хірургії. Зокрема, у своєму підручнику з основ військово-польової хірургії Микола Пирогов писав: «Я належу до ревних прихильників раціональної статистики та вірю, що додавання її до військової хірургії є безперечним прогресом».

На жаль, з другої половини 1920-х – початку 1930-х років, на відміну від цивілізованого світу, де біологічна статистика продовжувала свій активний розвиток і розширювала сфери застосування,

тоталітарна радянська система визначила біометрію як вияв шкідливих та ворожих буржуазних методологій. Біологічна статистика значною мірою впродовж багатьох років зазнавала утисків і нищівної критики від відданих злочинному режимові псевдонауковців, аналогічно із генетикою. Славнозвісний сталінський прихвостень товариш Лисенко активно виступав не лише проти справжньої генетики, а й з не меншими зусиллями нищив розвиток і блокував застосування статистичних методів у вітчизняній біології.

Згадані вище антинаукові процеси завдали неабиякої шкоди для української біологічної науки. Лише із другої половини XX ст. в Україні розпочалося відродження біологічної статистики і визнання незаперечної необхідності її застосування у науковій роботі вчених біологічного, медичного й аграрного спрямування. У Києві, Харкові, Одесі, інших українських університетських містах сформувалися сильні біометричні школи, з'явилися серйозні публікації та підручники, кафедри і навчальні курси. Один з найвідоміших наукових центрів біологічної статистики було створено у Львові завдяки зусиллям непересічного фізіолога і біофізика, доктора біологічних наук, професора, засновника кафедри біофізики та математичних методів у біології біологічного факультету Львівського національного університету ім. Івана Франка Мирона Пилиповича Деркача. Саме слова з його посібника «Елементи статистичної обробки результатів біологічного експерименту», вперше надрукованого видавництвом Львівського університету у 1963 р., ми винесли як епіграф нашої книги.

Підготоване нами видання не претендує на охоплення усіх біометричних методів, інструментів і підходів, а передусім покликане допомогти студентам, аспірантам, науковцям-біологам початківцям, медикам, аграріям і екологам зрозуміти основи біологічної статистики і навчитися їх застосовувати на практиці. Пропонований читачеві посібник побудований на основі курсу біологічної статистики, який викладається аспірантам нашого Інституту, але сподіваємося стане у пригоді для значно ширшої аудиторії.

Юрій Салига

ВСТУП

Наукові біологічні дослідження зазвичай передбачають проведення експерименту. Експеримент найчастіше є єдиним способом підтвердження справедливості гіпотези і результатів теоретичного дослідження, оскільки відсутність загальноприйнятої аксіоматики і адекватного формального апарата не дає змоги провести належного обґрунтування, не вдаючись до експерименту. Однак просто зібрати дані (об'єктивно та коректно) – недостатньо. Дослідник має вміти їх правильно обробити та проінтерпретувати, що неможливо без застосування математичних методів.

Математична обробка результатів, отриманих під час експериментальних досліджень, є одним з найважливіших етапів наукового пошуку. Висновки, яких дійшов дослідник у процесі інтерпретації даних, що засновані на первинному сприйнятті залежностей між явищами (навіть ґрунтуючись на логічних міркуваннях), не є істинними, якщо вони не підкріплені математичною статистикою, бо досліджувані факти мають бути перевірені з погляду їх статистичної значущості, тобто відповідати вимогам статистичної достовірності. Будь-яке серйозне наукове біологічне дослідження неможливе без кваліфікованого підкріплення у вигляді математичної обробки даних.

Перевагою цього посібника є те, що у ньому в лаконічній та доступній формі (без використання величезної кількості формул, які можуть лише заплутати та збити з пантелику дослідника) описано методологію та процедурні особливості методів статистичного аналізу даних з урахуванням специфіки біологіч-

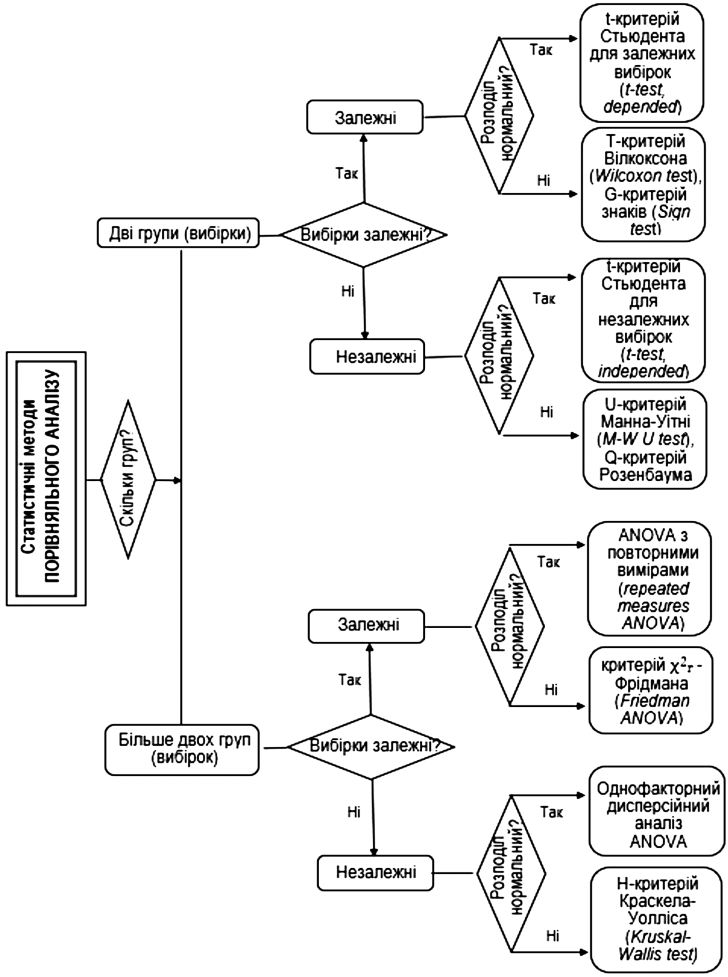
них об'єктів, розкрито сутність основних категорій та методів математичної статистики (описової статистики, порівняльного, кореляційного, кластерного, факторного, дискримінантного, множинного регресійного аналізів), окреслено основні вимоги та умови застосування цих методів, а також на конкретних прикладах досліджень біологічних явищ продемонстровано способи реалізації різних методів засобами універсального пакета статистичного аналізу, в якому реалізовані основні математичні методи аналізу даних.

Отож зміст навчального посібника охоплює такі методи статистичного аналізу:

- *методи дескриптивної* (*описової*) *статистики* – первинні методи статистичної обробки даних, що включають методи оцінювання міри центральної тенденції (середнього значення, медіани, моди) та методи оцінювання міри варіативності/мінливості (стандартне відхилення/дисперсія, мінімальне та максимальне значення змінної, розмах, ексцес та асиметрія). Основне призначення кожної з первинних описових статистик – заміна множини значень ознаки, виміряної на вибірці, одним числом (наприклад, середнім значенням як міри центральної тенденції). Компактний опис групи за допомогою первинних статистик дає змогу інтерпретувати результати вимірювань, зокрема, через порівняння первинних статистик різних груп. Описова статистика забезпечує короткий підсумок про вибірку та про спостереження, які були зроблені. Такі резюме можуть бути як кількісними, наприклад резюмувальна статистика, так і візуальна, наприклад прості графіки. Ці резюме можуть бути або основою початкового опису даних як частина більш комплексного статистичного аналізу, або вони можуть бути достатніми самі по собі для конкретного дослідження;
- *методи порівняльного аналізу.* Процедури порівняльного аналізу зумовлені завданнями конкретного дослідження. Виокремлюють два основні різновиди порівняльного аналізу: з'ясування істотних характеристик двох чи більше різних груп досліджуваних об'єктів через порівняння їхніх власти-

востей (незалежні вибірки); встановлення закономірностей розвитку тих самих досліджуваних об'єктів через порівняння їх станів і властивостей у різні періоди, наприклад до і після експериментального впливу (залежні вибірки). У випадку нормального розподілу емпіричних даних використовують параметричні методи для порівняльного аналізу, якщо ж розподіл даних не узгоджується з нормальним розподілом та/або дані «низької якості» з вибірок малого обсягу, – застосовують непараметричні методи. По суті, для кожного параметричного критерію є принаймні один непараметричний аналог. Алгоритм вибору статистичного методу для порівняльного аналізу для різних дослідницьких завдань подано нижче на рисунку;

- *методи аналізу зв'язку між змінними.* До цих методів належать:
  - кореляційний аналіз – сукупність статистичних прийомів, за допомогою яких досліджується зв'язок між ознаками. Існує два різні способи кореляційного аналізу: параметричний метод розрахунку коефіцієнта Пірсона і обчислення коефіцієнта кореляції рангів Спірмена, який є непараметричним;
  - регресійний аналіз – метод статистичного аналізу, що встановлює як кількісно змінюється одна ознака при зміні іншої (однофакторний регресійний аналіз) або як кількісно змінюється одна залежна змінна при зміні кількох незалежних змінних (багатофакторний/множинний регресійний аналіз);
- *багатовимірні методи опрацювання даних*, що розкривають особливості прихованих зв'язків між явищами і поділяються на: класифікаційні (кластерний аналіз), структурні (факторний аналіз), прогностичні (дискримінантний, множинний регресійний, багатофакторний дисперсійний аналізи). При кластерному аналізі на основі кількох досліджуваних параметрів (змінних) дослідник може класифікувати вибірку, тобто перевірити чи об'єднуються досліджувані об'єкти у групи, тобто у відносно однорідні (схожі, такі, що мають найбільший ступінь близькості досліджуваних параметрів) класи об'єктів.



Алгоритм вибору статистичного методу порівняльного аналізу для різних дослідницьких завдань

За дискримінантного аналізу дослідник шукає змінні, які найкраще розділяють (дискримінують) початково наявні групи. За факторного аналізу дослідник має змогу зменшити обсяг вихідних даних (змінних) для їх лаконічного опису при міні-

мальній втраті інформації, а також виявити внутрішню структуру взаємозв'язків між змінними. Множинний регресійний аналіз, як вже згадувалося вище, дає можливість передбачити певний результат за низкою інших показників.

Навчальний посібник містить також глосарій та тестові завдання, що дають змогу кожному читачу, за бажання, самостійно перевірити рівень засвоєних знань щодо методів статистичної обробки інформації.

Усі таблиці і рисунки, які не мають посилань, розраховано та сформовано авторами посібника.