

«Статистичні методи в охороні навколишнього середовища»

Питання і завдання для поточного контролю

Що таке статистика?

Що вивчає прикладна статистика в галузі охорони навколишнього середовища?

Розвиток уявлень про статистику, коротка історія статистичних методів.

Основні поняття (категорії) статистики.

Класифікація статистичних методів.

Що таке «Статистична сукупність»?

Що означає термін «Однорідна сукупність»?

Що означає термін «Різнорідна сукупність»?

Що таке «Одиниця сукупності»?

Що означає термін «Ознака», якісні та кількісні ознаки?

Що таке «Статистичні показники»?

Що таке «Статистичні дані»?

Що таке «Статистична закономірність»?

Які є види статистичних методів?

Особливості прикладної статистики.

Пояснення:

1. Прикладна статистика

Прикладна статистика — це наука про те, як обробляти дані довільної природи. Математичною основою прикладної статистики і статистичних методів аналізу є теорія ймовірностей й математична статистика.

Опис виду даних і механізму їхнього породження — початок будь-якого статистичного дослідження. Для опису даних застосовують як детерміновані, так й імовірнісні методи. За допомогою детермінованих методів можна проаналізувати тільки ті дані, які є в розпорядженні дослідника. Наприклад, з їхньою допомогою отримані таблиці, розраховані органами офіційної державної статистики на основі представлених підприємствами й організаціями статистичних звітів. Перенести отримані результати на ширшу сукупність, використати їх для проорокування й керування можна лише на основі ймовірнісно-статистичного моделювання. Тому в математичну статистику часто включають лише методи, що опираються на теорію ймовірностей.

У найпростішій ситуації статистичні дані — це значення деякої ознаки, властивої досліджуванім об'єктам. Значення можуть бути кількісними або являти собою вказівку на категорію, до якої можна віднести об'єкт. У другому випадку говорять про якісну ознаку.

При вимірі по декількох кількісних або якісних ознаках як статистичні дані про об'єкт одержуємо вектор. Його можна розглядати як новий вид даних. У такому випадку вибірка складається з набору векторів. Є частина координат — числа, а частина — якісні (категоризовані) дані, то говоримо про вектор різнотипних даних.

Одним елементом вибірки, тобто одним виміром, може бути й функція в цілому. Наприклад, що описує динаміку показника, тобто його зміна в часі, — електрокардіограма хворого або амплітуда биття валу двигуна. Або часовий ряд, що описує динаміку показників певної фірми. Тоді вибірка складається з набору функцій.

Елементами вибірки можуть бути й інші математичні об'єкти. Наприклад, бінарні відносини. Так, при опитуваннях експертів часто використовують упорядкування (ранжирування) об'єктів експертизи — зразків продукції, інвестиційних проектів, варіантів управлінських рішень. Залежно від регламенту експертного дослідження елементами вибірки можуть бути різні види бінарних відносин (упорядкування, розбивка, толерантності), множини, нечіткі множини і т. д.

Отже, математична природа елементів вибірки в різних завданнях прикладної статистики може бути найрізноманітнішою. Однак можна виділити два класи статистичних даних — числові й нечислові. Відповідно прикладна статистика розбивається на дві частини — числову статистику й нечислову статистику.

Числові статистичні дані — це числа, вектора, функції. Їх можна складати, множити на коефіцієнти. Тому в числовій статистиці велике значення мають різноманітні суми. Математичний апарат аналізу сум випадкових елементів вибірки — це (класичні) закони більших чисел і центральні граничні теореми.

Нечислові статистичні дані — це категоризовані дані, вектора різнотипних ознак, бінарні відношення, множини, нечіткі множини й ін. Їх не можна складати й множити на коефіцієнти. Тому не має змісту говорити про суми нечислових статистичних даних. Вони є елементами нечислових математичних просторів (множин). Математичний апарат аналізу нечислових статистичних даних заснований на використанні відстаней між елементами (а також мір близькості, показників різниці) у таких просторах. За допомогою відстаней визначаються емпіричні і теоретичні середні, доводяться закони більших чисел, будуються непараметричні оцінки щільності розподілу ймовірностей, вирішуються завдання діагностики й кластерного аналізу, і т. д.

У прикладних дослідженнях використовують статистичні дані різних видів. Це пов'язано, зокрема, зі способами їхнього одержання. Наприклад, якщо випробування деяких технічних

пристроїв тривають до певного моменту часу, то одержуємо так звані цензуровані дані, що складаються з набору чисел — тривалості роботи ряду пристроїв до відмови, і інформації про те, що інші пристрої продовжували працювати в момент закінчення випробування. Цензуровані дані часто використовуються при оцінці й контролю надійності технічних пристроїв.

2. Статистичний аналіз конкретних даних

Застосування статистичних методів і моделей для статистичного аналізу конкретних даних тісно прив'язане до проблем відповідної області. Хоч математичний апарат практично один і той же, проте при зборі даних та їх аналізі потрібно враховувати специфіку певної області і специфіку самого дослідження. Особливо це стосується початкових етапів постановки експерименту (чи спостереження) та збору даних, оскільки при їх неправильному здійсненні можна одержати непрезентабельні дані, які можуть призвести до хибних висновків. В результаті можна знайти в даних щось чого насправді не існує, або ж навпаки не помітити чогось істотного. В ідеалі будь-які результати статистичного дослідження повинні перепроверитись як шляхом неодноразової перевірки процесу аналізу статистичних даних так і шляхом здійснення повторного дослідження.

Теорія статистичних методів націлена на вирішення реальних завдань. Тому в ній постійно виникають нові постановки математичних завдань аналізу статистичних даних, розвиваються й обґрунтовуються нові методи. Обґрунтування часто проводиться математичними засобами, тобто шляхом доказу теорем. Більшу роль грає методологічна складова — як саме ставити завдання, які припущення прийняти з метою подальшого математичного вивчення. Велика роль сучасних інформаційних технологій, зокрема, комп'ютерного експерименту.

Актуальним є завдання аналізу історії статистичних методів з метою виявлення тенденцій розвитку й застосування їх для прогнозування.

3. Обчислювальна статистика

Розвиток обчислювальної техніки у другій половині ХХ століття вплинуло на статистику. Раніше статистичні моделі були представлені переважно лінійними моделями. Збільшення швидкодії ЕОМ і розробка відповідних чисельних алгоритмів послужило причиною підвищеного інтересу до нелінійних моделей таких, як штучні нейронні мережі, і привело до розробки складних статистичних моделей, наприклад узагальнена лінійна модель й ієрархічна модель.

Одержали широке поширення обчислювальні методи, засновані на повторній вибірці як критерій перестановок й бутстрепінг, а також методи як семплювання по Гібсу дозволили більш доступно використовувати баєсівські алгоритми. У цей час існує різноманітне статистичне програмне забезпечення загального й спеціалізованого призначення.

4. Некоректна інтерпретація статистичних досліджень

Існує думка, що дані статистичних досліджень все частіше навмисно спотворюють або неправильно інтерпретують, вибираючи тільки ті дані, які є сприятливими для ведучого конкретного дослідження. Неправильне використання статистичних даних може бути як випадковим, так і навмисним. У книзі Хаффа й Даррелла (1954) «Як обдурити за допомогою статистики?» (How to Lie With Statistics) викладається ряд міркувань із приводу використання й неправильного застосування статистичних даних. Деякі автори також проводять огляд статистичних методів, використовуваних у певних областях (наприклад, Варн, Лазо, Рамос, і Ріттер (2012)). Способи, що дозволяють уникнути неправильного тлумачення статистичних даних містять у собі використання належної схеми й виключення упередженості при проведенні досліджень. Зловживання відбувається тоді, коли такі висновки «замовляються» певними структурами, які навмисно або несвідомо виводять на добір упереджених даних або проб. При цьому гістограми, як найпростіший для використання й розуміння (сприйняття) вид діаграми,

можуть бути зроблені або із застосуванням звичайних програм для комп'ютера або просто намальовані. На жаль, більшість людей не роблять спроб шукати помилки або помиляються самі, тому й не бачать помилок. Таким чином, на думку авторів, статистичні дані, щоб бути правдою, повинні бути «не причесані» (тобто достовірні дані не повинні виглядати ідеальними). Для того, щоб отримані статистичні дані виявилися правдоподібно й точними, проба повинна бути репрезентативною в цілому.

Стентон Гланц, американський професор медицини, який викладає статистику студентам медичного профілю, автор ряду курсів та книг по статистиці та редактор *Journal of the American College of Cardiology*, зазначає, що близько 50% публікацій, які надходять в редакцію журналу містять статистичні помилки отримання та обробки медичних даних^[14]. Основними причинами є недостовірність отриманих даних (неправильно поставлені експерименти або ж взяті непрезентабельні дані), а також незнання та неправильне застосування статистичних методів. Також він зазначає, що часто дослідник та особи причетні до експерименту можуть підсвідомо видавати бажане за дійсне. Причому ненавмисне підтасування даних може відбуватися як на етапі постановки експерименту, збору даних так і на етапі аналізу даних. Виходом є максимальне врахування та усунення сторонніх факторів, які можуть вплинути на процес експерименту та на аналіз даних. Він пропонує якомога ширше використовувати «сліпий метод» чи навіть «подвійний сліпий метод», коли ні піддослідні, ні дослідники (чи помічники дослідників) достеменно не знають, що на якій групі хворих досліджується і навіть аналіз даних бажано щоб робила особа незацікавлена у некоректній інтерпретації даних, чи ще краще, якщо вона буде необізнана у конкретних деталях експерименту. В будь-якому разі в постановці, зборі та аналізі даних повинні брати участь особи, які добре володіють прикладними статистичними методами.

5. Експериментальні дослідження і спостереження

Загальна мета для статистичного дослідження проекту є вивчення причинно-наслідковий зв'язів, і, зокрема, зробити висновок про вплив зміни значень предикторів або незалежних змінних на залежні змінні або відповідь. Є два основних типи причинних статистичних досліджень: експериментальні дослідження та наглядові дослідження. Різниця між цими двома типами досліджень полягає в тому, як насправді дослідження проводилося. Кожне з цих досліджень може бути дуже ефективним. Експериментальне дослідження включає в себе прийом вимірювання даної системи, маніпулюючи системою, а потім приймати додаткові виміри використовуючи ту ж процедуру, щоб визначити, маніпуляції модифікованих значень вимірювань. На відміну від цього, наглядове дослідження не передбачає експериментальної маніпуляції. Замість цього, дані збираються і кореляції між предикторами і відповідями досліджуються. У той час як інструменти аналізу даних найкраще працюють на даних рандомізованих досліджень, вони також застосовуються до інших видів даних — як природних експериментів і наглядових досліджень — для яких статистика буде використовувати модифікований, більш структурований метод оцінки.

Література

- Статистика: навчальний посібник / С. О. Матковський, М. Л. Вдовин, Т. В. Панчишин. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2010. – 344 с.
- Статистика : навч. посіб. для студ. екон. спец. вищ. навч. закл. / С. О. Матковський, Л. І. Гальків, О. С. Гринькевич, О. З. Сорочак. – 2-ге вид., доповн. і виправл. – Л. : Новий Світ-2000, 2011. – 429, [3] с. : іл. – (Вища освіта в Україні). – Бібліогр. в кінці розділів. – ISBN 978-966-418-089-1
- Статистика : навч. посіб. / Р. В. Фещур, В. П. Кічор, А. Ф. Барвінський, М. Р. Тимошук ; М-во освіти і науки України, Нац. ун-т "Львів. політехніка". – 4-те вид., оновл. і доповн. – Л. : Бух. центр "Ажур", 2010. – 256 с. : іл., табл. – Бібліогр.: с. 251-252 (20 назв). – ISBN 978-966-1688-05-5
- Статистика ринку товарів та послуг : Навч. посіб. / Л. І. Крамченко; Укоопспілка. - Л., 2002. - 188 с. - Бібліогр.: 32 назви.
- Орлов А. И. Прикладная статистика. Учебник. - М.: Издательство "Экзамен", 2004. – 656 с.
- Учебно-методический комплекс дисциплины «Статистика» — часть 2 (рос.)