

Основи медичної статистики для науковців-початківців

Мета роботи: ознайомити науковців-початківців із можливостями медичної статистики.

У статті викладено основні, базові терміни та поняття, необхідні для правильного планування статистичних досліджень.

Ключові слова: статистика; відмінності; достовірність; кореляція; генеральна сукупність; вибірка; перемінні.

Сьогодні, мабуть, немає такого науковця, котрий заперечував би необхідність застосування методів статистики для оцінки результатів наукових досліджень. Але, на жаль, це часто відбувається лише на декларативному рівні, бо, насправді, серед молодих вчених часто існує інше, келійне, розуміння необхідності проведення статистичних досліджень. Часто, на їхню думку, перша і найголовніша причина, що спонукає їх до проведення таких досліджень, це обов’язкова умова вимог до публікацій.

З точки зору частини молодих науковців, результати їхніх досліджень настільки красномовно говорять про себе, що дані статистичного аналізу є лише формальною окрасою наукової роботи. Як не дивно, але у цьому існує певна “логіка”. Річ у тім, що часто статистичний аналіз зводиться до встановлення достовірності відмінності між результатами наукових досліджень або встановлення кореляції між досліджуваними явищами. Але, як правило, результати таких статистичних досліджень є очікуваними, бо вони апіорі закладені ще на етапі науково обґрунтованого планування. (Бо хто ж плануватиме роботу, що призведе до негативних результатів?) А коли результат передбачений, такі статистичні “дослідження” зводяться до формальних.

Другою, більш глибокою причиною формалізму в проведенні статистичного аналізу, є відсутність його як такого, бо встановлення факту кореляції чи достовірності відмінності можна назвати аналізом лише з певною “натяжкою”. Це радше констатація факту, який часто лежить на поверхні.

Існує й третя, важлива причина, яка відвертає науковця від проведення статистичних досліджень. Статистика – це один із методів математичного аналізу. І це досить складний розділ математики, який, як правило, не доступний для розуміння особам, що не мають відповідної математичної підготовки. З огляду на це, розробники тих чи інших методів статистичних досліджень намагались і намагаються подати запропонований той чи інший метод дослідження у своїй результу-

юючій частині в такій формі, яка є зрозумілою для пересічного користувача. Це щось на зразок “чорного ящика”. Ми знаємо, що вводимо і розуміємо, що отримали. А як це працює – “Terra Incognita”.

Може здатись, що це не суть важливим. Адже придбавши “крутий” “дивайс” його власник, як правило, не уявляє як це все працює. Його задовольняє результат. Але це уявлення не можна поширювати на предмет нашої розмови. Кожен метод дослідження є суто специфічний і призначений лише для застосування за певних умов. Отже, необхідні бодай елементарні уявлення що це і як працює.

Останнім часом з введенням цифрових технологій, програмного забезпечення ситуація ще більше погіршилась, бо нічого немає простішого, як натиснути на кнопку. Виникає ілюзія простоти.

Але найголовніше, що в силу наведених вище причин, часто молоді науковці, насправді, не розуміють всієї глибини та можливостей та потенціалу в проведенні наукових досліджень.

У рамках цієї публікації неможливо розглянути всі особливості та можливості статистичних методів дослідження, але на основних із них зосередимо свою увагу.

Що може запропонувати статистика для науковця, який проводить дослідження? Окрім вже розглянутих кореляції та достовірності, які використовують для підтвердження коректності досліджень та висновків по них, статистичні методи дозволяють провести поглиблений аналіз результатів досліджень і зокрема відповісти на питання чому існують якісь залежності, закономірності в результатах досліджень, які основні чинники впливають на них, яким чином їх корекція може вплинути на результати досліджень тощо.

Окрім того, методи статистичних досліджень дозволяють виявити невідомі, не передбачувані закономірності в досліджуваних явищах.

Також методи статистики є потужним інструментом для прогнозування процесів, явищ.

Часто існують ситуації, коли експеримент не можливо поставити, або отримані дані не мають розумного пояснення. Дати відповідь на поставлені питання дозволять методи математичного моделювання. Часто їх використовують в ситуації, коли необхідно змоделювати небезпечний процес, спрогнозувати його небажані наслідки і попередити їх.

Більше того, статистика дозволяє побачити явища, які залишаються непомітними при візуальному аналізі отриманих результатів.

Однак ми не будемо детально розглядати всі можливості математичної статистики. Для початку варто зупинитись на більш приземлених речах, бо саме з них починаються найбільш серйозні помилки, які зводять нанівець результати статистичного аналізу. Більше того, отримані результати можуть бути “фейковими” і при цьому на перший (і навіть не перший) погляд виглядати переконливими, достовірними. Це не означає, що помилку неможливо виявити. Можливо, але для цього потрібен спеціаліст, який володіє відповідними інструментами аналізу. Зрозуміло, що йдеться не про пересічного молодого науковця, далекого від складної математики та статистики. Тож який вихід? Як уникнути фатальних помилок, коли навіть не відомо чи результати статистичних досліджень є правдивими чи ні?

Головні і найнебезпечніші помилки з непередбаченими наслідками виникають, як правило, з перших кроків статистичного аналізу. Зупинимось на головних, важливих моментах. Це “ази”, без розуміння яких неможливо проводити дослідження.

Перше, що слід зробити, це розібратись у термінології. Почнемо з простого.

Генеральна сукупність – це вся множина однорідних за певною ознакою об’єктів чи подій, які є предметом дослідження. Прикладом є всі зірки на небі, люди на Землі, всі можливі комбінації предметів, явищ тощо. Тобто це те, чого більше не може бути.

Вибірка – це частина від цілого (від генеральної сукупності).

Перемінна – одна з характеристик елементів вибірки. Приклад. Населення Тернопільської області є вибіркою з генеральної сукупності жителів України. Колір очей, зріст чи якісь інші біометричні, фізіологічні досліджувані показники цього населення є перемінними.

Випадок – це окремо взятий показник з перемінної. (Приклад. Висота зросту окремої людини, що входить у досліджувану вибірку.) Випадки в перемінній можуть бути представлені як дискретні величини (температура тіла або частота пульсу в даний момент) або неперервні (тривалість людського життя, тривалість інкубаційного періоду –

величина, яка може приймати будь-яке значення в заданому відрізку виміру).

Перемінна будь-якої вибірки складається з сукупності окремих випадків, кожен з яких має своє кількісне або якісне значення. Приклад. У вибірці 100 чоловік, досліджується величина артеріального тиску (перемінна – артеріальний тиск) у кожного конкретного досліджуваного (випадку). При цьому виявляється, що різні значення артеріального тиску у різних осіб (випадків) в кількісному відношенні по-різному представлені у вибірці. Якісь значення артеріального тиску (випадки) кількісно представлені у вибірці більшою мірою, а якісь меншою мірою. Те, як кількісно у відсотках чи інших величинах представлені ці значення у вибірці – називається характером розподілу. У більшості випадків біологічні показники (випадки перемінних) розподіляються у вибірках таким чином, що найбільша їх частина зосереджується навколо середнього значення перемінної. І чим більше відхиляються значення окремих випадків в більшу або меншу сторону від середнього значення перемінної, тим їх менше у вибірці. Ступінь відхилення кількості випадків різних за своїм значенням від їх середнього значення описується правилом трьох сигм (рис. 1). Коли йдеться про неперервні величини, такий розподіл називається нормальним розподілом, а коли про дискретні – біноміальним (див. рис. 1 та 2).

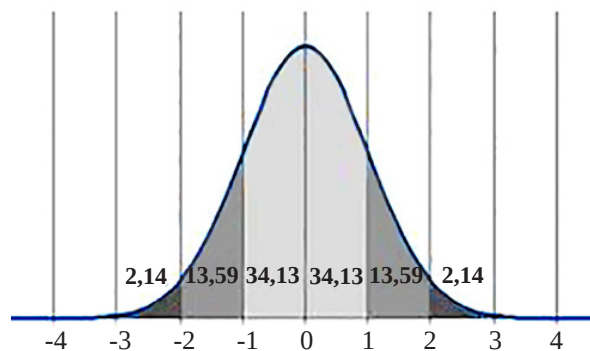


Рис. 1. Нормальний розподіл.

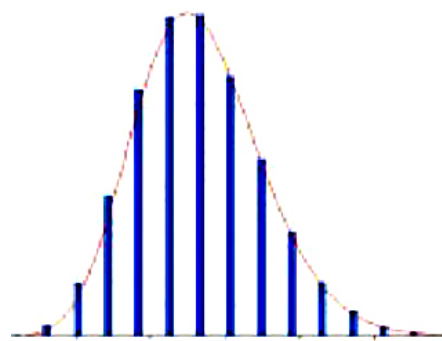


Рис. 2. Біноміальний розподіл.

Які ж існують ненормальні розподіли? Ненормальний розподіл неперервних величин найчастіше буває експоненціальним, лог – нормальним (рис. 3), гамма розподіл (рис. 4), Лапласа (рис. 5) тощо. Дискретні величини частіш за все можуть мати такі види ненормального розподілу, як Пуассона, геометричний, Бернуллі.

Досліджувані перемінні можуть належати до різних статистичних шкал. Існує 4 основні статистичні шкали:

- номінальна (чоловік – жінка, день – ніч);
- порядкова (1,2,3,4...n);
- інтервальна (характеризує явище за певний період часу чи якимось іншим інтервалом);
- шкала відношень (у скільки раз одна величина більша від іншої).

Наостанок, що необхідно знати, перш ніж розпочати статистичну обробку результатів досліджень. Що таке незалежна і залежна перемінна? Коротко і зрозуміло – це причина та наслідок.

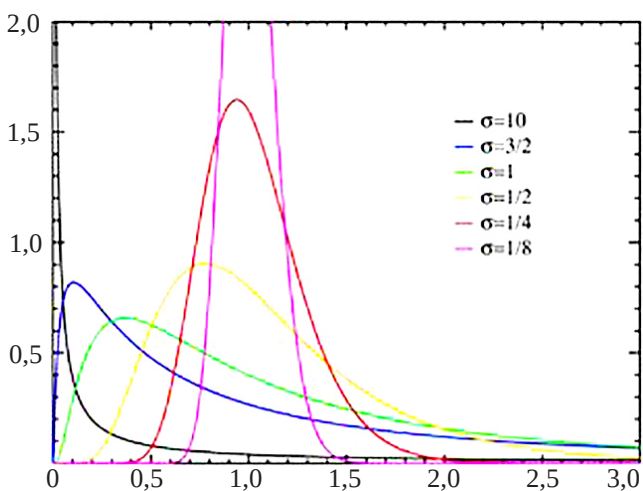


Рис. 3. Лог – нормальний розподіл.

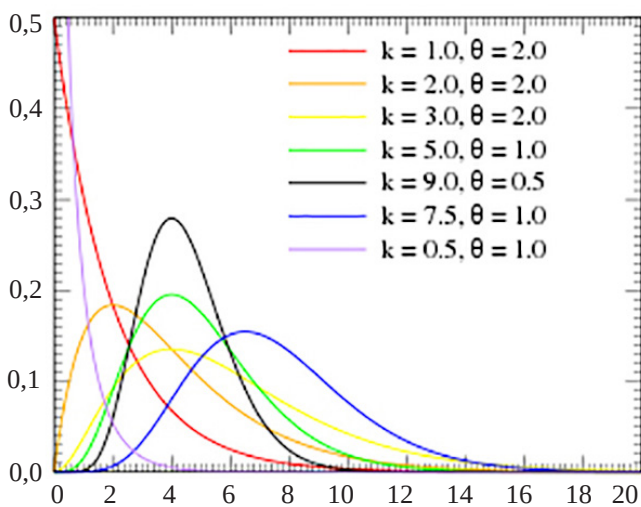


Рис. 4. Гамма розподіл.

Коли вивчають вплив куріння на розвиток раку легень, то кількість викурених цигарок буде незалежна перемінна, а кількість виявлених хворих – залежна. Однак кількість викурених цигарок може бути і залежною перемінною, коли ми вивчаємо вплив стресів на кількість викурених цигарок.

Ось ще один приклад. При зростанні фізичного навантаження зростає частота дихання та серцебиття. Останні є залежними від першої. А що вони є одна відносно іншої? На перший погляд, вони взаємопов'язані, взаємозалежні. Однак не все так просто. Адже, насправді, вони залежать від змін в організмі, а не напряму від навантаження.

А от як вони через ці механізми пов'язані, відповіді не просто. Скажімо, після гіпервентиляції легень звичний зв'язок між частотою дихання та частотою серцебиття буде порушено.

Отже, що є залежна, а що незалежна перемінна, слід визначитися лише після того, як буде зрозуміле явище і що в ньому досліджується.

З наведеного вище зрозуміло, що оцінювати властивості перемінної, аналізувати її параметри необхідно із врахуванням всіх характеристик випадків та їх характеру розподілу в перемінній.

Тож який перший крок необхідно зробити перш ніж розпочати статистичні дослідження (скажімо, виявлення кореляції чи встановлення достовірності відмінності)?

1. Встановити до якої статистичної шкали належить досліджувана перемінна.
2. В якому статусі вона розглядається (залежна чи незалежна)?
3. Який характер розподілу випадків у перемінній? (Як правило, характер розподілу визнача-

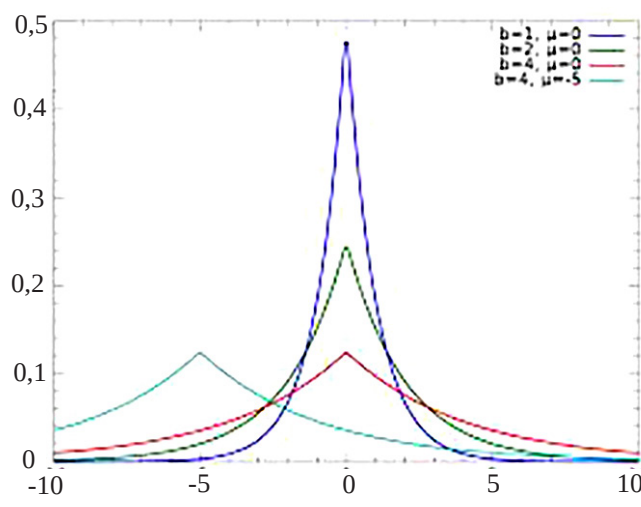


Рис. 5. Розподіл Лапласа.

ПОВІДОМЛЕННЯ

ється графічно по гістограмі, або за тестом Колмогорова – Смірнова).

На основі отриманої інформації проводиться вибір того чи іншого методу дослідження достовірності відмінності (рис. 6).

Щодо вибору методу виявлення кореляції. У випадку, коли досліджувані перемінні належить

до інтервальної чи номінальної шкали, коефіцієнт кореляції визначається Пірсоном (параметрична кореляція).

Якщо бодай одна з перемінних має порядкову шкалу або не має нормального розподілу – необхідно визначити рангову кореляцію за Спірменом або Кендалутау (непараметрична кореляція).

	• Порядкові перемінні	• Інтервальні перемінні
• Номінальні	не нормальний розподіл	• нормальний розподіл
• (дихтомна, категорійна)	2 Незалежні Манна- Уїтні	2 незалежні T тест для незалежних Стьюдента
• Порядкові	2 Залежні Уїлкоксона	2 залежні T тест для залежних
• Інтервальні	Більше 2 незалежні Крускала и Уолліса	Більше 2 незалежні Однофакторний дисперсійний аналіз
• Шкала відношень	Більше 2 залежні Фрідмана	Більше 2 залежні Однофакторний дисперсійний аналіз з повторними вимірами

Рис. 6. Умови, за яких слід вибирати для встановлення достовірності відмінності відповідний статистичний метод дослідження.

Якщо одна з двох перемінних є дихтомною – застосовується метод точкової дворядної кореляції.

Окрім достовірності відмінності та кореляції, в арсенал методів статистичних досліджень вхо-

дять регресійний аналіз, дискримінантний аналіз, факторний аналіз, аналіз придатності, класифікаційний аналіз, аналіз виживання та ряд інших. Однак це вже предмет іншої публікації.

СПИСОК ЛІТЕРАТУРИ

1. Бююль Ахим SPSS Анализ статистических данных и восстановление скрытых закономерностей / Ахим Бююль, Петер Цефель. – Москва. – Санкт-Петербург – Киев, 2002. – 601 с.
2. Боровиков В. STATISTICA. Искусство анализа данных на компьютере / В. Боровиков. – Москва – Киев – Минск, 2003.

– С. 688.

3. Майкрософт. Нейронные сети. STATISTICA Neural Network. – Москва, 2001. – С. 182.
4. Лакин Г. Ф. Биометрия / Г. Ф. Лакин. – М. : Высшая школа. – 1990. – С. 350.

REFERENCES

1. Byul Akhim, & Peter Tsefel (2002). SPSS *Analiz statisticheskikh dannykh i vosstanovleniye skrytykh zakonornostey* [Analysis of statistical data and restoration of hidden patterns]. Moscow – Saint-Petersburg – Kyiv [in Russian].
2. Borovikov, V. (2003). *STATISTICA Iskustvo analiza dannykh na kompyutere* [Art of data analysis on computer]. Moscow –

Kiev – Minsk [in Russian].

3. *Maykrosoft. Neyronnye seti. STATISTICA Neural Network* [Microsoft. Neuronal nets. STATISTICA Neural Network]. Moscow [in Russian].
4. Lakin, G.F. (1990). *Biometriya* [Biometry]. Moscow: Vysshaya shkola [in Russian].

Отримано 22.04.2019

Електронна адреса для листування: kashuba@tdmu.edu.ua

M. O. KASHUBA

I. Horbachevsky Ternopil State Medical University

FUNDAMENTALS OF MEDICAL STATISTICS FOR BEGINNER SCIENTISTS

The aim of the work: to familiarize the beginner scientists with the possibilities of medical statistics. The article outlines the main, basic terms and concepts necessary for the proper planning of statistical research.

Key words: statistics; differences; reliability; correlation; general population; sampling; variables.

Н. А. КАШУБА

ГВУЗ "Тернопольский государственный медицинский университет имени И. Я. Горбачевского МОЗ Украины"

ОСНОВЫ МЕДИЦИНСКОЙ СТАТИСТИКИ ДЛЯ НАЧИНАЮЩИХ-УЧЕНЫХ

Цель работы: ознакомление начинающих-ученых с возможностями медицинской статистики.

В статье изложены основные, базовые термины и понятия, необходимые для правильного планирования статистических исследований.

Ключевые слова: статистика; различия; достоверность; корреляция; генеральная совокупность; выборка; переменные.