

Подання результатів математичної та статистичної обробки даних медичних і біологічних досліджень у дисертаційних роботах

Сучасна наука не може існувати без математики. Тому представлення математичних методів, використаних для обробки, а також математичних доказів вірогідності отриманих висновків має стати стандартом викладу будь-яких наукових результатів. Такі вимоги висуваються до публікацій редакціями усіх провідних наукових журналів як за кордоном, так і в Україні. Вимогам коректного, з точки зору математики, викладу наукових результатів повинні задовольняти і дисертаційні роботи.

Аналіз кваліфікаційних робіт, що розглядалися на експертній раді з профілактичної медицини за останні роки, свідчить про те, що типовим є ставлення здобувачів до математики лише як до необхідного, але формального атрибуту подання матеріалу.

Найчастіше при викладі математичних і статистичних методів використовуються кліше (меми), переписані з інших робіт (наприклад, «для обробки даних використані загальновідомі методи варіаційної статистики»), часто у методичній частині дисертацій та авторефератів посилаються на методи, які насправді в роботі не використовуються («системний аналіз», «контент-аналіз», «багатовимірні методи математичної статистики» тощо). Типовою помилкою є використання некоректних і неадекватних математичних прийомів (зокрема кореляційного аналізу для якісних даних) або, навпаки, незастосування методів, обов'язкових для подальших розрахунків (наприклад, перевірка розподілу вибірки на «нормальність» як умова для використання критерію Ст'юдента). Зустрічаються стійкі, але методично невірні поняття (приміром, при застосуванні кореляційного аналізу відзначають «міцний» або «слабкий» зв'язок, не оцінюючи його достовірність). У регресійних моделях не аналізується достовірність коефіцієнтів і спрямованість дії факторів. Використовуються невірні або застарілі терміни типу «кореляційно - регресійний» (або і того гірше - «регресивний») аналіз тощо.

Це пояснюється, поряд з іншими загальновідомими причинами, недостатнім рівнем методичного забезпечення щодо коректного подання результатів математичної обробки результатів дисертаційної роботи. Часто здобувач просто не знає, в якому вигляді йому необхідно описувати математичні та статистичні методи.

У даній публікації надана спроба в деякій мірі заповнити цю прогалину і сформулювати основні побажання (критерії, вимоги) до викладу використаних математичних прийомів. Сподіваємося, що ці пропозиції будуть корисними не лише для дисертантів за профілем експертної ради з профілактичної медицини, а й пошукувачам з інших наукових дисциплін.

Розглянемо деякі аспекти опису результатів математичного аналізу у розділах дисертаційної роботи та авторефераті.

Матеріали і методи. Опис математичних і статистичних методів, використаних автором для аналізу даних, є обов'язковою частиною розділу «Матеріали і методи», нарівні з іншими методами дослідження, такими як біохімічні, імунологічні тощо.

Використані математичні методи повинні описуватись конкретно для кожного із завдань, які ставилися в дослідженні, і, бажано, у тому ж порядку. Якщо ті самі методи використовувалися для розв'язання не одного, а декількох завдань, досить згадати їх один раз.

При описі методів не рекомендується заглиблюватися в деталізацію математичних процедур, особливо якщо вони загальновідомі. Наприклад, не слід наводити формули розрахунку середніх та їх помилок, коефіцієнту Ст'юдента, коефіцієнтів кореляції тощо.

У тому випадку, якщо існують декілька варіантів використання окремого методу або декілька формул для розрахунків, необхідно конкретизувати варіант, обраний автором, і дати цьому відповідне пояснення. Наприклад, чому використані непараметричні методи порівняння та які саме. Або які критерії використані при перевірці вибірки на нормальність.

Специфічні або оригінальні математичні методики повинні викладатися більш докладно, в обсязі, достатньому для їх розуміння читачами дисертаційної роботи.

Якщо математичні формули використано для розрахунків, що визначають хід (і обсяг) подальших досліджень, варто наводити не тільки відповідну формулу (чи посилання на неї), але і конкретні результати розрахунків. Наприклад, при визначенні обсягів вибірок, досліджуваних у подальшій роботі, необхідно навести не тільки обрану формулу, але і конкретне значення, розраховане за нею.

Якщо математичні методики чи формули є результатом роботи здобувача, їх необхідно наводити у розділі власних досліджень.

Методи і формули, використані автором у своїй роботі, повинні містити посилання на відповідні літературні джерела, з яких вони були запозичені.

Звісно, у цьому розділі варто наводити **тільки ті методи і прийоми, які дійсно використані автором** і результати застосування яких представлені в даній роботі.

Не слід використовувати незрозумілі кліше типу «відповідні статистичні методи», «статистика за Ст'юdentом», «традиційна статистична обробка», «за допомогою стандартних статистичних пакетів» без пояснення, у чому саме ці методи полягають.

У розділі необхідно вказати, які прикладні програмні засоби були використані при комп'ютерній обробці даних. У випадку використання нестандартних програмних засобів необхідно навести посилання на джерела літератури, у якій опубліковано детальний опис цих математичних методів і алгоритмів.

Не слід вказувати назву і технічні характеристики комп'ютерів, на яких виконувалася математична обробка даних.

Власні дослідження. При початковому описі первинного масиву даних необхідно наводити результати дескриптивної (описової) статистики. Для кількісних даних – результати перевірки на нормальність розподілу (значення критеріїв і достовірності висновку).

У випадку відповідності нормальному вигляду розподілу слід навести значення обсягу вибірки (кількості спостережень), середніх арифметичних і хоча б одного з показників варіабельності (наприклад, помилки середніх, середньоквадратичні відхилення, довірчі інтервали).

Для опису розподілів, що не є «нормальними», крім обсягу вибірки, необхідно вказувати значення медіани та інтерквартильний розмах. При традиційному використанні 25% і 75% центилів, тобто верхньої границі 1-го і нижньої границі 4-го квартилей, цей запис може мати, наприклад, такий вигляд: $Me(25\%; 75\%) = 125(53; 182)$.

Для згрупованих кількісних даних, а також рангових, бінарних і номінальних змінних зведену інформацію слід представляти у вигляді таблиць частот зустрічальності. У випадку подання даних у вигляді відносних величин (відсотків), крім самих їх значень варто наводити і показники варіабельності (помилки, довірчі інтервали).

Рекомендується при описовому наведенні даних максимально використовувати графічні методи: стовпчасті та кругові діаграми, «ящики з вусами» тощо.

Слід звертати увагу на розрядність наведених у роботі цифрових значень. Дробову частину чисел (число знаків після коми), отриманих після комп'ютерних розрахунків, необхідно округлювати до мінімальної кількості, достатньої для сприймання та подальших розрахунків. Наприклад, при обробці цілих чисел середнє арифметичне варто вказувати тільки з одним знаком після коми, а помилку середнього - із двома.

Всі отримані наукові результати і висновки дисертаційної роботи, якщо вони можуть бути виражені кількісно, необхідно статистично підтверджувати. При цьому **вибір статистичних методик повинний визначатися** не простотою реалізації чи традицією, а **винятково адекватністю підходів** (наприклад, використовувати кореляційний аналіз виключно для кількісних змінних, для рангових - таблиці спряженості, критерій Ст'юдента – тільки при нормальному розподілі тощо).

При викладі матеріалів дослідження необхідно вказувати всі статистичні характеристики, які підтверджують висновки автора. Наприклад, при оцінці розходжень між «нормально розподіленими» вибірками - значення коефіцієнту Ст'юдента (t) і похибки висновку (p).

Значення вірогідності (помилки статистичного висновку - p) варто наводити в тексті дисертації або в традиційній формі порівняння на трьох рівнях значимості (наприклад, $p > 0.05$ чи $p < 0.001$), або що доцільніше, наводити його точне значення ($p = 0.031$). При цьому наведення повинне бути однаковим впродовж всієї роботи. Нагадуємо, що p не може дорівнювати нулю. Мінімальне його значення потрібно подавати у вигляді $p < 0.001$.

Якщо для доведення результату дослідження використано декілька математичних підходів (критеріїв, формул), у роботі рекомендується наводити результати тільки одного, найефективнішого підходу. Наприклад, при використанні критерію Ст'юдента і критерію Фішера – наводити те значення, яке більш вірогідне.

При використанні кореляційного аналізу, крім розрахунку власне коефіцієнту кореляції, необхідно **обов'язково** розраховувати значення його помилки і вірогідності. Варто пам'ятати, що **про наявність кореляційного зв'язку (не залежності!) свідчить** не величина коефіцієнту кореляції, а **тільки його вірогідність**. Тому результати треба представляти у вигляді,

наприклад, $r=0.56$, $p=0.028$. Поняття «сильна (слабка) кореляція» у загальному випадку – некоректні та можуть використовуватися тільки для порівняльного співставлення достовірних результатів в межах однієї задачі.

Якщо розраховується парна кореляція між декількома перемінними, у тексті дисертації варто наводити значення тільки достовірних коефіцієнтів. При цьому ті зв'язки, які не відповідають апріорним припущенням здобувача, необхідно також наводити та пояснювати. При трактуванні результатів варто звертати увагу на спрямованість кореляції (негативна вона чи позитивна), про що свідчить знак коефіцієнту кореляції.

При використанні лінійного регресійного аналізу (як і в кореляційному аналізі) у роботі слід наводити не тільки отримані математичні моделі (значення коефіцієнтів регресії), але і помилки коефіцієнтів та їх вірогідності. Недостовірні коефіцієнти регресії повинні вилучатися з моделі. Рекомендується наводити характеристики адекватності моделі в цілому (коефіцієнт Фішера, p тощо).

Всі **основні висновки** дисертаційної роботи повинні, по можливості, містити цифровий матеріал і бути підтверджені статистично. Наприклад, «були отримані зміни ... на 34%, $p<0.05$ ».

Автореферат повинен відображати цифрові результати використання всіх математичних і статистичних методів, про застосування яких заявлено в методичному розділі. Наприклад, якщо вказується, що в роботі застосовувався кореляційний аналіз, необхідно навести його результати (значення коефіцієнтів кореляції та їх вірогідності).

Доцільно у математичних виразах для латинських та грецьких літер використовувати курсивний шрифт.

*А.М.Сердюк, голова експертної ради з
профілактичної медицини ВАК
України, академік НАМН України,*

*Члени експертної
ради: М.Ю. Антомонов, професор,
доктор біологічних наук, В.Г. Бардов,
член-кор. НАМН, професор, доктор
медицини наук, В.М.Лехан, професор,
доктор медичних
наук, О.С.Прилуцький професор,
доктор медичних наук.*