

УДК 612.7771.12014

## ОСНОВНІ СКЛАДОВІ ЕЛЕМЕНТИ СТАТИСТИЧНОГО АНАЛІЗУ ДИСЕРТАЦІЙНИХ ДОСЛІДЖЕНЬ

**Р.А. Абизов, С.С. Самойленко, Я.В. Шкоба**

*Національна медична академія післядипломної освіти імені П.Л. Шупика*

Для об'єктивізування та стандартизації наукових досліджень, а також з метою допомоги молодим науковцям-оториноларингологам у виборі математичних методів у даній роботі розглядаються непараметричні методи оцінки достовірності досліджень, які найчастіше використовуються в дисертаційних дослідженнях. Особлива увага приділена дослідженням з цензурованими (неповними) даними.

**Ключові слова:** статистичний аналіз, довірчі інтервали, похибка, показник виживання.

## ОСНОВНЫЕ СОСТАВЛЯЮЩИЕ ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДИССЕРТАЦИОННЫХ ИССЛЕДОВАНИЙ

**Р.А. Абызов, С.С. Самойленко, Я.В. Шкоба**

*Национальная медицинская академия последипломного образования имени П.Л. Шупика*

Для объективизации и стандартизации научных исследований, а также с целью помощи молодым ученым-оториноларингологам в выборе математических методов в данной работе рассматриваются наиболее используемые в диссертационных исследованиях непараметрические методы оценки достоверности исследований. Особое внимание уделено исследованиям с цензурированными (неполными) данными.

**Ключевые слова:** статистический анализ, доверительные интервалы, погрешность, показатель выживаемости.

## MAIN COMPONENT ELEMENTS OF STATISTICAL ANALYSIS APPLIED IN THESES

**R.A.Abyzov, S.S.Samoylenko, Ya.V.Shkoba**

*National Medical Academy of Postgraduate Education named P.L.Shupik*

The lecture presents the most commonly utilized nonparametric methods of statistical investigations with the view of objectivization and standardization of research and to aid defenders of theses and applicants in selecting appropriate techniques of numerous existing mathematical methods. Special attention is paid to investigations with censored (incomplete) data.

**Key words:** statistical analysis, confidence intervals, measure of inaccuracy, survival index.

**Вступ.** Статистичний аналіз досліджень дисертаційної роботи необхідно проводити відповідно до методологічних вимог до статистичних методів і підходів, прийнятих в 1998 році на Міжнародному конгресі з гармонізації статистики клінічних досліджень [9].

Всебічне вивчення картини дії досліджуваного виду лікування вимагає виділення порівнюваних параметрів. Ця задача вирішується, наприклад, якщо з усієї маси досліджень використовувати для статистичного аналізу лише спостереження за виходами, наприклад, отримано ефект від терапії - так чи ні; виявлено побічні ефекти - так чи ні тощо. Всі ці при-

клади зображують спосіб обліку реакції в альтернативній формі, тобто реакції, яка або настає, або ні.

**Основна частина.** Альтернативний розподіл - це розподіл елементів сукупності на 2 частини (2 альтернативи) за якою-небудь ознакою, частіше за якісною. Ці ознаки не пов'язані між собою ніякими арифметичними співвідношеннями. Єдиний спосіб опису якісних ознак - це підрахунок числа об'єктів, що мають одне і те саме значення, або частки від загального числа об'єктів, яка відповідає тому чи іншому значенню [5].

У випадку альтернативної класифікації неможливо ввести такі загальноприйняті кількісні параметри як

математичне очікування, дисперсія тощо. Але можна вказати певний числовий параметр, який має цілком точний і об'єктивний зміст: частка варіант одного з двох типів. Так **часткар** може бути оцінена як:

$$p = n_1 : (n_1 + n_2) = n_1 : n, \quad (1)$$

де  $n_1, n_2$  - чисельності альтернатив,  $n = n_1 + n_2$  - чисельність всієї сукупності. Окрім того, частка може бути визначена в відсотках:

$$p = n_1 : (n_1 + n) \times 100 \% = n_1 : n \times 100 \%. \quad (2)$$

У відношенні частки варіант в альтернативному розподілі виникають ті ж статистичні задачі, що і для параметрів, які представлені в кількісній формі:

1) оцінка частки  $p$  в генеральній сукупності по вибіркових даних, заходження довірчого інтервалу для  $p$ ;  
2) виявлення відмінності між генеральними частками  $p_1$  і  $p_2$  двох сукупностей за вибірковими даними, тобто порівняння двох вибіркових часток варіант.

Довірчі інтервали на практиці визначаються різними наближеними методами (точно ця задача вирішується з використанням біноміального розподілу) [3, 4, 6, 7]. Як правило, використовується нормальне наближення (заміна біноміального розподілу нормальним). З центральної межової теореми виходить, що при досить великому об'ємі вибірки вибіркова оцінка частки наближено підпорядковується нормальному закону розподілу, який має генеральне середнє і стандартне відхилення, яке дорівнює стандартній похибці частки  $p$ . Тоді **довірчі інтервали** для генеральної частки визначаються як:

$$p - z_\alpha \times \sigma_p < p < p + z_\alpha \times \sigma_p, \quad (3)$$

де  $\alpha$  - рівень значущості,  $z_\alpha$  - двостороннє критичне значення для стандартного нормального розподілу (визначається за значенням довірчої вірогідності),  $\sigma_p$  - **стандартна похибка частки**, яка задається співвідношенням:

$$\sigma_p \approx \sqrt{p \times (1 - p) : n} \quad (4)$$

**Оцінка різниці між частками (тестова статистика  $T$ )** є відношенням різниці вибіркових часток до стандартної похибки часток:

$$Z = (p_1 - p_2) : \sigma_{dp}$$

де  $p_1, p_2$  - порівнювані вибіркові частки,  $\sigma_{dp}$  - стандартна похибка різниці вибіркових часток. Величина  $Z$  має наближено нормальний розподіл.

Нульова гіпотеза в даному випадку полягає в тому, що  $p_1 = p_2$ . Її відкидають, якщо розрахована статистика  $Z$  перевищує або дорівнює табличному значенню, яке обране в відповідності з заданим рівнем значущості  $\alpha$ .

**Похибка різниці між частками**, взятими з приблизно рівновеликих вибірок (об'єм вибірок

відрізняється менш ніж на 25 %), вираховується за формулою:

$$\sigma_{dp} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2} = \sqrt{(p_1 \times (1 - p_1) : n_1) + (p_2 \times (1 - p_2) : n_2)}. \quad (6)$$

В випадку порівняння часток з нерівновеликих вибірок і для підвищення чутливості тесту похибку різниці часток визначають за формулою:

$$\sigma_{dp} = \sqrt{p_{cp} \times (1 - p_{cp}) \times (n_1 + n_2) : (n_1 \times n_2)}, \quad (7)$$

де  $p_{cp}$  - визначають як середньовиважену з  $p_1$  і  $p_2$  часток:

$$p_{cp} = (p_1 \times n_1 + p_2 \times n_2) : (n_1 + n_2). \quad (8)$$

Якщо частки визначають у відсотках від  $n$ , то в наведених формулах замість  $1 - p$  необхідно брати  $100 - p$ .

Описаний вище критерій перевірки тотожності часток в двох вибірках застосовується при не дуже великих і не дуже малих значеннях  $p$  ( $25 \% < p < 75 \%$ ).

Певна річ, що, якщо будь-який чинник спостерігався в 100 % випадків в групі, що складається з 10 варіант, такий результат може пояснюватися випадковим збігом і не відтворюватися в подальшому при продовженні дослідження. Той же чинник, що спостерігається в 100 % випадків в групі, яка складається з 1000 варіант, має, певно, більшу достовірність. Тому навіть в випадку появи в результаті дослідження 0 або 100 % часток ці результати повинні бути скореговані.

Для статистичної обробки наближених до межових результатів (також нульового або 100 % ефекту) необхідно застосовувати наступний підхід: розглядати **можливий скорегований відсоток ефекту** при подальшому збільшенні числа подібних спостережень і **його стандартну похибку** [5]:

$$p = (a+1) : (n+2) \times 100, \quad (9)$$

$$\sigma_{dp} = \sqrt{p \times (100 - p) : (n+3)}, \quad (10)$$

де  $a$  - отриманий в дослідженні узагальнений показник (%),  $p$  - скореговане значення цього показника (%).

**Довірчий інтервал для скорегованого значення частки** задається співвідношенням:

$$a - t \times \sigma_p < a < a + t \times \sigma_p, \quad (11)$$

де  $t$  - критичне значення двостороннього І критерію Стьюдента, визначається для обраного рівня довірчої вірогідності об'єму вибірки за таблицями [4-

б]. Якщо ж один з довірчих інтервалів виходить за крайню межу (0 або 100 %), то його корегують вручну, приймаючи крайнє значення, так як частка не може бути від'ємною або більшою за ціле.

Статистичні методи аналізу віддалених результатів лікування (тривалості життя) використовують поняття показника виживання і представлення даних в вигляді таблиць і кривих відносного виживання [1, 2, 5].

Показник виживання  $S(t)$  - це вірогідність прожити інтервал часу більше  $t$  з початку спостереження. А крива відносного виживання, що використовується для опису показника виживання, задає вірогідність пережити будь-який з моментів часу  $t$  після деякої початкової події. Цю вірогідність звичайно називають просто виживання.

При вивченні показника виживання, ситуація, коли дослідження повинно бути завершено до настання виходу (кінця) хвороби у всіх хворих, скоріше правило, ніж виняток. Звичайно хворих включають в дослідження на всьому його протязі. Тому воно завжди закінчується до смерті останнього хворого. Істинна тривалість хвороби пацієнтів, що вижили до закінчення дослідження, залишається невідомою. Окрім того, дослідник може втратити хворого з поля зору до завершення дослідження, якщо той, наприклад, переїхав в інше місто. Хворий також, може померти від чинників, не пов'язаних з захворюванням, що вивчається, наприклад загинути в автокатастрофі. В усіх цих випадках тривалість захворювання залишається невідомою, ми знаємо, що вона перевищує певний термін часу.

Аналіз показника виживання необхідно проводити за умови, що всі дослідження виживання відповідають наступним вимогам: 1) вибір хворих проведено випадково; 2) для всіх хворих відомо час початку спостереження; 3) для всіх хворих відомо час закінчення спостереження, а також - помер він чи вибув (хворі, які померли від чинників не пов'язаних з захворюванням, що вивчається, а також хворі, які з якихось причин вибули з дослідження, вважаються тими, що вибули (тобто дані цензуровані<sup>1</sup> [8, 10]). В усіх цих випадках тривалість захворювання залишається невідомою, ми знаємо лише, що вона перевищує деякий термін часу).

Необхідно відзначити, що підхід і математичний апарат, які відносяться до аналізу показника виживання, можуть успішно застосовуватися для аналізу

інших показників ефекту і тривалості різноманітних процесів, наприклад, можна вивчати термін лікування якого-небудь захворювання (вихід - ремісія), тривалість лікування безплідності або ефективність контрацепції (вихід в обох випадках - настання вагітності), довголіття протеза (вихід - поломка) тощо.

На типовій кривій в початковий момент показник виживання дорівнює 1, а потім крива поступово знижується і наближається до 0. В випадку, коли померли не всі хворі, що спостерігалися, лінія нуля не досягає. Хоча найповніша характеристика показника виживання - це крива відносного виживання, однак існують і узагальнені показники, що характеризують виживання у вигляді одного числа. Розподіл за тривалістю життя, як правило, асиметричний, тому найкраще тут підходить медіана. Для вибірки **медіана виживання** визначається як **найменший час, для якого показник виживання менше 0,5**, тобто час, до якого доживає половина сукупності. Для визначення медіани виживання необхідно побудувати криву відносного виживання і подивитись, де вона вперше знижується нижче 0,5.

Як і у випадку вибіркового середнього, в даній задачі існує поняття кривої відносного виживання для сукупності та її вибіркової оцінки за результатами дослідження. Якби не можливе вибуття в процесі досліджень, вибіркова оцінка показника виживання  $S(t)$  визначалась би як **відношення числа хворих, що пережили момент  $t$ , до об'єму вибірки  $n$** .

Для обліку вибуття при побудові таблиць виживання використовують **моментальний метод Каплана-Мейєра**.

Математичний вираз моментального методу:

$$S(t) = \prod (1 - (d_{t_i} : n_{t_i})), \quad (12)$$

Де  $t$  - час,  $t_i$  - певний момент часу,  $S(t)$  - виживання хворих на момент  $t$ ,  $d_{t_i}$  - число померлих в момент  $t_i$ ,  $n_{t_i}$  - число хворих, що спостерігались на момент  $t_i$ ,  $(1 - (d_{t_i} : n_{t_i}))$  частка хворих, що пережили момент  $t_i$ ,  $\prod$  (велика грецька буква "пі") - символ добутку, який означає, що необхідно перемножувати значення  $(1 - (d_{t_i} : n_{t_i}))$  для всіх моментів часу, коли виникла хоча б одна смерть, за період від 0 до  $t$ .

Для кожного моменту часу, коли виникла хоча б одна смерть, оцінюється вірогідність *пережити цей момент*. Такою оцінкою буде відношення числа хво-

*Примітка 1.* В дослідженнях показника виживання неповні дані називають цензурованими. Дані, коли відомо момент початку спостереження, але невідомо, коли хворий помер, називаються цензурованими з правого боку. Якби в дослідженні брали участь хворі, в яких невідомо час початку спостереження, то ми могли б отримати дані, цензуровані з лівого боку, а також дані, цензуровані з обох боків.

рих, що пережили цей момент, до числа хворих, що спостерігались до даного моменту. Тоді, згідно з правилом множення вірогідностей, вірогідність пережити деякий момент часу для кожного хворого, що вступив в дослідження, буде дорівнювати добутку цих оцінок від нульового до даного моменту.

Такий підхід дозволяє навіть в випадку вибування пацієнта з дослідження використовувати для статистичного аналізу раніше зібрану інформацію про нього.

Отримані результати розрахунків представляються в вигляді таблиці, рядки якої відповідають моментам часу, в які відбувалась хоча б одна смерть, а також в вигляді графіка. Точки на графіку також відповідають моментам, коли помер хоча б один з хворих, що спостерігались. Ці точки з'єднуються лінією. Отриманий графік і буде вибірковою оцінкою кривої виживання. Окрім того, побудовану криву можна охарактеризувати і узагальненим показником, наприклад медіаною. Для цього потрібно знайти точку, в якій крива виживання вперше знижується нижче 0,5. Необхідно зазначити, що якщо число померлих менше половини числа хворих, що спостерігаються, медіану визначити неможливо. При цьому узагальненим показником може бути будь-який інший перцентиль (менше 50 %). За таблицями виживання більш точно, ніж при використанні загальноприйнятого розрахунку середнього значення, може бути визначений і показник, що називається **середньою тривалістю життя**.

Як завжди при дослідженні вибірки, вибіркова крива виживання являє собою оцінку кривої виживання для сукупності. При цьому для кожної точки на кривій можна за формулою Грінвуда [8] визначити оцінку точності наближення або **стандартну похибку показника виживання**:

$$\sigma_{S(t)} = S(t) \times \sqrt{\sum (d_{t_i} : (n_{t_i} \times (n_{t_i} - d_{t_i})))}, \quad (12)$$

де ( $\Sigma$  ("сигма") - знак суми) сума береться по всіх моментах часу  $t_i$  з таблиці виживання від 0 до і включно.

**Довірчі інтервали для виживання** в момент і можна будувати подібно будованню такого інтервалу для вибіркової частки. Такий довірчий інтервал для кожного моменту і задається співвідношенням:

$$S(t) - z_{\alpha} \times \sigma_{S(t)} < S(t) < S(t) + z_{\alpha} \times \sigma_{S(t)}, \quad (14)$$

де  $\alpha$  - рівень значущості,  $z_{\alpha}$  - двостороннє критичне значення для стандартного нормального розподілу (в випадку 95 % інтервалу  $z_{\alpha} = 1,960$ ).

Відклавши на графіку довірчі інтервали ми бачимо "рукав", що розширюється - довірчу область для

показника виживання. Причина розширення довірчої області зрозуміла: чим менше залишається хворих, що спостерігаються, тим більша похибка. Наведена вище формула дає симетричну оцінку довірчих інтервалів, яка може вийти за граничні значення 1 або 0. Найпростіший спосіб підправити таку похибку - це замінити значення, що більші за одиницю, на одиницю, а значення менші нуля - на нуль.

В клінічних дослідженнях (особливо в онкології [2]) часто виникає необхідність порівнювати виживання різних груп або, іншими словами, порівнювати криві відносного виживання. Якщо, скажімо, в порівнюваних групах отримано однакове виживання, але при цьому в одній з них більша частина пацієнтів помирає в більш ранні строки, порівняно з другою, ці відмінності будуть наочно продемонстровані при побудові кривих відносного виживання і приховані при оцінці ефекту прямим методом.

Існують методи множинного порівняння кривих відносного виживання, але вони потребують великої кількості складних обчислень [2, 11]. Деяко простішою є математична задача, коли необхідно порівняти дві групи. Нульова гіпотеза  $H_0$  будується на тому, що в обох групах виживання однакове. Якби не було вибування і всі хворі спостерігались рівний час, нам підійшов би аналіз таблиць сполученості. Якби всі хворі спостерігались впритул до смерті, можна було б порівняти виживання в обох групах за допомогою рангового критерію Манна-Вітні або метода Крускала-Волліса. В реальному житті такі ситуації рідкісні, тому що вибуття практично неминуче. Серед непараметричних методів для порівняння кривих відносного виживання, побудованих моментальним методом, найбільш відомий **лонгранговий критерій**, який вживається, коли число спостережень в кожній групі не менше 10.

Лонгранговий критерій базується на наступних допущеннях: 1) дві порівнювані вибірки незалежні і випадкові; 2) вибування в обох вибірках однакове; 3) функції показника виживання пов'язані тотожністю:

$$S_2(t) = [S_1(t)]^{\Psi}. \quad (15)$$

Перевірити, чи виконується співвідношення, можна намалювавши графіки  $\ln[-\ln S_1(t)]$  і  $\ln[-\ln S_2(t)]$ . Якщо вони паралельні, умова виконується. Перетин свідчить про порушення даної умови. Величина  $\Psi$  ("пси") називається відношенням смертності. Якщо  $\Psi = 1$ , то криві відносного виживання співпадають. Якщо  $\Psi < 1$ , то люди в другій вибірці помирають пізніше, ніж в першій. І навпаки, якщо  $\Psi > 1$ , пізніше помирають в першій вибірці.

Для перевірки лонгрангового критерію для кожного моменту часу в таблиці виживання розраховується очікуване число померлих по обох таблицях сумісно (подібно аналізу таблиць сполученості). При цьому визначаємо можливий процент смертності по двох групах разом при умові вірності нульової гіпотези. Далі розраховуємо очікуване число померлих в кожній групі при умові відсутності відмінностей між групами. Критерій перевіряється за ступенем близькості реальних значень в таблиці виживання до очікуваних, оцінених при умові дотримання нульової гіпотези.

**Очікуване число померлих** в першій групі в кожний момент часу  $t$ , коли виникла хоча б одна смерть, хоча б в одній групі (обрахунки проводяться для однієї з груп і не має значення для якої):

$$E_{it} = n_{1t} \times d_{обт} : n_{обт} \quad (16)$$

де  $E_{it}$  - очікуване число померлих в першій групі момент часу  $t$ ,  $n_{1t}$  - число хворих, що спостерігались в першій групі до моменту  $t$ ,  $d_{обт}$  - загальне число смертей в момент  $t$  в обох групах,  $n_{обт}$  - загальне число хворих, що спостерігались на момент  $t$ . В даній формулі хворі, що вибули, враховуються побічно - впливаючи на число хворих, що спостерігаються.

**Сума різниць** спостережуваного і очікуваного числа померлих береться за всіма моментами часу, коли наступала хоча б одна смерть в якій-небудь групі, в підсумку визначається **статистика  $U_L$** :

$$U_L = \sum (d_{it} - E_{it}), \quad (17)$$

де  $d_{it}$  - число померлих в час  $t$ ,  $(d_{it} - E_{it})$  - доданок для  $U_L$ . Вона наближено підлягає нормаль-

ному закону розподілу з **стандартною похибкою**:

$$\sigma U_L = \sqrt{\sum (n_{1t} \times n_{2t} \times d_{обт} \times (n_{обт} - d_{обт}) : (n_{обт}^2 \times (n_{обт} - 1)))} \quad (18)$$

**Статистика  $Z$**  для перевірки критерію виходить при діленні значення  $U_L$  на його стандартну похибку:

$$z = U_L : \sigma U_L. \quad (19)$$

Ця статистика розподілена наближено до нормального закону, тому отримане значення порівнюється з критичним значенням для стандартного розподілу.

В зв'язку з тим, що криві відносного виживання є дискретними величинами, а дискретний розподіл лише наближено описується нормальним, який по суті своїй безперервний, це призводить до надлишкового пом'якшення критерію: ми дещо частіше, ніж належить, відкидаємо нульову гіпотезу. Для того, щоб скорегувати статистику, в випадку лонгрангового критерію необхідно застосувати **поправку Йейтса**:

$$z = (|U_L| - 0,5) : \sigma U_L. \quad (20)$$

**Висновок.** Метою даної роботи було ознайомлення з основними аспектами статистичного аналізу, що використовуються при обробці медичної інформації, зокрема в оториноларингології. В зв'язку з численними та різноманітними методами статистичних досліджень перед науковцем виникають певні труднощі у виборі доказових критеріїв. Сконцентрований у даній роботі комплекс методів статистичного аналізу, на нашу думку допоможе об'єктивізувати та стандартизувати дисертаційні дослідження.

## Література

1. Гланц С. Медико-биологическая статистика / С. Гланц - М.: Практика, 1999. - 459 с.
2. Двойрин В.В. Методика контролируемых клинических испытаний / В.В. Двойрин, А.А. Клименков - М.: Медицина, 1985.
3. Лапач С.Н. Статистические методы в медико-биологических исследованиях с использованием Excel / С.Н. Лапач, А.В. Чубенко, П.Н. Бабич - К.: Морион, 2000. - 320 с.
4. Мюллер П. Таблицы по математической статистике / П. Мюллер, П. Нойман, Р. Шторм - М.: Финансы и статистика, 1982.
5. Сергиенко В. И. Математическая статистика в клинических исследованиях / В.И. Сергиенко, И.Б. Бондарева - М.: Гэотар Медицина, 2000. - 256 с.
6. Справочник по теории вероятностей и математической статистике; под ред. В.С. Корольюка и др. - М.: Наука, 1985.
7. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях - М., 1976.
8. Collett D. Modeling survival in medical research - London: Chapman and Hall, 1994.
9. ICH Guidelines // Good Clin. Pract. J. - 1998. - Vol. 5, № 4. - P 27-37.
10. Lee E.T. Statistical methods for survival data analysis - New York: Wiley, 2<sup>nd</sup> ed., 1992.
11. Peto R., Pike M.C., Armitage P. Et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient // Br. J. Cancer. - 1976. - Vol. 34. - P. 585-612.