



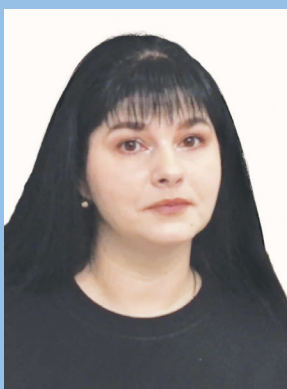
Володимир Володимирович Бабієнко, академік НАН ВО України, доктор медичних наук, професор, завідувач кафедри гігієни та медичної екології Одеського національного медичного університету. Заслужений діяч науки і техніки України. Лікар вищої кваліфікаційної категорії, фахівець у галузі профілактичної медицини.

Автор понад 180 наукових публікацій, монографій, співавтор підручника для вищих учбових закладів.



Андрій Вікторович Мокієнко, доктор медичних наук, старший науковий співробітник, старший викладач кафедри гігієни та медичної екології Одеського національного медичного університету. Сфера наукових інтересів включає, зокрема, гігієнічні та медико-екологічні аспекти водопостачання населення, в тому числі аналіз і дослідження проблем знезараження води.

Автор понад 600 наукових праць, серед них 16 монографій, 4 фрагментів монографій та учбового посібника.



Левковська Вікторія Юрїївна, кандидат медичних наук, доцент кафедри гігієни та медичної екології Одеського національного медичного університету. Магістр державного управління. Фахівець у галузі профілактичної медицини. Сфера наукових інтересів, зокрема, вивчення рівня еколого-гігієнічної безпеки морського середовища в зоні рекреаційного водокористування.

Автор понад 50 наукових публікацій.

БІОСТАТИСТИКА

В. В. Бабієнко, А. В. Мокієнко, В. Ю. Левковська

В. В. Бабієнко, А. В. Мокієнко,

В. Ю. Левковська

БІОСТАТИСТИКА





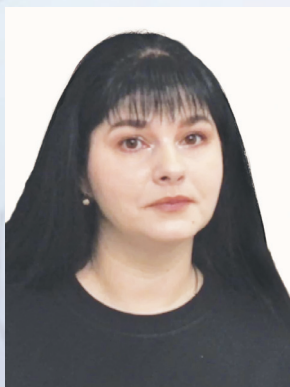
Володимир Володимирович Бабієнко, академік НАН ВО України, доктор медичних наук, професор, завідувач кафедри гігієни та медичної екології Одеського національного медичного університету. Заслужений діяч науки і техніки України. Лікар вищої кваліфікаційної категорії, фахівець у галузі профілактичної медицини.

Автор понад 180 наукових публікацій, монографій, співавтор підручника для вищих учбових закладів.



Андрій Вікторович Мокієнко, доктор медичних наук, старший науковий співробітник, старший викладач кафедри гігієни та медичної екології Одеського національного медичного університету. Сфера наукових інтересів включає, зокрема, гігієнічні та медико-екологічні аспекти водопостачання населення, в тому числі аналіз і дослідження проблем знезараження води.

Автор понад 600 наукових праць, серед них 16 монографій, 4 фрагментів монографій та учбового посібника.



Левковська Вікторія Юріївна, кандидат медичних наук, доцент кафедри гігієни та медичної екології Одеського національного медичного університету. Магістр державного управління. Фахівець у галузі профілактичної медицини. Сфера наукових інтересів, зокрема, вивчення рівня еколого-гігієнічної безпеки морського середовища в зоні рекреаційного водокористування.

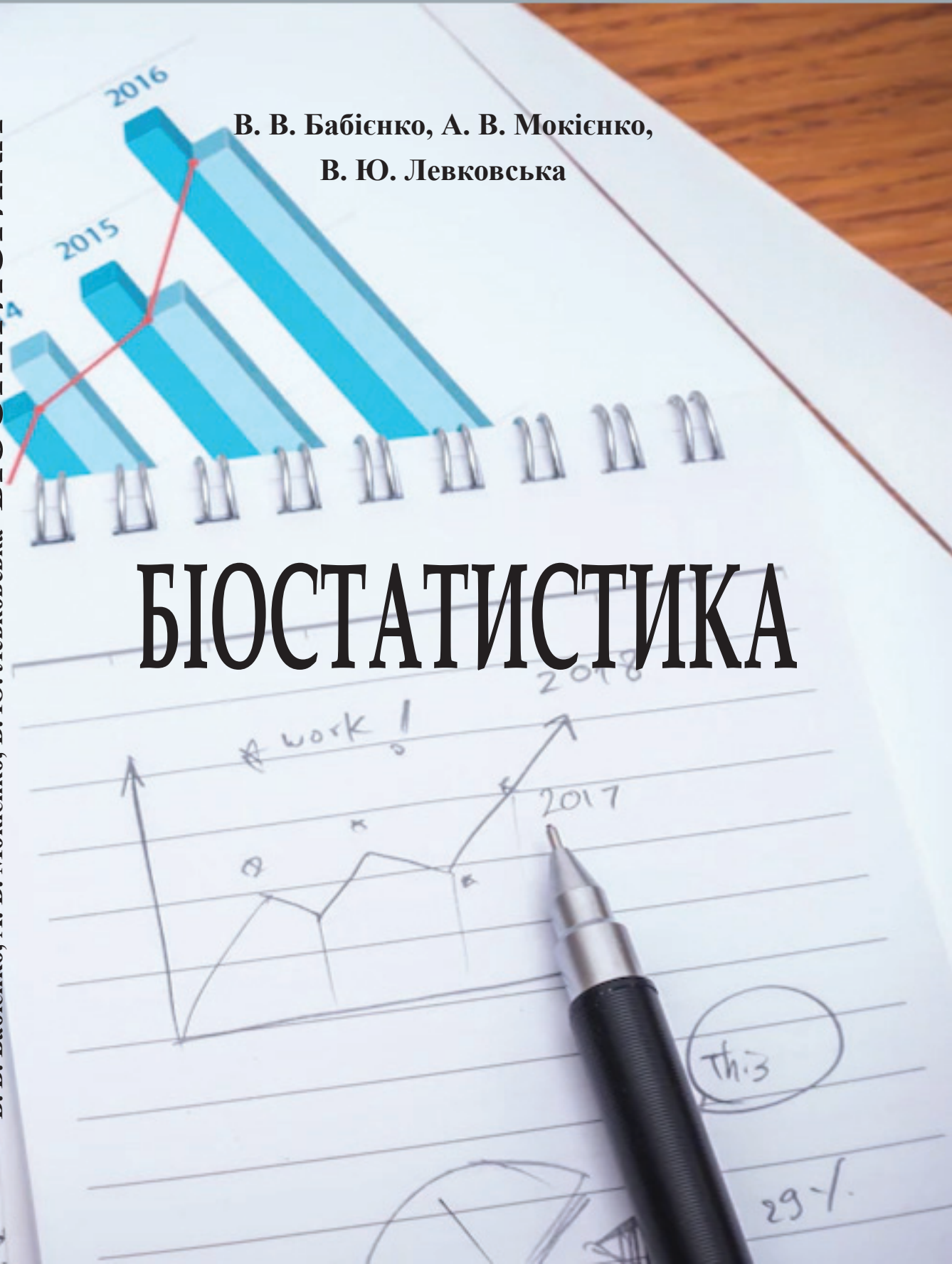
Автор понад 50 наукових публікацій.

БІОСТАТИСТИКА

В. В. Бабієнко, А. В. Мокієнко, В. Ю. Левковська

В. В. Бабієнко, А. В. Мокієнко,
В. Ю. Левковська

БІОСТАТИСТИКА



**В. В. Бабієнко, А. В. Мокієнко,
В. Ю. Левковська**

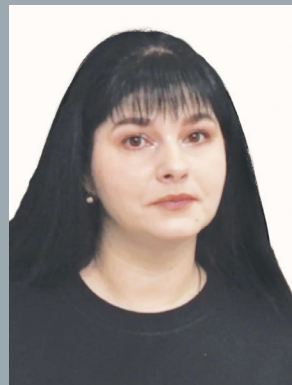
БІОСТАТИСТИКА



Володимир Володимирович Бабієнко,
академік НАН ВО України, доктор медичних наук, професор, завідувач кафедри гігієни та медичної екології Одеського національного медичного університету. Заслужений діяч науки і техніки України. Лікар вищої кваліфікаційної категорії, фахівець у галузі профілактичної медицини.
Автор понад 180 наукових публікацій, монографій, співавтор підручника для вищих учбових закладів.



Андрій Вікторович Мокієнко,
доктор медичних наук, старший науковий співробітник, старший викладач кафедри гігієни та медичної екології Одеського національного медичного університету. Сфера наукових інтересів включає, зокрема, гігієнічні та медико-екологічні аспекти водопостачання населення, в тому числі аналіз і дослідження проблем знезараження води.
Автор понад 600 наукових праць, серед них 16 монографій, 4 фрагментів монографій та учбового посібника.



Левковська Вікторія Юрївна,
кандидат медичних наук, доцент кафедри гігієни та медичної екології Одеського національного медичного університету. Магістр державного управління. Фахівець у галузі профілактичної медицини. Сфера наукових інтересів, зокрема, вивчення рівня еколого-гігієнічної безпеки морського середовища в зоні рекреаційного водокористування.
Автор понад 50 наукових публікацій.

В. В. Бабієнко, А. В. Мокієнко, В. Ю. Левковська
БІОСТАТИСТИКА





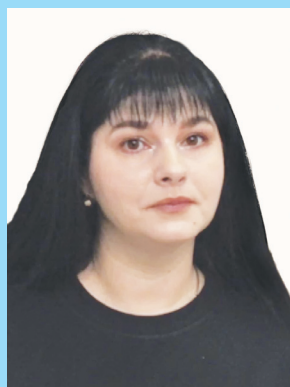
Володимир Володимирович Бабієнко,
академік НАН ВО України, доктор медичних наук, професор, завідувач кафедри гігієни та медичної екології Одеського національного медичного університету. Заслужений діяч науки і техніки України. Лікар вищої кваліфікаційної категорії, фахівець у галузі профілактичної медицини.

Автор понад 180 наукових публікацій, монографій, співавтор підручника для вищих учбових закладів.



Андрій Вікторович Мокієнко,
доктор медичних наук, старший науковий співробітник, старший викладач кафедри гігієни та медичної екології Одеського національного медичного університету. Сфера наукових інтересів включає, зокрема, гігієнічні та медико-екологічні аспекти водопостачання населення, в тому числі аналіз і дослідження проблем знезараження води.

Автор понад 600 наукових праць, серед них 16 монографій, 4 фрагментів монографій та учбового посібника.



Левковська Вікторія Юріївна,
кандидат медичних наук, доцент кафедри гігієни та медичної екології Одеського національного медичного університету. Магістр державного управління. Фахівець у галузі профілактичної медицини. Сфера наукових інтересів, зокрема, вивчення рівня еколого-гігієнічної безпеки морського середовища в зоні рекреаційного водокористування.

Автор понад 50 наукових публікацій.

БІОСТАТИСТИКА

В. В. Бабієнко, А. В. Мокієнко, В. Ю. Левковська

В. В. Бабієнко, А. В. Мокієнко,
В. Ю. Левковська

БІОСТАТИСТИКА



В. В. Бабієнко, А. В. Мокієнко,
В. Ю. Левковська

БІОСТАТИСТИКА

*Допущено Вченою радою Одеського національного
медичного університету як навчально-методичний
посібник для студентів вищих навчальних
медичних закладів II-IV рівнів акредитації*

Одеса
«Прес-кур'єр»
2022

УДК 614.1(072)
Б125

*Рекомендовано до друку Вченою радою
Одеського національного медичного університету
Міністерства охорони здоров'я України
(протокол № 5 від 28.12.2021 р.)*

Рецензенти:

Г. С. Сатурська - завідувачка кафедри громадського здоров'я та управління охороною здоров'я Тернопільського національного медичного університету імені І. Я. Горбачевського МОЗ України, доктор медичних наук, професор;

В. А. Огнєв - завідувач кафедри громадського здоров'я та управління охороною здоров'я Харківського національного медичного університету МОЗ України, доктор медичних наук, професор.

В. В. Бабієнко, А.В. Мокієнко, В. Ю. Левковська

Б125 Біостатистика : навчально-методичний посібник /
Бабієнко В. В., Мокієнко А. В., Левковська В. Ю. -
Одеса : Прес-кур'єр, 2022. 180 с.
ISBN 978-617-7797-34-9

Навчально-методичний посібник присвячено узагальненню даних літератури щодо методичних аспектів викладання практичних занять з біостатистики для студентів вищих навчальних медичних закладів. Тематика відповідає діючій програмі з предмету. Конспективно викладено теоретичний матеріал щодо відносних величин та графічних методів аналізу; динамічних рядів і їх аналізу; характеристики і аналізу статистичних даних; методу стандартизації; параметричних і непараметричних методів оцінки і аналізу статистичних гіпотез; аналізу взаємозв'язку між параметрами статистичних сукупностей; методичних основ організації статистичних досліджень у системі охорони здоров'я; чинників ризику і методик розрахунку і оцінки ризиків. В кожній темі докладно представлено навчальні завдання і методики їх виконання, рішення типових завдань, контрольні питання.

Для студентів вищих навчальних медичних закладів II-IV рівнів акредитації.

ISBN 978-617-7797-34-9

© Бабієнко В. В., Мокієнко А. В.,
Левковська В. Ю., 2022

Зміст

<i>Вступ</i>	5
Тема № 1 <i>Відносні величини (статистичні коефіцієнти).</i> <i>Графічні методи аналізу.</i>	9
Тема № 2 <i>Динамічні ряди і їх аналіз.</i>	30
Тема № 3 <i>Характеристика і аналіз статистичних даних.</i> <i>Середні величини і показники варіації.</i>	39
Тема № 4 <i>Метод стандартизації</i>	59
Тема №5 <i>Параметричні методи оцінки і аналізу</i> <i>статистичних гіпотез</i>	72
Тема №6 <i>Непараметричні методи оцінки і аналізу</i> <i>статистичних гіпотез</i>	85
Тема № 7 <i>Аналіз взаємозв'язку між параметрами</i> <i>статистичних сукупностей</i>	106

Тема № 8

Методичні основи організації статистичних досліджень у системі охорони здоров'я. 124

Тема № 9

Чинники ризику. Методика розрахунку і оцінки ризиків 157

Література..... 178

Вступ

Статистика як наука має давню і багату історію. Однак, статистичні методи в медицині стали використовуватися відносно недавно — з 30-х років ХХ ст. При описі і аналізі клінічних явищ панував кількісно статистичний підхід.

Біостатистика зіграла ключову роль в переході медичних досліджень від опису окремих спостережень і серій випадків до проведення експериментальних робіт із застосуванням контрольних груп і масштабних рандомізованих контрольованих випробувань, які стали новим стандартом якості наукових досліджень.

У середині ХХ ст. глибоке знання принципів біостатистики було доступним далеко не для усіх учених-медиків. Кількість клінічних досліджень збільшувалася, використання статистичних методів швидко поширювалося.

За кордоном почали з'являтися публікації, які оцінювали якість статистичного аналізу в медичних статтях. Частота виявлених статистичних погрешностей коливалась від 50 до 80 %. Переважна частина цих погрешностей була пов'язана з нерозумінням найпростіших принципів, викладених навіть в елементарних посібниках із статистики, наслідком чого були недостовірні висновки.

Чому ж статистика настільки складна для розуміння лікарями? Тому, що розуміння статистичних принципів вимагає розвиненого абстрактного мислення, медицина ж — справа практична. Лікарям незвично мислити такими категоріями, як популяція, вірогідність, розподіл, нульова гіпотеза. В той же час вони вірять у всемогутність ста-

тистичних методів без розуміння їх ролі і особливостей застосування.

За останні десятиліття спостерігається швидкий розвиток доказової медицини — напряму, який перевіряє ефективність медичних втручань в клінічній практиці. Лікарі все частіше звертаються до подібних досліджень для того, щоб прийняти правильне рішення і з'ясувати, чим нові методики відрізняються від старих. Проте оцінити якість сучасних досліджень неможливо без знання методів обробки інформації, які ще недавно більшості лікарів були невідомі. Ці методи включають статистичний аналіз, на результати якого дослідники спираються у своїх висновках.

На жаль, іноді методи статистичного аналізу недостатньо обгрунтовані і часто неправильно використовуються. Одна з поширених помилок полягає в тому, що статистичні методи дозволяють отримати кількісний показник, який сам по собі, без урахування природи досліджуваного явища і результатів попередніх досліджень, упереджує помилкові висновки. Цей обман заважає правильним висновкам, ускладнюючи розуміння зв'язку між доказовістю результатів окремого дослідження і переконливістю інших доказів (дані клінічних і експериментальних досліджень, накопичений практичний досвід). Ось чому, результати багатьох досліджень не витримують перевірки часом.

Саме методи прикладної статистики є найкращим інструментом для отримання науково обгрунтованих достовірних результатів дослідження, іноді за умов невизначеності багатьох характеристик і при невеликій кількості спостережень.

Біостатистика — наука прикладна, усі статистичні методи ґрунтуються на складному математичному апараті, але її принципи неможливо розглядати окремо від методології медичних досліджень. Адже статистика — це тільки

частина дослідження, яке варто розглядати в контексті усіх проблем, які визначають якість дослідження. Формулювання основної мети дослідження, вибір відповідного методу і способу організації дослідження, особливості відбору хворих, і характер отриманих даних, — усе це визначає вибір адекватного методу статистичного аналізу і впливає на достовірність отриманих результатів. Формування сучасної методології досліджень відбувається паралельно з удосконаленням способів кількісного аналізу статистичних даних. Біостатистика, використовуючи статистичну інформацію і статистичні методи, вивчає питання охорони здоров'я і соціальні проблеми, розробляє спеціальні рекомендації.

Тема 1

Відносні величини (статистичні коефіцієнти). Графічні методи аналізу

Об'єктом статистичного дослідження є статистична сукупність — група, яка складається з великого числа відносно однорідних елементів (одиниць спостереження), узятих разом у відомих межах часу і простору. Тобто статистична сукупність — це загальне число одиниць спостереження, узятих для дослідження. Чисельність одиниць спостереження в сукупності визначає обсяг дослідження і позначається буквою «*n*». Об'єктом спостереження є сукупність осіб або явищ, певні контингенти осіб, певне явище, предмети.

Одиниця спостереження — первинний елемент статистичної сукупності, відібраний відповідно до мети і завдань дослідження. За одиницю спостереження може служити окрема особа або окреме явище.

Кожній одиниці спостереження притаманні певні характеристики (стать, вік, діагноз тощо). У дослідженні враховуються (реєструються) тільки ті з них, що необхідні для досягнення мети і вирішення завдань дослідження. Ці ознаки називаються обліковими.

Облікові ознаки поділяють на атрибутивні (описові), виражені словами та кількісні (виражені числом). Кожна величина кількісної ознаки називається варіантою і позначається буквою «*V*».

При вивченні впливу окремих ознак явища, що вивчається, облікові ознаки, що враховуються, поділяють на факторні і результативні.

Факторними називаються такі ознаки, під впливом яких змінюються інші, залежні від них (результативні) ознаки.

Сукупності, що використовуються як об'єкт статистичного дослідження, можуть бути генеральними або вибірковими.

Генеральна сукупність складається з усіх одиниць спостереження, які можуть бути до неї віднесені відповідно до мети дослідження, тобто вона включає всі одиниці спостереження.

Вибіркова сукупність — частина генеральної сукупності, відібрана певним спеціальним методом і призначена для характеристики генеральної сукупності. На основі аналізу вибіркової сукупності можна отримати достатньо повне уявлення про закономірності, притаманні всій генеральній сукупності.

Основними вимогами до вибіркової сукупності є:

а) репрезентативність (здатність представляти) — у вибірці повинні бути представлені пропорційно всі елементи генеральної сукупності; це досягається за допомогою використання спеціальних методів, що забезпечують об'єктивність відбору;

б) достатній обсяг (кількість спостережень).

Статистика має в своєму розпорядженні спеціальні формули або ж готові таблиці, за якими можна визначити необхідну кількість спостережень у вибірковій сукупності.

Теоретичне обґрунтування вибіркового методу дає математична теорія вірогідності і закон великих чисел.

Статистична сукупність, на відміну від окремих одиниць спостереження (індивідуумів), має особливі, тільки для неї характерні властивості, до яких відносять:

а) характер розподілу явища, що вивчається;

б) його середній рівень, який дає узагальнюючу характеристику явищу, що вивчається;

в) різноманітність (варіабельність, коливання, мінливість) одиниць спостереження, складових сукупності;

г) репрезентативність вибіркової сукупності по відношенню до генеральної;

д) взаємозв'язок (кореляція) між ознаками, що вивчаються.

За допомогою спеціальних статистичних методів виявляють ці властивості і отримують узагальнені характеристики.

Метою вивчення будь-якої статистичної сукупності є виявлення загальних властивостей, загальних закономірностей різних явищ, оскільки властивості не можуть бути виявлені при аналізі одиничних явищ.

Для оцінки явищ, що вивчаються і є складовими статистичної сукупності, використовують статистичні величини — абсолютні, відносні і середні.

Абсолютні величини несуть важливу інформацію про розмір того чи іншого явища (наприклад, кількість хворих, кількість новонароджених, кількість ліжок в стаціонарі тощо), проте вони мають обмежене пізнавальне значення, оскільки у більшості випадків вони неспівставні внаслідок відмінностей у чисельності груп, для яких були отримані показники.

Відносні величини застосовують, головним чином, для характеристики розподілу ознак у сукупності, а також для порівняння в ході аналізу різних сукупностей.

Відносні величини розраховуються шляхом ділення однієї величини на іншу та подальшого множення отриманого дробу на 100 (1000, 10000 і т.д.). Відповідно цьому величини можуть бути виражені у відсотках (%), проміле (‰) або продецимілі (‱). При цьому підбір того чи іншого множника пов'язаний з тим, що відносні величини доцільно подавати в цілих числах, які легко сприймаються при аналізі.

Розрізняють наступні види відносних величин:

а) інтенсивні;

- б) екстенсивні;
- в) показники співвідношення;
- г) показники наочності;
- д) показники відносної інтенсивності.

Вихідною інформацією для розрахунку відносних величин є абсолютні величини, що характеризують середовище, для якого розраховується відповідний показник, а також величини, які характеризують абсолютний розмір явища, що вивчається.

Середовище — це населення або його окрема група, сформована за певною ознакою або за комплексом ознак (наприклад, населення окремої території, робітники певної галузі промисловості, діти певного віку, хворі, що знаходяться на стаціонарному лікуванні з приводу певного захворювання в певному ЛПЗ тощо).

Явище, що вивчається, визначається метою дослідження. Якщо метою дослідження є вивчення захворюваності, то явищем, яке вивчається буде захворюваність. Прикладом таких явищ в практиці біостатистичних досліджень є народжуваність, смертність, інвалідність, результати лікування. Відповідно, абсолютним розміром явищ в даних прикладах будуть кількість випадків захворювань, кількість випадків народження, смерті, інвалідності, певних результатів захворювання.

Інтенсивний показник або показник частоти поширеності вказує на частоту явища в середовищі, яке безпосередньо «продукує» дане явище (наприклад, рівні народжуваності, смертності, захворюваності). Інтенсивні показники застосовуються в наступних випадках:

- а) для визначення рівня, частоти, поширеності того чи іншого явища;
- б) для порівняння ряду різних сукупностей за ступенем частоти того чи іншого явища;

в) для виявлення в динаміці змін у частоті явища в спостережуваній сукупності.

Екстенсивний показник, або показник структури розподілу, вказує на відношення частини до цілого, на долю частини в цілому. Дає уявлення про кількісний розподіл складових (структурних) частин в будь-якій одній сукупності (наприклад, структура захворюваності, смертності, інвалідності).

Показник співвідношення, як і інтенсивний показник, характеризує рівень, частоту, поширеність явища, яке вивчається Але, на відміну від інтенсивного показника, в цьому випадку середовище і явище не пов'язані між собою, тобто середовище не «продукує» явище, а визначає відношення між різнорідними сукупностями. До показників співвідношення належать, наприклад, рівень забезпеченості населення ліжками, лікарями.

Показник наочності — це відношення кожної з величин, що зіставляються, до початкового рівня, прийнятого за 1 або за 100 %. У показниках наочності можна представити абсолютні величини, інтенсивні показники, показники співвідношення, а також середні величини.

Показники відносної інтенсивності являють собою чисельне співвідношення двох структур. Вони використовуються при вивченні структурних особливостей різних статистичних сукупностей, які мають відношення до одного середовища.

Можливими помилками у використанні відносних величин є:

а) використання екстенсивних показників замість інтенсивних;

б) помилковий вибір середовища при розрахунку групових інтенсивних показників;

в) недооблік чинника часу при порівнянні отриманих показників,

г) зіставлення показників, які мають різні одиниці вимірювання (% , ‰).

Для наочного зображення статистичних величин (абсолютних, середніх, відносних) і можливості глибшого їх аналізу використовують графічні зображення. Вид графічного зображення має чітко відповідати змісту показника, який зображується.

Розрізняють наступні основні види графічних зображень:

а) діаграми (лінійні, радіальні, стовпчикові, фігурні, внутрішньо-стовпчикові, секторні);

б) картограми;

в) картодіаграми.

Лінійна і радіальна діаграми застосовуються для зображення динаміки того чи іншого явища або процесу, вираженого у вигляді абсолютних середніх величин, показників інтенсивності, наочності. При цьому радіальна діаграма використовується для зображення динаміки явища в замкнутому (що повторюється) циклі часу (доба, тиждень, рік тощо). Прикладами лінійної діаграми можуть бути температурна крива, динаміка народжуваності, смертності тощо. За допомогою радіальної діаграми можуть бути зображені сезонність захворюваності і смертності населення, динаміка кількості звернень по швидку медичну допомогу протягом доби та інші подібні показники.

Стовпчикова і фігурна діаграми можуть бути використані для графічного зображення тих же статистичних величин, що і лінійні, проте явище, яке вивчається, може бути зображене не тільки в динаміці, але і в статистиці, наприклад, рівні захворюваності, смертності, забезпеченості населення лікарями в різних містах, країнах тощо. Стовпчики на діаграмі можуть бути вертикальними і горизонтальними. У останньому випадку діаграма є різновидом стовпчикової діаграми

і називається стрічковою. Фігурна діаграма використовується за тим же призначенням, що і стовпчикова, відрізняючись від неї лише більшою наочністю, тобто замість стовпчиків використовуються фігури, які символізують явище, що вивчається (наприклад, чисельність населення, забезпеченість ліжками тощо).

Внутрішньо-стовпчикова і секторна діаграми застосовуються для зображення екстенсивних показників, тобто структури явища, наприклад, структура захворюваності, структура причин смерті населення.

Картограма — це зображення статистичних величин на географічній карті. Можуть наноситися будь-які статистичні величини (абсолютні, середні, відносні). Ділянки територій, які відрізняються значенням зображуваних показників, забарвлюються певним кольором або штрихуванням відповідної інтенсивності

Картодіаграма — це зображення на географічній карті різного роду діаграм. Діаграми (найчастіше стовпчикові) наносяться на карту і ставляться на тій території, яку вони представляють.

Методика розрахунку, аналізу і графічного зображення відносних величин

Розрахунок відносних величин здійснюється за формулами:

а) інтенсивний показник

$$\frac{\text{Абс.розмір явища} \times 100 (1000, 10\ 000, 100\ 1000)}{\text{Абс. розмір середовища (що показує дане явище)}}$$

б) екстенсивний показник

$$\frac{\text{Абс. розмір частини явища} \times 100}{\text{Абс. розмір явища в цілому}}$$

в) показник співвідношення

$$\frac{\text{Абс. розмір явища} \times 100 (1000, 10\ 000, 100\ 000)}{\text{Абс. розмір середовища (яке не продукує дане явище)}}$$

г) показник наочності

$$\frac{\text{Абс. розмір явища} \times 100}{\text{Абс. розмір також явища (за характером) з ряду порівнюваних прийнятий за 100\%}}$$

д) показник відносної інтенсивності

$$\frac{\text{Структура явища}}{\text{Структура середовища}}$$

Методика розрахунку і аналізу відносних величин приводиться на прикладі типового завдання.

Типове завдання

На основі даних табл. 1 обчислити показники: а) інтенсивні; б) екстенсивні; у) співвідношення; г) наочності; д) коефіцієнти відносної інтенсивності.

Таблиця 1

Чисельність населення різного віку в районі діяльності поліклініки, чисельність лікарів, звернень і відвідувань (у абсолютних числах)

Вік (роки життя)	Чисельність населення	Кількість звернень	Кількість відвідувань	Чисельність лікарів
15-19	5000	3000	-	-
20-59	25000	30000	-	-
60-69	10000	21000	-	-
70 і більше	10000	6000	-	-
Разом	50000	60000	500000	50

Показники наочності (у відсотках по відношенню до інтенсивного показника звернень осіб віком 15-19 років, прийнятого за 100 %).

Таблиця 2

Показники	Вік (роки життя)			
	15-19	20-59	60-69	70 і більше
Інтенсивний (на 1000 населення)	600	1200	2100	600
Наочності (у % до віку 15-19 років)	100	200	350	100

Зразок виконання завдання.

1. Інтенсивні показники

Загальний рівень відвідування (на 1000 осіб)

$$\frac{\text{Кількість відвідувань} \times 1000}{\text{Чисельність населення}} = \frac{500\,000 \times 1000}{50\,000} = 10\,000\text{‰}$$

Загальний рівень звернень (на 1000 осіб)

$$\frac{\text{Кількість звернень} \times 1000}{\text{Чисельність населення}} = \frac{60\,000 \times 1000}{50\,000} = 1200\text{‰}$$

Рівень звернень осіб 15 – 19 років

$$\frac{\text{Кількість звернень осіб 15–19 років} \times 1000}{\text{Чисельність населення 15–19 років}} = \frac{3000 \times 1000}{5000} = 600\text{‰}$$

Рівень звернень осіб 20 – 59 років

$$\frac{\text{Кількість звернень осіб 20–59 років} \times 1000}{\text{Чисельність населення 20–59 років}} = \frac{30\,000 \times 1000}{25\,000} = 1200\text{‰}$$

Рівень звернень осіб 60 – 69 років

$$\frac{\text{Кількість звернень осіб 60–69 років} \times 1000}{\text{Чисельність населення 60–69 років}} = \frac{21\,000 \times 1000}{10\,000} = 2100\text{‰}$$

Рівень звернень осіб 70 і більше років

$$\frac{\text{Кількість звернень осіб 70 і більше років} \times 1000}{\text{Чисельність населення}} = \frac{6000 \times 1000}{10\,000} = 600\text{‰}$$

2. Екстенсивні показники

Склад населення за віком:

Частка осіб у складі 15-19 років у складі населення

$$\frac{\text{Кількість осіб віком 15–19 років} \times 100}{\text{Чисельність населення}} = \frac{5000 \times 100}{50\,000} = 10\%$$

Частка осіб у 20-59 років у складі населення

$$\frac{\text{Кількість осіб віком 20–59 років} \times 100}{\text{Чисельність населення}} = \frac{25\,000 \times 100}{50\,000} = 50\%$$

Частка осіб 60-69 років у складі населення

$$\frac{\text{Кількість осіб віком 60–69 років} \times 100}{\text{Чисельність населення}} = \frac{10\,000 \times 100}{50\,000} = 20\%$$

Частка осіб 70 і більше років у складі населення

$$\frac{\text{Кількість осіб віком 70 років і більше} \times 100}{\text{Чисельність населення}} = \frac{10\,000 \times 100}{50\,000} = 20\%$$

Структура звернень за віком:

Частка звернень осіб віком 15-19 років серед всіх звернень

$$\frac{\text{Кількість звернень осіб віком 15–19 років} \times 100}{\text{Кількість всіх звернень}} = \frac{3000 \times 100}{60\,000} = 5\%$$

Частка звернень осіб віком 20-59 років серед всіх звернень

$$\frac{\text{Кількість звернень осіб віком 20–59 років} \times 100}{\text{Кількість всіх звернень}} = \frac{30\,000 \times 100}{60\,000} = 50\%$$

Частка звернень осіб віком 60-69 років серед всіх звернень

$$\frac{\text{Кількість звернень осіб віком 60–69 років} \times 100}{\text{Кількість всіх звернень}} = \frac{21\,000 \times 100}{60\,000} = 35\%$$

Частка звернень осіб віком 70 років і більше серед всіх звернень

$$\frac{\text{Кількість звернень осіб 70 і більше років} \times 100}{\text{Кількість всіх звернень}} = \frac{6000 \times 100}{60\,000} = 10\%$$

3. Показник співвідношення

Число лікарів на 10 000 населення

$$\frac{\text{Чисельність лікарів} \times 10\,000}{\text{Чисельність населення}} = \frac{50 \times 10\,000}{50\,000} = 10$$

4. Показник відносної інтенсивності

Коефіцієнт відносної частоти звернень у віковій групі
15-19 років –

$$\frac{\text{Частка звернень осіб 15–19 років серед всіх звернень}}{\text{Частка осіб віком 15–19 років у складі населення}} = \frac{5\%}{10\%} = 0,5$$

Коефіцієнт відносної частоти звернень у віковій групі
20-59 років –

$$\frac{\text{Частка звернень осіб 20–59 років серед всіх звернень}}{\text{Частка осіб віком 20–59 років у складі населення}} = \frac{50\%}{50\%} = 1,0$$

Коефіцієнт відносної частоти звернень у віковій групі
60-69 років –

$$\frac{\text{Частка звернень осіб 60–69 років серед всіх звернень}}{\text{Частка осіб віком 60–69 років у складі населення}} = \frac{35\%}{20\%} = 1,75$$

Зводимо отримані показники в таблицю. Аналізуючи і оцінюючи деякі показники діяльності поліклініки (табл. 3), можна зробити такі висновки:

Рівень відвідувань поліклініки склав 10000 ‰, що відповідає середньому по країні рівню відвідувань населенням міст.

Рівень звернень до поліклініки склав 1200%, що також відповідає середньому по країні рівню звернень міського населення.

Таблиця 3

Деякі показники діяльності поліклініки залежно від віку населення, яке обслуговується

Вік (роки життя)		Структура звернень за віком у %	Число звернень на 1000 осіб певного віку	Показники наочності (за числом осіб певного віку)	Показники відносної інтенсивності
15-19	10,0	5,0	600	100	0,5
20-59	50,0	50,0	1200	200	1,0
60-69	20,0	35,0	2100	350	1,75
70 і більш	20,0	10,0	600	100	0,5
Разом	100,0	100,0	1200	200	1,0
Відвідувань на 1000 осіб			10000		
Чисельність лікарів на 10000 населення			10		

На рівні середнього показника по всій групі населення (1200‰) знаходиться частота звернень осіб у віці 20-59 років.

Найбільш високий рівень звернень (2100 ‰) спостерігається у віковій групі населення 60-69 років. Показник наочності в цій групі складає 350% по відношенню до вікової групи 15-19 років.

Найбільш низький показник звернень (600‰) відзначається в групах осіб віком 15- 19 та 70 років і більше.

У структурі звернень населення за віком найбільший відсоток (50%) складають відвідування середньої вікової групи 20-59 років, оскільки вона найчисленніша (25000 осіб), і найнижчий відсоток складають відвідування наймолодшої

вікової групи — 15- 19 років у зв'язку з тим, що ця група найменш численна (5000 осіб).

Коефіцієнти відносної інтенсивності свідчать про те, що частка звернень осіб вікової групи 60-69 років у 1,75 раз вища за питому вагу осіб цієї групи у складі населення. У той же час у вікових групах 15-19 років та 70 років і старше питома вага звернень удвічі нижча, ніж питома вага цих груп у складі населення. Таке становище підтверджує висновки про вікові особливості звернень, зроблені на основі аналізу вікових інтенсивних показників.

Коефіцієнт відносної частоти звернень у віковій групі 70 і більше років

$$\frac{\text{Частка звернень осіб 70 і більше років серед всіх звернень}}{\text{Частка осіб віком 70 і більше років у складі населення}} = \frac{10\%}{20\%} = 0,5$$

Графічне зображення статистичних показників здійснюється за умови дотримання наступних правил:

- а) графічне зображення повинно бути адекватним показнику, що зображується;
- б) кожне графічне зображення повинно мати назву (зазвичай ставиться під зображенням);
- в) повинно будуватись за певним масштабом;
- г) для кожного графічного зображення повинні даватися пояснення (у вигляді умовних позначень) про застосований колір або штрихування.

Основою для побудови лінійної діаграми найчастіше є прямокутна система координат. На осі абсцис (x) відкладаються, наприклад, рівні за масштабом проміжки часу, а на осі ординат (y) — показники чисельності населення, захворюваності, смертності тощо. Значення статистичних величин наносять у вигляді точок на систему координат і сполучають їх лініями. При побудові лінійної діаграми необхідно враховувати пропорції в масштабі між величиною осі

абсцис (x) і ординат (y). При відношенні $x : y = 4 : 3$ не буде спостерігатися спотворень кривої. У тих випадках, коли на одній діаграмі зображують декілька явищ, лінії наносять різного кольору або різним штрихуванням. Не рекомендується на одній діаграмі використовувати більше 4 ліній.

При побудові радіальної діаграми за вісь абсцис використовують коло, яке ділять на однакову кількість частин відповідно відрізка часу того чи іншого циклу. Віссю ординат служить радіус кола або його продовження. За радіус кола зазвичай беруть середню величину явища, яке вивчається. Кількість радіусів відповідає кількості інтервалів періоду, що вивчається: 12 радіусів — при вивченні явища за рік, 7 — при вивченні явища за тиждень тощо. На кожному радіусі робиться позначка, відповідна інтервалу часу, і відкладаються у відповідному масштабі показники. Кінцеві точки сполучають — отримують багатокутник, який наочно показує динаміку явища.

При побудові стовпчикової діаграми малюють стовпчики, висота яких повинна відповідати величині показників, які зображуються, з урахуванням масштабу. Ширина стовпчиків, а також відстань між ними повинні бути однаковими та довільними. Нижче за вісь абсцис, на якій розташований стовпчик, прийнято вказувати назву статистичної сукупності, для якої обчислено представлений показник (стать, вік, місто, область тощо), а під стовпчиком — значення показників. При горизонтальному (стрічковому) розташуванні стовпчикової діаграми використовується аналогічна методика її побудови.

Побудова фігурної діаграми здійснюється за тими ж правилами, що і стовпчикової з тією лише різницею, що замість стовпчиків для зображення статистичних величин використовуються різні фігури. Наприклад, зростання числа ліжок зображують у вигляді схематичних лікарняних ліжок,

зростання чисельності населення — у вигляді людських фігур тощо.

При побудові секторної діаграми все коло приймається за 100% (якщо екстенсивні показники виражені у відсотках), при цьому 1% відповідає $3,6^\circ$ кола. Потім $3,6^\circ$ перемножують на кількість відсотків кожного показника, отримуючи тим самим розмір кожного сектора у градусах. За допомогою транспортира на колі відкладають відрізки (відлік ведеться від 0°), відповідні величині кожного показника. Знайдені точки кола сполучають з центром кола. Окремі сектори кола (у відсотках або в проміле) зображують складові частини явища, що вивчається. Кожний сектор позначається певним кольором або штрихуванням. У середині сектора (або зовні круга, якщо сектор маленький) наводиться значення екстенсивного показника. Замість секторної діаграми можна застосовувати внутрішньо-стовпчикову, в якій ширина і висота стовпчика беруться довільно. Висота береться за 100%, у відповідному масштабі перераховуються екстенсивні показники (у %), які в сумі дають єдине ціле.

Навчальне завдання

Користуючись методикою розрахунку, аналізу і графічного зображення відносних величин, початковими даними таблиць свого варіанту, розрахувати, проаналізувати і зобразити графічно відомі вам відносні величини, дотримуючись правил побудови діаграм.

Варіант 1

Таблиця

**Чисельність населення різних вікових категорій
в районі діяльності поліклініки, чисельність лікарів,
кількість звернень в поліклініку і її відвідування
(в абсолютних числах)**

Вік (роки життя)	Чисельність населення	Кількість звернень	Кількість відвідувань	Чисельність лікарів
15-19	8000	4000	-	-
20-59	40000	48000	-	-
60-69	12000	12000	-	-
70 і більше	20000	16000	-	-
Разом	80000	80000	720000	88

Варіант 2

Таблиця

**Чисельність населення різних вікових категорій
в районі діяльності поліклініки, чисельність лікарів,
кількість звернень в поліклініку і її відвідування
(в абсолютних числах)**

Вік (роки життя)	Чисельність населення	Кількість звернень	Кількість відвідувань	Чисельність лікарів
15-19	6000	4800	-	-
20-59	24000	24000	-	-
60-69	5000	9600	-	-
70 і більше	5000	9600	-	-
Разом	40000	48000	400000	48

**Чисельність осіб, які пройшли комплексні
профілактичні огляди, та виявлених хворих
гіпертонічною хворобою в різних групах населення
(в абсолютних числах)**

Групи населення	Кількість осіб, які пройшли комплексні профілактичні огляди	Кількість виявлених хворих
Працівники промислових підприємств	40000	1600
Працівники харчових і комунальних установ	20000	200
Працівники дитячих і лікувально-профілактичних установ	10000	100
Учні шкіл, технікумів, ВНЗ	30000	100
Разом	100000	2000

Примітка. У лікувально-профілактичних установах міста працює 250 лікарів і 700 середніх медичних працівників. Чисельність населення — 200000 осіб.

Чисельність осіб, які пройшли комплексні профілактичні огляди та виявлених хворих шлунково-кишковими захворюваннями в різних групах населення (у абсолютних числах)

Групи населення	Кількість осіб, які пройшли комплексні профілактичні огляди	Кількість виявлених хворих
Працівники промислових підприємств	40000	200
Працівники харчових і комунальних установ	20000	150
Працівники дитячих і лікувально-профілактичних установ	10000	50
Учні шкіл, технікумів, ВНЗ	30000	600
Разом	100000	1000

Примітка. У лікувально-профілактичних установах міста працює 250 лікарів і 700 середніх медичних працівників. Чисельність населення — 200000 осіб.

Чисельність осіб, які пройшли комплексні профілактичні огляди та виявлених хворих ішемічною хворобою серця в різних групах населення (у абсолютних числах)

Групи населення	Кількість осіб, які пройшли комплексні профілактичні огляди	Кількість виявлених хворих
Працівники промислових підприємств	40000	300
Працівники харчових і комунальних установ	20000	100
Працівники дитячих і лікувально-профілактичних установ	10000	100
Учні шкіл, технікумів, ВНЗ	30000	-
Разом	100000	500

Примітка. У лікувально-профілактичних установах міста працює 250 лікарів і 700 середніх медичних працівників. Чисельність населення — 200000 осіб.

Контрольні запитання

- Що таке об'єкт статистичного дослідження та одиниця спостереження?
- Як розрізняються облікові ознаки за характером і роллю в сукупності?

- Які є види статистичних сукупностей і чим вони відрізняються?
- Для чого вивчають статистичну сукупність?
- На чому засновано виявлення загальних закономірностей різних явищ?
- Як виявити розподіл кількісних і якісних ознак в статистичній сукупності?
- Які величини використовують для характеристики статистичної сукупності?
- Для чого застосовуються відносні величини? Назвіть види відносних величин.
- Що характеризує показник інтенсивності?
- Які початкові дані необхідні для розрахунку інтенсивного показника?
- Що характеризує показник екстенсивності?
- Які початкові дані необхідні для розрахунку екстенсивного показника?
- Як обчислити екстенсивний та інтенсивний показники?
- Охарактеризуйте сутність показника співвідношення.
- Які початкові дані необхідні для розрахунку показників співвідношення?
- Як обчислити показник співвідношення?
- Що характеризує показник наочності?
- Які показники і величини можуть бути використані при розрахунку показників наочності?
- Як обчислити показник наочності?
- Для чого застосовують графічні зображення?
- Як вибирають вид графічного зображення?
- Коли застосовують лінійні, радіальні, стовпчикові, фігурні, секторні, внутрішньо-стовпчикові діаграми, і як вони будуються?
- Коли застосовують картограми, картодіаграми?

- Назвіть вимоги до вибіркової сукупності.
- Які види відносних величин можна використовувати для порівняння явищ у динаміці, по територіях, установах?
 - Що є показником відносної інтенсивності і для чого він використовується?
 - Які можливі помилки при застосуванні відносних величин?
 - Які вимоги пред'являються до графічних зображень у статистиці?
 - Які види графічного зображення можуть бути використані для зображення інтенсивних показників?
 - Які види графічного зображення можуть бути використані для зображення екстенсивних показників?
 - Які дані необхідні для визначення частоти хвороб системи кровообігу у населення віком 60 років і більше?
 - Які дані необхідні для визначення частки дифтерії серед всіх інфекційних хвороб у дітей віком до 7 років?
 - Які дані необхідно мати для визначення забезпеченості району лікарями окремих спеціальностей?
 - У порівнянні з минулим роком рівень захворюваності грипом підвищився на 30%. Який показник був використаний для ілюстрації?
 - Які види графічних зображень можна використовувати для зображення динаміки показника загальної смертності?
 - Які графічні зображення можуть бути використані для ілюстрації рівня народжуваності в різних регіонах?
 - Які графічні зображення застосовуються при ілюстрації даних про розподіл захворювань за класами хвороб?

Тема 2

Ряди динаміки і їх аналіз

Динамічним рядом називається ряд однорідних статистичних величин, які показують зміну якогось явища в часі. Динамічні ряди використовуються для аналізу динаміки будь-яких явищ і для оцінки тенденцій їх розвитку.

Числа, що складають динамічний ряд, називаються рівнями ряду. Вони можуть бути абсолютними, відносними і середніми величинами.

Динамічний ряд, що складається з абсолютних величин, називається простим. Динамічний ряд, що складається з відносних або середніх величин, називається складним або похідним.

Прості динамічні ряди бувають моментні та інтервальні.

Моментний динамічний ряд складається з чисел, які характеризують ознаку на певні моменти — дати (на кінець року, місяця, декади тощо) Рівні моментного ряду не можуть дробитись.

Інтервальний динамічний ряд характеризує ознака, яка вивчається, за певний інтервал часу (добу, тиждень, місяць, рік). Інтервальний динамічний ряд можна розділити на більш дрібні періоди або, навпаки, збільшити інтервали. Вибір величини інтервалу певною мірою визначається ступенем мінливості явища: чим повільніше змінюється явище в часі, тим більшими можуть бути періоди спостереження.

Прості ряди (як моментні, так і інтервальні) є початковими даними для побудови складних (похідних) рядів.

Динамічний ряд, як правило, дозволяє прослідкувати основну закономірність динаміки явища — його зниження або зростання, якщо рівні ряду змінюються послідовно. Іноді

ця послідовність порушується, рівні динамічного ряду значно коливаються, що утруднює виявлення основної закономірності. Існують певні прийоми вирівнювання динамічного ряду: збільшення інтервалу, обчислення групової середньої, обчислення ковзної середньої, метод найменших квадратів. Проте вирівнювання динамічного ряду необхідно здійснювати тільки після глибокого та всебічного аналізу причин, що зумовили коливання його рівнів. Механічне вирівнювання може штучно згладити рівні і завуалювати причинно-наслідкові зв'язки.

Укрупнення інтервалу проводять шляхом підсумовування даних за ряд суміжних періодів. Якщо ряд складається з величин якогось явища за місяць, наприклад, числа захворювань за місяцями, то можна підсумовувати числа за три місяці і дати числа захворювань поквартально. Тоді закономірність може виявлятися більш значущою.

Обчислення групової середньої для кожного збільшення інтервалу проводять шляхом підсумування суміжних рівнів сусідніх періодів з подальшим діленням отриманої суми на число доданків.

Обчислення ковзної середньої дозволяє кожний рівень замінити на середню величину з даного рівня та двох суміжних з ним.

Сенс вирівнювання динамічного ряду методом найменших квадратів полягає у визначенні теоретичних значень ряду, квадрати відхилень фактичних значень яких є найменшими з можливих. Щоб зробити вирівнювання слід перш за все виконати графічне зображення ряду, і за його виглядом підібрати відповідну математичну модель. Найчастіше зустрічаються наступні види графіків: параболи 1-го, 2-го, 3-го порядку, гіпербола, логістична функція.

Приклад вирівнювання динамічного ряду чисел згідно параболи першого ряду наводиться в табл. 1.

**Вирівнювання динамічного ряду чисел
згідно параболи першого порядку**

Роки	Зайнятість ліжка (у днях), Y	t	tY	t ²	Y _T	Y _T = a+bt
2002	340,1	-3	-1020,3	9	341,4	a =
2003	340,9	-2	-681,8	4	341,9	$\frac{\Sigma Y}{n} = \frac{23865}{7} = 340,9$
2004	338,0	-1	-338,0	1	342,4	
2005	341,2	0	0	0	342,9	b =
2006	343,0	1	343	1	343,4	
2007	339,1	2	678,2	4	343,9	
2008	344,2	3	1032,6	9	344,4	$\frac{\Sigma tY}{\Sigma t^2} = \frac{13,7}{28} = 0,5$
Σ (сума)	2386,5		13,7	28		
2009					344,9	
2010					345,4	

Y_T — теоретична лінія;

T — точка часу (серединне число ряду приймається за 0, якщо ряд непарний, за 1, якщо ряд парний);

a — розмір початкового теоретичного рівня,

b — розмір приросту в часі теоретичної лінії або кут її нахилу (тенденція).

Таким чином, з наведеного прикладу видно, що зайнятість ліжка має тенденцію до зростання, що дозволило скласти прогноз на 2009 і 2010 роки.

Основними показниками, якими можна охарактеризувати динамічний ряд, є:

а) абсолютний приріст (спад) — різниця між наступним і попереднім рівнями;

б) темп приросту (спаду) — відсоткове відношення абсолютного приросту (спаду) до попереднього рівня;

в) значення 1 % приросту (спаду) — відношення абсолютного приросту (спаду) до темпу приросту (спаду);

г) темп зростання (спаду) — відсоткове відношення наступного рівня до попереднього.

Аналіз динамічного ряду включає:

а) графічне зображення;

б) обчислення і оцінку основних показників.

Методика виконання навчального завдання

1. Основні показники, що характеризують динамічний ряд, обчислюються за формулами:

$$a = \Pi_2 - \Pi_1$$

$$T_n = \frac{a \times 100}{\Pi_1}$$

$$\Pi\% = \frac{a}{T_n}$$

$$T_{зр.} = \frac{\Pi_2 \times 100}{\Pi_1}$$

де a — абсолютний приріст (спад), Π_1 — попередній рівень; Π_2 — наступний рівень; T_n — темп приросту (спаду); $\Pi\%$ — значення 1 % приросту (спаду); $T_{зр.}$ — темп зростання (спаду).

2. Темп приросту : $T_n = \frac{a \times 100}{\Pi_1}$

2004 – 2005 рр.

$$0,3 \times 100 / 14,5 = 2,07\%$$

2005 – 2006 рр.

$$-1,2 \times 100 / 14,8 = -8,11\%$$

2006 – 2007 рр.

$$2,5 \times 100 / 13,6 = 18,38\%$$

2007 – 2008 рр.

$$0,5 \times 100 / 16,1 = 3,11\%$$

2008 – 2009 рр.

$$0,3 \times 100 / 16,6 = 1,81\%$$

2004 – 2009 рр.

$$2,4 \times 100 / 14,5 = 16,55\%$$

3. Значення 1% приросту : $\Pi\% = \frac{a}{T_{\Pi}}$

2004 – 2005 рр.

$$0,3 / 2,07 = 0,14$$

2005 – 2006 рр.

$$-1,2 / -8,11 = 0,15$$

2006 – 2007 рр.

$$2,5 / 18,38 = 0,14$$

2007 – 2008 рр.

$$0,5 / 3,11 = 0,16$$

2008 – 2009 рр.

$$0,3 / 1,81 = 0,17$$

2004 – 2009 рр.

$$2,4 / 16,55 = 0,15$$

4. Темп зростання : $T_{Зр.} = \frac{\Pi_2 \times 100}{\Pi_1}$

Типове завдання

Середня тривалість перебування хворих у стаціонарі міської лікарні (у днях) складала:

2004 — 14,5

2005 — 14,8

2006 — 13,6

2007 — 16,1

2008 — 16,6

2009 — 16,9

Зразок виконання завдання:

Абсолютний приріст дорівнює різниці між наступним і попереднім рівнями: $a = \Pi_2 - \Pi_1$

2004-2005 р. $a=14,8-14,5=0,3$

2005-2006 р. $a=13,6-14,8=-1,2$

2006 -2007 р. $a=16,1-13,6=2,5$

2007-2008 р. $a=16,6-16,1=0,5$
 2008-2009 р. $a=16,9-16,6=0,3$;
 за 6 років $a=16,9-14,5=2,4$
 Темп зростання, %:
 2004-2005 рр. $14,8 \cdot 100 / 14,5 = 102,07\%$;
 2005-2006 рр. $13,6 \cdot 100 / 14,8 = 91,81\%$;
 2006-2007 рр. $16,1 \cdot 100 / 13,6 = 118,38\%$;
 2007-2008 рр. $16,6 \cdot 100 / 16,1 = 103,11\%$;
 2008-2009 рр. $16,9 \cdot 100 / 16,6 = 101,81\%$;
 2004-2009 рр. $16,9 \cdot 100 / 14,5 = 116,55\%$.

Таблиця 2

**Динаміка середньої тривалості перебування хворих
в стаціонарі міської лікарні (кількість днів)**

Показники	2004	2005	2006	2007	2008	2009	За 6 років
Середня тривалість перебування хворих в стаціонарі (кількість днів)	14,5	14,8	13,6	16,1	16,6	16,9	15,42
Абсолютний приріст (кількість днів)	-	0,3	-1,2	2,5	0,5	0,3	2,4
Темп приросту (%).	-	2,07	-8 11	18,38	3,11	1,81	16,55
Значення 1% приросту (кількість днів)	-	0,14	0,15	0,14	0,16	0,17	0,15
Темп зростання, %	-	102,07	91,89	118,38	103,11	101,81	116,55

Аналізуючи динаміку середньої тривалості перебування хворих в стаціонарі міської лікарні за період з 2004 по 2009 рр., можна зробити наступні висновки:

- За 6 років середня тривалість перебування хворих в стаціонарі міської лікарні збільшилася більш, ніж на 16% або на 2,4 дні і склала в 2009 році 16,9 днів.

- Темп зростання і темп приросту мали істотні коливання по роках. Найбільш низькі значення цих показників відмічені в 2006 р., найбільш високі — в 2007 р.

- Разом з цим значення 1% приросту середньої тривалості перебування хворих в стаціонарі мали відносно стабільний рівень з тенденцією до деякого підвищення показника.

Навчальні завдання

На основі наведених у варіантах вихідних даних обчислити показники динамічного ряду:

- а) абсолютний приріст;
- б) темп приросту;
- в) темп зростання;
- г) значення 1% приросту.

Проаналізувати результати і зробити висновки

Варіант 1

Рівень летальності в стаціонарі міської лікарні (%)

2004 р.	0,6
2005 р.	0,5
2006 р.	0,7
2007 р.	0,75
2008 р.	1,2
2009 р.	1,3

Варіант 2

Чисельність лікарів в районі (на 01.01. кожного року)

2004 р.	56
2005 р.	58
2006 р.	60
2007 р.	65
2008 р.	57
2009 р.	59

Варіант 3

Кількість випадків втрати працездатності (на 100 робітників підприємства).

2004 р.	56
2005 р.	58
2006 р.	60
2007 р.	65
2008 р.	57
2009 р.	59

Варіант 4

Середня тривалість перебування хворих в стаціонарі міської лікарні (кількість днів)

2004 р.	13,8
2005 р.	14,0
2006 р.	14,3
2007 р.	14,1
2008 р.	14,5
2009 р.	14,6

Варіант 5

Рівні первинної захворюваності за даними профоглядів на підприємстві (кількість на 100 оглянутих)

2004 р.	5,0
2005 р.	6,1
2006 р.	6,8
2007 р.	7,2
2008 р.	7,6
2009 р.	5,0

Контрольні запитання

- Дайте визначення динамічного ряду.
- З якою метою використовуються динамічні ряди?
- Які існують види динамічних рядів?
- Що таке рівні динамічного ряду і якими величинами вони виражаються?
 - Який динамічний ряд називають простим?
 - Який динамічний ряд називають похідним?
 - Що являють собою моментні та інтервальні динамічні ряди?
- З якою метою проводиться вирівнювання динамічних рядів?
 - Назвіть методи вирівнювання динамічних рядів. У чому полягає сутність кожного з них?
 - Що включає аналіз динамічного ряду?
 - Які показники використовуються для аналізу динамічного ряду?
 - Як розрахувати абсолютний приріст, темп приросту, значення 1 % приросту, темп зростання?

Тема 3

Характеристика і аналіз статистичних даних. Середні величини і показники варіації

Збір, реєстрація і аналіз даних в процесі будь-якого дослідження завершується формуванням статистичної сукупності (Statistical aggregate), яку можна визначити як сукупність об'єктів або явищ одного виду, об'єднаних за певною ознакою. Наприклад, хворі з певним діагнозом, визначеним методом лікування тощо. При цьому для усіх явищ, які вивчаються в медицині, характерні мінливість і варіабельність. Кожна людина має кількісну оцінку певного набору фізіологічних і клінічних параметрів, які є індивідуальними. Але в групі людей будь-який клінічний параметр може змінюватися і набувати значень в певному діапазоні.

Статистична сукупність (Statistical aggregate) — сукупність об'єктів або явищ одного виду, об'єднаних за певною ознакою.

Перш ніж давати характеристику варіабельній сукупності, яка має різні значення ознак в окремих її одиницях, необхідно мати єдину типову для сукупності величину (показник), що дозволяє дати їй узагальнену характеристику. Для цього застосовуються середні величини, які розраховуються тільки за кількісними ознаками, тобто визначення середньої для атрибутивних ознак неможливо.

Середні величини дають узагальнену кількісну характеристику певної ознаки в статистичній сукупності за певних умов місця і часу.

У практиці охорони здоров'я середні величини використовуються досить широко:

- для характеристики організації роботи установ охорони здоров'я (середня зайнятість ліжка, середній термін перебування в стаціонарі, ін.);

- для характеристики показників фізичного розвитку (довжина, маса тіла, обводи голови немовлят, ін.);

- для аналізу клініко-фізіологічних показників (частота пульсу, дихання, рівень артеріального тиску, ін.);

- для оцінки медико-соціальних і санітарно-гігієнічних досліджень (середнє число лабораторних досліджень, середні норми раціону харчування, середній рівень радіаційного забруднення, ін.).

Обов'язковою умовою розрахунку середніх величин для досліджуваної сукупності є її однорідність. Якщо окремі елементи сукупності мають занадто великі або занадто малі кількісні значення ознаки, які істотно відрізняються від інших, такі елементи впливатимуть на розмір середньої величини для цієї сукупності і середня об'єктивно не виражатиме узагальнюючу характеристику сукупності.

Одним з варіантів вирішення проблеми може бути виключення окремих варіант з подальшого аналізу (що вимагає використання відповідних методик оцінки) або проведення розрахунку погрупових середніх з визначенням максимальних і мінімальних коливань.

Рівні ряду — це величини, з яких складається динамічний ряд — розмір того або іншого явища, досягнутий впродовж певного періоду або на певний момент часу. Властивістю середньої величини є її узагальнена характеристика.

Середня величина розраховується шляхом зіставлення абсолютних або відносних величин. При цьому якісно однорідна сукупність і достатнє число спостережень є основними вимогами для розрахунку середніх величин. Змішування сукупності, яка визначається різними якісними ознаками,

призводить до розрахунку нетипових середніх величин, які не можуть бути основою наукового аналізу.

Середня величина має подвійний характер: з одного боку вона характеризує сукупність в цілому, з другого — вона є основою для оцінки окремих одиниць сукупності, їх різноманітності і мінливості.

1. За формою розрахунку можна виділити:

- а) середню арифметичну величину;
- б) середню гармонійну величину;
- в) середню геометричну величину;
- г) середню квадратичну, кубічну і інші величини.

2. По охопленню сукупності існують:

- а) групова середня величина;
- б) загальна середня величина.

Середня арифметична є найпоширенішим видом середніх величин. Вона відображається як \bar{X} . Проте, часто застосовують букву M (лат. Media). За характером даних вона може бути простою або зваженою. Середня арифметична проста визначається як сума варіант варіаційного ряду, розділена на їх число. При цьому варіаційний ряд — це сукупність числових значень ознак (варіант), які можуть бути не систематизовані по своїх абсолютних значеннях (неранговий ряд), систематизовані в порядку зростання або зменшення — (ранговий ряд).

Окремі елементи (значення) сукупності однорідних за якісним складом предметів, явищ, параметрів є варіантами, а усю їх сукупність можна представити у вигляді варіаційного ряду, який є основою для визначення середніх величин.

Варіаційний ряд — це ряд варіант і частот, що відповідають їм. Варіаційний ряд може бути простим, де кожна варіанта представлена окремо, тому частота кожної з них дорівнює одиниці. Такий згрупований ряд є неінтервальним, тому

що групування проведено по абсолютних значеннях кожної варіанти. Варіаційні ряди, де значення варіант представлено у вигляді інтервалів, називаються інтервальними. У вигляді інтервального ряду часто представляють ознаки зі значною кількістю варіант. При цьому значення кожної варіанти подані у вигляді інтервалу. Згрупований інтервальний варіаційний ряд можна отримати шляхом об'єднання варіант в групи. При цьому необхідно пам'ятати, що:

а) розмір варіаційних груп повинен залежати від природи явища;

б) має сенс визначати однакові інтервали;

в) межі варіаційних груп не повинні повторюватися.

Усі варіаційні ряди за якісною характеристикою розподіляються на дискретні, в яких варіанти можуть бути представлені тільки цілими числами, або отримані в результаті підрахунків (розподіл по частоті пульсу, числу ліжкових днів, відвідувань) і інкретні (безперервні), де варіанти можуть бути представлені як цілими, так і дробовими числами або є результатом вимірів. Клінічні параметри є здебільшого прикладом інкретних варіант.

Іншим варіантом, гнучкішим з практичної точки зору, є метод визначення амплітуди ряду (різниця між максимальним і мінімальним значенням варіанти). Для вирішення питання про число груп необхідно подати статистичну сукупність у вигляді рангового ряду, тобто розмістити її одиниці в певному порядку. При чисельності сукупності менше 100 одиниць недоцільно планувати більше 10 груп.

Етапи складання інтервального варіаційного ряду:

- визначення амплітуди ряду;

- визначення числа груп;

- визначення величини інтервалу.

Розрахунок середніх величин базується на значеннях варіант. Якщо варіанту представлена у вигляді інтервалу,

величиною в кожній з них приймають центральну варіанту, тобто середину інтервалу. Для дискретного ряду центральна варіанта визначається як напівсума одного інтервалу.

Середніми величинами, які мають відносний характер, є медіана і мода.

Медіана (Me) — це серединна, центральна варіанта, яка ділить варіаційний ряд на дві рівні частини. Наприклад, коли число спостережень складає 29, медіаною буде 15-а за рахунком, тому що по обидві сторони від неї стоїть по 14 спостережень. У ряду з парним числом спостережень центральне положення мають дві величини. Якщо числові значення цих двох величин різні, то за медіану береться їх напівсума.

Розглядаючи моду не можна не розглянути поняття кватилей, які також застосовуються для характеристики сукупності. Кватилі — це значення, які ділять дві половини сукупності (розділені медіаною) ще раз навпіл (від слова кварта — чверть). Виділяють верхній кватиль, який часто позначають символом 75 % (це означає, що 75 % варіант менше верхнього кватіля). Нижній кватиль 25 % (25 % варіант менше нижнього кватіля). Таким чином, три точки — нижній кватиль, медіана і верхній кватиль — ділять вибірку на 4 рівних частини.

Мода (Mo) — величина, яка найчастіше зустрічається або найчастіше повторюється. Відповідає на графічному зображенні максимальній ординаті, тобто найвищому значенню графічної кривої. Таким чином, при наближеному знаходженні моди в простому (незгрупованому) ряду вона визначається як варіанта з найбільшою кількістю частот (наприклад: 2, 6, 6, 8, 9, 9, 9, 10 — мода = 9). При цьому ряд може мати бімодальний (два значення трапляються з однаковою частотою) і полімодальний характер.

Відмінність медіани і моди від середньої арифметичної полягає в тому, що ці величини визначаються досить легко

і не залежать від крайніх варіант або від міри розсіювання ряду.

Середня арифметична часто має обмежене значення тому, що вона не відбиває розміри коливання кількісних варіант ряду (варіабельність ряду). Важливою характеристикою ряду є оцінка різноманітності (мінливості, варіабельності) варіант досліджуваної сукупності. Варіацією є мінливість тільки тих ознак, на які впливають зовнішні чинники. Явища, які змінюються внаслідок своєї природи, не можна характеризувати як варіацію, наприклад зміна зростання дитини від народження до повноліття.

В першу чергу важлива саме випадкова варіація, оскільки це поняття є основою формування і оцінки норми і патології в медицині.

Однією з характеристик різноманітності варіант ряду є його амплітуда — різниця між крайніми значеннями. Проте, амплітуда не враховує характер розподілу варіант, тому вона має обмежене використання в медицині. Наприклад, два варіаційні ряди з різною амплітудою можуть мати однакове значення середньої арифметичної.

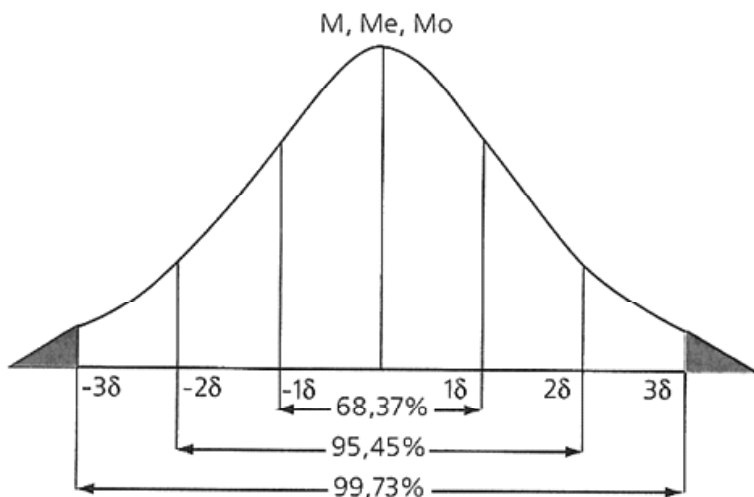
Для аналізу варіації ознак досліджуваної сукупності необхідно використовувати показники, які враховують значення усіх одиниць спостереження цієї сукупності. Таким показником є середнє квадратичне відхилення (стандартне відхилення — *standard deviation*), яке позначається символом σ (сигма). Середнє квадратичне відхилення враховує міру різноманітності усіх варіант сукупності відносно середньою арифметичною і визначається по формулі:

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}} \text{ для простого варіаційного ряду;}$$

$$\sigma = \sqrt{\frac{\sum d^2 f}{n-1}} \text{ для сгрупованого варіаційного ряду.}$$

де n — число спостережень в досліджуваній сукупності (при досить великому числі спостережень — $n > 30$ — у формулі замість $n - 1$ можна використати n); f — частота варіант; $d = x - X$ — відхилення кожної варіанти від середньої арифметичної; x — значення варіанти.

Правило нормального розподілу варіант у варіаційному ряду можна визначити як правило трьох сигм ($X \pm 3\sigma$) — відносно середньої арифметичної практично усі варіанти сукупності знаходяться в діапазоні $\pm 3\sigma$. Схематично це виражається так:



Згідно теорії статистики, яка має як арифметичне, так і геометричне вираження (по площі фігур) в межах $X \pm 1\sigma$ знаходиться не менше 68,37 % усіх варіант сукупності.

За межами цього інтервалу може бути до 31,63 % усіх спостережень.

В межах $X \pm 2\sigma$ будуть розташовані близько 95,45 % усіх варіант.

Практично увесь варіаційний ряд — 99,7 % варіант знаходитиметься в діапазоні $X \pm 3\sigma$.

Окремі варіанти — до 0,3 % досліджуваної сукупності можуть не відповідати загальному характеру розподілу і випадати з нього в результаті занадто низького або високого рівня (варіанти, що “вискакують”).

Узагальнення представленого матеріалу дозволяє зробити висновок про можливість практичного використання середнього квадратичного відхилення:

- для визначення амплітуди ряду;
- відновлення крайніх його значень;
- визначення вірогідного числа спостережень в певних інтервалах.

Для вищезгаданих умов значення середньої арифметичної моди і медіани будуть однаковими. Середнє квадратичне відхилення (як міра варіації) у квадраті називається дисперсією. По суті дисперсія — це середня величина квадратів відхилень варіант від середньої арифметичної. Цей показник також використовується в медицині і біології для характеристики однорідності досліджуваної сукупності.

При малих значеннях σ середня арифметична досить повно характеризує сукупність (є типовою), тоді як велике значення σ свідчить про неоднорідність варіаційного ряду (нетиповість середньої). У медицині при розробці критеріїв норми часто приймається діапазон $X \pm 1\sigma$ (рідше $X \pm 1,5\sigma$). На практиці іноді виникає потреба порівняти міру однорідності (різноманітність) різних параметрів сукупності. Середня арифметична часто має обмежене значення тому, що вона не відбиває розміри коливання кількісних варіант ряду (варіабельність ряду). Саме тому виникає потреба в розрахунку інших параметрів варіаційного ряду.

Для забезпечення зіставлення необхідно визначити для кожного ряду коефіцієнт варіації (С) — відношення середнього квадратичного відхилення (σ) до середньої арифметичної у відсотках. Цей показник є відносною мірою варіабельності, що виражається у відсотках:

$$\frac{\sigma}{\bar{X}} \cdot 100\%$$

Чим вище коефіцієнт варіації, тим більша варіабельність цієї ознаки. Критерієм в даному випадку є така оцінка: $C < 10\%$ — низький рівень варіабельності; $C = 10 - 25\%$ — середній рівень варіабельності; $C > 25\%$ — високий рівень варіабельності свідчить про неможливість використання цієї сукупності для аналізу, необхідність розрахунку по групових середніх або виключення окремих варіант, що “вискакують”.

При проведенні вибіркового дослідження можливі загальні погрішності і погрішності вибірки. Загальні погрішності можуть мати як систематичний характер (методичні, недоліки вимірювальної апаратури), так і випадковий (помилки дослідника). Погрішності вибіркового спостереження пов’язані з відбором його одиниць. Це погрішності типовості, репрезентативності.

В процесі аналізу розраховані показники (середня тривалість лікування, частота ускладнень, рівень летальності і тому подібне) розглядають як узагальнювальні величини. Якщо результати отримані на основі достатнього по кількості і якості однорідного матеріалу, то можна вважати, що вони досить точно характеризують досліджувані явища. Для оцінки достовірності результатів будь-яких вибірових досліджень визначають середню помилку відносною (m_p) або середньої величини (m_M).

Середня помилка для відповідних показників при значному числі спостережень ($n > 30$) може бути розрахована по таких формулах:

$$m_x = \frac{\sigma}{\sqrt{n}} \text{ — середня помилка середньої величини;}$$

$$m = \sqrt{\frac{Pq}{n}} \text{ — середня помилка відносної величини;}$$

де σ — середнє квадратичне відхилення; n — число спостережень у вибірковій сукупності; P — відносний показник; q — величина, зворотна показнику, тобто достовірність того, що це явище не буде зареєстровано. Сума протилежної вірогідності дорівнює одиниці: $P + q = 1$. Якщо показник розрахований на 100 (%), тоді $q = 100 - P$, якщо на 1000 (‰), то $q = 1000 - P$ і так далі.

При малому числі спостережень ($n < 30$) в знаменнику замість n використовується $n - 1$. Вірогідність безпомилкового прогнозу і довірчий критерій визначають на етапі планування статистичного дослідження.

При заданих мірах вірогідності довірчий критерій (t) має незмінну величину, а довірчий інтервал залежить від величини середньої погрішності (m), значення якої зменшується при збільшенні числа і якісного складу спостережень.

У медико-біологічних дослідженнях часто виникають ситуації, коли при порівнянні окремих параметрів необхідно оцінити істотність різниці між ними. Істотна різниця між окремими показниками вибіркового дослідження свідчить про можливість перенесення отриманих виведень на генеральну сукупність. Критерієм оцінки істотності різниці є коефіцієнт достовірності — критерій Госсета (Стьюдента), який визначають по формулі:

$$\frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{m_1^2 + m_2^2}}$$

для середніх величин;

$$\frac{|P_1 - P_2|}{\sqrt{m_1^2 + m_2^2}}$$

для відносних величин.

При великому числі спостережень ($n > 30$) різниця між показниками є істотною, якщо:

$t > 2$ (відповідає достовірності безпомилкового прогнозу 95,5).

$t > 3$ (відповідає достовірності безпомилкового прогнозу 99,7).

За умови $t < 2$ міра достовірності безпомилкового прогнозу складає менше 95 %. В цьому випадку не можна стверджувати, що різниця між показниками є істотною.

Часто при клінічних або експериментальних дослідженнях доводиться мати справу з малими числами спостережень (30 і менше) : 5-6 лабораторних щурів, 10-12 хворих і тому подібне. Якщо дослідження правильно організоване і відібрані однорідні групи, їх можна використати як вибіркові з малим числом спостережень. При малому числі спостережень ($n < 30$) оцінка достовірності різниці між параметрами окремих груп проводиться на основі порівняння результату і значень критерію Госсета (Стьюдента) з підсумковим табличним значенням для відповідних ступенів свободи ($n = n_1 + n_2 - 2$) Якщо певний t — критерій перевищує табличне значення, різницю між показниками слід вважати статистично доведеною.

Методика виконання навчального завдання

Завдання 1

Складання простого варіаційного ряду. Обчислення простої середньої арифметичної і середнього квадратичного відхилення при малій кількості спостережень.

Для побудови простого варіаційного ряду при малій кількості спостережень ($n < 30$) варіанти розташовують у ранговому порядку (у порядку зростання або спаду).

Проста середня арифметична (M) обчислюється за формулою: $M = \sum v/n$

Середнє квадратичне відхилення (σ) при малій кількості спостережень визначається за формулою:

$$m_m = \sigma \sqrt{\sum d^2 / n - 1}$$

у якій d — арифметичне відхилення варіант від середньої ($d = v - M$), n — кількість спостережень. Оскільки дослідження проводиться на вибірковій сукупності, потрібно представити середню арифметичну із значенням похибки репрезентативності (середньої похибки середньої арифметичної). Ця похибка (m_m) при малій кількості спостережень визначається за формулою:

$$m_m = \sigma / \sqrt{n - 1}$$

Типове завдання

Завдання 1

Результати вимірювання частоти пульсу (кількість ударів за хвилину) у 9 осіб: 64, 69, 63, 67, 74, 66, 62, 65, 73.

Зразок виконання завдання.

а) будуємо простий варіаційний ряд, розташовуючи варіанти за ранговим порядком, графа 1 табл. 1.

б) за формулами встановлюємо значення середньої арифметичної, середнього квадратичного відхилення. Послідовність дій представлена в табл. 1.

Таблиця 1

**Побудова простого варіаційного ряду,
визначення простої середньої арифметичної,
середнього квадратичного відхилення
при малій кількості спостережень**

v (уд/хв)	d	d ²
62	-5	25
63	-4	16
64	-3	9
65	-2	4
66	-1	1
67	0	0
69	2	4
73	6	36
74	7	49
$\sum v = 603$ $n = 9$		$\sum d^2 = 144$

$$M = 603 / 9 = 67 \text{ уд/хв.}$$

$$\sigma = \sqrt{144(9 - 1)} = \sqrt{1152} = 33,94 \text{ уд/хв}$$

в) похибку середньої арифметичної визначаємо за формулою:

$$m = 33,94 / \sqrt{8} = 33,94 / 2,83 = 12,00 \text{ уд/хв}$$

Висновок: середня частота пульсу обстеженої групи осіб становить $67 \pm 12,00$ уд/хв. $\sigma = 33,94$ уд/хв

Завдання 2

Складання згрупованого варіаційного ряду. Визначення моди і медіани. Обчислення зваженої середньої арифметичної та середнього квадратичного відхилення при великій кількості спостережень.

Згрупований варіаційний ряд складається шляхом розміщення варіант у ранговому порядку із зазначенням частоти кожної варіанти.

Мода відповідає числовому значенню ознаки, яка частіше за інших зустрічається у варіаційному ряду. Для визначення медіани треба знайти середину ряду. При парному числі спостережень за медіану приймають середню величину з двох центральних варіант. При непарному числі спостережень медіана відповідає значенню центральної варіанти. Для визначення порядкового номера медіани у варіаційному ряду, парне число спостережень ділять навпіл, до непарного числа спостережень додають одиницю, після чого отримане число також ділять навпіл.

Зважена середня арифметична (M) обчислюється за формулою:

$$M = \sum vp/n$$

Середнє квадратичне відхилення (σ) при великій кількості спостережень ($n > 30$) визначається за формулою:

$$\Sigma = \Sigma = \sqrt{\Sigma d^2 p / n}$$

у якій d — відхилення варіант від середньої ($d = v - M$); p — частота відповідних варіант; n — кількість спостережень.

Середня похибка середньої арифметичної m_m при великій кількості спостережень обчислюється за формулою:

$$m_m = \sigma / \sqrt{n}$$

Типове завдання

Отримано наступні дані про тривалість лікування в поліклініці 45 хворих ангіною (у днях): 20, 18, 19, 16, 17, 16, 14, 13, 15, 14, 15, 13, 12, 13, 3,4, 12, 11, 12, 11, 10, 12, 11, 10, 11, 8, 7, 11, 11, 10, 10, 9, 8, 8, 9, 4, 5,6, 9, 5, 9, 6, 7,7.

Зразок виконання завдання:

а) будуємо згрупований варіаційний ряд, послідовно розташовуючи варіанти в порядку зростання з відповідними їм частотами (графи 1, 2 табл. 2);

б) знаходимо моду (M_0): з найбільшою частотою зустрічається варіанта, яка дорівнює 11 дням.

Отже, $M_0 = 11$ дням;

в) знаходимо порядковий номер медіани (M_e) за формулою: $(n+1):2 = (45+1):2 = 23$.

Отже, 23-я за порядком варіанта є медіаною. У нашому прикладі такою варіантою є 11, таким чином, $M_e = M_0 = 11$ дн.

Таблиця 2

**Побудова згрупованого варіаційного ряду,
визначення зваженої середньої арифметичної,
середнього квадратичного відхилення
при великій кількості спостережень**

v (дні)	p	vp	d	d^2	d^2p
3	1	3	-7,7	59,49	59,49
4	2	8	-6,7	44,89	89,78
5	2	10	-5,7	32,49	64,98
6	2	12	-4,7	22,09	44,18
7	3	21	-3,7	13,69	41,07
8	3	24	-2,7	7,29	21,87
9	4	36	-1,7	2,89	11,56
10	5	50	-0,7	0,49	2,45
11	6	66	0,3	0,09	0,54
12	4	48	1,3	1,69	6,76
13	3	39	2,3	5,29	15,87
14	2	28	3,3	10,89	21,78
15	2	30	4,3	18,49	36,98
16	2	32	5,3	28,09	56,18
17	1	17	6,3	39,69	39,69

18	1	18	7,3	53,29	53,29
19	1	19	8,3	68,89	68,89
20	1	20	9,3	86,49	86,49
	n=45	$\sum vp = 482$			$\sum d^2 p = 721,85$

г) отримуємо добутки кожної варіанти на свою частоту (vp) та їх суму ($\sum vp$) – графа 3 табл.2;

д) за формулою $M = \sum vp/n = 482/45 = 10,7$ дн. отримуємо середню арифметичну зважену;

е) визначаємо відхилення (d) кожної варіанти від середньої ($d = v - M$) і їх квадрати (d^2) графі 4, 5 табл. 2;

ж) визначаємо добутки квадрата відхилення (d^2) кожної варіанти на свою частоту (p) — $d^2 p$ і їх суму ($\sum d^2 p$) – графа 6 табл. 2;

з) за формулою

$\sigma = \sqrt{\sum d^2 p/n} = \sqrt{721,85/45} = \sqrt{16,04} = 4,01$ дн. визначаємо середнє квадратичне відхилення

і) за формулою : $m_m = \sigma / \sqrt{n}$ визначаємо середню похибку середньої арифметичної:

$$m_m = 4,01 \sqrt{45} = 4,01/6,71 = 0,60 \text{ дн.}$$

Висновок: тривалість лікування групи хворих ангіною в поліклініці складає $10,7 \pm 0,60$ дн. $M_o = M_e = 11$ дн., $\sigma = 4,01$ дн.

Навчальні завдання

Завдання 1

На основі приведених початкових даних необхідно:

- скласти простий варіаційний ряд;
- обчислити просту середню арифметичну (M);
- визначити середнє квадратичне відхилення (σ);
- розрахувати середню похибку середньої арифметичної;

д) констатувати отримані результати у висновках.

Варіант 1.

Чисельність населення (кількість чоловік) на 5 територіальних лікарських дільницях: 2100, 2350, 2120, 2600, 2150.

Варіант 2.

Число хворих, які перебувають на диспансерному обліку з хронічними захворюваннями у 9 дільничних лікарів: 148, 130, 151, 141, 114, 123, 136, 143, 120.

Варіант 3.

Частота пульсу (кількість ударів на хвилину) у 10 осіб після проведення атропінової проби: 82, 92, 100, 96, 90, 102, 88, 80, 86, 84.

Варіант 4.

Число звернень за першу добу кожного календарного року протягом 12 міс. по швидку медичну допомогу у місті Н.: 165, 161, 167, 165, 164, 163, 142, 143, 137, 156, 151, 147.

Варіант 5.

Число виробничих травм серед працівників ТОВ «Гарант» протягом 12 міс. календарного року: 85, 82, 85, 124, 96, 107, 137, 151, 82, 83, 59, 56.

Завдання 2

На основі наведених початкових даних необхідно:

- а) побудувати згрупований варіаційний ряд;
- б) знайти моду (M_0) і медіану (M_e);
- в) обчислити зважену середню арифметичну (M);
- г) визначити середнє квадратичне відхилення (σ);

д) розрахувати значення середньої похибки середньої арифметичної;

е) констатувати отримані результати у висновках.

Варіант 1.

Тривалість лікування в стаціонарі 45 хворих на пневмонію (у днях): 25, 11, 12, 13, 24, 23, 23, 24, 21, 22, 21, 23, 22, 21, 14, 14, 22, 20, 20, 15, 15, 16, 20, 20, 20, 16, 16, 20, 17, 17, 19, 19, 19, 18, 18, 18, 18, 19, 19, 17, 17, 18, 18, 19, 26.

Варіант 2.

Частота дихання (кількість дихальних рухів на хвилину) у 47 осіб віком 40-45 років: 12, 14, 13, 15, 16, 16, 16, 19, 19, 20, 20, 20, 19, 13, 15, 12, 15, 13, 15, 12, 17, 12, 17, 17, 13, 16, 17, 18, 14, 15, 16, 18, 14, 15, 14, 17, 18, 14, 18, 20, 17, 18, 19, 20, 21, 22.

Варіант 3.

Частота пульсу (кількість ударів на хвилину) у 55 студентів-медиків перед ліцензійним іспитом «КРОК-2»: 64, 66, 60, 62, 64, 68, 70, 66, 70, 68, 62, 78, 70, 72, 60, 70, 74, 62, 70, 72, 72, 64, 70, 72, 66, 76, 68, 70, 58, 76, 74, 76, 76, 82, 76, 72, 76, 74, 79, 78, 74, 78, 74, 78, 74, 74, 78, 76, 78, 76, 80, 80, 80, 78, 78.

Варіант 4.

Тривалість непрацездатності (у днях) у 35 хворих з гострими респіраторними захворюваннями, які лікувалися у дільничного лікаря-терапевта: 6, 7, 5, 3, 9, 8, 7, 5, 6, 4, 9, 8, 7, 6, 6, 9, 6, 5, 10, 8, 7, 5, 11, 13, 6, 7, 12, 4, 3, 5, 2, 5, 6, 6, 7

Варіант 5.

Кількість хворих, які перебувають на диспансерному обліку у 33 невропатологів в поліклініках великого міста: 85, 87, 90, 91, 89, 91, 90, 93, 94, 90, 93, 88, 98. 92. 94, 88,96, 90, 92, 95, 87, 90, 91,86, 92, 89, 97, 89, 99, 100, 82, 93, 88.

Контрольні запитання

1. Що таке варіаційний ряд?
2. Які є характеристики варіаційного ряду?
3. У яких випадках складається простий варіаційний ряд?
4. У яких випадках складається згрупований варіаційний ряд?
5. Наведіть приклади варіаційних рядів.
6. У чому полягають особливості рангованого ряду?
7. У чому полягають особливості дискретного ряду, інкретного ряду?
8. Назвіть основні вимоги до складання варіаційних рядів. У чому ці вимоги відрізняються для простих і згрупованих варіаційних рядів.
9. Що таке середня величина?
10. Яке практичне використання середніх величин?
11. Які є види середніх величин?
12. Що таке мода (Mo) і медіана (Me)?
13. У яких випадках визначається проста середня арифметична і за якою формулою?
14. У яких випадках обчислюється зважена середня арифметична і за якою формулою?
15. Яким вимогам повинен відповідати первинний статистичний матеріал для розрахунку середніх величин?
16. Які властивості притаманні середнім величинам?

17. Які є критерії різноманітності ознаки у варіаційному ряду?

18. Як визначається ліміт, амплітуда, середнє квадратичне відхилення, коефіцієнт варіації? Які початкові дані необхідні для їх розрахунку?

19. З якою метою використовується середнє квадратичне відхилення?

20. Яким чином середнє квадратичне відхилення пов'язано із структурою ряду розподілу ознаки? Як цей зв'язок використовується для з'ясування питання про типовість середньої величини?

21. Для чого використовується коефіцієнт варіації?

22. Якому значенню коефіцієнта варіації відповідає сильна, середня і слабка відносна різноманітність ознаки?

23. Що таке середня похибка середньої арифметичної, як її розрахувати, і про що вона дозволяє судити?

Тема 4

Метод стандартизації

Об'єктивне співставлення загальних інтенсивних показників можливе лише за умови якісної однорідності порівнюваних груп. Так, наприклад, показники летальності в двох опікових відділеннях можна порівнювати між собою за умови, що обидва стаціонари мають приблизно однаковий склад хворих по ряду основних параметрів — віку, статі, важкості патології, термінам госпіталізації і тому подібне. Якщо їх склад відрізняється, порівняння загальних інтенсивних показників, що характеризують силу і поширеність явища, ускладнене. При цьому на величину загального інтенсивного показника впливає склад оцінюваної клініко-статистичної групи.

Статистичний метод, який дозволяє виключити вплив неоднорідності складу порівнюваних груп на досліджувані загальні показники, називається методом стандартизації.

При проведенні клінічних досліджень при вивченні ефективності певного методу лікування також необхідно формувати однорідні в порівнянні групи. При використанні методу стандартизації розраховують стандартизовані (умовні) показники, які могли б бути за умови однакового складу населення в порівнюваних групах. Практична значущість методу стандартизації:

- дозволяє порівняти частоту однотипних явищ в неоднорідних групах;
- дозволяє оцінити вплив досліджуваного чинника на величину загальних показників.

Оцінка впливу певного чинника на величину загальних інтенсивних показників базується на динаміці співвідношен-

ня цих показників за умови змін у складі досліджуваних груп. Якщо умовна зміна складу порівнюваних груп за певними критеріями призводить до зміни співвідношення загальних інтенсивних показників (зміна знаку між ними), то це дає можливість зробити висновок про значущість (вплив) цього чинника для оцінювання рівнів досліджуваних показників.

Існує три методи стандартизації:

- прямий;
- опосередкований;
- зворотній.

Вибір будь-якого з методів визначається формою представлення первинного матеріалу, зручністю і швидкістю розрахунків, даними попередніх досліджень. Прямий метод використовують за наявності даних про склад населення і досліджуваного явища за точними параметрами (вік, професія, терміни госпіталізації, важкість захворювання і тому подібне).

Відсутність даних про розподіл певного явища або невелика чисельність груп при цьому розподілі знижує достовірність групових показників і є основою для використання опосередкованого методу стандартизації. Відсутність даних про склад населення зумовлює необхідність використання зворотнього методу.

Найпоширенішим в медико-біологічних дослідженнях є прямий метод стандартизації. Основними етапами обчислень при прямому методі стандартизації є:

I етап — розрахунок загальних і спеціальних (по кожній групі — віковій, статевій та ін.) показників інтенсивності (або співвідношення, або середніх величин) для двох порівнюваних сукупностей,

II етап — вибір і розрахунок стандарту;

III етап — розрахунок «очікуваних» величин для кожної групи стандарту;

IV етап — визначення стандартизованих показників;

V етап — порівняння груп за загальними і стандартизованими показниками (показниками інтенсивності, співвідношення або середніми величинами).

Висновки.

Стандартизований показник демонструє, якою була б захворюваність (летальність, смертність, інвалідність) в порівнюваних контингентах, якби їх склад був однорідним, тобто розбіжності у складі контингентів були б нівельованими.

Стандартизовані показники застосовуються для порівняння двох показників, але не відображають справжні розміри явищ, тому що величина їх може змінюватися при зміні стандарту.

У такий спосіб можна встановити факт впливу будь-якого чинника на згадані показники і виявити причину змін показників, які розраховуються звичайними методами.

Методика виконання навчального завдання

Обчислення стандартизованих показників прямим методом.

На початку роботи слід вирішити питання про доцільність застосування методу стандартизації при виконанні відповідного варіанту завдання.

Метод стандартизації необхідно застосовувати за наявності наступних умов: а) при значних відмінностях у рівнях групових показників; б) при значній неоднорідності складів порівнюваних сукупностей.

Якщо показники окремих груп майже однакові, то навіть при різному складі сукупностей, що вивчаються, по цих групах стандартизацію проводити не потрібно. Якщо показники в групах різні, а їх склад однаковий, то проведення стандартизації також не має сенсу. При проведенні стандар-

тизації прямим методом необхідно побудувати таблиці для розміщення в них початкових даних і результатів розрахунків за етапами.

На першому етапі обчислюють загальні і спеціальні (тобто для кожної групи кожної з сукупностей, що вивчаються) інтенсивні показники (тут і далі: або показники співвідношення, або середні величини).

На другому етапі проводиться вибір і розрахунок стандарту. За стандарт слід визначати той склад сукупностей, в якому відбилися б всі особливості складу порівнюваних груп. Як стандартний склад може бути вибрана сума (використовується найчастіше), напівсума груп порівнюваних сукупностей, або склад однієї з них. При виборі за стандарт сумарного складу груп, отримують суму осіб двох сукупностей окремо по кожній однаковій групі, а також загальну суму всіх одиниць спостереження порівнюваних сукупностей.

Загальна чисельність колективу, взятого за стандарт, береться рівною основі, відносно якої взято показник. Якщо показник виражається на 100 населення, то за стандарт береться розподіл, наприклад, за віком, 100 осіб.

На третьому етапі визначаються «очікувані» числа. При цьому як початкові дані служать результати, отримані на двох попередніх етапах. «Очікувані» числа отримують шляхом складання і обчислення пропорції. Приймається умова, що розподіл обох статистичних сукупностей, що зіставляються за істотною ознакою, що вивчається, наприклад, віком, однаковий і відповідає стандарту. Ставиться питання: якою буде очікувана кількість явищ, наприклад, захворювань, при новому, умовному (стандартному) розподілі. Послідовно кожен груповий інтенсивний показник, отриманий на першому етапі стандартизації, множать на груповий стандарт і ділять на основу, відносно якої виражений показник (тобто на 100,

1000, 10000 і таке інше), отримуючи, таким чином, «очікувані» числа.

Визначення стандартизованих показників (четвертий етап) здійснюється шляхом знаходження суми «очікуваних» чисел для кожної із статистичних сукупностей, що співставляються.

На п'ятому етапі проводиться порівняння груп за загальними інтенсивними і стандартизованими показниками. Висновок роблять приблизно за наступною схемою: якщо співвідношення стандартизованих показників виходить інше, ніж реальне, то робиться висновок про те, що на рівень реальних показників вплинув факт різного складу порівнюваних сукупностей за певною ознакою. І навпаки, якщо стандартизовані показники повторюють співвідношення реальних, робиться висновок, що на рівень реальних показників факт різного складу порівнюваних сукупностей за цією ознакою не вплинув.

Типове завдання

Розрахунок стандартизованих показників захворюваності з тимчасовою втратою працездатності (ТВП) прямим методом.

Зразок виконання завдання.

I етап методу стандартизації: розрахунок інтенсивних показників (у даному прикладі — частоти випадків захворюваності з ТВП на 100 працівників) у двох порівнюваних сукупностях. Показники розраховуються для кожної вікової групи окремо і по рядку «Всього» для кожного цеху:

$$\frac{\text{Кількість випадків} \times 100}{\text{Кількість працюючих}}$$

У нашому прикладі:

$$80 \times 100 / 120 = 67,0 \quad 380 \times 100 / 400 = 105,3 \text{ тощо.}$$

Таблиця 1

**Розподіл працюючих і випадків захворювань з ТВП
в цехах №1 і №2 (дані умовні)**

Вік (роки життя)	Цех №1			Цех №2		
	Кіль- кість праців- ників	Кількість випадків захворю- вань	Кількість випадків захворю- вання на 100 праців- ників	Кіль- кість праців- ників	Кількість випадків захворю- вань	Кількість випадків захворю- вання на 100 праців- ників
До 20	120	80		200	140	
21-39	380	400		500	575	
41-59	400	680		80	132	
60 і старше	100	185		20	34	
Всього	1000	1345		800	881	

Результати розрахунків по першому етапу надаються в табл. 2.

II етап — визначення і розрахунок стандарту.

За стандарт прийнятий середній віковий склад працівників в обох цехах. Проводяться наступні обчислення (табл. 3).

Таблиця 2

**Показники захворюваності з ТВП по вікових групах
і по цехах №№1 і 2 в цілому, число випадків
на 100 працівників (I етап стандартизації)**

Вік (роки життя)	Цех № 1	Цех №2
До 20	67,0	70,0
21-39	105,3	115,0
41-59	170,0	165,0
60 і старше	185,0	170,0
Всього	134,5	110,0

Таблиця 3

Розрахунок стандарту (II етап стандартизації)

Вік (роки життя)	Стандарт (кількість працівників)
До 20	120 (цех №1) +200 (цех №2) = 320
21-39	380+500 = 880
41-59	400 +80 = 480
60 і старше	100 +20 = 120
Всього	1000 +800=1800

Оскільки в нашому прикладі показник захворюваності з ТВП обчислений на 100 працівників, за стандарт береться розподіл за віком 100 чоловік, тобто в %.

При виконанні навчального завдання результати розрахунків вносяться безпосередньо до табл. 4.

Таблиця 4

Результати розрахунку стандарту (II етап стандартизації)

Вік (роки життя)	Кількість працівників		Загальна кількість працівників в обох цехах	Стандарт Середній віковий склад працівників обох цехів (у %)
	Цех №1	Цех №2		
До 20	120	200	320	18
21-39	380	500	880	49
41-59	400	80	480	26
60 і старше	100	20	120	7
Всього	1000	800	1800	100

III етап. Розрахунок очікуваних величин (у даному прикладі — кількості випадків захворювань) в кожній групі стандарту.

Прийmemo умовно, що розподіл робітників в обох цехах за віком однаковий і відповідає стандарту, і розраховуємо очі-

кувану кількість захворювань при новому умовному розподілі робітників за віком (табл. 5).

Числа таблиці отримані таким чином: у віці до 20 років в цеху №1 показник захворюваності на 100 працівників складає 67,0, а в цеху №2 — 70,0. Отже, серед 18 працівників (за стандартом) захворювань в цеху №1 буде: $67,0 \times 18 / 100 = 12,0$, а в цеху №2: $70,0 \times 12 / 100 = 12,6$

Таким же чином визначаються числа інших рядків у двох останніх графах. Склавши очікувані числа випадків захворювань за віком (IV етап стандартизації), отримаємо стандартизовані показники захворюваності на 100 працівників для кожного цеху.

Таблиця 5

Розрахунок очікуваних величин (числа випадків захворювань з ТВП) в кожній групі стандарту за віком і по цехах №№ 1 і 2 в цілому, розрахунок стандартизованих показників (III і IV етапи стандартизації)

Результати розрахунку стандарту			Число захворювань в стандарті		
Вік (роки життя)	Захворюваність на 100 прац.		Стандарт	Цех №1	Цех №2
	Цех №1	Цех №2			
До 20	67,0	70,0	18	12,0	12,6
21-39	105,3	115,0	49	51,6	56,3
41-59	170,0	165,0	26	44,2	42,9
60 і старше	185,0	170,0	7	12,9	11,9
Всього	134,5	110,0	100	120,7	123,7

На V етапі стандартизації проводиться інтерпретація результатів, і формулюються висновки. Так, аналіз захворюваності з ТВП по цехах №№1 і 2 виявив наступне:

Захворюваність з ТВП в цеху №1 в цілому вища, ніж в цеху №2 ($134,5 > 123,7$).

Захворюваність в окремих вікових групах у цехах, які зіставляються, має відмінності: у вікових групах до 20 років і 21-39 років вона вища в цеху №2, у реалі груп — в цеху №1.

Оскільки стандартизовані показники захворюваності вищі в цеху №2, можна констатувати, що на рівень загальних показників спричиняють вплив неоднаковий за віком склад працюючих (переважають старші вікові групи в першому цеху). Якби віковий склад тих, що працюють в цих цехах був однаковим, то в цеху №2 захворюваність була б вища ніж в цеху №1. Можна припустити, що санітарно-гігієнічні умови праці і побуту в цьому цеху були гірші.

Навчальне завдання

На основі початкових даних за варіантами необхідно:

Обчислити стандартизовані показники, використовуючи прямий метод стандартизації.

Порівняти:

- загальні показники, розраховані звичайним способом;

- спеціальні показники в окремих групах;

- стандартизовані і звичайні показники.

Зробити висновок, що витікає із зіставлення звичайних і стандартизованих показників у порівнюваних групах.

*Варіант 1**Таблиця***Розподіл населення міст А і Б за віком
і кількістю померлих (у абсолютних числах)**

Вік (роки життя)	Місто А		Місто Б	
	Кількість жителів	З них по- мерло	Кількість жителів	З них по- мерло
0-14	3000	зо	1000	10
15-49	5000	10	5000	10
50 і старше	2000	60	4000	120
Всього	10000	100	10000	140

*Варіант 2**Таблиця***Розподіл міського і сільського населення області за віком
і кількістю померлих (у абсолютних числах)**

Вік (роки життя)	Міське населення		Сільське населення	
	Кількість жителів	З них померло	Кількість жителів	З них померло
0-14	50000	600	100000	1400
15-49	100000	400	400000	1600
50 і старше	50000	1000	300000	5400
Всього	200000	2000	800000	8400

Варіант 3

Таблиця

Розподіл осіб, що мали травми на промисловому підприємстві, за статтю за два періоди часу (у абсолютних числах)

Стать	Перший період		Другий період	
	Кількість працівників	Кількість травм	Кількість працівників	Кількість травм
Чоловіки	200	32	600	72
Жінки	400	28	200	16
Всього	600	60	800	88

Варіант 4

Таблиця

Розподіл обстежених і хворих на гіпертонічну хворобу за статтю в районах А і Б (у абсолютних числах)

Стать	Район А		Район Б	
	Кількість обстежених	Кількість хворих	Кількість обстежених	Кількість хворих
Чоловіки	200	3	600	9
Жінки	800	48	400	24
Всього	1000	51	1000	33

Розподіл хворих і померлих у двох лікарнях залежно від терміну госпіталізації при апендициті (у абсолютних числах)

Термін госпіталізації в днях	Лікарня №1		Лікарня №2	
	Кількість хворих	Кількість померлих	Кількість хворих	Кількість померлих
1-2	400	1	100	-
3-4	150	3	200	2
5-6	50	6	300	10
Всього	600	10	600	12

Контрольні запитання

- У чому полягає сутність і практична цінність методу стандартизації?
- З якою метою застосовується метод стандартизації?
- Чому порівняння загальних показників (смертності, захворюваності тощо) не завжди об'єктивне?
- Які існують методи стандартизації?
- Які початкові дані необхідні для застосування прямого методу стандартизації?
- Які початкові дані необхідні для застосування непрямого методу стандартизації?
- Які початкові дані необхідні для застосування зворотного методу стандартизації?
- Назвіть етапи прямого методу стандартизації.
- У чому полягає сутність кожного з етапів стандартизації?
- Які є способи вибору і розрахунку стандарту?
- Як з «очікуваних» чисел отримати стандартизований показник?

- 12 Чи відображає стандартизований показник дійсну частоту явища?
- Як інтерпретувати стандартизовані показники?
- Наведіть приклади використання стандартизованих показників у практичній лікарській діяльності.

Тема 5

Параметричні методи оцінки і аналізу статистичних гіпотез

У більшості медичних статистичних досліджень використовується вибірковий метод, який дозволяє за частиною явища судити про явище в цілому. Вибіркова сукупність повинна бути репрезентативною, тобто представницькою по відношенню до генеральної сукупності за кількістю і за якістю.

Науковим підґрунтям вибіркового методу є теорія вірогідності та закон великих чисел.

Основними положеннями закону великих чисел є наступні:

а) при збільшенні кількості спостережень дані вибіркової сукупності прагнуть відтворити дані генеральної сукупності;

б) досягши певної, достатньої кількості спостережень дані вибіркової сукупності відтворюють дані генеральної сукупності.

При проведенні вибірових статистичних досліджень необхідне проведення оцінки достовірності їх результатів.

Під достовірністю статистичних показників розуміють ступінь їх відповідності дійсності, що відображається ними. Достовірними результатами вважаються ті, які не спотворюють і правильно відображають об'єктивну реальність.

Оцінити достовірність результатів дослідження означає визначити, з якою вірогідністю безпомилкового прогнозу можливо перенести результати, отримані на вибірковій сукупності, на всю генеральну сукупність.

Оцінка достовірності (істотності) результатів статистичного дослідження проводиться для визначення і оцінки:

а) відмінностей між чисельними характеристиками генеральної і вибіркової сукупності;

б) відмінностей між чисельними характеристиками даної і аналогічної вибіркової сукупностей;

в) достовірності різниці між двома відносними або середніми величинами;

г) достовірності похідних величин (середніх і відносних);

д) достовірності різниці результатів досліджень без обчислення похідних величин.

Існують параметричні і непараметричні методи оцінки достовірності результатів статистичного дослідження.

Оцінка достовірності результатів дослідження параметричним методом передбачає визначення:

а) похибок репрезентативності (середніх похибок середніх арифметичних і відносних величин) — m_M , m_p ($m\%$).

б) довірчих меж середніх (або відносних) величин;

в) достовірності різниці середніх (або відносних) величин (за критерієм Стьюдента t).

Параметричні критерії (коефіцієнти) оцінки результатів досліджень застосовують для визначення і оцінки:

а) достовірності (істотності) похідних величин (середніх і відносних);

б) достовірності різниці між двома середніми або відносними величинами.

Похибка репрезентативності (m) — найважливіша статистична величина, необхідна для оцінки результатів дослідження. Похибки репрезентативності виникають з суті вибіркового дослідження: генеральна сукупність може бути охарактеризована за вибірковою сукупністю тільки з деякою погрішністю, вимірюваною похибкою репрезентативності.

Похибку репрезентативності не можна плутати із звичайним уявленням про похибки методичні, точності вимірювання, арифметичні та ін. Це єдиний вид похибок, що враховуються статистичними методами, які не можуть бути усунені, якщо не здійснено перехід на суцільне вивчення. Похибки репрезентативності можна звести до достатньо малої величини, тобто до величини допустимої погрішнос-

ті. Робиться це шляхом залучення у вибірку достатньої кількості спостережень. Кожна середня величина — M (середня тривалість лікування, середній зріст, середня маса тіла, середній рівень білка в крові тощо), а також кожна відносна величина — P (рівень захворюваності, летальності і ін.) повинні бути представлені зі своєю середньою похибкою — m .

Середня арифметична величина вибіркової сукупності (M) має похибку репрезентативності, яка називається середньою похибкою середньої арифметичної (m_M).

Відносні величини показники інтенсивності, співвідношення, екстенсивності (P), отримані при вибіркового дослідженні, також мають свою похибку репрезентативності, яка називається середньою похибкою відносної величини (m_p або $m_{\%}$).

За величиною похибки репрезентативності визначають, наскільки результати, отримані при вибіркового спостереженні, відрізняються від результатів, які могли б бути отримані при проведенні суцільного дослідження всіх без винятку елементів генеральної сукупності. Іншими словами, похибка репрезентативності є мірою достовірності середньої (або відносної) величини, отриманої при проведенні вибіркового дослідження.

Визначаючи для середньої арифметичної (або відносної) величини два крайні значення: мінімально можливе і максимально можливе, знаходять межі, в яких може бути невідома величина генерального параметра. Ці межі називають довірчими межами.

Довірчі межі — межі крайніх значень похідних (середніх або відносних величин), вихід за які внаслідок випадкових коливань має незначну вірогідність.

Довірчі межі залежать від розміру довірчого інтервалу. Довірчим інтервалом є інтервал, в межах якого при заданому рівні ймовірності знаходяться значення похідної величини. Довірчим межах відповідає крайнє значення довірчого інтервалу із заданим рівнем ймовірності.

Довірчі межі можуть бути визначені за допомогою похибки репрезентативності середньої (або відносної) величини.

Оцінити достовірність різниці результатів дослідження означає вирішити питання, чи є відмінність показників істотною, зумовленою дією різних чинників, або вона викликана випадковими коливаннями

У медицині і охороні здоров'я за різницею параметрів оцінюють середні і відносні величини, отримані для різних груп населення за статтю, зростом, а також груп хворих і здорових тощо. У всіх випадках при зіставленні двох порівнюваних величин виникає необхідність не тільки визначити їх різницю, але і оцінити її достовірність.

Достовірність різниці величин, отриманих при вибіркових дослідженнях, означає, що висновок про їх відмінності може бути перенесений на відповідні генеральні сукупності.

Разом з параметричними критеріями в практиці статистичної обробки даних застосовуються непараметричні критерії, незалежні від форм розподілу. Непараметричними вони називаються тому, що при їх використанні не потрібне обчислення параметрів розподілу: середніх величин, середнього квадратичного відхилення та ін. Вивченню непараметричних критеріїв присвячена окрема тема.

Методика виконання навчальних завдань

Завдання 1

Визначення похибки (m_M) і довірчих меж середньої арифметичної або відносної величини при малому числі спостережень.

При малому числі спостережень ($n < 30$) похибки репрезентативності визначаються за формулами, — для середніх величин — середня похибка середньої арифметичної:

$$m_M = \sigma / \sqrt{n - 1}$$

де σ — середнє квадратичне відхилення, n — число спостережень; для відносних величин — похибка відносної величини (m_p):

$$m_p = \sqrt{p \cdot q / n - 1}$$

де p — величина показника, для якого визначається m_p , а q — альтернатива ($q = 100 - p$, якщо величина p виражена у відсотках, або $1000 - p$, якщо величина p виражена в проміллі).

За допомогою похибки можна визначити довірчі межі середньої або відносної величини. При цьому використовуються наступні формули:

а) для середніх величин: $M_{\text{ген}} = M_{\text{виб}} \pm tm_M$

де $M_{\text{ген}}$ — середня величина ознаки в генеральній сукупності, $M_{\text{виб}}$ — середня величина, отримана в результаті дослідження вибіркової сукупності, t -довірчий коефіцієнт, — величина, на яку потрібно помножити m для того, щоб з певною ймовірністю безпомилкового прогнозу (p) отримати межі коливань середньої величини в генеральній сукупності; tm_M — довірчий інтервал (або максимальна похибка), позначається також символом Δ ;

б) для відносних величин: $P_{\text{ген}} = P_{\text{виб}} \pm tm_p$

де $P_{\text{ген}}$ — показник (відносна величина) в генеральній сукупності, $P_{\text{виб}}$ — показник, отриманий в результаті дослідження вибіркової сукупності, m_p — середня похибка, t — довірчий коефіцієнт; tm_p (Δ) — довірчий інтервал.

Поняття «Вірогідність безпомилкового прогнозу» (p) — це вірогідність, з якою можна стверджувати, що в генеральній сукупності M знаходиться в межах $M_{\text{виб}} \pm tm_M$ (або $P_{\text{ген}}$ в межах $P_{\text{виб}} \pm tm_p$). Якщо $n < 30$, критерій t для відповідної вірогідності безпомилкового прогнозу знаходиться за таблицею Стюдента. Якщо $n > 30$, то при $p = 95\%$ $t = 2$, при $p = 99\%$ $t = 3$.

Для абсолютної більшості медичних досліджень ступінь вірогідності безпомилкового прогнозу (p) повинен бути не менше 95%.

Типове завдання

Обчислити похибку (m_m) і визначити довірчі межі середньої арифметичної (M) при $p = 95\%$ і $p = 99\%$ для наступних початкових даних: середня частота пульсу у 9 чоловік склала 67 ударів за хвилину. Середнє квадратичне відхилення = 4 уд./хв.

Таблиця значень критерію t (за Стьюdentом)

n-1	Рівень вірогідності безпомилкового прогнозу (у %)			n-1	Рівень вірогідності безпомилкового прогнозу (у %)		
	95	99	99,9		95	99	99,9
1	12,7	63,6	636,6	16	2,1	2,9	4,0
2	4,3	9,9	31,6	17	2,1	2,8	3,9
3	3,1	5,8	12,9	18	2,1	2,8	3,9
4	2,7	4,6	8,6	19	2,0	2,8	3,8
5	2,5	4,0	6,8	20	2,0	2,8	3,8
6	2,4	3,7	5,9	21	2,0	2,8	3,8
7	2,3	3,5	5,4	22	2,0	2,8	3,7
8	2,3	3,3	5,1	23	2,0	2,8	3,7
9	2,2	3,2	4,7	24	2,0	2,7	3,7
10	2,2	3,1	4,6	25	2,0	2,7	3,7
11	2,2	3,1	4,4	26	2,0	2,7	3,7
12	2,2	3,0	4,3	27	2,0	2,7	3,6
13	2,1	3,0	4,2	28	2,0	2,7	3,6
14	2,1	2,9	4,1	29	2,0	2,7	3,6
15	2,1	2,9	4,0	30	2,0	2,7	3,6

Зразок виконання завдання.

Оскільки в даному випадку $n < 30$, похибка середньої визначається за формулою $m_m = \sigma / \sqrt{n - 1} = 4,2 / \sqrt{8} = 1,5$ уд./хв.

Довірчий інтервал ($tm=\Delta$) середньої величини ($M_{\text{ген}}$) визначаємо шляхом знаходження довірчого коефіцієнта t за таблицею Стьюдента:

а) при $p = 95\%$ і при $n = 9$ $t = 2,3$, отже, $tm(\Delta) = 2,3 \cdot 1,5 = 3,45$ уд./хв. $M_{\text{ген}} = 67 \pm 3,45$ уд./хв., тобто в генеральній сукупності при $p = 95\%$ середня величина пульсу коливається від 63,55 до 70,45 ударів за хвилину,

б) при $p = 99\%$ $t = 3,3$, $tm(\Delta) = 3,3 \cdot 1,5 = 4,9$ ударів за хвилину; $M_{\text{ген}} = 67 \pm 4,9$ ударів за хвилину, тобто в генеральній сукупності при $p = 99\%$ середня частота пульсу знаходиться в межах від 62,1 до 71,9 ударів за хвилину.

Завдання 2

Визначення похибки (m) і довірчих меж середньої арифметичної або відносної величини при великому числі спостережень.

При великому числі спостережень ($n > 30$) похибки репрезентативності визначаються за формулою

$$m_m = \sigma / \sqrt{n}$$

для відносних величин — похибка відносної величини (m_p):

$$m_p = \sqrt{p \cdot q / n}$$

Позначення символів — див. вище.

Типове завдання

Обчислити похибку (m_m) і визначити довірчі межі середньої арифметичної (M) $p = 95\%$ і $p = 99\%$ для наступних початкових даних:

Середня тривалість лікування в поліклініці 45 хворих ангіною склала 10,7 днів. Середнє квадратичне відхилення = 4,02 дня.

Зразок виконання завдання.

1. Оскільки в даному випадку $n > 30$, похибка середньої визначається за формулою

$$m_m = \sigma / \sqrt{n} = 4,02 / \sqrt{45} = 4,02 / 6,7 = 0,6 \text{ дня}$$

2. Довірчий інтервал ($tm = \Delta$) знаходимо таким чином:
при $n > 30$ і при $p = 95\%$ $t = 2$, при $n > 30$ і при $p = 99\%$ $t = 3$, отже, Δ відповідно буде: $2 \times 0,6 = 1,2$ дня і $3 \times 0,6 = 1,8$ дня.

Таким чином, при $p = 95\%$ $M_{\text{ген}} = 10,7 \pm 1,2$ дня. Це означає, що в генеральній сукупності середня тривалість лікування коливається від 9,5 до 11,9 днів. При $p = 99\%$ $M_{\text{ген}} = 10,7 \pm 1,8$ днів, тобто в генеральній сукупності середня тривалість лікування перебуватиме в межах від 8,9 до 12,5 днів

Завдання 3

Визначення вірогідності різниці між середніми і відносними величинами за критерієм t (Ст'юдента)

Вірогідність різниці між двома середніми величинами (M_1 і M_2) або між двома відносними величинами (P_1 і P_2) визначається за формулами:

$$t = \frac{M_1 - M_2}{\sqrt{m_{m1}^2 + m_{m2}^2}} \quad \text{і} \quad t = \frac{P_1 - P_2}{\sqrt{m_{p1}^2 + m_{p2}^2}}$$

Величина t повинна бути рівна або більше 2. Тільки в цьому випадку із ймовірністю безпомилкового прогнозу, рівною 95% і вище, можна стверджувати, що є істотні відмінності між порівнюваними середніми або відносними величинами.

Типове завдання

Потрібно визначити, чи є достовірне зниження частоти пульсу і наближення його до норми в групі студентів після

іспиту, якщо відомо, що середня частота пульсу (M_1) до іспиту склала 98,8 удару в хвилину ($m_{M1} = 4$ удари за хвилину); після іспиту (M_2) — 84 удари за хвилину ($m_{M2} = 5$ ударів за хвилину).

Зразок виконання завдання.

Достовірність різниці між середніми величинами визначається за формулою:

$$t = \frac{M_1 - M_2}{\sqrt{m_{M1}^2 + m_{M2}^2}} \quad \text{і} \quad t = \frac{P_1 - P_2}{\sqrt{m_{P1}^2 + m_{P2}^2}}$$

Оскільки $t > 2$, можна із вірогідністю безпомилкового прогнозу (p), більше 95% стверджувати, що після іспиту частота пульсу у студентів знижується і наближається до норми.

Навчальні завдання

Завдання 1

Визначення похибки (m_M) і довірчих меж середньої арифметичної при малому числі спостережень.

На основі наведених початкових даних потрібно обчислити похибку (m_M) і визначити довірчі межі середньої (M) при $p = 95\%$ і $p = 99\%$.

Варіант 1.

Середній вміст жиру в материнському молоці у 7 жінок складав 4,4 гр. Середнє квадратичне відхилення $\sigma = 0,33$ гр.

Варіант 2.

Середній рівень систолічного артеріального тиску у 11 дітей віком 7 років, що страждають на хворобу нирок, скла-

дав 110 мм. рт. ст. Середнє квадратичне відхилення $\sigma = 8,0$ мм. рт. ст.

Варіант 3.

Середня кількість зцідженого і висмоктаного молока у 8 жінок під час годування дитини однією груддю складала 100 г. Середнє квадратичне відхилення $\sigma = 7,0$ г.

Варіант 4.

Середня температура тіла у 7 новонароджених складала $36,9$ °С. Середнє квадратичне відхилення $\sigma = 2,7$ °С.

Варіант 5.

Середній зріст 10 хлопчиків віком 2 роки складав 93 см. Середнє квадратичне відхилення $\sigma = 6,5$ см.

Завдання 2

Визначення похибки (m_M) і довірчих меж середньої арифметичної при великому числі спостережень.

На основі наведених початкових даних потрібно обчислити похибку (m_M) і визначити довірчі межі середньої (M) при $p = 95\%$ і $p = 99\%$.

Варіант 1.

Середня довжина тіла при народженні у 32 хлопчиків складала 52 см. Середнє квадратичне відхилення $\sigma = 3,8$ см.

Варіант 2.

Середня маса тіла у 35 новонароджених хлопчиків складала 3,8 кг, Середнє квадратичне відхилення $\sigma = 0,27$ кг.

Варіант 3.

Середнє число дітей, хворих на ревматизм, віком до 15 років, які перебували на диспансерному обліку у 40 лікарів-ревматологів, складало 34 осіб. Середнє квадратичне відхилення $\sigma = 2,2$ чол.

Варіант 4.

Середнє число дітей віком до 1 року, що перебувають на обліку у 34 дільничних педіатрів, складало 47 чол. Середнє квадратичне відхилення $\sigma = 3,8$ чол.

Варіант 5.

Середня частота пульсу у 40 дівчат — першокласниць складала 74 удари за хвилину. Середнє квадратичне відхилення $\sigma = 5,0$ уд. / хв.

Завдання 3

Визначення достовірності різниці між середніми і відносними величинами за критерієм t (Ст'юдента). На основі наведених початкових даних потрібно оцінити достовірність різниці між двома середніми величинами (M_1 і M_2) або між двома відносними величинами (P_1 і P_2).

Варіант 1.

При вивченні успішності студентів медичного університету, які не працюють, та тих, що поєднують навчання з роботою, були отримані наступні дані: у непрацюючих середній бал (M_1) = 4,1 ($m_{M_1} = \pm 0,09$), у тих, що поєднують навчання з роботою (M_2) = 3,65 ($m_{M_2} = \pm 0,05$).

Варіант 2.

При вивченні працездатності у хворих, які перенесли інфаркт міокарду за наявності гіпертонічної хвороби та без

неї, були отримані наступні дані: число тих, хто повернувся до праці, серед тих, хто переніс інфаркт міокарду з гіпертонічною хворобою (P_1) = 61,0 % ($m_{p1} = \pm 4,0$ %), а без гіпертонічної хвороби (P_2) = 75,0% ($m_{p1} = \pm 3,0$ %).

Варіант 3.

При вивченні частоти ускладнень після апендектомії в 2 групах хворих, в одній з яких застосовувалися антибіотики, а в іншій не застосовувалися, були отримані наступні дані: у першій групі (P_1) ускладнення мали 30,0% хворих ($m_p = \pm 5,1$ %), в другій групі (P_2) = 40,0% ($m_p = \pm 5,4$ %).

Варіант 4.

У студентів-медиків досліджували частоту пульсу (за хвилину) до і після складання іспиту. Частота пульсу в середньому до іспиту (M_1) склала 94,2 удари за хвилину ($m_{M1} = \pm 3,9$ ударів за хвилину), після іспиту $M_2 = 82$ удари в хвилину ($m_{M2} = \pm 4,1$ удари в хвилину).

Варіант 5.

У групах хворих коронарним атеросклерозом досліджували вплив холіну на вміст холестерину сироватки крові. Вміст холестерину сироватки до застосування холіну в середньому (M_1) склав 231 мг% ($m_{M1} = \pm 4,0$ міліграм %), після застосування холіну — $M_2 = 204,0$ мг% ($m_{M2} = \pm 3,0$ мг%).

Контрольні запитання

- Що є науковим підґрунтям вибіркового методу?
- Назвіть основні положення закону великих чисел.
- Що таке достовірність, оцінка достовірності результатів статистичного дослідження?

- Для чого проводиться оцінка достовірності результатів статистичного дослідження?
- Які існують методи оцінки достовірності результатів статистичного дослідження?
- Які методи оцінки достовірності результатів статистичного дослідження належать до параметричних, для чого вони застосовуються?
- Які елементи включає оцінка достовірності результатів статистичного дослідження параметричним методом?
- Що таке вірогідність безпомилкового прогнозу?
- Як розрахувати похибки репрезентативності для середніх і відносних величин (при малому і великому числі спостережень)? Які початкові дані для цього необхідні?
- З якою метою використовуються похибки репрезентативності?
- Що є довірчими межами середніх і відносних величин? Яким чином вони визначаються, які початкові дані для цього необхідні?
- З яких етапів складається розрахунок довірчих меж середніх і відносних величин?
- Як визначити достовірність різниці між похідними величинами (середніми, відносними)? Які початкові дані для цього потрібні?

Тема 6

Непараметричні методи оцінки і аналізу статистичних гіпотез

Непараметричні критерії оцінки — це сукупність статистичних методів, які дозволяють оцінити результати досліджень без обчислення загальноприйнятих параметрів (M , σ , m тощо)

Переваги непараметричних методів (критеріїв):

- не вимагають знання характеру (типу) розподілу, тому можуть застосовуватися за будь-яким розподілом;
- можуть застосовуватися при будь-якому, навіть невеликому числі спостережень;
- можуть застосовуватися для ознак, що мають кількісне вираження, і ознак напівкількісного характеру (наприклад, ступінь тяжкості захворювання, результати лікування);
- відносно прості і не вимагають проведення складних розрахунків, відповідно економлять час при обчисленні;
- мають достатню потужність (чутливість) параметрів.

В основу розрахунків непараметричних критеріїв покладено впорядкування (ранжування) наявних значень по відношенню один до одного, типу «більше — менше» або «краще — гірше». Це розмежування значень не припускає точних кількісних співвідношень, отже обмежень на параметри і вид розподілу. Тому для використання непараметричних критеріїв треба менше інформації, ніж для параметричних критеріїв.

В якості оцінок при непараметричних методи використовуються відносні характеристики — ранги, серії, знаки та ін.

Якщо в ситуації можливе застосування параметричних критеріїв (нормальний розподіл ознаки і незначна різноманітність ознаки в сукупності) то їм, як більш інформативним, слід віддати перевагу, вони є більш потужними, ніж непараметричні критерії, хоча й більш трудомісткими.

Використання непараметричних критеріїв пов'язано з такими поняттями, як «нульова гіпотеза», рівень значущості, достовірність статистичних відмінностей.

«Нульова гіпотеза» — це припущення про те, що в порівнюваних групах відсутня різниця у розподілі частот.

Рівень значущості — це така ймовірність, яку приймають за основу при оцінці статистичної гіпотези. В якості максимального рівня значущості, при якому «нульова гіпотеза» ще відхиляється, приймається 5%. При рівні значущості більше 5% «нульова гіпотеза» приймається, відмінності між порівнюваними сукупностями приймаються статистично недостовірними, незначними.

Особливу увагу заслуговує питання про потужність (чутливість) критеріїв. Кожен з досліджуваних критеріїв має характерну для себе потужність. Оцінку значущості розходжень необхідно починати з найменш потужного критерію. Якщо цей критерій спростовує нульову гіпотезу, то на цьому аналіз закінчується. Якщо ж нульова гіпотеза цим критерієм не спростовується, то треба перевірити досліджувану гіпотезу більш потужним критерієм. Однак, якщо значення характеристики, обчисленої для менш потужного критерію, виявилось дуже далеким від критичного значення, то мало надії, що більш потужний критерій спростує нульову гіпотезу.

Для вибору того чи іншого непараметричного критерію, оцінки результатів медичних досліджень (для визначення суттєвості відмінностей двох сукупностей) необхідно визначити наступні моменти:

- в якому вигляді отримані результати: в кількісному або альтернативному (атрибутивному), тобто, представлені числом або альтернативною (атрибутивною, двохваріантною) оцінкою «є ознака» — «нема ознаки», «є симптом» — «нема симптому»;

- чи пов'язані між собою порівнювані вибіркові сукупності або вони взаємно незалежні;

- порівнюються дві або декілька вибірових сукупностей.

До пов'язаних між собою відносяться вибіркові сукупності з попарно пов'язаними варіантами, наприклад при вивченні кількості гемоглобіну в крові одних і тих самих хворих до і після лікування, різних фізіологічних показників у спортсменів в нормі, перед стартом і після закінчення змагань і т. ін.

Взаємно незалежні сукупності не пов'язані між собою і можуть мати різну чисельність, наприклад, результати дослідження крові у декількох груп хворих з різними стадіями хвороби, результати спостережень над піддослідній і контрольній групами дослідження і т. ін.

Методика виконання навчальних завдань

Застосування непараметричних критеріїв для визначення суттєвості відмінностей взаємопов'язаних (парних) сукупностей.

Критерій знаків на відміну від критерію Стьюдента (t) при оцінці парних спостережень (наприклад, до і після лікування) враховує не величину змін які відбулись, а тільки їх спрямованість. Тому характер цих змін враховується в альтернативній формі (збільшення — зменшення, погіршення — поліпшення і т.д.), що для кратності звичайно позначається знаками «+» або «-», звідки і відбулася назва критерію.

рію). Випадки, коли парні спостереження не мають різниці (що можна позначити $=$ або 0), з подальшого порівняння виключаються. Якщо число позитивних змін близько до числа негативних змін, то очевидно, що відмінності між ними не можуть бути визнані статистично значущими. Навпаки, вірогідність значущого розходження зростає у випадках помітної зміни спрямованості в одну із сторін, тобто у випадках переважання одного із знаків.

Завдання 1.

Визначити та оцінити критерій знаків (Z) за даними, наведеними в табл. 1.1-1.3, використовуючи типовий приклад і критичні значення Z -числа знаків, що рідше зустрічаються.

Типовий приклад

Зміна маси тіла після дієти у 10 осіб, кг

До дієти	Після дієти	Спрямованість змін
75	74	-
82	82	=
73	75	+
85	83	-
80	78	-
78	80	+
87	85	-
88	90	+
90	88	-
94	83	-

Зразок виконання завдання

Спрямованість змін парних спостережень позначають знаками “+”, “—” та “=”. Варіанти без змін (“=”) не беруть до уваги в підрахунку загальної кількості спостережень (у нашому прикладі $n = 9$).

Окремо підраховують кількість спостережень з позитивними (3) та негативними (6) результатами (це числа знаків Z).

Порівнюють менше число знаків з критичними значеннями. Якщо одержана величина менша за табличну, то різниця достовірна (істотна). У прикладі при $n = 9$, $Z = 3$, що перевищує табличне значення для цієї кількості спостережень ($Z = 2$) Отже, різниця недостовірна, вірогідність похибки більша за 5 % ($p > 0,05$).

Критичні значення Z — кількості знаків, що рідше зустрічаються (за В.Ю. Урбахом)

N	P		n	P		n	P	
	0,05	0,01		0,05	0,01		0,05	0,01
7	1	-	21-22	6	5	32	10	9
8	1	1	23	7	5	33	11	9
9-11	2	1	24	7	6	34	11	10
12-14	3	2	25	8	6	35-36	12	10
15 -16	4	3	26 -27	8	7	37-38	13	11
17	5	3	28	9	7	39	13	12
18-19	5	4	29	9	8	40-41	14	12
20	6	4	30-31	10	8			

Завдання 1

Таблиця 1.1.

Зміни кількості лейкоцитів порівняно з нормою після лікування новим препаратом (характер змін)

Лікарня	Хворий										
	1	2	3	4	5	6	7	8	9	10	11
1	+	-	-	=	-	-	-	+	-	-	-
2	-	=	-	+	-	-	-	-	-	-	-
3	-	-	-	-	-	-	+	-	-	-	=
4	-	=	-	-	+	-	-	-	+	-	-
5	-	+	-	-	-	+	-	-	-	-	-

Таблиця 1.2.

Зміни ШОЕ на 7-й день від початку протизапального лікування у хворих на пневмонію (спрямованість змін)

Лікарня	Хворий										
	1	2	3	4	5	6	7	8	9	10	11
6	+	-	-	=	-	-	-	+	-	-	-
7	-	=	-	+	-	-	-	-	-	-	-
8	-	-	-	-	-	-	+	-	-	-	=
9	-	=	-	-	+	-	-	-	+	-	-
10	-	+	-	-	-	+	-	-	-	-	-

Таблиця 1.3.

**Вплив гемосорбції на загально токсичні симптоми
у дітей з площею опіків 10 -15 % поверхні тіла
(спрямованість змін)**

Лікарня	Дитина										
	1	2	3	4	5	6	7	8	9	10	11
11	+	-	-	=	-	-	-	+	-	-	-
12	-	-	-	+	-	-	-	-	-	-	-
13	-	=	-	-	-	-	+	-	-	-	=
14	-	=	-	-	+	-	-	-	+	-	-
15	=	+	-	-	-	+	-	-	-	-	-

Завдання 2

Визначити й оцінити Т-критерій Вілкоксона за даними, наведеними в табл. 2.1-2.3, використовуючи типовий приклад і критичні значення Т-критерію Вілкоксона для взаємопов'язаних сукупностей.

**Рівень артеріального тиску у хворих на гіпертонічну
хворобу до та після лікування, мм. рт. ст.**

Хворий	Рівень артеріального тиску		Різниця	Ранг різниці	Значення рангів зі знаком ” — ”	Значення рангів зі знаком ” +”
	До лікування	Після лікування				
В	210	175	-35	6,5	6,5	4
К	185	140	-35	6,5	6,5	
Р.	160	185	+ 25	4		1
Н.	175	145	-30	5	5	
П.	190	150	-40	8	8	

А	155	160	+ 5	1		
С.	180	160	-20	3	3	
Ю.	200	155	-45	9	9	
Т.	170	155	- 15	2	2	
					T = 40	T = 5

Зразок виконання завдання

Визначають різницю в порівнюваних парах між кінцевим і початковим рівнями артеріального тиску.

Визначають ранг різниці для кожної з порівнюваних пар результатів (порядкові номери абсолютних значень, не враховуючи знаки “+” чи “-” — “у порядку зростання”). Ранги однакових значень дорівнюють півсумі двох порядкових місць.

Обчислюють окремо суму рангів для різниць із знаками “+” та “-”. Для подальшої оцінки використовують меншу суму рангів (у нашому прикладі $T = 5$).

Визначену меншу суму рангів T порівнюють із табличним значенням T для відповідної кількості спостережень. Якщо одержаний критерій T менший за табличний, різниця достовірна.

У нашому прикладі при $n = 9$, $T = 5$, що не перевищує табличного значення для рівня вірогідності $P < 0,05$ ($T = 7$).

Висновок. Різницю між рівнями артеріального тиску можна вважати достовірною.

Граничні значення T-критерію Вілкоксона для взаємопов'язаних сукупностей (за В. Ю. Урбахом)

n	P		n	P		n	P	
	0,05	0,01		0,05	0,01		0,05	0,01
6	1	-	13	18	11	20	53	39

7	3	-	14	22	14	21	60	44
8	5	1	15	26	17	22	67	50
9	7	3	16	31	21	23	74	56
10	9	4	17	36	24	24	82	62
11	12	6	18	41	29	25	90	69
12	15	8	19	47	33	-	-	-

Завдання

Таблиця 2.1

Вміст гемоглобіну в крові хворих з глибокими опіками через 1 та 2 доби після отримання опіку, г/л

Лікарня	Термін (діб)	Вміст гемоглобіну в крові досліджуваних хворих							
		1	1	130	125	120	110	130	115
	2	120	110	112	100	110	90	125	88
2	1	135	120	110	115	118	130	120	110
	2	120	110	112	120	100	110	95	98
3	1	120	115	100	130	125	125	118	114
	2	110	105	105	110	115	122	116	120
4	1	128	118	130	126	116	125	130	128
	2	120	126	120	110	100	110	118	112

Таблиця 2.2

Зміни бактерицидної активності сироватки крові у хворих на затяжну пневмонію на 5-й та 10-й день від початку захворювання, %

Лікарня	Період	Бактерицидна активність сироватки крові							
5	5-й день	96	95	93	94	90	95	90	89
	10-й день	86	96	87	86	91	84	83	92
6	5-й день	93	96	92	95	90	89	91	95
	10-й день	88	87	83	86	91	84	84	85

7	5-й день	90	92	94	96	88	95	93	91
	10-й день	85	88	87	90	89	92	83	93
8	5-й день	95	86	90	92	90	93	95	88
	10-й день	88	91	83	82	86	87	86	84

Таблиця 2.3

Зміни діастолічного артеріального тиску у хворих на гіпертонічну хворобу після лікування новим препаратом, мм. рт. ст.

Лікарня	Період	Артеріальний тиск							
9	До лікування	90	100	105	85	105	95	110	100
	Після лікування	85	80	90	90	115	85	85	75
10	До лікування	85	105	110	95	90	90	85	95
	Після лікування	80	110	100	100	80	85	80	80
11	До лікування	110	100	95	105	100	85	90	95
	Після лікування	105	110	80	100	90	90	80	80
12	До лікування	110	90	105	90	85	110	100	95
	Після лікування	100	85	110	80	80	100	75	90

Застосування непараметричних критеріїв для визначення суттєвості відмінностей незалежних сукупностей

Завдання

Визначити та оцінити критерій Колмогорова—Смирнова за даними, наведеними в табл. 1, використовуючи типовий приклад і граничні значення $\chi^2_{0,05} = 1,84$ та $\chi^2_{0,01} = 2,65$.

Типовий приклад

Таблиця 1

Збільшення маси тіла новонароджених доношених дівчаток за різних методів вигодовування (протягом 4 міс після народження), кг:

Метод I	(X)	1,6	1,8	2,0	2,0	2,1	2,3
Метод II	(y)	1,9	2,1	2,2	2,3	2,5	2,5

Варіанти x і y у висхідному порядку	Частота варіант по групах		Накопичені частоти по групах		Накопичені частки		Різниця $S_x/n_x - S_y/n_y$
			S_x	S_y	S_x/n_x	S_y/n_y	
1,6	1	0	1	0	0,166	0	0,166
1,8	1	0	2	0	0,333	0	0,333
1,9	0	1	2	1	0,333	0,166	0,167
2,0	2	0	4	1	0,666	0,166	0,500
2,1	1	1	5	2	0,833	0,333	0,500
2,2	0	1	5	3	0,833	0,555	0,333
2,3	1	1	6	4	1,000	0,666	0,334
2,5	0	2	6	6	1,000	1,000	0

$$n_x = 6, n_y = 6.$$

Зразок обчислення

1. Об'єднують в один ряд усі варіанти обох груп у висхідному порядку (графа 1).

2. Записують частоту варіант окремо в групах x і y (графи 2, 3).

3. Визначають накопичену частоту шляхом послідовного додавання до частоти поточної варіанти частот усіх попередніх варіанту відповідній групі (графи 4, 5). Сума накопичених частот не повинна перевищувати загальної кількості варіант по кожному ряду.

4. Накопичені частоти ділять на кількість спостережень у відповідних групах і визначають накопичені частки (графи 6, 7) Ці дані можна отримати з таблиць Колмогорова — Смирнова.

5. Обчислюють різницю накопичених часток груп x і y (не враховуючи знаків) (графа 8). Для подальших розрахунків використовують максимальну різницю

6. Визначають критерій χ^2 за формулою:

$$\chi^2 = \chi^2 = \frac{D^2 \cdot n_x \cdot n_y}{n_x \cdot n_y}$$

де D — найбільша різниця (за графою 8 — у прикладі 0,500).

7. Якщо χ^2 більше $0,05 = 1,84$ або $\chi^2 0,01 = 2,65$, то різниця істотна.

У нашому прикладі $\chi^2 = 0,75$.

Висновок. Різницю не можна вважати достовірною.

**Значення накопичених часток S/n за різних величин
накопичених частот S і кількості n
(для критерію Колмогорова — Смірнова)**

№	1	2	3	4	5	6	7	8	9
5	0,200	0,400	0,600	0,800	1,000				
6	0,166	0,333	0,500	0,666	0,833	1,000	-	-	-
7	0,143	0,286	0,482	0,571	0,714	0,857	1,000	-	-
8	0,125	0,250	0,375	0,500	0,625	0,750	0,875	1,000	-
9	0,111	0,222	0,333	0,444	0,555	0,666	0,777	0,888	1,000
10	0,100	0,200	0,300	0,400	0,500	0,600	0,700	0,800	0,900
11	0,091	0,182	0,273	0,364	0,455	0,545	0,636	0,727	0,818
12	0,083	0,166	0,250	0,333	0,417	0,500	0,583	0,666	0,750
13	0,077	0,154	0,231	0,307	0,385	0,461	0,538	0,615	0,692
14	0,071	0,143	0,214	0,286	0,357	0,428	0,500	0,571	0,643
15	0,066	0,133	0,200	0,266	0,333	0,400	0,466	0,533	0,600
16	0,062	0,125	0,187	0,250	0,312	0,375	0,437	0,500	0,562
17	0,059	0,118	0,176	0,235	0,235	0,353	0,412	0,471	0,529
18	0,055	1,112	0,166	0,222	0,222	0,333	0,388	0,444	0,500

Завдання

Таблиця 1.

**Вміст лімфоцитів у крові хворих на виразкову хворобу,
які мали ускладнення у вигляді пенетрації (група X),
і без ускладнень (група V), %**

Лікарня	Група	Вміст лімфоцитів							
		12	15	14	14	22	20	20	24
	X								
	V	20	25	25	23	29	28	30	27

2	X	14	18	15	15	19	20	21	21
	V	19	24	24	23	27	26	28	28
3	X	17	16	14	14	18	21	23	20
	V	24	18	26	26	23	25	29	27
4	X	18	17	15	14	19	21	24	21
	У	24	19	25	24	23	24	28	26
5	X	21	16	14	18	14	21	22	20
	У	24	18	26	23	26	25	28	27

Таблиця 2

**Активність фагоцитозу у хворих на гостру (група X),
та затяжну (група V), пневмонію, %**

Лікарня	Група	Активність фагоцитозу							
6	X	60	62	66	67	70	64	63	
	У	58	60	61	57	57	62	59	
7	X	61	63	64	67	65	68	60	
	У	55	60	62	56	57	62	58	
8	X	59	63	65	67	64	68	66	
	V	60	58	55	56	59	57	56	
9	X	60	62	63	66	64	67	60	
	V	56	60	61	57	58	64	57	
10	X	60	63	64	67	67	62	63	
	У	58	60	65	56	57	60	59	

Застосування непараметричних критеріїв для визначення суттєвості відмінностей будь-яких сукупностей.

Завдання

Визначити й оцінити критерій відповідності χ^2 (критерій Пірсона), використовуючи типовий приклад і таблицю значень критерію χ^2

Типовий приклад

Наслідки застосування трьох методів неспецифічної профілактики грипу

Метод	Кількість піддослідних	Із них		Очікувана кількість		P-P'		(P - P') ² /P'	
		захворіли	не захворіли	захворіють	не захворіють	захворіли	не захворіли	захворіли	не захворіли
		(P)		(P')					
I	140	80	60	42	98	38	-38	1444/42=34,4	14,7
II	232	60	172	70	162	-10	10	100/70=1.42	0,6
III	160	20	140	48	112	-28	28	784/48 = 16,3	7,0
Разом абсолютне число	532	160	372	160	372				
%	100	30,0	70,0						

1. Обирають «нульову» гіпотезу, в основу якої покладено припущення про відсутність різниці за ефективністю досліджуваних методів. У цьому прикладі це сумарний поділ піддослідних (у %) на тих, що захворіли (30 %), й тих, які не захворіли (70 %), за трьома методами.

2. Обчислюють «очікувану» абсолютну кількість (P') заданими — «нульової» гіпотези: скільки було б тих, що захворіли та не захворіли, відповідно до обраного поділу піддослідних (графа 5, 6).

Для I методу: із 140 піддослідних 30% хворих становлять 42 особи, відповідно не хворіли 98 осіб (70 %) і т. д.

3. Обчислюється різниця між фактичними та «очікуваними» числами в кожній групі ($P - P'$) (графа 7-8).

4. Визначений результат (графа 7-8) підноситься до квадрату й обчислюється частка від ділення квадратів різниць на «очікувані» числа по групах серед тих, що захворіли та не хворіли (графа 9-10).

У нашому прикладі $\chi^2 = 34,4 + 1,42 + 16,3 + 14,7 + 0,6 + 7,0 = 74,42$.

5. Обчислюється критерій відповідності за формулою

$$\Sigma = \frac{(p - p^1)}{p^1}$$

6. Оцінка χ^2 проводиться за таблицею для відповідної кількості «ступенів свободи» ($p1$). Для будь-якої сукупності кількість ступенів свободи є на 1 меншим від кількості варіант(груп) за відповідним критерієм оцінки. Для нашого прикладу p' визначається з кількості досліджуваних методів ($r = 3$) та кількості груп результатів ($s = 2$):

$$p' = (S - 1) \cdot (r - 1),$$

де S — кількість граф вихідної таблиці (без графи «Разом»),

r — кількість рядків без рядка «Разом», тобто $p' = (2-1) - (3-1) = 2$.

Для кількості ступенів свободи $p' = 2$ значення χ^2 0,05=5,991, χ^2 0,01= 9,210.

Одержаний результат $\chi^2 = 74,42$, що набагато перевищує табличне значення χ^2 0,01. Отже, «нульова гіпотеза» не під-

тверджується та різниця різних методів профілактики грипу є істотною.

Таблиця значень критерію χ^2

Кількість ступенів свободи n'	Ймовірність похибки (P)		
	0,05	0,01	0,001
1	3,8	6,6	9,5
2	6,0	9,2	12,4
3	7,8	11,3	14,8
4	9,5	13,3	16,9
5	11,1	15,1	18,9
6	12,6	16,8	20,7
7	14,1	18,5	22,6
8	15,5	20,1	24,3
9	16,9	21,7	26,1
10	18,3	23,2	27,7

Завдання

Задача 1

**Дані про загострення хронічних радикулоневритів
серед робітників ТЕЦ залежно від умов праці**

Умови праці	Разом хворих	Кількість загострень на рік	
		2	3 і більше
Вимушена поза	92	10	17
Різкі зміни температур	110	5	12
Сприятливі	320	4	9

Задача 2

**Поділ обстежених,
що мають знижений артеріальний тиск, за віком**

Вік, роки	Обстежено	Виявлено знижений тиск
18-25	250	80
26-35	300	70
35 і більше	115	15
Разом	665	165

Задача 3

**Дані про ускладнення серед прооперованих хворих,
які потребували екстреної допомоги,
залежно від терміну доставки**

Термін доставки	Кількість прооперованих	Мали ускладнення
До 24 годин	1400	42
Пізніше за 24 годин	340	24
Разом	1710	66

Задача 4.4

**Поділ осіб, що хворіли на грип,
залежно від застосованих методів профілактики**

Метод профілактики	Обстежено	Захворіли
Приймання ремантадину	120	12
Застосування оксолінової мазі	152	28
Щеплення протигрипозною вакциною	685	103

Задача 5

Поділ вагітних, що хворіють на анемію та хвороби нирок, залежно від проживання в умовно чистому чи забрудненому районі

Район	Обстежено жінок	Хворіли на анемію	Діагностовано хвороби нирок
Умовно чистий	560	140	17
Забруднений	1485	594	89

Задача 6

Поділ дітей, що народилися хворими, залежно від проживання їхніх сімей в умовно чистому чи забрудненому районі

Район	Кількість новонароджених	Народилися хворими
Умовно чистий	600	120
Забруднений	785	355

Задача 7

Дані про реакцію ШОЕ у робітників різних цехів

Цех	Обстежено	Була прискорена реакція ШОЕ
Ливарний	420	70
Інструментальний	320	30
Сортувальний	280	8
Разом	1020	108

Задача 8

**Розподіл чоловіків, що померли від раку легень
в залежності від тютюнокуріння**

	Не курили	Курили люльку	Курили цигарки	Разом
Померли від раку легень	480	315	1290	2085
	40	55	380	475

Задача 9

**Поділ робітників цементного заводу, що хворіють
на хронічний бронхіт, залежно від стажу роботи**

Тривалість стажу, роки	Обстежено	Хворіють на хронічний бронхіт
До 5	185	25
5-9	130	35
10-14	85	40
Разом	400	100

Задача 4.10

**Дані про виявлення порушень постави
під час обстеження школярів різного віку**

Вік, роки	Обстежено	Виявлено сколіоз
7	200	18
10	400	65
14	300	77
Разом	900	160

Контрольні запитання

- Умови для використання параметричних критеріїв оцінки вірогідності різниці результатів дослідження.

- Можливості та переваги використання непараметричних критеріїв.

- Характеристика незалежних і взаємопов'язаних сукупностей.

- Непараметричні критерії, які використовують для оцінки різниці результатів дослідження у взаємопов'язаних сукупностях, методика їхнього визначення та оцінка.

- Непараметричні критерії, які використовують для оцінки різниці в незалежних сукупностях, методика їхнього визначення та оцінка.

- Практичне значення критерію відповідності, методика визначення та оцінка.

Тема 7

Аналіз взаємозв'язку між параметрами статистичних сукупностей

Кореляційно-регресійний аналіз — спеціальний статистичний метод, який дозволяє встановити напрямок і силу зв'язку між ознаками, а також судити за величиною якої-небудь ознаки про середні розміри іншої, пов'язаної з нею ознаки, без її вимірювання.

Існує 2 види (форми прояву) кількісних зв'язків між ознаками: функціональний і кореляційний.

Функціональним називається зв'язок, при якому кожному значенню однієї ознаки відповідає строго певне значення іншої ознаки. Наприклад, із збільшенням радіусу обов'язково збільшується довжина і площа кола. Цей вид зв'язку характерний для фізико-хімічних явищ

Кореляційний зв'язок (кореляція) — зв'язок, при якому кожному значенню однієї ознаки відповідає декілька значень іншої ознаки. Вона виявляється лише при масовому зіставленні явищ, що вивчаються. У кожному окремому випадку вона може не виявлятися. Прикладами кореляційного зв'язку можуть служити: залежність маси тіла від зросту, залежність рівня захворюваності на дифтерію від ступеня охоплення населення щепленнями проти цього захворювання, залежність між термінами проведення операції при гострих захворюваннях черевної порожнини і частотою післяопераційних ускладнень тощо. В цілому такий вид зв'язку характерний для соціально-гігієнічних процесів, клінічної медицини і біології.

Кореляційний зв'язок є статистичним зв'язком. Перш ніж застосувати статистичний метод для вивчення кореля-

ційного зв'язку необхідно заздалегідь вирішити питання про те, чи є насправді який-небудь зв'язок між явищами, шляхом проведення конкретного аналізу матеріальної природи явища, що вивчається, тієї області знань, в якій передбачається застосувати метод кореляції. Шляхом же статистики визначається тільки розмір (сила, тіснота) зв'язку і встановлюється ступінь залежності між явищами, що вивчаються.

За своїм характером (напрямком) кореляційний зв'язок може бути прямим і зворотнім, а за силою — сильним, середнім і слабким. Крім того, зв'язок може бути відсутнім або бути повним.

Прямим (позитивним) називається такий зв'язок, при якому із збільшенням одного явища (ознаки) збільшується інше.

При зворотному (негативному) зв'язку — із збільшенням одного явища (ознаки) інше зменшується.

Кореляційний зв'язок може бути прямолінійним і криволінійним.

Прямолінійний зв'язок характеризується відносно рівномірною зміною середніх значень однієї ознаки при рівних змінах іншої.

У випадках криволінійної залежності при рівномірній зміні однієї ознаки можуть спостерігатися зростаючі і убуваючі середні значення іншої ознаки.

Кореляційний аналіз застосовується з метою:

- а) визначення сили і напрямку зв'язку між ознаками;
- б) визначення провідних чинників, що формують рівні ознак.

Кореляція може бути представлена у вигляді таблиць, графіків або коефіцієнтів: коефіцієнта кореляції r , ρ (якщо зв'язок прямолінійний) або кореляційного відношення η (при криволінійному характері зв'язку).

Характер і розмір прямолінійного зв'язку встановлюється за допомогою коефіцієнта кореляції.

Існують наступні методи визначення коефіцієнтів кореляції:

а) метод визначення лінійного зв'язку способом квадратів (Пірсона);

б) ранговий метод (Спірмена).

Формули для розрахунку надаються нижче.

Метод лінійної кореляції застосовується для оцінки зв'язку в тих випадках, коли результати представлені точними кількісними характеристиками.

Основою методу лінійної кореляції є:

а) визначення середніх величин;

б) розрахунок відхилень варіант від середніх величин.

Метод рангової кореляції використовується у випадках, коли результати надані:

а) описово,

б) у напівкількісному вигляді,

в) точними кількісними характеристиками.

Основою методу рангової кореляції є ранжування (встановлення порядкових номерів) рядів, що співставляються.

Щоб переконатися в тому, що коефіцієнт кореляції, обчислений за даними вибіркового дослідження, відповідає розміру зв'язку в генеральній сукупності, необхідно визначити середню похибку коефіцієнта кореляції (m) і критерій t , методика розрахунку і оцінки яких наводяться нижче.

Коефіцієнт кореляції указує лише на напрямок і силу зв'язку між двома граничними величинами, але не дає можливості припустити, як кількісно зміняться величини внаслідок зміни величин іншої ознаки.

Відповідь на це питання дає застосування методу регресії.

Регресія — функція, що дозволяє за величиною однієї ознаки визначити середні величини іншої ознаки. За допомогою регресії ставиться завдання з'ясувати, як кількісно зміниться одна величина при зміні іншої величини на одиницю. Для визначення розміру цієї зміни застосовується коефіцієнт регресії.

Коефіцієнт регресії R_{xy} — абсолютна величина, на яку в середньому змінюється ознака, при зміні іншої ознаки на одиницю.

За допомогою коефіцієнта регресії без спеціальних вимірювань можна визначити величину однієї з ознак (наприклад, маси тіла), знаючи значення іншої (зросту). Для цих цілей служить рівняння лінійної регресії.

У практиці дослідження фізичного розвитку дітей і підлітків є дуже поширеним метод оцінки показників зросту, маси тіла кола грудної клітини за шкалою регресії. Індивідуальні значення окремих ознак дуже різноманітні: у людей з однаковим зростом показники маси тіла і кола грудної клітки можуть коливатися в широких межах.

Міру різноманітності індивідуальних ознак характеризує сигма регресії σR_{xy}

$$\sigma R_{xy} = \sigma_y \cdot \sqrt{1 - r_{xy}^2}$$

де σ_y — середнє квадратичне відхилення вимірюваної ознаки (наприклад, маси); r_{xy} — коефіцієнт кореляції. Чим менше значення сигми регресії, тим більш вузькими будуть межі коливань індивідуальних ознак щодо середньої. Знаючи коефіцієнт регресії, використовуючи рівняння регресії і сигму регресії, можна побудувати шкалу регресії.

Приклад розрахунку шкали регресії:

Виміряно зріст (x) і вагу (y) хлопчиків 6 років.

$M_x = 118,8$ см; $\sigma_x = \pm 5,8$ см; $M_y = 24,1$ кг; $\sigma_y = \pm 2,6$ кг; $r_{xy} = \pm 0,7$

Коефіцієнт регресії маси тіла за зростом $R_{xy} = 0,7 \cdot 2,6 / 5,8 = 0,3$ кг/см.

За значенням знаходимо масу тіла за допомогою рівняння регресії $y = (R_{xy} \cdot x) + a$ де $a = M_y - (R_{xy} \cdot M_x)$; $a = 24,1 - (0,3 \cdot 118,8) = -11,54$.

Знаходження маси тіла за зростом:

Середній зріст, см	Середня маса тіла, кг
$x_1 = 100$	$y_1 = 0,3 \cdot 100 - 11,54 = 18,46$
$x_2 = 110$	$y_2 = 0,3 \cdot 110 - 11,54 = 21,46$
$x_3 = 120$	$y_3 = 0,3 \cdot 120 - 11,54 = 24,46$

Для визначення меж коливань індивідуальної маси тіла по середньому зросту використовують сигму регресії:

$$\sigma R_{xy} = \sigma_y \cdot \sqrt{1 - r^2_{xy}} = 2,6 \cdot \sqrt{1 - 0,7^2} = \pm 1,3 \text{ кг}$$

Таблиця 1

Зріст, x	Маса, y ±	Межі коливань
100	18,46 ± 1,3 кг	19,76 + 17,16 кг
110	21,46 ± 1,3 кг	22,76 + 20,16 кг
120	24,46 ± 1,3 кг	25,76 + 23,16 кг

Застосовуючи регресію в дослідженні, можна за величиною якої-небудь ознаки судити про середні розміри іншої, взаємопов'язаної з нею ознаки, при цьому не проводити кожного разу нових дослідів і нових вимірювань. Наприклад, за середнім зростом можна судити про можливий рівень маси, за рівнем максимального тиску крові — про мінімальний рівень, за кількістю захворювань — про середню кількість бацилоносіїв тощо. На цьому принципі будують номограму — спеціальний графік — сітку, на якій відкладають значення ознаки на одній лінії, а на іншій — відповідні значення іншої ознаки.

Методика виконання навчальних завдань

Завдання 1

Обчислення рангової кореляції (Спірмена) і оцінка його достовірності.

Коефіцієнт кореляції рангів визначається за формулою:

$$\rho_{xy} = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)}$$

де x і y — ознаки, між якими визначається зв'язок; b — постійний коефіцієнт, d — різниця рангів, n — число парних спостережень.

Для визначення імовірності коефіцієнта кореляції обчислюється його похибка (m_p):

$$m_p = \sqrt{\frac{1 - \rho_{xy}^2}{n - 1}}$$

Достовірність коефіцієнта кореляції визначається за значенням критерію t :

$$t = \frac{\rho_{xy}}{m}$$

Критерій t має дорівнювати або бути більше 3, що відповідає вірогідності безпомилкового прогнозу (p) >99 %.

Послідовність розрахунку коефіцієнта рангової кореляції:

1. Скласти ряди з парних ознак (x і y). Спочатку проаналізувати зв'язок графічним методом.

2. Кожну величину ознаки замінити ранговим (порядковим) номером — x_1 і y_1 . У тих випадках, коли є декілька однакових за величиною чисел, порядковий номер позначають середнім числом з суми їх чергових порядкових номерів.

3. Визначити різницю рангів $d = x_1 - y_1$

4. Звести в квадрат різницю рангів d^2 .

5. Отримати суму квадратів різниці рангів $\sum d^2$.

6. Визначити ρ за формулою.
7. Визначити напрямок і силу зв'язку (за табл. 2).
8. Визначити похибку m_p , критерій t і оцінити достовірність ρ .
9. Зробити висновки.

Таблиця 2

Схема оцінки характеру і сили кореляційного зв'язку за коефіцієнтом кореляції

Сила зв'язку	Характер зв'язку	
	Прямий	Зворотній (-)
Повний	+1	-1
Сильний	від + 1 до +0,7	від -1 до -0,7
Середній	від +0,69 до +0,3	від -0,69 до -0,3
Слабкий	від +0,29 до 0	від -0,29 до 0
Зв'язок відсутній	0	0

Типове завдання

Відомо, що із збільшенням віку збільшується маса тіла. Слід перевірити величину цього зв'язку шляхом рангової кореляції.

Таблиця 3

Результати вимірювання маси тіла у студентів медичного університету відповідно до їх віку

Порядковий номер студента	Вік, роки життя (x)	Маса тіла в кг (y)
1	22	56
2	22	57
3	21	59
4	24	60
5	23	63
6	23	65
7	24	67

8	23	72
9	24	79
10	24	82

Зразок виконання типового завдання

Для цього складається таблиця за нижченаведеною формою і обчислення проводиться таким чином:

Таблиця 4

Обчислення коефіцієнта кореляції методом рангів

Ранги за віком і масою тіла		Різниця рангів (d)	Квадрат різниці рангів d ²
Вік в роках (x)	Маса тіла в кг (y)		
2,5	1	-1,5	2,25
2,5	2	+0,5	0,25
1	3	-2	4,0
8,5	4	+4,5	20,25
5	5	0	0
5	6	-1	1
8,5	7	+1,5	2,25
5	8	-3	9,0
8,5	9	-0,5	0,25
8,5	10	-1,5	2,25
			Σd ² = 41,5

Підставляємо отримані дані у формулу розрахунку коефіцієнта кореляції методом рангів:

$$r_{xy} = 1 - \frac{6 \cdot 41,5}{990} = 1 - 0,25 = + 0,75$$

Коефіцієнт кореляції, що дорівнює + 0,75, свідчить про наявність прямого сильного зв'язку між віком студентів і масою їх тіла.

Визначаємо достовірність коефіцієнта кореляції:

а) обчислюємо його похибку:

$$m_p = \sqrt{\frac{1-0,64}{8}} = \sqrt{0,055} = 0,23$$

б) визначаємо довірчий коефіцієнт (t) і ступінь вірогідності безпомилкового прогнозу (p):

$$t = \frac{0,75}{0,23} = 3,26 \text{ при } t = 3,26 \text{ } p > 99\%$$

Завдання 2

Обчислення коефіцієнта лінійної кореляції (Пірсона) і оцінка його достовірності. Розрахунок коефіцієнта і рівняння лінійної регресії.

Коефіцієнт лінійної кореляції r_{xy} визначається за формулою:

$$r_{xy} = \frac{\Sigma(d_x \times d_y)}{\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}}$$

де x і y — ознаки, між якими визначається зв'язок; d_x і d_y — відхилення кожної варіанти від середньої величини, обчисленої у ряді ознаки x і у ряді ознаки y ; Σ — знак суми.

Для визначення достовірності коефіцієнта кореляції обчислюється його похибка (m_r):

$$m_r = \frac{1 - r_{xy}^2}{n - 2}$$

Достовірність коефіцієнта кореляції визначається за формулою:

$$t = \frac{r_{xy}}{m_r}$$

Критерій t повинен має дорівнювати або бути більшим 3, що відповідає вірогідності безпомилкового прогнозу (p) > 99%.

Послідовність розрахунку:

1. Побудувати варіаційні ряди з парних ознак x і y .
2. Визначити їх середні величини M_x і M_y .

3. Знайти d — відхилення кожної варіанти від середньої для ряду x ($d_x = x - M_x$) і для ряду y ($d_y = y - M_y$).

4. Отримані відхилення перемножити ($d_x \cdot d_y$) і підсумувати $\Sigma (d_x \cdot d_y)$.

5. Кожне відхилення звести в квадрат і підсумовувати по ряду x : Σd_x^2 і по ряду y : Σd_y^2

6. Визначити добуток $\Sigma (d_x \cdot \Sigma d_y)$ і з добутку обчислити квадратний корінь.

7. Вирахувати r_{xy} .

8. Визначити напрямок і силу зв'язку (за табл. 2).

9. Визначити похибку m_p , критерій t і оцінити достовірність g .

10. Зробити висновки.

Типове завдання

Потрібно визначити, чи є залежність між температурою тіла і частотою пульсу (табл. 5).

Таблиця 5.

Залежність між температурою тіла і частотою серцевих скорочень

Температура тіла, x	Частота пульсу в хвилину, y	d_x	d_y	$d_x \cdot d_y$	d_x^2	d_y^2
36	60	-2	-20	40	4	400
36	70	-2	-10	20	4	100
38	80	0	0	0	0	0
40	90	+2	+10	20	4	100
40	100	+2	+20	40	4	400
$x = 190$	$y = 400$	$d_x = 0$	$d_y = 0$	$\Sigma (d_x \cdot d_y) = 120$	$\Sigma d_x^2 = 16$	$\Sigma d_y^2 = 1000$
$M_x = 190/5 = 38^\circ\text{C}$	$M_y = 400/5 = 80$ уд. в хв.					

$$r_{xy} = \frac{(d_x \cdot d_y)}{\sqrt{d_x^2 \cdot d_y^2}} = \frac{120}{\sqrt{16 \times 100}} = +0,952$$

Обчислення середньої похибки коефіцієнта кореляції (m_r) і критерію t :

$$m_r = \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \sqrt{\frac{1 - 0,952^2}{5 - 2}} = \pm 0,19$$

$$t = \frac{r_{xy}}{m_r} = \frac{0,952}{0,19} = 5,0$$

Висновок: Результати вибіркового дослідження дозволяють стверджувати з вірогідністю безпомилкового прогнозу (p) > 99%, що в генеральній сукупності між температурою тіла і частотою пульсу є прямий і сильний зв'язок.

Для розрахунку коефіцієнта регресії R користуються наступною формулою:

$$R_{xy} = r_{xy} \cdot \frac{\sigma_y}{\sigma_x}$$

Якщо у позначимо вагу, x — зріст, то $R_{y/x}$ — коефіцієнт регресії ваги за зростом, σ_y — середнє квадратичне відхилення ряду зростання, σ_x — середнє квадратичне відхилення ряду ваги, % — коефіцієнт кореляції ваги і зросту.

За допомогою коефіцієнта регресії без спеціальних вимірювань можна визначити величину однієї з ознак (наприклад, маси тіла), знаючи значення іншої (зросту). Для цих цілей служить рівняння лінійної регресії:

$$y = M_y + R_{y/x} (x - M_x)$$

де y — шукана величина маси тіла; x — відома величина зросту; $R_{y/x}$ — коефіцієнт регресії маси тіла за зростом; M_y — середнє значення маси тіла, характерне для даного віку; M_x — середнє значення зросту.

Послідовність розрахунку:

За формулою $\sigma = \sqrt{\frac{\sum d^2}{n - 1}}$ отримати значення середніх квадратичних відхилень для ряду x і ряду y ($\sigma_{x,y}$).

За формулою визначити коефіцієнти регресії.

Типове завдання

Визначити коефіцієнт регресії $R_{y/x}$ маси тіла 7-річних дівчаток за їх зростом, якщо відомо, що середнє квадратичне відхилення зросту (σ_x) дівчаток цього віку дорівнює $\pm 3,4$, середньоквадратичне відхилення маси тіла (σ_y) дорівнює $2,6$; коефіцієнт кореляції зросту і ваги $r_{y/x} = +0,60$.

Зразок вирішення типового завдання.

$$R_{y/x} = +0,60 \cdot 2,6 / 3,4 = +0,44 \text{ кг}$$

Висновок: при зміні середнього зросту 7-річних дівчаток на 1 см середня вага їх зміниться на 0,44 кг. Таким чином, якщо група 7-річних дівчаток, що навчається, має середній зріст на 2 см вище середнього зросту одноліток, то можна очікувати, що вага їх буде на $2 \times 0,44 \text{ кг} = 0,88 \text{ кг}$ вище середнього ваги їх одноліток.

Типове завдання

Якою буде в середньому вага (y) у групі 7-річних дівчаток, якщо їхній зріст дорівнює 120 см (x).

З матеріалів фізичного розвитку 7-річних дівчаток певної території відомо, що їх середній зріст дорівнює 118,20 см (M_x), а середня вага — 22,14 кг (M_y). Коефіцієнт регресії ваги за зростом ($R_{y/x}$) дорівнює 0,44.

Зразок вирішення типового завдання.

Для вирішення завдання необхідно застосувати рівняння регресії: $y = M_y + R_{y/x} (x - M_x)$, тобто y (маса тіла) = $22,14 + 0,44 (120 - 118,2) = 22,93 \text{ кг}$

Таким чином, при зрості 120 см середня маса тіла групи дівчаток віку 7 років складе 22,93 кг.

Навчальні завдання

Завдання 1

Обчислення коефіцієнта рангової кореляції (Спірмена) і оцінка його достовірності.

На основі початкових даних по своєму варіанту завдання потрібно:

- а) обчислити коефіцієнт рангової кореляції;
- б) визначити характер і силу зв'язку між відповідними ознаками;
- в) визначити достовірність коефіцієнта кореляції.

Варіант 1

Таблиця

При проведенні комплексних медичних оглядів у осіб різного віку число хронічних захворювань (на 1000 оглянутих даного віку) склало:

Вік, роки	Число хронічних захворювань	Вік, роки	Число хронічних захворювань
0-4	748,6	25-29	1340,9
5-9	903,8	30-39	1944,8
10-14	982,4	40-49	1679,6
15-19	1281,6	50-59	2635,8
20-24	1010,6	60-69	3564,7
		70-79	4171,8

Варіант 2

Таблиця

**Післяопераційні ускладнення і час, що пройшов
від моменту гострого нападу холецистититу
до початку операції:**

Час, год	До 3	3-5	6-8	9-11	12-14	15-17	18-20	21-23	24 і більше
Число ускладнень	6	8	12	19	24	20	28	34	46

Варіант 3

Таблиця

**Рівні систолічного і діастолічного тиску (у мм. рт. ст.)
у 12 здорових хлопців у віці 18 років:**

Систола	105	115	115	110	110	120	120	120	125	110	125	120
Діастола	65	70	65	65	70	75	75	70	75	70	80	80

Варіант 4

Таблиця

**Результати вимірювання зросту і маси тіла студентів
у віці 20 років:**

Зростання, см	157	158	160	165	167	162	171	174	168	176	170	180
Маса тіла, кг	56	55	57	57	58	60	65	63	67	72	79	82

Варіант 5

Таблиця

Вірогідність смерті від судинних уражень головного мозку на 10000 жінок залежно від віку:

Вік, роки	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	70-74	75-79	80 і старше
Вірогідність смерті	5,0	5,6	5,7	5,7	5,6	7,6	7,7	9,3	10,7	10,5	14,1	15,0	23,2

Завдання 2

Обчислення коефіцієнта лінійної кореляції (Пірсона) і оцінка його достовірності. Розрахунок коефіцієнта і рівняння лінійної регресії.

На основі початкових даних вибраного варіанту завдання потрібно:

- обчислити коефіцієнт лінійної кореляції;
- визначити характер і силу зв'язку між ознаками;
- визначити достовірність коефіцієнта кореляції;
- розрахувати коефіцієнт регресії першої з вказаних у варіанті ознак по другій, заздалегідь розрахувавши значення середніх квадратичних відхилень (σ_{xy});
- розрахувати рівняння лінійної регресії;
- інтерпретувати результат, відповісти на питання завдання.

Варіант 1

Таблиця

Вміст водорозчинного фтору в орному шарі ґрунту і у вирощеній капусті (міліграм/кг)

Ґрунт	5	10	15	20	25	30	35	40	45	50
Капуста	13,5	16,4	17,2	18,8	21,1	22,3	24,8	25,0	26,4	28,6

Питання: яким буде в середньому вміст водорозчинного фтору в капусті, якщо в орному шарі ґрунту воно складе 23 міліграми/кг

Варіант 2

Таблиця

Рівень артеріального тиску (АТ систолічного і діастолічного, мм.рт.ст.) при навантаженні у хлопців 17 років

АТ сист.	101	102	103	104	105	106	107	108	109	110
АТ діаст.	45	47,8	50,1	51,2	51,4	51,4	51,6	52,8	53,0	54,5

Питання: Яким буде в середньому рівень діастолічного артеріального тиску при навантаженні у хлопців 17 років, якщо систолічний артеріальний тиск складе 100 мм.рт. ст.

Варіант 3

Таблиця

Частота серцевих скорочень в хвилину (ЧСС) у хлопців 16 років в стані спокою і при навантаженні

В стані спокою	67	70	73	76	79	82	85	87	90	93
При навантаженні	91,5	93,0	99,1	104,2	108,6	112,5	116,3	120,1	127,7	134,5

Питання: Якою буде ЧСС в середньому при навантаженні у хлопців 16 років, якщо в стані спокою вона рівна 75 в хвилину.

Варіант 4

Таблиця

**Систолічний артеріальний тиск (АТ сист.)
і частота серцевих скорочень (ЧСС)
при навантаженні у хлопців 18 років:**

АТ сист.	101	102	103	104	105	106	107	108	109	110
ЧСС	91	92,6	94,3	95,1	97,5	98,0	100,3	101,6	103,2	105,5

Питання: Якою буде в середньому ЧСС при навантаженні у хлопців 18 років, якщо систолічний артеріальний тиск складе 111 мм. рт. ст.

Варіант 5

Таблиця

**Рівень шуму (дБА) на робочому місці і частота скарг на
зниження слуху у робочих прокатних цехів (у %):**

Рівень шуму	94	99	104	108	112	114	116
Частота скарг	33,3	17,2	25,0	34,0	33,3	50,0	61,5

Питання: Якою буде в середньому частота скарг на зниження слуху, якщо рівень шуму на робочому місці складе 105 дБА.

Контрольні запитання

1. У чому полягає сутність кореляційно-регресійного аналізу?
2. Які види зв'язку існують між явищами і ознаками?
3. Що таке кореляційний зв'язок (кореляція)?
4. Для чого використовується кореляційний аналіз?
5. Який може бути кореляційний зв'язок за своїм характером і силою?

6. Що є критерієм оцінки характеру і сили кореляції?
7. Як встановити розмір зв'язку між явищами або ознаками, що вивчаються?
8. Які існують методи для визначення коефіцієнтів кореляції?
9. У яких випадках використовують метод рангової кореляції і в чому полягає його сутність?
10. Як розрахувати коефіцієнт рангової кореляції і які початкові дані для цього потрібні?
11. У яких випадках використовується метод лінійної кореляції і в чому його сутність?
12. Як розрахувати коефіцієнт лінійної кореляції і які початкові дані для цього потрібні?
13. Як оцінюється достовірність коефіцієнта кореляції?
14. У чому полягає сутність регресії?
15. У яких випадках використовується регресійний аналіз?
16. Як розрахувати коефіцієнт регресії і які початкові дані для цього необхідні?
17. Як визначити розмір ознаки за величиною іншої взаємопов'язаної з нею ознаки, наприклад, ваги за зростом?
18. У чому полягає методика розрахунку рівняння регресії?

Тема 8

Методичні основи організації статистичних досліджень у системі охорони здоров'я

Статистичне спостереження (дослідження) проводиться поетапно. Розподіл клінічного дослідження по етапах має назву дизайн клінічного дослідження. Етапами статистичного дослідження є:

- а) перший — складання плану і програми;
- б) другий — збір матеріалу (реєстрація даних, статистичне спостереження);
- в) третій — розробка і зведення даних;
- г) четвертий — статистичний аналіз, висновки рекомендації;
- д) п'ятий — впровадження результатів дослідження в практику.

Робота над програмою і планом дослідження починається з побудови робочої гіпотези — формулювання у формі припущення основної ідеї дослідження, яку потрібно довести або спростувати в ході подальшої роботи.

План статистичного дослідження містить в собі:

- а) формулювання мети і завдань дослідження;
- б) об'єкт і одиницю спостереження;
- в) період спостереження;
- г) місце спостереження;
- д) термін виконання дослідження і окремих його етапів;
- е) вид (одночасне, поточне) і метод (суцільне, несучільне, в тому числі вибіркоче) статистичного спостереження;
- ж) джерела (офіційні облікові і звітні документи, спеціально розроблені облікові документи) і засоби отримання

інформації (анкетний опит, інтерв'ю, безпосереднє спостереження, вкопювання);

- з) перелік методів статистичного дослідження;
- и) склад виконавців роботи;
- к) організаційне і методичне керівництво роботою;
- л) інструкції для виконавців;
- м) фінансові розрахунки.

Програма статистичного дослідження включає відомості, які необхідно зібрати, і вказує головні напрями дослідження. Вона складається з трьох підпрограм:

- а) програми збору матеріалу (статистичного спостереження);
- б) програми розробки (табличного зведення);
- в) програми аналізу.

Мета соціально-медичного статистичного дослідження впливає з побудованої робочої гіпотези і передбачає:

- а) визначення закономірностей, властивих явищу, що вивчається, і зв'язків цього явища з іншими;
- б) розробку заходів щодо зниження впливу несприятливих чинників на здоров'я;
- в) визначення нових напрямів розвитку медичної науки і практики;
- г) розробку заходів щодо впровадження результатів роботи в практику охорони здоров'я;
- д) розробку заходів щодо поліпшення медичної допомоги.

Формулювання мети має відповідати на питання: для чого проводиться дослідження.

Завдання дослідження — це дії, спрямовані на досягнення мети. Формулювання завдань дослідження має відповідати на питання: яким чином буде досягнута мета, тобто якими конкретними шляхами (вони і є завданнями) буде забезпечено досягнення мети. Іншими словами, кінцева мета

розбивається на підцілі, виходячи з того, що здійснення кожної з них — ступінь до досягнення кінцевої мети, кінцевого результату.

Прикладами завдань в соціально-медичних дослідженнях можуть служити такі:

а) визначення рівня і структури досліджуваного впливу в конкретних групах населення;

б) визначення частоти явищ в групах, на які впливають різноманітні чинники (наколишнього середовища, біологічні, соціальні);

в) визначення обсягу і якості медичної допомоги, наданої окремим групам населення.

Відповідно до мети і завдань визначається об'єкт дослідження (статистична сукупність) і одиниця спостереження (первинний елемент). Об'єктом дослідження можуть бути мешканці певної території, певної соціальної або вікової групи, статі, професії, хворі певними захворюваннями тощо. Найчастіше статистична сукупність, яку використовують як об'єкт спостереження, характеризується комбінацією декількох ознак. За охопленням статистичної сукупності дослідження може бути суцільним або не суцільним. При суцільному статистичному дослідженні група спостереження формується шляхом повного охоплення всіх одиниць явища, що вивчається.

До методів не суцільного спостереження відносяться монографічний метод, метод основного масиву і власне вибірковий метод.

Монографічний метод застосовується для детального опису об'єкту, який має які-небудь яскраві особливості, наприклад, соціально-гігієнічний опис промислового центру. Висновки, отримані шляхом таких досліджень, стосуються або тільки конкретного об'єкту досліджень, або можуть бути поширені на вельми обмежену групу аналогічних об'єктів.

Метод основного масиву передбачає обстеження контингентів, які можуть бути зосереджені на конкретному об'єкті.

Власне вибіркове дослідження охоплює вибірккову сукупність або просто вибірку з генеральної сукупності.

У плані дослідження треба зазначити метод спостереження: суцільний (якщо об'єкт дослідження — генеральна сукупність) або вибіркковий (якщо вивчення проводиться на вибіркковій сукупності). При використанні в якості дослідження вибірккової статистичної сукупності необхідно забезпечити її репрезентативність щодо генеральної статистичної сукупності, частиною якої вона є, і достатнє число спостережень.

Репрезентативність вибірккової статистичної сукупності забезпечується застосуванням спеціальних методик відбору одиниць спостереження, а достатнє число спостережень — попереднім розрахунком необхідного обсягу вибірки.

Способами відбору одиниць спостереження для вибірккової сукупності є наступні: випадковий, механічний, типологічний, серійний, попарно-сполучений, спрямований в т.ч. когортний), гніздовий (районований), комбінований.

Випадковий відбір — це такий відбір, коли всі одиниці спостереження генеральної сукупності мають однакову вірогідність потрапити у вибірку. Найбільш поширеним методом такого відбору є метод жеребкування.

Механічний відбір — відбір, коли відбирається, наприклад, кожен п'ятий, десятий або двадцятий за списком.

Типологічний відбір — відбір, при якому генеральна сукупність заздалегідь розбивається на типи (за віком, статтю, соціальним і професійним групам і так далі), а потім відбирається пропорційна частина одиниць спостереження кожного типу.

Серійний відбір — це, по суті, поєднання типологічного і випадкового або механічного відбору, коли вся сукупність ділиться на приблизно однорідні серії, потім вибираються певні серії і в них досліджуються всі одиниці спостереження.

Попарно-сполучений відбір, або метод пара-копій — це формування, наприклад контрольної групи, коли кожній одиниці спостереження основної досліджуваної групи спостереження підбирається максимально схожа на неї одиниця контрольної групи; при цьому розбіжності між ними повинні бути тільки по досліджуваному чиннику, наприклад, методу лікування. При проведенні досліджень з малим числом спостережень формування груп спостереження і контролю проводиться відповідно з певними правилами. Одним з основних правил є рандомізація.

Рандомізація — це метод випадкового розподілу хворих на експериментальну і контрольну групи. Єдиний спосіб урівноважити всі сторонні чинники полягає в тому, щоб розподілити пацієнтів по групах випадковим чином так, щоб кожен пацієнт мав однаковий шанс бути віднесеним або до групи з впливом певного чинника, або до контрольної групи. Рандомізація врівноважує вірогідність впливу на пацієнта не тільки тих чинників, які, за припущенням, можуть вплинути на прогноз, але і тих, про яких нічого невідомо.

Спрямований відбір за однією, двома або декількома ознаками використовується для поглибленого вивчення взаємозв'язку між менш дослідженими ознаками, при включенні ознак, вплив яких відомий.

Різновидом спрямованого відбору є когортний відбір, який дає можливість виявити найбільш яскраві закономірності в групах осіб, об'єднаних тією або іншою ознакою.

Гніздовий або районований відбір — відбір, коли береться певний район міста (гніздо), який за попередніми даними є типовим для всього міста.

При комбінованому методі поєднуються вищезгадані способи відбору одиниць спостереження у вибірккову сукупність.

При використанні вибіркового методу спостереження виникає необхідність визначення обсягу вибірки (достатнього числа спостережень).

За часом проведення спостереження може бути поточним і одночасним. При поточному спостереженні статистичні дані збирають шляхом реєстрації кожного окремого випадку у міру його виникнення за певний проміжок часу. Таким шляхом вивчають народжуваність, смертність, захворюваність, навантаження лікарів і інші явища

Одночасне спостереження відображає стан досліджуваного явища на певний момент часу, збір всіх відомостей здійснюється одночасно. Прикладом одночасного спостереження є перепис населення, санітарні обстеження об'єктів і так далі.

Відібрані статистичні величини і спеціальні методи (стандартизація, оцінка достовірності, кореляційно-регресійний аналіз тощо) повинні бути адекватними меті і завданням дослідження. Статистичні величини, — абсолютні і похідні (середні і відносні), відібрані з числа відомих, або запропоновані автором, повинні максимально відображати суть досліджуваного явища. Обґрунтуванню запропонованих показників може бути присвячено одне із завдань дослідження.

У плані статистичного дослідження мають бути також відбиті форми практичної реалізації результатів дослідження. Ними можуть бути звіт, методичні рекомендації, стаття в журналі або збірці, монографія тощо.

Програма збору матеріалу є первинним обліковим документом (карта, анкета, бланк, перфокарта, таблиця, журнал і так далі) з переліком ознак, що підлягають реєстрації в процесі спостереження. У заголовку кожного облікового документа (програми) повинно бути зазначено, на кого він заповнюватиметься (формулювання одиниці спостереження). Як джерело інформації в соціально-медичних дослідженнях використовуються дані офіційних облікових і звітних документів. Крім того, програма збору матеріалу може бути розроблена авторами дослідження. При розробці такої програми повинні враховуватися наступні правила:

а) всі ознаки, які враховуються, безумовно повинні відповідати меті і завданням дослідження (тобто не повинні включатися питання «про всяк випадок»);

б) ознаки, що враховуються, повинні бути сформульовані чітко і коротко;

в) не повинні включатися питання, на які важко відповісти або отримати точну відповідь;

г) рекомендується на кожне питання передбачити можливі варіанти відповідей, при цьому кожній групі «питання-відповідь» дається умовний номер — шифр (код);

д) необхідно передбачити раціональне, компактне розташування на карті всіх ознак, що підлягають реєстрації.

Дотримання цих правил створює зручні умови для роботи, економію часу і сил, тому що дозволяє на другому етапі при отриманні первинної інформації (збір матеріалу) одночасно здійснювати реєстрацію, групування і кодування.

Одна з найбільш поширених методик збору інформації — соціологічні і маркетингові опити. Опит забезпечує отримання інформації зі слів респондента шляхом інтерв'ю або заочним шляхом (поштою, телефонні і інші опити). Реєстрація такої інформації здійснюється на спеціальні опитні листи або анкети.

Анкетування, у свою чергу, є найбільш поширеним способом опиту. Його перевагою є те, що в результаті обробки відповідей може бути отримана кількісна статистична характеристика явища, що вивчається, можуть бути з'ясовані і змодельовані причинно-наслідкові зв'язки.

Існує два способи анкетування: експедиційний, коли особа, що проводить опит, сама роздає і збирає анкети; і кореспондентський, коли анкети розсилаються і повертаються поштою.

Анкета є переліком питань, на які відповідає опитувана особа (респондент). Зазвичай анкета будується за наступною схемою:

а) введення, де містяться мета опиту і відомості про установу, яка його проводить;

б) перелік питань, які характеризують предмет опиту; потрібен розумний компроміс між бажанням зробити анкету максимально повною і реальною можливістю отримати відповіді;

в) відомості про опитувану особу.

Питання анкети розрізняють за ступенем свободи відповіді, характером відповідей, формою питань.

За ступенем свободи всі питання розділяються на відкриті і закриті. Відкрите або вільне, питання передбачає відповідь у вільній формі, без яких-небудь обмежень. Закриті питання передбачають перелік варіантів відповідей, зокрема альтернативних (так, ні, не знаю). До них належать питання — перелік варіантів відповідей, з яких потрібно вибрати один або декілька варіантів. Відповіді можуть бути надані у вигляді шкальної (бальної) оцінки, тобто числової величини.

Фільтруючі питання, ставляться, якщо частина питань стосується не всіх опитуваних осіб. Наприклад, чи користуєтесь Ви даною медичною послугою? (так, ні). Якщо так, то в якому лікувальному закладі? Зрозуміло, що це питання і

наступні адресуються лише тим, хто позитивно відповів на перше питання.

У будь-якій анкеті містяться контрольні питання, які використовуються для перевірки імовірності відповідей.

Питання повинні ставитися в логічній послідовності. Необхідно враховувати особливості тих груп населення, яким адресована анкета.

Обробка анкет проводиться, як правило, у формі групування з кожної відповіді або їх комплексу, тобто за потрібними для аналізу ознаками. Якщо обсяг вибірки невеликий, анкети можна обробити ручним способом.

Для якісного проведення опитування рекомендують привертати фахівців з розробки опитувальних листів. Найбільш точним, але і найбільш вартісним є опитування методом інтерв'ю. Для отримання оптимальних результатів опиту необхідно дотримуватися наступних основних вимог до змісту опитувального листа (анкети):

а) простота і доступність викладу питань і варіантів відповідей з урахуванням особливостей контингенту населення, серед якого проводитиметься опитування;

б) зручність для подальшого введення отриманої інформації в комп'ютер і її обробки;

в) по можливості, анкети мають містити контрольні питання;

г) ідентифікаційні дані (вік, місце проживання, телефон, і ін.) бажано надати в кінці опитування;

д) оптимальний обсяг опитування.

Дані, отримані в результаті клінічних спостережень, незалежно від методу їх отримання, відносяться до одного з трьох типів: якісні, порядкові і кількісні.

Якісні дані — це такі ознаки, які не можна виразити кількісно. Прикладами якісних даних можуть бути стать, діа-

гноз, професія, район проживання тощо. Якісні дані, які можуть бути віднесені тільки до двох категорій (є — немає, так — ні, живий — помер тощо), називаються дихотомічними.

Порядкові дані — це величини, які можуть бути розташовані в природному порядку або розташуванні, наприклад, від меншого до більшого або від хорошого до поганого, але розмір інтервалу між такими категоріями не може бути виражений кількісно (наприклад, стадії хвороби; оцінки рівня — «високий, середній, низький» або «відсутній, слабкий, помірний, важкий»).

Кількісні дані — це такі величини, яким властивий природний порядок розташування з інтервалами між послідовними значеннями, незалежно від їх місця на шкалі. Приклади: маса тіла, концентрація глюкози в крові. Кількісні дані можуть бути безперервними (інкретними) і переривистими (дискретними). Безперервні дані можуть приймати будь-яке значення на безперервній шкалі. Як приклади можна навести більшість біохімічних показників крові, масу тіла, кров'яний тиск, парціальний тиск кисню в артеріальній крові. Дискретні дані завжди виражаються цілими числами. Приклади дискретних даних — кількість вагітностей і кількість народжень живих немовлят, кількість судом у хворого протягом місяця тощо.

Програма розробки даних є набором макетів статистичних таблиць. Статистична таблиця — найбільш зручний вид статистичного зведення матеріалу. У статистичних таблицях прийнято виділяти такі основні елементи:

а) заголовок, який повинен повністю відображати зміст таблиці (включати у відповідному порядку всі ознаки і вказувати, в яких показниках наведені дані — абс. числа %, ‰ і т. д.);

б) статистичний підмет — основна ознака, яка аналізується в таблиці. Формулювання цієї ознаки указується у

верхньому рядку першої графи, а його градації — в наступних рядках тієї ж графи;

в) статистичний присудок (один або декілька) — облікові ознаки, які доповнюють і розкривають підмет, дозволяють дати глибшу і повнішу характеристику досліджуваної сукупності. Формулювання і градації присудка надаються в графах першого рядка таблиці;

г) підсумкові результати по горизонталі і вертикалі.

Розрізняють 3 види статистичних таблиць:

а) проста — має один статистичний підмет, дозволяє провести зведення матеріалу за однією ознакою (табл. 1)

Таблиця 1

Розподіл травматологічних хворих за характером пошкоджень (макет простої таблиці)

Характер пошкоджень	Число хворих
Ізольовані	
Множинні	
Поєднані	
Разом	

б) групова — має один статистичний підмет і декілька не пов'язаних між собою статистичних присудків, тобто підмет поєднується кожного разу з одним присудком (табл. 2):

Таблиця 2

Розподіл травматологічних хворих за характером пошкоджень, зростом і статтю (макет групової таблиці)

Характер пошкоджень	Стать		Вік (в роках)			Всього
	Чол.	Жін.	до 30	30-49	≥ 50	
Ізольовані						
Множинні						
Поєднані						
Разом						

в) комбінаційна — має один або декілька взаємопов'язаних присудків (табл. 3). У комбінаційних таблицях підмет поєднується відразу з 2-3 і більш присудками, що збільшує можливості аналізу. Кожна група першого присудка включає всі групи другого і всіх наступних присудків.

Таблиця 3

Розподіл травматологічних хворих за характером пошкоджень, зростом і статтю (макет комбінаційної таблиці)

Характер пошкоджень	Чоловіки			Жінки			Всього
	до 30	30-49	≥50	до 30	30-49	≥50	
Ізольовані							
Множинні							
Поєднані							
Разом							

Методичний прийом для побудови комбінаційної таблиці: для кожного присудка відводиться рядок. Перший рядок — для груп першого присудка, другий рядок — для груп другого присудка, третій рядок — для груп третього присудка. Більше 3-4 присудків в таблицях розміщувати не рекомендується, оскільки це призведе до формування дуже дрібної групи.

Одночасно з складанням макетів статистичних таблиць, або навіть до нього, здійснюється групування за обліковими ознаками. Під статистичним групуванням розуміють вибір (визначення) груп за ознаками на основі якісного аналізу. Його мета — виділення якісно однорідних груп для встановлення тих або інших закономірностей в досліджуваному явищі. Сформовані групи повинні бути якісно однорідними. У клінічних дослідженнях результати аналізу даних надаються за підгрупами пацієнтів з подібними характеристиками

(стратифікація). Від правильного групування ознак залежать результати дослідження.

Розрізняють 2 види угруповань: типологічні і варіаційні.

Типологічне угруповання — це угруповання атрибутивних, описових ознак, які виражаються словесно (стать, види захворювань, професія).

Варіаційне угруповання — це угруповання кількісних ознак, що мають числові вирази (вік, зріст, рівень доходу і так далі).

Вибір методу або способу угруповання в значній мірі визначається видом облікових ознак. Для угруповання якісних ознак використовуються альтернативна шкала і шкала рангів.

Альтернативна шкала, шкала рангів, шкала номіналів, шкала категорій, характеризується тим, що для визначення відмінності одного “вимірювання” від іншого використовуються найменування, мітки, ярлики (номер телефону, поштовий індекс тощо). Ці мітки можуть бути дихотомічними, тобто допускати розподіл на два, або декілька варіантів, наприклад стать, діагноз захворювання, місце народження. Вимірювання в такій шкалі не містить ніяких вказівок на величину ознаки або її якісну характеристику щодо іншої ознаки.

Шкала рангів (балів) або порядкова шкала (шкала ординарів) є альтернативною шкалою, але додатково вводить ранговий порядок взаємного розташування. Це проводиться за певним правилом, наприклад, від більшого до меншого, або навпаки. У такій шкалі кожне конкретне значення ознаки може бути вищим, чи нижчим або відповідати іншому значенню. Наприклад: стадія онкологічного захворювання, розподіл територій по рівню забруднення (високий рівень, середній, низький). Значення рангів (бали, класи) при цьому зачитуються якісними. Перевагою бальних шкал є мож-

ливість отримання інтегрованих оцінок. Оптимальне число рангів при розподілі медико-біологічної інформації перебуває в межах 6 — 14.

Інтервальні шкали дозволяють отримувати кількісні оцінки об'єктів дослідження (облікових ознак), упорядковувати об'єкти дослідження, виражати їх характеристики числами і проводити порівняння. Наприклад, збільшення при деякому дозованому навантаженні діастолічного тиску в легеневій артерії на 4,7 мм.рт.ст. (з 10,5 до 15,2 мм.рт.ст.).

Відносні шкали (шкали відношень) дуже схожі на інтервальні шкали. Додатково до всіх властивостей змінних, вимірних в інтервальній шкалі, їх характерною особливістю є наявність нульової точки. Типовий приклад шкали відношень — температура за Кельвіном. Можна достатньо упевнено стверджувати, що температура 200° удвічі вища ніж 100°. Інтервальні шкали (шкала Цельсія) не володіють даною властивістю шкали відношень.

Угрупування даних за своєю суттю є процесом класифікації, тобто встановлення приналежності явищ і об'єктів до певних класів. У державній статистиці для цього використовуються класифікатори — спеціальні довідники, інструкції і вказівки, у вигляді алфавітних і систематичних словників, які доповнюються стандартним переліком об'єктів і їх груп. Використання класифікаторів в офіційній статистичній звітності є обов'язковим і має силу державного стандарту. Наприклад, зведення даних про захворюваність з тимчасовою недієздатністю проводиться за спеціальними інструкціями, які регламентують розподіл нозологічних форм захворювань за певними групами діагнозів. Аналогічним чином, за спеціальними інструкціями формується статистика даних про загальну, госпіталізовану захворюваність, інвалідність, смертність, і так далі. Основні класифікатори розраховані на тривале використання. Проте з часом вони переглядають-

ся, доповнюються, до них вносяться необхідні корективи. У медико-біологічних наукових дослідженнях використання державних і міжнародних класифікаторів хвороб і причин смерті не є суворо обов'язковим. Проте тільки їх пряме використання або можливість шляхом вторинної обробки даних угруповань привести дані досліджень в рамки, які визначаються загальноприйнятими класифікаторами, гарантує співставність статистичних матеріалів про захворюваність, причини смерті різних категорій населення на різних територіях країни і за кордоном.

Під програмою аналізу розуміють перелік і послідовність застосування статистичних методик, які необхідні для виявлення закономірностей досліджуваного явища. Основні напрями аналізу визначаються програмою розробки, проте, необхідно також визначити і перелічити, які статистичні критерії і показники потрібно буде розраховувати на третьому етапі дослідження.

Програма аналізу визначається виходячи з 5 основних властивостей статистичної сукупності:

а) розподіл ознак може бути охарактеризований абсолютними числами і екстенсивними показниками;

б) середній рівень ознак характеризується різноманітними середніми величинами, інтенсивними показниками;

в) варіабельність — різноманітність ознак характеризують такі показники як ліміт, амплітуда, середнє квадратичне відхилення, коефіцієнт варіації;

г) репрезентативність — характеризують за допомогою похибки — відмінності середньої величини вибіркової сукупності від середнього розміру генеральної сукупності, визначають межі коливання середніх розмірів вибіркової сукупності в генеральній сукупності з певною вірогідністю;

д) взаємозв'язок ознак характеризують за допомогою коефіцієнтів кореляції і інших методів.

Всі величини і показники розраховуються на певні моменти часу або в динаміці, порівнюються між собою по групах спостереження, знову ж таки в динаміці, з даними інших авторів, територій, з нормативами, стандартами і так далі.

Зміст другого етапу складає:

а) збір матеріалу (заповнення облікових документів);

б) поточний контроль реєстрації.

Методика збору інформації містить в собі:

а) заповнення різноманітних облікових документів або викопіювання відомостей з них (з історій хвороби, різноманітних журналів, карт, записів в комп'ютерних базах даних);

б) безпосереднє спостереження;

в) опит (анкетування, інтерв'ю);

г) поєднання зазначених методик в одному дослідженні.

На другому етапі дослідник повинен зібрати матеріал за складеною програмою дослідження, не відхиляючись від неї і не змінюючи її.

У зміст третього етапу входить:

а) перевірка зібраного матеріалу з погляду правильності і повноти відповідей на всі питання,

б) шифрування (кодування) матеріалу за градаціями облікових ознак;

в) розподіл всіх спостережень по однорідних групах відповідно до програми розробки — макетів таблиць;

г) підрахунок по групах і внесення даних в таблиці (порожніх граф в таблицях не повинне залишатися, у разі відсутності даних ставиться «0» або прочерк);

д) розрахунок похідних величин (середніх, відносних);

е) застосування спеціальних методів статистичної обробки матеріалу;

ж) графічне зображення даних.

Четвертий етап статистичного дослідження містить в собі:

а) статистичний аналіз за програмою, складеною на першому етапі дослідження, в тому числі осмислення різноманітних отриманих статистичних величин (абсолютних, середніх, відносних, різноманітних коефіцієнтів і так далі) і графічних зображень;

б) оцінку отриманих показників порівняно з наявними нормативами, середніми рівнями статистичних величин, стандартами фізичного розвитку, з даними інших лікувальних установ і територій, в динаміці, з іншими критеріями оцінки;

в) написання і літературне оформлення роботи;

г) формулювання висновків;

д) перевірка відповідності отриманих висновків прийнятій гіпотезі;

е) розробка практичних рекомендацій.

При статистичному аналізі мають місце три групи помилок:

а) методичні помилки (недостатнє число спостережень, неправильне визначення одиниці спостереження, неправильне групування);

б) неправильна оцінка статистичних величин (висновок про рівень і динаміку явища на основі екстенсивних показників, оцінка темпу зростання без урахування початкового рівня, використання середніх величин в неоднорідних групах);

в) логічні помилки аналізу (порівняння даних без урахування їх якісної характеристики, помилкова інтерпретація причин і наслідків, недооблік зв'язків явищ).

П'ятий етап — впровадження результатів дослідження в практику. Форми впровадження:

а) у клініці — застосування розробленого методу дослідження, лікування, операції і так далі, в різних лікувально-профілактичних закладах;

б) в організації охорони здоров'я — розробка методичних рекомендацій, інструкцій і посібників;

в) в учбовому процесі — в лекціях, семінарах, практичних заняттях;

г) у пресі — публікації, виступи на конференціях, симпозіумах, з'їздах.

Методика виконання навчальних завдань

Завдання 1

Розрахунок числа спостережень, необхідного для отримання достовірних результатів вибіркового статистичного дослідження.

Розраховують число спостережень, необхідне для отримання достовірних результатів вибіркового статистичного дослідження, використовуючи початкові дані за варіантами, по наступній методиці:

Якщо в результаті дослідження кінцевий висновок повинен бути виражений середніми величинами (наприклад, належить визначити середній рівень артеріального тиску в міліметрах ртутного стовпчика у хворих, в лікуванні яких був застосований новий препарат), то необхідний обсяг вибірки розраховується за формулою:

$$n = \frac{t^2 \sigma^2}{\Delta^2}$$

де n — число спостережень; t — довірчий коефіцієнт, який при вірогідності безпомилкового прогнозу (p) рівній 95%, дорівнює 2, а при $p = 99\%$ дорівнює 3 і більше; σ — середнє квадратичне відхилення, яке береться з опублікованих аналогічних досліджень або (якщо аналогічні дослідження не

проводилися) розраховується при проведенні пробного дослідження з малим числом спостережень; Δ — максимально допустима похибка, яка задається дослідником відповідно до специфіки вимірюваної ознаки, мети і завдань дослідження.

Якщо кінцевий результат має бути виражений у відносних величинах (наприклад, у відсотках), необхідний обсяг вибірки розраховується за формулою:

$$n = \frac{t^2 pq}{\Delta^2}$$

де n — число спостережень; t — довірчий коефіцієнт, який при вірогідності безпомилкового прогнозу (p) рівній 95%, дорівнює 2, а при $p = 99\%$ дорівнює 3 і більше; p — показник у відсотках, який береться з аналогічних опублікованих досліджень або (якщо аналогічні дослідження відсутні) вибирається з розрахунком, щоб добуток pq був максимальним, що можливо, якщо $p = q = 50\%$; $q = 100\% - p$; Δ — максимально допустима похибка у відсотках, яка задається відповідно до явища що вивчається, мети і завдань дослідження. Зазвичай максимальна похибка не перевищує 5%

Завдання 2

Складання макетів статистичних таблиць

Побудувати макети статистичних таблиць, використовуючи початкові дані за варіантами і дотримуючись правил їх побудови, викладених вище.

Завдання 3

Аналіз статистичних даних.

Провести аналіз, сформулювати висновки і відповіді на поставлені питання за аналогією із зразком виконання типового завдання.

Типове завдання

Розподіл обстежених осіб (п=872) за рівнем артеріального тиску

Артеріальний тиск, мм рт. ст.	Виявлено	
	Абс.	%
Діастолічний		
норма <90	454	51,8
погранична зона 90-94	135	15,4
Підвищений 95-104	177	20,2
Підвищений 105-114	77	8,7
Підвищений 115-124	21	2,3
>125	8	1,6
Систолічний	Абс.	%
Норма 159	112	81,7
Підвищений 160-189	134	15,4
>190	26	2,9

Запитання 1. Який відсоток обстежених має підвищений рівень діастолічного артеріального тиску ?

Запитання 2. Яку частину обстежених осіб можна віднести до групи ризику (особи, які знаходяться за рівнем діастолічного артеріального тиску в “пограничній” зоні)?

Зразок виконання завдання

Аналіз даних таблиці дозволяє відповісти на поставлені питання таким чином:

Підвищений діастолічний артеріальний тиск (95 мм рт. ст. і вище) має 32,8% обстежених: $20,2\% + 8,7\% + 2,3\% + 1,6\% = 32,8\%$.

Підвищений рівень систолічного артеріального тиску (160 мм рт. ст. і вище) має 18,3% обстежених осіб: $15,4\% + 2,9\% = 18,3\%$.

До групи ризику слід віднести 15,4% обстежених, тому що рівень тиску у них складає 90-94 мм рт. ст.

Таким чином, тільки 51,8% від числа обстежених мали нормальний рівень діастолічного артеріального тиску. Систолічний артеріальний тиск знаходився в межах нормальних показників у 81,7% обстежених.

Завдання 4.

Оцінка правильності висновків, що витікають з аналізу статистичних даних.

Типове завдання

Структура захворюваності дітей в районі обслуговування дитячої поліклініки у 2008 — 2009 рр.

Захворювання	Абс. число захворювань		Питома вага %	
	2008	2009	2008	2009
Кір	22	22	2,2	4,4
Скарлатина	14	14	1,4	2,8
Дизентерія	34	34	3,4	6,8
Грип	830	330	83,0	66,0
Інші захворювання	100	100	10,0	20,0
Всього	1000	500	100,0	100,0

Головний лікар поліклініки зробив висновок: захворюваність дітей на кір, скарлатину і дизентерію в районі обслуговування в 2009 р. збільшилася в порівнянні з 2008 р. в 2 рази.

Зразок виконання завдання.

Висновок невірний (неправильна оцінка показників: за екстенсивними показниками не можна робити висновок про

частоту явища). З абсолютних даних очевидно, що число захворювань на кір, скарлатину і дизентерію в 2009 р. залишилося тим же, що і в 2008 р. Зростання питомої ваги зазначених хвороб в 2009 р. пояснюється значним зниженням загального числа захворювань (з 1000 до 500) за рахунок грипу. Це призвело до того, що якщо в 2008 р. 22 випадки кору склали 2,2%, то в 2009 р. 22 випадки кору (при зменшенні удвічі загального числа захворювань) склали вже 4,4%.

Навчальні завдання

Завдання 1

Розрахунок числа спостережень, необхідного для отримання достовірних результатів вибіркового статистичного дослідження.

Варіант 1.

Розрахувати число спостережень, необхідне в дослідженні з вивчення середньої тривалості перебування хворих на гострий холецистит в хірургічному відділенні стаціонару, виходячи з того, що:

необхідна вірогідність безпомилкового прогнозу (p) повинна бути 95%;

середнє квадратичне відхилення (a) пробного дослідження з малим числом спостережень складає 1,9 дні;

максимально допустима похибка (t) складає 0,5 дня.

Варіант 2.

Розрахувати число спостережень, необхідне в дослідженні з вивчення результатів лікування хворих на виразкову хворобу шлунку (за показниками відсотка хворих, що видужали), виходячи з того, що:

необхідна вірогідність безпомилкового прогнозу (p) має бути 95%;

показник відсотка хворих (P), що видужали, пробного дослідження з малим числом спостережень складає 72%;
максимально допустима похибка (m) складає 5%.

Варіант 3.

Розрахувати число спостережень, необхідне в дослідженні з вивчення ефективності диспансеризації хворих на ревматизм (за показником відсотка хворих, що видужали), виходячи з того, що:

необхідна вірогідність безпомилкового прогнозу (p) повинна бути 95%;

показник відсотка хворих (P), що видужали, пробного дослідження з малим числом спостережень складає 60%;
максимально допустима похибка (m) складає 5%.

Варіант 4.

Розрахувати число спостережень, необхідне в дослідженні з вивчення організації диспансерного спостереження хворих на бронхіальну астму (за показниками регулярності диспансерного спостереження — відсотку хворих, які спостерігалися регулярно), виходячи з того, що:

необхідна вірогідність безпомилкового прогнозу (p) повинна бути 95%:

показник відсотка хворих, які спостерігалися регулярно (P) пробного дослідження з малим числом спостережень складає 32%;

максимально допустима похибка (m) складає 5%.

Варіант 5.

Розрахувати число спостережень, необхідне в дослідженні з вивчення результатів стаціонарного лікування дітей,

хворих на лімфогранулематоз (за показником відсотка хворих, виписаних з поліпшенням стану), виходячи з того, що: необхідна вірогідність безпомилкового прогнозу (р) має бути 95%;

показник відсотку хворих, які виписані з поліпшенням стану (Р) пробного дослідження з малим числом спостережень складає 85%;

максимально допустима похибка (m) складає 5%.

Завдання 2.

Складання макетів статистичних таблиць

Побудувати макети статистичних таблиць, використовуючи вихідні дані за варіантами.

Варіант 1.

Скласти макети статистичних таблиць (простої, групової, комбінаційної) для надання результатів вивчення показників лікарняної летальності з урахуванням профілю відділення стаціонару, віку і статі хворих.

Варіант 2.

Скласти макети статистичних таблиць (простої, групової, комбінаційної) для надання результатів вивчення захворюваності за даними зверненості в поліклініку за класами хвороб міжнародної статистичної класифікації хвороб, травм, причин смерті, з урахуванням статі, віку, стажу роботи.

Варіант 3.

Скласти макети статистичних таблиць (простої, групової, комбінаційної) для надання результатів вивчення показників фізичного розвитку дітей дошкільного віку з урахуванням району їх місця проживання, статі, відвідування або невідвідування дитячої дошкільної установи.

Варіант 4.

Скласти макети статистичних таблиць (простої, групової, комбінаційної) для надання результатів вивчення показників тривалості перебування хворих на гострий холецистит в хірургічному відділенні стаціонару з урахуванням термінів доставки і віку хворих.

Варіант 5.

Скласти макети статистичних таблиць (простої, групової, комбінаційної) для надання результатів вивчення ефективності диспансеризації (одужання, поліпшення, погіршення стану, без змін, частота ускладнень) з урахуванням регулярності спостереження і віку хворих.

Завдання 3

Аналіз статистичних даних.

Варіант 1.

Таблиця

**Рівень захворюваності на гіпертонічну хворобу (ГХ)
у чоловіків, що зловживають алкоголем
і всього чоловічого населення залежно від віку
(на 1000 чоловік відповідної вікової групи)**

Вік (років)	Рівень захворюваності ГХ	
	Чоловіки що зловживають алкоголем	Все чоловіче населення
20-39	21,9	14,6
40-49	40,1	30,7
50-59	93,0	15,6
60 і старше	86,9	4,7
Всього	38,5	15,3

Запитання 1. У скільки разів рівень захворюваності на гіпертонічну хворобу вищий у чоловіків, що зловживають алкоголем, порівняно з частотою цього захворювання у всього чоловічого населення?

Запитання 2. У яких вікових групах найбільш високі розбіжності в рівнях захворюваності чоловіків, що зловживають алкоголем, і всього чоловічого населення?

Варіант 2.

Таблиця

**Поширеність хронічних захворювань серед студентів
(на 1000 студентів відповідного курсу)**

Нозологічна форма	Курс				
	1	2	3	4	5
Виразкова хвороба	2,0	4,0	5,4	9,1	10,3
Хронічний гастрит	2,8	4,0	4,3	11,7	12,3
Хронічний холецистит	1,4	1,4	1,7	1,9	3,0
Гіпертонічна хвороба	2,1	2,3	3,2	3,4	3,9
Ожиріння	4,5	2,8	1,5	0,4	0,4
Всього:					

Запитання 1. Яка динаміка поширеності хвороб органів травлення (виразкової хвороби, хронічного гастриту, хронічного холециститу) спостерігається у студентів різних курсів?

Запитання 2. Як змінилися показники поширеності ожиріння і гіпертонічної хвороби серед студентів різних курсів?

Варіант 3.

Таблиця

**Захворюваність з тимчасовою непрацездатністю робочих
машинобудівного заводу за нозологічними формами
(на 100 осіб)**

Нозологічна форма	Число випадків	Число днів
Гострі респіраторні захворювання	32,7	185,5
Грип	7,4	56,8
Побутові травми	1,2	18,6
Гострі фарингіти, тонзіліти	1,7	13,8
Інфекції шкіри і підшкірної клітковини	1,0	6,9
Хвороби нервової системи	0,6	9,3
Захворювання органів дихання	0,6	7,6

Запитання 1. Які захворювання займають провідні місця в структурі тимчасовий непрацездатності?

Запитання 2. При яких захворюваннях непрацездатність є найбільш тривалою?

Варіант 4.

Таблиця

Рівень госпіталізації в 2001-2009 рр. (на 100 осіб)

	2001 р.	2005 р.	2009 р.	2005 р % до 2001 р.	2009 р % до 2005 р.
Все населення	21,5	22,7	23,1	106,0	101,8
Міське населення	22,4	22,9	23,4	102,6	102,2
Сільське населення	20,4	22,5	22,8	111,3	101,3

Запитання 1. Яка динаміка рівня госпіталізації спостерігалася за період 2001-2009 гг?

Запитання 2. Серед яких мешканців рівень госпіталізації вищий — міських або сільських?

Запитання 3. На скільки відсотків в 2009 р. зріс рівень госпіталізації сільського населення в порівнянні з 2001 г ?

Варіант 5

Таблиця

**Захворюваність міських і сільських мешканців
по окремих класах хвороб**

Діагноз	Місце проживання		
	Місто	Село	В середньому
Хвороби нервової системи і органів чуття	176,9	82,0	108,5
Хвороби органів дихання	39,1	19,7	25,2
Хвороби органів травлення	4,5	0,4	1,6
Хвороби системи кровообігу	54,5	35,4	40,8
Хвороби сечостатевого органів	2,9	0,4	1,1
Хвороби кістково-м'язової системи і сполучної тканини	7,3	3,1	4,3
Хвороби ендокринної системи, розлади харчування і порушення обміну речовин	2,1	0,8	1,1
Вроджені аномалії	4,8	2,4	3,1
Інші	2,8	0,8	1,3
Разом	294,9	145,0	187,6

Запитання 1. Які хвороби переважають в структурі захворюваності міських і сільських мешканців?

Запитання 2. У яких мешканців (міських або сільських) вище показник захворюваності в цілому і за окремими класами хвороб?

Завдання 4

Оцінка правильності висновків, що витікають з аналізу статистичних даних.

Оцінити правильність висновків, зроблених на основі аналізу статистичних даних. Якщо висновки зроблені невірно, відзначити характер помилок і вказати напрями аналізу для отримання відповідних висновків.

При вивченні діагностичних помилок лікарів лікарні на основі зіставлення клінічних і патологоанатомічних діагнозів були отримані наступні дані.

Варіант 1.

Таблиця

**Число діагностичних помилок,
виявлених при патологоанатомічному розтині**

Причина смерті	Число розтинів	Число діагностичних помилок
Туберкульоз легень	102	12
Рак легень	40	8
Кардіосклероз	78	8
Виразкова хвороба шлунку і 12-палої кишки	100	10

В результаті аналізу лікар зробив висновок про те, що найчастіше діагностичні помилки мають місце при туберкульозі легень, а найрідше — при раку легень. Чи правильний цей висновок?

Варіант 2.

У клініці були отримані наступні дані про ускладнення у хворих, що лікувалися від опіків

Таблиця

Ступінь опіків	Число хворих	Число ускладнень	
		Абс.	%
I	146	10	3,6
II	320	45	16,2
III	378	211	75,9
IV	12	12	4,3
Всього	856	278	100,0

В результаті аналізу лікар зробив висновок, що найбільша частота ускладнень відзначається у хворих з опіками III ступеня. Чи правильний цей висновок?

Варіант 3.

Аналізуючи захворюваність з тимчасовою втратою працездатності за рік, лікар, що обслуговує 1-й механічний цех, при обчисленні показника «число випадків тимчасової непрацездатності на 100 працюючих» врахував всі лікарняні листи (сплачені і несплачені), внаслідок чого показник склав 120 випадків на 100 працюючих. Лікар, що обслуговує 2-й механічний цех, при обчисленні аналогічного показника врахував тільки сплачені лікарняні листи, внаслідок чого показник склав 110 випадків на 100 працюючих.

На підставі цих даних було зроблено висновок про те, що рівень захворюваності в 2-му механічному цеху нижчий, ніж в 1-му. Чи правильний цей висновок ?

Варіант 4.

У лікарні був застосований новий препарат для лікування хворих на певне захворювання. В результаті лікування показник летальності знизився до 5%. У аналогічній групі хворих, яка лікувалася відомими препаратами, показник летальності складав 10%. На підставі цих даних було зроблено висновок про перевагу нового медикаментозного засобу. Чи правильний цей висновок?

Варіант 5.

У гінекологічному відділенні лікарні було застосовано новий метод лікування раку шийки матки. Групи хворих, що лікувалися новими і відомими методами, були однакові за своїм складом. В результаті застосування нового методу лікування відсоток тих, що видужали збільшився на 5. На підставі цих даних лікар, що проводив лікування, зробив висновок про перевагу нового методу. Чи правильний цей висновок?

Контрольні запитання

1. Назвіть етапи статистичного дослідження.
2. Що таке робоча гіпотеза?
3. Яка роль робочої гіпотези в статистичному дослідженні?
4. Які елементи включає план статистичного дослідження?
5. Що являє собою мета статистичного дослідження?
6. Що являє собою завдання статистичного дослідження?
7. Що таке одиниця спостереження?
8. Що є об'єктом дослідження?
9. Що таке генеральна статистична сукупність?
10. Що таке вибіркова статистична сукупність?

11. Які розрізняють методи дослідження статистичної сукупності?

12. Які вимоги пред'являються до вибіркової статистичної сукупності?

13. Як розрахувати необхідний обсяг спостереження, якщо результат повинен бути отриманий у відносних величинах, в середніх величинах?

14. Які можуть бути види статистичного спостереження за часом?

15. Які джерела інформації використовуються в статистичному дослідженні?

16. Які способи збору інформації використовуються в статистиці?

17. Які види облікових документів використовуються в статистичних дослідженнях?

18. Які існують форми практичної реалізації результатів дослідження?

19. Які програми складаються на I етапі статистичного дослідження?

20. Що являє собою програма статистичного спостереження (збору даних)?

21. Що являє собою програма статистичної розробки (зведення)?

22. Що розуміють під статистичним групуванням?

23. Які існують види групування ознак і чим вони відрізняються?

24. Які існують види статистичних таблиць?

25. З яких елементів складається макет таблиці?

26. Які вимоги пред'являються до статистичних таблиць?

27. Що являє собою програма аналізу?

28. У чому полягає зміст II етапу статистичного дослідження?

29. У чому полягає зміст III етапу статистичного дослідження?

30. У чому полягає зміст IV етапу статистичного дослідження?

31. Які умови необхідні для правильного статистичного аналізу?

32. Перелічіть основні елементи статистичного аналізу.

33. Назвіть найбільш типові помилки, що зустрічаються при проведенні статистичного аналізу.

34. Назвіть види помилок статистичного аналізу, обумовлені помилками групування.

35. Наведіть приклади помилок, що допускаються дослідником при оцінці статистичних показників.

36. Наведіть приклади логічних помилок, обумовлених формальним аналізом.

Тема 9

Чинники ризику.

Методика розрахунку і оцінки ризиків

Показники ризику завдяки своїй наочності знайшли широке застосування в епідеміологічних дослідженнях

Під ризиком розуміють вірогідність негативних наслідків або чинник, що підвищує таку вірогідність Згідно іншому визначенню, ризик це вірогідність і правдоподібність розвитку шкідливого ефекту при конкретній дії конкретного чинника на конкретну популяцію.

Популяцію, схильну до ризику, складають всі особи, у яких може виникнути даний несприятливий ефект, наприклад, дане захворювання. Рівень захворюваності характеризує вірогідність ризику виникнення захворювання в популяції (по суті, це міра ризику). І якщо одна популяція має вищий показник захворюваності, чим інша, то можна сказати, що у цієї популяції вищий ризик виникнення хвороб. Термін «правдоподібність» відображає та обставина, що ризик може бути виражений не тільки кількісно в масштабі від 0 до 1, але і напівкількісно, або навіть якісно.

Ризик для здоров'я — вірогідність розвитку загрози життю або здоров'ю людини чи загрози життю або здоров'ю майбутніх поколінь, обумовленої дією чинників середовища існування. Про виникнення ризику доцільно говорити, якщо: а) встановлений науково аргументований зв'язок між порушенням здоров'я і аналізованим чинником (тобто коли досліджуваний чинник має небезпечні властивості); б) якщо є дія даного чинника на чутливий до нього організм, оскільки ризик без дії не існує.

Розрізняють поняття індивідуального і популяційного ризику.

Хоча індивідуальний ризик вводиться як «індивідуальний» показник, він не може вимірюватися безпосередньо у окремої особи, а виявляється на основі спостереження за популяцією, схильною до несприятливої дії (експонована група).

Чинник ризику — це особливість організму або зовнішня дія, які самі по собі не є причинами захворювання, але окремо або в комбінації один з одним збільшують вірогідність ризику виникнення окремих хвороб або іншого несприятливого результату.

Для оцінки ризику виникнення захворювання у зв'язку з дією конкретного чинника використовуються два типи аналітичних епідеміологічних досліджень — дослідження типу «випадок — контроль» і когортне дослідження.

У дослідженнях типу «випадок — контроль» дію чинників порівнюють для індивідів, що мали дане захворювання (досліджувана група), і тих, у кого даного захворювання немає (група контролю). Мета полягає в тому, щоб визначити, чи розрізняються обидві групи за тим, яку частку складають особи, схильні до дії певного чинника.

У когортних дослідженнях в учасників зазвичай немає досліджуваного захворювання. Досліджувану групу складають особи, що піддаються дії певного чинника ризику, контрольну — особи, які не піддаються дії чинника ризику.

Кількісно ефект дії можна виразити через показники ризику, які підрозділяються на:

1) показники індивідуального ризику, у тому числі абсолютний ризик (додатковий, атрибутивний) і відносний ризик;

2) показники ризику популяції, у тому числі додатковий ризик популяції, додаткова частка ризику популяції.

Використання показників індивідуального ризику дозволяє з різних сторін охарактеризувати вплив чинників ризику. Ці показники ризику доповнюють один одного.

Абсолютний ризик, додатковий ризик — це різниця показників захворюваності або результатів у осіб, схильних і не схильних до впливу чинника, що вивчається. Додатковий ризик вказує на додаткове число випадків захворювань або їх результатів, обумовлене дією чинників ризику і підтверджує, що є фактичні шанси виникнення якоїсь події протягом заданого періоду часу (вірогідність хвороби, смерті). Абсолютний ризик дає можливість виявити, до якого абсолютного підвищення захворюваності (смертності, інвалідності) приводить дія чинника, що підкреслює важливість проблеми з погляду суспільного здоров'я.

Достатньо наочним і найбільш простим показником, що характеризує ступінь ризику виникнення захворювання, вважається відносний ризик.

Відносний ризик — це відношення показника захворюваності (або результатів) у осіб, схильних до впливу чинника, що вивчається, до величини показника захворюваності (або результатів) осіб, не схильних до такого впливу. Велика величина відносного ризику указує на важливу роль даної дії в етіології патології, що вивчається. Величина відносного ризику дозволяє відповісти на питання: «У скільки разів вище захворюваність серед осіб, що піддаються дії чинника ризику, порівняно з особами, що не піддаються цій дії?» Відношення ризику, що дорівнює 1,0, свідчить: ризик захворювання однаковий як в групі експонованих, так і неекспонованих, тобто виникнення захворювання не пов'язане з даним чинником. Відношення ризику більше 1,0 дає підставу говорити про підвищений ризик виникнення захворювання в експонованій до даного впливу групі. Відношення ризику менше 1,0 свідчить про менший ризик в експонованій групі (можна припус-

тити, що в цій групі діє якийсь захисний чинник). Одним з істотних недоліків показника відносної ризику є те, що його значення змінюється залежно від періоду часу, до якого він віднесений. При збільшенні періоду спостереження ризик для будь-якого захворювання наближається до одиниці. Відносний ризик характеризує силу зв'язку між дією і результатами, тобто біологічний аспект.

Поняття відносного і додаткового (абсолютного) ризиків відрізняються за клінічним змістом. Оскільки додатковий ризик відображає додаткову вірогідність захворювання, його застосування як індивідуального показника ризику в більшості ситуацій є більш інформативним в порівнянні з відносним ризиком. З іншого боку, відносний ризик краще показує силу причинного зв'язку. Відносний ризик не несе інформації про величину абсолютного ризику захворюваності або її результатів. Навіть при високих значеннях відносного ризику абсолютний ризик може бути зовсім невеликим, якщо захворювання достатньо рідкісне.

Для повної оцінки тієї або іншої небезпечної дії з погляду суспільного здоров'я необхідно брати до уваги також поширеність чинника ризику. Менш небезпечний чинник ризику (з низьким відносним ризиком), але з високою поширеністю в даній популяції може спровокувати більш значну захворюваність, ніж небезпечний чинник ризику, що рідко зустрічається. Виникає питання: «Який внесок чинника ризику в загальну захворюваність групи людей, а не окремих індивідуумів?» Така інформація допомагає визначити, які чинники ризику дійсно істотні, а які не мають особливого значення для здоров'я населення даної території. Це допомагає керівникам системи охорони здоров'я на науковій основі визначати пріоритети при розподілі ресурсів.

Додатковий популяційний ризик розраховується як добуток додаткового ризику на поширеність чинника ризику в

популяції. Цей показник дає уявлення про додаткову захворюваність в популяції, пов'язану з чинником ризику. Крім того, можна визначити частку захворюваності в популяції, обумовлену даним чинником ризику, тобто додаткову частку популяційного ризику. Вона розраховується як відношення показника додаткового популяційного ризику до величини загальної захворюваності в популяції.

Для оцінки відносного ризику в дослідженнях типу «випадок-контроль» використовується спеціальний показник, який називається відношенням шансів. Річ у тому, що методика організації і проведення ретроспективного епідеміологічного дослідження типу «випадок-контроль» відрізняється від методики організації і проведення когортного дослідження. Групи спостереження (основна — хворі конкретним захворюванням і контрольна — особи без патології), що вивчаються, формуються не природним чином, а дослідником. Розрахувати показники частоти захворювання, а на їх основі — показники ризику, як в когортних дослідженнях, неможливо. Проте в дослідженнях типу «випадок-контроль» відомі частоти дії чинників ризику в основній і контрольній групах. Порівняння цих частот дає показник ризику, який і за своєю суттю, і математично рівнозначний відносному ризику і визначається як відношення шансів події в одній групі до шансів події іншої групи.

Шанс — відношення вірогідності того, що подія відбудеться, до вірогідності того, що подія не відбудеться. Шанси і вірогідність містять одну і ту ж інформацію, але по-різному її виражають. Якщо вірогідність того, що подія відбудеться, позначити p , то шанси цієї події будуть рівні $p / (1-p)$. Значення відношення шансів (ВШ) в межах від 0 до 1 відповідає зниженню ризику. ВШ, рівне 1, свідчить про відсутність ефекту. Якщо частота дії вища в основній групі, то ВШ буде більше 1, що вказує на підвищений ризик. Отже, чим силь-

ніший зв'язок між чинником, що впливає, і захворюванням, тим вище ВШ, і навпаки.

Для оцінки ефективності лікувальних і профілактичних втручань проводяться експериментальні дослідження, результати цих досліджень мають дати відповіді на питання, які зрештою можуть стати основою для ухвалення рішень про впровадження результатів досліджень в практику: Чи ефективно втручання, чи достатня клінічна значущість ефекту (не плутати із статистичною значущістю)? Якщо втручання ефективно, то наскільки воно ефективно в порівнянні з плацебо або відсутністю лікування, або в порівнянні з іншими втручаннями, використовуваними в даний час? Чи безпечно втручання? Чи буде воно безпечним і ефективним для конкретного хворого? Як співвідноситься корисний ефект з можливим ризиком і побічними діями втручання? Чи доцільне нове втручання з економічної точки зору?

Оцінити ефективність втручань можна якісно і кількісно. Це може бути зроблено з використанням як непрямих критеріїв, так і клінічно значущих результатів. Непрямий критерій оцінки в клінічному випробуванні — це лабораторний показник або симптом, який замінює клінічно значущий результат, що прямо характеризує самопочуття хворого, його функціональний стан і виживання. Прикладами непрямих критеріїв можуть бути: усунення шлуночкової аритмії, зниження рівня холестерину або артеріального тиску є кардіологічних дослідженнях, зменшення розмірів пухлини в дослідженнях, присвячених лікуванню раку тощо. Зміна непрямого критерію оцінки ефективності, що викликається лікуванням, повинна відбивати зміни клінічно значущого критерію (результату). Прикладами клінічно значущих результатів є зниження або зростання захворюваності, смертності, інвалідності, втрата зору і інші явища, які істотно впливають на якість життя).

Для оцінки і представлення ефекту втручання найчастіше використовуються три основні кількісні параметри:

а) відносне зниження частоти несприятливих результатів — зниження відносного ризику;

б) абсолютне зниження частоти несприятливих результатів зниження абсолютного ризику;

в) число хворих, яких необхідно лікувати конкретним методом протягом певного часу, щоб досягти сприятливого ефекту або запобігти несприятливому результату у одного хворого.

Відносне зниження частоти несприятливих результатів виражається показником «Зниження відносного ризику» (ЗВР), який є відношенням різниці в частотах результату, що вивчається, між контрольною групою і групою лікування до частоти цього результату в групі контролю, вираженим зазвичай у відсотках. Значення ЗВР більше 50% завжди відповідає клінічно значущому ефекту, від 25 до 50% — відповідає дуже часто.

Абсолютне зниження частоти несприятливих результатів виражається через показник зниження абсолютного ризику (ЗАР), який відбиває різницю в частоті результатів між контрольною групою та групою втручань (експериментальна група). Значення ЗАР, що дорівнює 50%, свідчить про клінічно значущий ефект втручання.

Показники відносного ризику і його зниження кількісно відображають результати втручання у відносних величинах, однак не дають ніякої інформації про інтенсивність впливу в абсолютному виразі. І навпаки, абсолютний ризик менше відображає відносну ефективність, але показує, чи буде ефект втручання клінічно значущим. Але навіть визначення абсолютного ризику не вирішує всіх проблем, оскільки це абстрактна величина, що не має прямого зв'язку з конкретною клінічною ситуацією, в якій перебувають хворий і лікар. Та-

кий зв'язок може забезпечити показник числа хворих, яких необхідно лікувати протягом певного часу, щоб запобігти одному несприятливому результату – ЧХНР. ЧХНР є специфічним показником ефективності певного втручання. Він показує відмінність (різницю) між досліджуваним та контрольним методами лікування або профілактики з точки зору досягнення певного клінічного ефекту (результату) і тому може бути використаний при виборі методу лікування конкретного хворого, профілактичного втручання, а також при проведенні фармако-економічного аналізу.

ЧХНР для певного виду лікування — величина, зворотна зниженню абсолютного ризику для цього ж лікування. При вивченні ефективності профілактичних втручань, які призводять до зниження частоти ускладнень в основній групі, показник ЧХНР приймає негативне значення. В цьому випадку від'ємний знак можна ігнорувати або в знаменнику поміняти місцями показники основної і контрольної груп. ЧХНР можна також розрахувати на основі відношення шансів або зниження відносного ризику, які приводяться при публікації результатів систематизованих оглядів клінічних досліджень. Не існує чітких меж значень ЧХНР, що свідчать про ефективність або, навпаки, неефективність лікування.

В ідеальному випадку значення ЧХНР дорівнює 1, тобто сприятливий ефект спостерігається у кожного хворого, що одержує певне лікування, і відсутній у всіх хворих контрольної групи. Це можливо лише теоретично і практично ніколи не відбувається. Будь-яке лікування рідко дає ефект у 100% хворих, в той час, як в контрольній групі, як правило, спостерігається певний ефект, навіть при прийомі плацебо або відсутності лікування.

Зазвичай вважають, що чим нижче значення ЧХНР, тим краще. Для високоефективних методів лікування ЧХНР складає 2-4, за винятком, наприклад, клінічних випробувань,

метою яких є порівняння ефективності антибіотиків і плацебо при інфекції *Helicobacter pylori* (ЧХНР приймає значення 1,1) або оцінюючих ефективність інсектицидів при педикульозі (ЧХНР — 1,1). Профілактичні методи лікування ефективні в частині популяції, тому в подібних випадках значення ЧХНР буде високим. Наприклад, при використуванні аспірину для профілактики смерті протягом 5 тижнів після інфаркту міокарду ЧХНР складає 40, проте ефект препарату вважається високим. Те ж стосується тромболітичної терапії після інфаркту міокарду (ЧХНР = 100, якщо тромболітичну терапію починають протягом перших 5 годин). Метод ЧХНР пропонується використовувати і для аналізу небажаних ефектів лікування — число хворих, яких необхідно пролікувати, щоб з'явився небажаний ефект.

Методика виконання навчальних завдань.

Завдання 1

Методика розрахунку показників ризику, отриманих в когортних дослідженнях.

Значення показників ризику визначаються за формулами:

Абсолютний ризик (додатковий ризик, різниця ризиків):

$$AR = Z_1 - Z_2,$$

де AR — Абсолютний ризик; Z_1 — коефіцієнт захворюваності у осіб, що піддавалися дії чинника ризику; Z_2 — коефіцієнт захворюваності у осіб, що не піддавалися дії чинника ризику.

Відносний ризик, відношення ризиків:

$$RR = Z_1/Z_2,$$

де RR — відносний ризик; Z_1 Z_2 — див. вище

Додатковий популяційний ризик:

$$AR_p = AR \times P,$$

де AR_p — додатковий популяційний ризик; AR — абсолютний ризик; P — поширеність чинника ризику в популяції.

Додаткова частка популяційного ризику:

$$A_{fp} = AR_p / 2,$$

де A_{fp} — додаткова частка популяційного ризику; AR_p — додатковий популяційний ризик; 2 — коефіцієнт поширеності захворювань.

Типове завдання

За наслідками наукового дослідження встановлено, що захворюваність на виразкову хворобу серед чоловіків, що палять, складає 25,44‰, серед чоловіків, що не палять — 10,07‰. Частка осіб, що палять, серед чоловіків складає 45%, загальна захворюваність на виразкову хворобу серед чоловіків складає 17,75‰. Розрахувати показники абсолютного і відносного ризику, додаткового популяційного ризику, додаткової частки популяційного ризику.

Зразок виконання завдання

Абсолютний (додатковий) ризик захворюваності на виразкову хворобу серед чоловіків, що палять, складає: $25,44 - 10,07 = 15,37$ на 1000 осіб, що палять.

Відносний ризик захворюваності на виразкову хворобу серед чоловіків, що палять, складає: $25,44 : 10,07 = 2,5$.

Додатковий популяційний ризик захворюваності на виразкову хворобу серед чоловіків, що палять, складає: $15,37 \times 0,45 = 6,92$

Додаткова частка популяції ризику захворюваності на

виразкову хворобу серед чоловіків, що палять, складає: $6,92:17.75 = 0,39 = 39\%$

Завдання 2

Методика розрахунку показників відносного ризику (відношення шансів), отриманих в дослідженнях типу «випадок-контроль».

Відношення шансів можна розрахувати на підставі даних таблиці зв'язаності.

Таблиця зв'язаності

	Захворювання		Всього
	Основна група	Контрольна група	
Була дія	a	b	a + b
Не було дії	c	d	c + d
Всього	a + c	b+d	a + b + c + d

Типове завдання

Метою епідеміологічного дослідження типу «випадок-контроль» було вивчення впливу особливостей харчового раціону на виникнення раку підшлункової залози. Дані проведеного дослідження наведені в таблиці. Необхідно розрахувати показник відносного ризику (відношення шансів).

**Споживання хворими і пацієнтами контрольної групи
смаженого м'яса частіше 1 разу на тиждень**

Група	Експоновані (споживають смажене м'ясо частіше 1 разу на тиждень)		Всього
	Так	Ні	
Хворі на рак підшлункової залози	53	43	96
Контрольна група	53	85	138
Всього	106	128	234

$$\text{Відношення шансів (ВШ)} = \frac{\frac{a / (a + c)}{c / (a + c)}}{\frac{b / (b + d)}{d / (b + d)}} = \frac{a / c}{b / d} = \frac{ad}{bc}$$

Зразок виконання завдання

Відносний ризик (відношення шансів) для цього типу досліджень розраховується за формулою:

$$\text{OR(ВШ)} = \frac{ad}{bc} = \frac{53 \cdot 85}{53 \cdot 43} = 1,98$$

Значення відносного ризику 1,98 свідчить про наявність виразної залежності між споживанням смаженого м'яса частіше 1 разу на тиждень і виникненням раку підшлункової залози.

Завдання 3

Методика розрахунку і оцінки показників ефекту втручання.

Значення показників ефекту втручання визначаються за формулами:

Зниження відносного ризику (ЗВР):

$$\text{ЗВР} = \frac{\text{ЧРК} - \text{ЧРВ}}{\text{ЧРК} \times (\text{ЧРК} - \text{ЧРВ})} \times 100\%,$$

де ЗВР — зниження відносного ризику; ЧРК — частота результатів в групі контролю; ЧРВ — частота результатів в групі втручань.

Зниження абсолютного ризику:

$$\text{ЗАР} = \text{ЧРК} - \text{ЧРВ},$$

де ЗАР — зниження абсолютного ризику; ЧРК-частота результатів в групі контролю; ЧРВ — частота результатів в групі втручань.

Число хворих, яких потрібно лікувати протягом певного часу, щоб запобігти одного несприятливого результату:

$$\text{ЧХНР} = 1/\text{ЗАР},$$

де ЧХНР — число хворих, яких потрібно лікувати протягом певного часу, щоб запобігти одного несприятливого результату; ЗАР — зниження абсолютного ризику.

Типове завдання

Проведено проспективне дослідження, метою якого було вивчити вплив гіпохолестерінемічних препаратів на ризик розвитку інфаркту міокарду.

Об'єкт втручання — пацієнти з високим ризиком цього захворювання. Протягом 5 років 1000 пацієнтів експериментальної групи щодня приймали досліджуваний препарат, 1000 пацієнтів з контрольної групи одержували плацебо. За період спостереження інфаркт міокарду розвинувся у 500 пацієнтів з контрольної групи і лише у 250 — з експериментальної. Частота результатів в групі втручання складатиме $250/1000 = 0,25$ або 25%; у групі контролю — $500/1000 = 0,5$ або 50%. Визначити показники, що характеризують ефект втручання: зниження відносного ризику, зниження абсолютного ризику, число хворих, яких потрібно лікувати протягом певного часу, щоб запобігти одного несприятливого результату.

Зразок виконання завдання.

$$\text{ЗВР} = 50-25 / 50 \times 100\% = 50 \%,$$

$$\text{ЗАР} = 50-25 = 25 \%$$

$$\text{ЧХНР} = 1/0,25 = 4$$

Навчальні завдання

Завдання 1

Методика розрахунку показників ризику, отриманих в когортних дослідженнях.

На основі наведених вихідних даних потрібно обчислити показники абсолютного і відносного ризику, додаткового популяційного ризику, додаткової частки популяційного ризику.

Варіант 1.

За наслідками наукового дослідження встановлено, що смертність від раку легень у осіб, що палять складає 0,96 %, у осіб, що не палять — 0,07 %. поширеність паління — 56%, загальна смертність від раку легень складає 0,56 %.

Варіант 2.

За наслідками наукового дослідження встановлено, що захворюваність на рак молочної залози у жінок, що мали в анамнезі 1-2 пологів складає 1,07%; у жінок, що мали в анамнезі > 2 пологів — 0,11%; частка жінок, що мають в анамнезі 1-2 пологів складає 50%; загальна захворюваність на рак молочної залози жінок складає 0,59 %.

Варіант 3.

За наслідками наукового дослідження встановлено, що захворюваність на бронхіальну астму дітей, що проживають в

районі з високим рівнем забруднення атмосферного повітря, складає 1,16‰, дітей, що проживають в умовно «чистих» районах — 0,22‰. Частка дітей, що проживають в районі з високим рівнем забруднення атмосферного повітря, серед всіх дітей складає 25%, загальна захворюваність на бронхіальну астму дітей складає 0,93‰.

Варіант 4.

За наслідками наукового дослідження встановлено, що захворюваність на туберкульоз серед осіб з низьким рівнем доходів складає 2,08‰, серед осіб з високим і середнім рівнем доходів — 0,07‰. частка осіб з низьким рівнем доходів серед населення складає 20%, загальна захворюваність на туберкульоз складає 0,47‰.

Варіант 5.

За наслідками наукового дослідження встановлено, що смертність від інфаркту міокарду серед осіб, що мають надмірну вагу, складає 16,8‰, серед осіб, що мають вагу в межах нормальних показників — 10,2‰ Частка осіб, що мають надмірну вагу, серед населення складає 5%, загальна смертність від інфаркту міокарду складає 5,6‰.

Завдання 2

Методика розрахунку показників відносного ризику (відношення шансів), отриманих в дослідженнях типу «випадок-контроль».

На основі наведених вихідних даних потрібно розрахувати показник відношення шансів (ВШ), дати інтерпретацію отриманих результатів.

Варіант 1.

Метою епідеміологічного дослідження типу «випадок-контроль» було вивчення вплив антибіотиків, що вживались жінками під час вагітності, на народження дітей з вродженими аномаліями. Були отримані наступні результати:

**Вживання антибіотиків під час вагітності матерями,
діти яких мали вроджені аномалії,
і вагітними контрольної групи.**

Вживання антибіотиків	Матері, що народили дітей з вродженими аномаліями	Контрольна група	Всього
Так	300	60	360
Ні	200	280	480
Всього	500	340	840

Варіант 2.

Метою епідеміологічного дослідження типу «випадок-контроль» було вивчення впливу паління батьків на виникнення бронхіту і пневмонії у дітей на першому році життя.

Були отримані наступні результати:

**Паління батьків, діти яких мали на першому році життя
бронхіти і пневмонії.**

Куріння батьків	Діти, що мали на першому році життя бронхіти і пневмонії	Контрольна група	Всього
Так	250	50	300
Ні	100	300	400
Всього	350	350	700

Варіант 3.

У проведеному в Україні епідеміологічному дослідженні смертності від інфаркту міокарду порівнювалися одружені і самотні чоловіки.

Були отримані наступні результати:

**Сімейний стан чоловіків 40-64 років,
померлих від інфаркту міокарду**

Сімейний стан	Померли від інфаркту міокарду	Контрольна група	Всього
Одружені	601	300	901
Самотні	888	210	1098
Всього	1489	510	1999

Варіант 4.

У дослідженні «випадок-контроль» опитано 110 заміжніх жінок з інфарктом міокарду і 220 заміжніх жінок без інфаркту міокарду. Запитання були спрямовані на виявлення жінок з чинниками ризику виникнення інфаркту міокарду (зокрема, артеріальної гіпертензії).

Результати зведено в таблицю.

Поширеність артеріальної гіпертензії серед заміжніх жінок, що перенесли і не перенесли інфаркт міокарду

Наявність гіпертензії	Жінки, що перенесли інфаркт міокарду	Контрольна група	Всього
Так	51	66	117
Ні	59	154	213
Всього	110	220	330

Варіант 5.

У дослідженні «випадок-контроль» опитано 110 заміжніх жінок з інфарктом міокарду і 220 заміжніх жінок без інфаркту міокарду. Запитання були спрямовані на виявлення жінок з чинниками ризику виникнення інфаркту міокарду (зокрема, цукрового діабету). Результати зведено в таблицю.

Поширеність цукрового діабету серед заміжніх жінок, що перенесли і не перенесли інфаркт міокарду

Наявність цукрового діабету	Жінки, що перенесли інфаркт міокарду	Контрольна група	Всього
Так	26	26	52
Ні	84	194	178
Всього	110	220	330

Завдання 3

Методика розрахунку і оцінки показників ефекту втручання

На основі наведених вихідних даних потрібно визначити показники, що характеризують ефект втручання: зниження відносного ризику, зниження абсолютного ризику, число хворих, яких потрібно лікувати протягом певного часу, щоб запобігти одному несприятливому результату.

Варіант 1.

З п'яти Скандинавських країн 4444 пацієнти (2223 особи, що приймали плацебо — група контролю і 2221 пацієнт, що приймав симвастатін — група втручання) взяли участь в рандомізованому подвійному сліпому, плацебо-контрольованому дослідженні з вивчення впливу симвастаті-

на на поліпшення клінічного перебігу ІХС і збільшення тривалості життя хворих з цю патологію. Отримані результати свідчать, що кількість інсультів в групі плацебо складає 98. а в групі втручань — 70. Розрахуйте ефективність симвастатіна для профілактики інсульту

Варіант 2.

Проведено епідеміологічне рандомізоване плацебо-контрольоване дослідження з оцінки ефективності трьохкомпонентної схеми лікування виразкової хвороби. У групі контролю (традиційна схема лікування), що складається з 1860 хворих, лікування інфекції спостерігалось у 577 пацієнтів, в групі втручання (використання сучасної трьохкомпонентної схеми лікування) що складається з 2010 пацієнтів, позитивний результат визначено у 1898 хворих. Розрахуйте показники ефективності трьохкомпонентної схеми лікування хвороби на виразкову хворобу.

Варіант 3.

Проведено епідеміологічне рандомізоване плацебо-контрольоване дослідження з оцінки ефективності нової схеми лікування гіпертонічної хвороби. У групі контролю (традиційна схема лікування), що складається з 1524 хворих, позитивний результат лікування спостерігався у 407 пацієнтів, в групі втручання (використання нової сучасної схеми лікування), що складалася з 1820 пацієнтів, позитивний результат визначено у 1750 хворих. Розрахуйте показник ефективності нової схеми лікування гіпертонічної хвороби.

Варіант 4.

Проведено епідеміологічне рандомізоване плацебо-контрольоване дослідження з оцінки ефективності нової

схеми лікування хронічного бронхіту. У групі контролю (традиційна схема лікування), що складається з 2200 хворих, позитивний результат лікування спостерігався у 68 пацієнтів, в групі втручання (використання сучасної трьохкомпонентної схеми лікування), що складалася з 1950 пацієнтів, позитивний результат відзначався у 1870 хворих. Розрахуйте показники ефективності нової схеми лікування хронічного бронхіту.

Варіант 5.

Проведено епідеміологічне рандомізоване плацебо-контрольоване дослідження з оціни ефективності трьохкомпонентної схеми лікування хронічного холециститу. У групі контролю (традиційна схема лікування), що складається з 1520 хворих, позитивний результат лікування спостерігався у 478 пацієнтів, в групі втручання (використання сучасної трьохкомпонентної схеми лікування), що складалася з 1850 пацієнтів, позитивний результат відзначався у 178 хворих. Розрахуйте показники ефективності нової схеми лікування хронічного холециститу.

Контрольні запитання

1. Поняття «ризик», «ризик для здоров'я», «популяція, що піддається ризику»
2. Поняття індивідуального і популяційного ризику.
3. Що таке чинник ризику?
4. Які типи епідеміологічних досліджень використовуються при вивченні і оцінці ризиків?
5. Охарактеризуйте епідеміологічні дослідження типу «випадок — контроль» і когортне дослідження.
6. Які показники належать до показників ризику?

7. Що являють собою показники індивідуального ризику і як вони обчислюються?

8. Що являють собою показники популяційного ризику і як вони обчислюються?

9. Чим відрізняються поняття відносного і абсолютного ризиків за клінічним змістом?

10. Який спеціальний показник використовується для оцінки відносного ризику в дослідженнях типу «випадок-контроль»?

11. Як розрахувати і інтерпретувати показник «відношення шансів»?:

12. Які дослідження проводяться для оцінки ефективності лікувальних і профілактичних втручань?

13. Які показники найчастіше використовуються для оцінки ефекту втручання?

14. Що являє собою показник «Зниження відносного ризику» (ЗВР)?

15. Що являє собою показник «Зниження абсолютного ризику» (ЗАР)?

16. Що являє собою показник числа хворих, яких потрібно лікувати протягом певного часу, щоб запобігти одному несприятливому результату — ЧХНР?

17. Яким чином оцінюються показники ЗВР, ЗАР, ЧХНР?

Література

Біостатистика. За загальною редакцією В.Ф. Москаленко. К.: Книга плюс, 2009. 184 с

Основи доказової медицини. За редакцією М.П. Скакун. Тернопіль, Укрмедкнига. 2005. 244 с.

Голяченко О. М., Сердюк А.М., Приходський А. А. Соціальна медицина, організація та економіка охорони здоров'я. Тернопіль: Джура, 1997. 327 с.

Програмні тестові питання з соціальної медицини та охорони здоров'я. За редакцією Ю.В. Вороненка. Тернопіль: Укрмедкнига, 2001. 316 с.

Біостатистика: навчально-методичний посібник. За редакцією В.І. Клименко. Запоріжжя, 2011. 106 с.

Навчальне видання

**БАБІЄНКО Володимир Володимирович,
МОКІЄНКО Андрій Вікторович,
ЛЕВКОВСЬКА Вікторія Юріївна**

Біостатистика

Навчально-методичний посібник

Головний редактор
Комп'ютерна верстка

Й. О. Бурчо
О. В. Замойська

Формат 60 x 84 1/16. Ум. друк. арк. 10,46
Наклад 100 прим. Зам.

Оригінал-макет виготовлено
в редакційно-видавничій фірмі «Прес-кур'єр».
Свідоцтво про внесення видавця до Державного реєстру видавців,
виготівників і розповсюджувачів видавничої продукції
серія ДК № 3764 від 22.04.2010 р.
(65076, м. Одеса, пл. Б. Дерев'янка, 1, оф. 717,
тел./факс (0482) 64-96-58, e-mail: pk.gazeta.odessa@gmail.com

Віддруковано з готового оригінал-макета.
Видавництво і друкарня «Астропринт»
65091, м.Одеса, вул. Разумовська, 21. Тел. 7-855-855.
www.astroprint.ua e-mail: astro_print@ukr.net
Свідоцтво суб'єкта видавничої справи
ДК №1373 від 28.05.2003 р.