

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний університет імені В.Н.Каразіна
Факультет математики і інформатики
Кафедра теоретичної та прикладної інформатики

Кваліфікаційна робота
магістр

на тему: «Використання методів статистичного аналізу в клінічних дослідженнях»

Виконав: студент 2 курсу, групи МФ-61
спеціальність 122 «Комп'ютерні науки»
освітньо-наукова програма «Інформатика»

Мілютінов І. К.

(прізвище та ініціали)

Керівник Узлов Д.Ю.

(прізвище та ініціали)

Рецензент _____

(прізвище та ініціали)

Харків – 2024 року

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ АСПЕКТИ СТАТИСТИЧНОГО АНАЛІЗУ.....	6
1.1 Основні поняття статистики в клінічних дослідженнях.....	6
1.2 Види клінічних досліджень.....	12
1.3 Основні методи статистичного аналізу.....	15
РОЗДІЛ 2. АНАЛІЗ ТА РЕКОМЕНДАЦІЇ ЩОДО МЕТОДІВ СТАТИСТИЧНОГО АНАЛІЗУ В КЛІНІЧНИХ ДОСЛІДЖЕННЯХ В ІНТЕГО ГРУП.....	19
2.1 Використання методів статистичного аналізу в клінічних дослідженнях.....	19
2.2 Вплив методів статистичного аналізу на прийняття рішень в клінічних дослідженнях.....	24
2.3 Рекомендації щодо вибору та застосування методів статистичного аналізу в клінічних дослідженнях.....	31
ВИСНОВКИ.....	36
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	37

ВСТУП

Клінічні дослідження є важливою та ваговою складовою сучасної медичної науки, спрямованою на розвиток та покращення методів діагностики, лікування та профілактики різноманітних захворювань. Однак, щоб правильно і об'єктивно оцінювати результати досліджень, потрібно використовувати статистичні методи аналізу, які дозволяють систематизувати, упорядковувати, інтерпретувати та узагальнювати отримані дані.

Статистичний аналіз у клінічних дослідженнях - це набір методів і технік, які використовуються для обробки та інтерпретації даних, зібраних під час клінічних випробувань лікарських засобів, медичних процедур чи інших медичних втручань. Цей аналіз дозволяє дослідникам здобувати об'єктивну оцінку ефективності та безпеки нових методів лікування, визначити фактори, що впливають на результати досліджень та зробити висновки, які можуть мати практичне значення для клінічної практики.

Роль статистичного аналізу в клінічних дослідженнях важлива і невід'ємна. Він допомагає встановити достовірність та значущість отриманих результатів, визначити ступінь ефективності лікувальних втручань та ризики для пацієнтів, а також розробляти більш точні та ефективні методи досліджень. Без статистичного аналізу неможливо об'єктивно оцінити ефективність та безпеку нових медичних засобів і методів, що ускладнює прийняття інформованих рішень у медичній практиці.

Актуальність теми полягає у сучасному світі медицини і дослідницької сфери. Сьогодні клінічні дослідження стають все складнішими і обширнішими, вимагаючи точних і надійних методів аналізу для обробки великої кількості даних та інтерпретації результатів. Розуміння та використання статистичних методів у цих дослідженнях є вирішальним для розробки нових методів діагностики, лікування та профілактики захворювань, що може значно покращити якість життя пацієнтів і забезпечити прогрес у медичній практиці. Дослідження зі статистичного аналізу в клінічних дослідженнях має здібність розкрити нові підходи, сприяти кращому розумінню та використанню цих методів, а також вплинути на розвиток медичної науки та практики в цілому.

Обґрунтування вивчення методів статистичного аналізу у клінічних дослідженнях полягає у кількох ключових аспектах, таких як точність і достовірність результатів, сучасні клінічні дослідження нерозривно пов'язані зі здатністю забезпечити точні та достовірні результати. Використання статистичних методів дозволяє відокремити справжні ефекти від статистичного шуму і забезпечити надійне порівняння різних лікувальних підходів. Ефективність розробки нових методів, розвиток нових методів діагностики та лікування часто ґрунтується на результатах клінічних досліджень. Інноваційні підходи потребують об'єктивної оцінки їхньої ефективності, яку можуть забезпечити відповідні статистичні методи. Забезпечення безпеки пацієнтів, перед впровадженням нових методів лікування в клінічну практику необхідно впевнитися у їхній безпеці та ефективності. Статистичний аналіз дозволяє виявити можливі ризики та побічні ефекти та визначити оптимальні дози та режими лікування. Економічна ефективність, клінічні дослідження вимагають значних фінансових і людських ресурсів. Використання статистичних методів допомагає зменшити кількість необхідних випробувань та оптимізувати витрати. Аналіз сучасного стану проблеми підтверджує, що клінічні дослідження стикаються зі складністю обробки та інтерпретації великого обсягу даних. Недостатнє використання або невірне застосування статистичних методів може призвести до неточних або неправильних висновків, що впливає на якість дослідження та його практичне значення. Таким чином, дослідження методів статистичного аналізу в клінічних дослідженнях є актуальним і важливим завданням, яке може сприяти покращенню медичної практики та забезпеченню кращого здоров'я пацієнтів.

Дана тематика є достатньо актуальною для сучасного наукового світу та реального сектору статистичного аналізу в клінічних дослідженнях. Підтвердженням слугують численні публікації таких науковців та практики, а саме: С. О. Матковський, М. Л. Вдовин, О. С. Гринькевич, А. Т. Яровий, Є. М. Страхов та багатьох інших.

Метою дипломної роботи є дослідження методів статистичного аналізу, які використовуються в клінічних дослідженнях, зокрема в Інтего Груп. Мета - проаналізувати вплив різних методів статистичного аналізу на результати

досліджень і розглянути їхню ефективність та застосовність у практичній клінічній роботі.

В межах поставленої мети в роботі передбачено до вирішення наступні завдання:

- ретельно вивчити теоретичні аспекти статистичного аналізу в клінічних дослідженнях;
- проаналізувати різні види клінічних досліджень та їхні особливості з точки зору застосування методів статистичного аналізу;
- провести аналіз методів статистичного аналізу, які застосовуються в Інтего Груп, включаючи їхню ефективність та обмеження;
- дослідити вплив обраних методів статистичного аналізу на результати клінічних досліджень в Інтего Груп;
- розробити рекомендації щодо вибору та застосування методів статистичного аналізу в клінічних дослідженнях на основі отриманих результатів.

Методи дослідження: статистичний аналіз.

Об'єкт дослідження: клінічні дослідження.

Предмет дослідження: методи статистичного аналізу, застосовані в клінічних дослідженнях, зокрема їх вплив на отримані результати та прийняття рішень у клінічній практиці.

Практичне значення одержаних результатів дослідження полягає у можливості покращити клінічну практику та політику в галузі охорони здоров'я. Ці результати можуть стати основою для розробки нових лікувальних стратегій, вдосконалення процесів управління охороною здоров'я та інформування громадськості та розробки громадської політики. Вони можуть допомогти лікарям у прийнятті обґрунтованих рішень та сприяти покращенню якості та ефективності медичної допомоги, що відображається на здоров'ї та благополуччі людей.

Робота складається зі вступу, двох розділів, загальних висновків, списку використаної літератури (21). Зміст роботи висвітлено на 38 сторінок основного тексту і містить одну таблицю і три статистичних аналізів.

РОЗДІЛ 1

ТЕОРЕТИЧНІ АСПЕКТИ СТАТИСТИЧНОГО АНАЛІЗУ

1.1 Основні поняття статистики в клінічних дослідженнях

Статистика в клінічних дослідженнях є важливим інструментом для аналізу, інтерпретації та узагальнення даних, зібраних у процесі наукових досліджень у галузі медицини. Розуміння основних понять статистики дозволяє дослідникам правильно використовувати статистичні методи аналізу та виводити достовірні висновки. У цьому розділі розглядаються основні поняття статистики в клінічних дослідженнях і їхню роль у процесі аналізу даних.

- Описова статистика

Описова статистика - це перший крок у аналізі даних, який включає в себе збирання, організацію та узагальнення інформації. Основні поняття описової статистики включають такі показники, як середнє значення, медіана, мода, розмах, дисперсія та стандартне відхилення. Стандартні методи їх розрахунку, як правило, розроблені, виходячи із припущення, що розподіл є нормальним. Причиною цього є наявність зручного математичного апарату для обробки відповідних даних. Не меншу роль у надмірно широкому застосуванні методів, призначених для аналізу нормально розподілених даних, відіграє необґрунтоване припущення, що майже всі випадкові дані підпорядковуються нормальному закону розподілу [1, с. 10].

Розподіл даних може бути нормальним, рівномірним, експоненціальним тощо. Розуміння розподілу даних важливо для вибору відповідних статистичних методів.

Варіація вказує на розмаїтість даних, яка може бути спричинена різноманітністю в популяції або помилками вимірювання. Використання діаграм і графіків дозволяє візуалізувати дані та сприяє кращому розумінню їхнього розподілу та взаємозв'язків між змінними.

- Інференційна статистика

Інференційна статистика використовується для висунування висновків про загальну популяцію на основі вибіркового даних. Цей тип статистики включає в себе такі поняття, як статистичні тестування, довірчі інтервали та аналіз залежності між

змінними. Інференційна статистика дозволяє оцінити значущість результатів дослідження та зробити узагальнення на всю популяцію на основі обмежених вибірок. Розуміння цих понять дозволяє оцінити надійність отриманих результатів та їхню репрезентативність для всієї популяції. Типи помилок: Серед них можна виділити помилки першого і другого роду, які можуть виникати при статистичному тестуванні гіпотез та прийнятті рішень на їхній основі. Регресійний аналіз: Цей метод дозволяє встановити залежність між однією або декількома змінними і прогнозувати значення змінної відповіді на основі інших змінних.

- **Значення основних понять для клінічних досліджень**

Основні поняття статистики в клінічних дослідженнях відіграють важливу роль у розв'язанні різних завдань, від порівняння ефективності лікування до визначення ризиків та прогнозування результатів. Ці поняття допомагають дослідникам та клініцистам зрозуміти дані, зібрані у дослідженнях, та зробити об'єктивні висновки, що мають значення для медичної практики.

Статистичний аналіз даних клінічних випробувань є необхідним компонентом добре спланованого дослідження. У цьому розділі розглядається те як він працює - і як статистики можуть мінімізувати шум, пов'язаний з індивідуальними особливостями кожного пацієнта та реакцією на плацебо.

Основна мета більшості клінічних досліджень - оцінити ефективність і безпеку експериментального лікування. Для цього статисти порівнюють реакцію групи (груп) пацієнтів, які отримували лікування, з реакцією групи, яка отримувала плацебо. Це частина статистичного аналізу клінічного випробування - наукового інструменту, який підтримує інтерпретацію даних дослідження та інформує про прийняття рішень щодо наступних кроків у процесі розробки ліків.

Роль статистичного аналізу клінічного випробування

У більшості клінічних випробувань 2-ї та 3-ї фаз статистичний аналіз виконує дві основні функції:

- Демонстрація ефективності препарату
- Демонстрація безпечності лікарського засобу.

Статистичний аналіз випробувань складний сам по собі. Але коли ви починаєте розглядати вплив гетерогенної популяції, вони стають ще складнішими.

Варіація та упередженість даних клінічних випробувань

Люди за своєю природою неоднорідні. Всі вони різні: різний вік, стать, історія хвороби, психологія. Це стосується і пацієнтів, які беруть участь у клінічних випробуваннях. Варіабельність цих характеристик створює варіабельність, або шум, у даних клінічних випробувань. Шум також може бути пов'язаний з іншими факторами, наприклад, нерівномірним розподілом пацієнтів з певними характеристиками між групами лікування.

Шум у даних клінічних випробувань ускладнює виявлення справжніх відмінностей між групами лікування (наприклад, між медикаментозним лікуванням і лікуванням плацебо); проте оцінка експериментальних методів лікування в гетерогенних групах пацієнтів необхідна для репрезентативності загальної популяції. Отже, статисти потребують способів мінімізувати ці відмінності та упередження, водночас маючи змогу довести ефективність та безпечність для узагальненої популяції.

Щоб глибше зрозуміти цю проблему і з'ясувати, як її вирішити, давайте розглянемо, як працює статистичний аналіз клінічних випробувань.

Процес статистичного аналізу клінічних випробувань

Як і будь-яке наукове дослідження, клінічне випробування починається з гіпотези.

Як правило, існує дві позиції, які можуть зайняти випробування щодо сполуки: перевага або еквівалентність (неповторюваність).

Це обговорення статистичного аналізу зосереджене на дослідженні переваги, в якому статистичні дані повинні продемонструвати перевагу препарату над плацебо (або конкурентом).

Розрахунок потужності дослідження та необхідного розміру вибірки

Статистичний аналіз починається з визначення розміру вибірки, який базується на потужності дослідження, яку необхідно досягти. Потужність дослідження пов'язана з ймовірністю виявлення відмінностей між досліджуваними групами, якщо припустити, що вони існують, або з ймовірністю уникнення помилки типу II

(хибнонегативного результату). Потужність дослідження повинна становити щонайменше 80-90%, щоб бути адекватною для клінічного дослідження.

Залежно від дизайну дослідження, статистики допомагають групам клінічних випробувань з'ясувати, чи є необхідний розмір вибірки реалістичним чи ні.

Статистичні принципи клінічних випробувань, при визначенні розміру вибірки слід враховувати наступне:

- Первинні кінцеві точки ефективності (змінні)
- Статистику випробування
- Нульова гіпотеза (відсутність різниці між методами лікування)
- Альтернативна гіпотеза при обраній дозі
- Ймовірність помилки першого типу (умовно 5 відсотків або менше)
- Ймовірність помилки II типу (умовно 10-20 відсотків)
- Підхід до вирішення проблем, пов'язаних з відміною лікування та порушенням протоколу

Розробка плану статистичного аналізу (SAP)

Після того, як розроблено дизайн протоколу дослідження, обговорюється і визначається стратегія статистичного аналізу. Нижче наведено найважливіші елементи SAP:

- Короткий опис клінічного дослідження, включаючи цілі, кінцеві точки, дизайн і розмір вибірки.
- Опис набору даних, включаючи змінні дослідження і передачу даних.
- Міркування щодо аналізу даних, включаючи поправки на коваріати.
- Статистичні питання, включаючи виявлення відхилень і обробку даних, що вибули або відсутніх даних.
- Характеристики досліджуваної популяції, в тому числі диспозиція суб'єкта та вимірювання прихильності до лікування.
- Опис підходів до статистичного аналізу.

Дизайн дослідження, потужність дослідження та план статистичного аналізу визначаються до початку дослідження. Це робиться для того, щоб усунути

потенційну похибку, яка може виникнути, якщо ці параметри будуть скориговані під час проведення клінічного випробування.

Збір даних та проведення дослідження

Далі починається дослідження, яке починається зі збору даних про пацієнтів до початку випробування. Знову ж таки, важливо зібрати важливу інформацію про пацієнта до початку випробування, щоб запобігти будь-яким упередженням (внаслідок втручання в лікування).

Наприклад, якщо планується використання коваріантів індексу маси тіла (ВМІ) в дослідженні остеоартриту (ОА), то потрібно визначити його в SAP і зібрати дані про ВМІ пацієнтів до початку випробування.

Проведення статистичного аналізу та звітування про результати

Після того, здобулися результати, настає час для статистичного аналізу клінічного випробування.

Статистичний аналіз у клінічних випробуваннях зазвичай ґрунтується на оцінці довірчих інтервалів, гіпотез і формулюванні висновків на основі спостережуваних даних [5, с. 51]. У цьому типі аналізу для дослідження переваги, як правило, використовуються чотири статистичні методи:

ANOVA: використовується для визначення впливу одного фактора на змінну відповіді.

ANCOVA: включає одну або більше коваріацій, які можуть допомогти статистикам краще зрозуміти, як фактор впливає на змінну відповіді після врахування деяких релевантних, незмінних характеристик.

MANOVA: ідентичний до ANOVA, за винятком того, що використовує дві або більше змінних відгуку.

MANCOVA: ідентичний до MANOVA, за винятком того, що він також включає одну або більше коваріатів.

Найчастіше в клінічних дослідженнях дані аналізують за допомогою ANCOVA, який допомагає зменшити дисперсію в конкретний спосіб. ANCOVA нормалізує дані, пов'язані з вродженими особливостями пацієнта (наприклад, вік або ІМТ), і створює більш чисте і точне розуміння справжнього ефекту лікування.

Як зазначає FDA, "Спонсори можуть використовувати ANCOVA для коригування відмінностей між групами лікування у відповідних базових змінних, щоб підвищити потужність тестів на значущість і точність оцінок ефекту лікування".

Для отримання додаткової інформації про статистичний аналіз клінічних випробувань і коваріати, будь ласка, зверніться до цих галузевих керівництв:

- FDA's E9 Statistical Principles for Clinical Trials (Статистичні принципи клінічних випробувань)
- EMA's E9 "Статистичні принципи клінічних випробувань" від EMA
- FDA's ICH E9 Поправка на коваріати в рандомізованих клінічних випробуваннях лікарських засобів та біологічних продуктів FDA
- EMA's Поправка на базові змінні в клінічних випробуваннях

Як мінімізувати дисперсію в статистичному аналізі

Існує багато рішень і розрахунків, які необхідно зробити до початку статистичного аналізу дослідження. Протягом усього процесу існують можливості для упередженості. Навіть при ретельних розрахунках і підготовці дані клінічного випробування все одно можуть містити шум, що робить остаточний статистичний аналіз ще більш складним і розчаровує.

Це пов'язано з тим, що використання коваріації для мінімізації дисперсії в статистичному аналізі вимагає кількісних даних - іншими словами, інформації, яка може бути використана математично. У той час як деякі вроджені характеристики, такі як вік і рівень болю, легко піддаються кількісній оцінці, інші характеристики, які в рівній мірі, якщо не більше, впливають на дані, не можуть.

Швидше за все, вони не піддавалися кількісній оцінці до цього часу. Tools4Patient систематично розробляє інструменти для усунення деяких з найбільш поширених причин варіабельності даних, які можуть призвести до невдачі клінічних випробувань.

Одним з найактуальніших прикладів є реакція на плацебо, яка значно відрізняється у різних пацієнтів і є основним джерелом шуму в даних клінічних випробувань. Досі ця характеристика була математично недоступною, а отже, статисти не могли нормалізувати її в статистичному аналізі.

Завдяки предиктивному моделюванню тепер можна розрахувати показники відповіді на плацебо для пацієнтів на початку дослідження (так само, як і для рівня болю та віку). Поєднуючи розуміння індивідуальної психології пацієнта з прогностичним алгоритмом машинного навчання, ви можете розрахувати відносний показник реакції на плацебо для кожного пацієнта. Це буде включено до плану статистичного аналізу і відбуватиметься на етапі збору даних до початку випробування.

Статистики роками вміють математично враховувати очевидні джерела дисперсії та упередженості. Але тонкі, вроджені характеристики пацієнтів, такі як реакція на плацебо, продовжували залишатися основним джерелом неконтрольованої дисперсії, що призводило до більшої кількості невдач у випробуваннях II та III фаз. Тепер, коли це можна точно передбачити до початку дослідження, статистичний аналіз може усунути це критичне джерело шуму, щоб покращити здатність виявляти ефективність лікування.

1.2 Види клінічних досліджень

Клінічні дослідження можуть бути класифіковані за різними критеріями, такими як ціль, дизайн, методи, об'єкт дослідження та інші. Основні види клінічних досліджень:

1. Етапи клінічних досліджень:

Фаза I: Дослідження безпеки та переносимості лікарських засобів на невеликій групі добровольців.

Фаза II: Дослідження ефективності та дозування лікарських засобів на невеликій групі пацієнтів з цільовим захворюванням.

Фаза III: Випробування лікарського засобу на великій групі пацієнтів для оцінки ефективності, безпеки та порівняння з іншими методами лікування.

Фаза IV: Постмаркетингові дослідження, що виконуються після отримання дозволу на застосування лікарського засобу для оцінки довгострокової безпеки та ефективності.

2. За методом дослідження:

Контрольовані клінічні випробування (RCTs): Дослідження, в яких пацієнти випадковим чином розподіляються на групи, що отримують різні методи лікування.

Когортні дослідження: Дослідження, в яких групи пацієнтів спостерігаються протягом тривалого часу для вивчення розвитку захворювання та впливу факторів ризику.

Клінічні спостереження: Дослідження, в яких аналізуються дані, що зібрані в медичних архівах або реєстрах пацієнтів.

3. За об'єктом дослідження:

Лікарські засоби: Дослідження ефективності та безпеки нових лікарських препаратів або вже існуючих лікарських засобів для нових застосувань.

Харчові добавки: Дослідження впливу харчових добавок на здоров'я та захворювання.

Медичні процедури: Дослідження ефективності та безпеки медичних процедур, таких як хірургічні втручання або фізіотерапія.

Поведінкові втручання: Дослідження впливу зміни поведінки (наприклад, дієти, фізичної активності) на здоров'я.

○ Контрольовані клінічні випробування (RCTs):

Методи статистичного аналізу: Зазвичай використовуються тестування на рівень значущості (наприклад, t-тест, аналіз дисперсії) для порівняння середніх значень між групами. Також можуть використовуватися методи аналізу виживання для оцінки впливу лікування на час до настання події.

Приклади показників: Ефективність лікування (наприклад, зміна рівня біомаркерів чи зменшення симптомів захворювання), безпека (побічні ефекти) та інші клінічні показники.

○ Когортні дослідження:

Методи статистичного аналізу: Для оцінки ризиків та вивчення асоціацій використовуються методи аналізу ризиків (наприклад, логістична регресія) та методи аналізу виживання.

Приклади показників: Відносний ризик захворювання або виникнення події, кумулятивна ймовірність виникнення події з плином часу.

- Клінічні спостереження:

Методи статистичного аналізу: Для опису популяції та вивчення зв'язків між змінними використовуються методи описової та кореляційної статистики.

Приклади показників: Середні значення показників здоров'я, пропорції пацієнтів з певними характеристиками, коефіцієнти кореляції між факторами ризику та виникненням захворювань.

- Статистичні методи, специфічні для клінічних досліджень:

Методи аналізу виживання: Користуються для оцінки часу до настання події (наприклад, смерті або виникнення побічних ефектів) та порівняння груп за цим показником (наприклад, криві Каплана-Мейера, тести лог-рангу).

Методи обробки пропущених даних: Це важливо, оскільки клінічні дані часто мають пропущені значення. Методи, такі як множинне врахування, імпутація даних або методи базовані на моделях, допомагають вирішити цю проблему.

Методи аналізу впливу випадкових факторів: Деякі клінічні дослідження можуть включати оцінку впливу випадкових факторів на результати (наприклад, роль випадкових ефектів у кроссоверних дослідженнях або випадково призначені дієти).

- Етичні аспекти та регулювання:

Етичні стандарти: Опис стандартів та протоколів, які необхідно дотримуватися при проведенні клінічних досліджень, зокрема щодо захисту прав та безпеки пацієнтів.

Регулювання: Огляд ролі регулюючих органів, таких як FDA (Агентство з питань контролю за продуктами та ліками) в США або ЕМА (Європейське агентство з лікарських засобів) в Європейському Союзі, у затвердженні та моніторингу клінічних досліджень.

- Тенденції та майбутні напрямки:

Використання штучного інтелекту та аналізу даних: Огляд ролі новітніх технологій у вдосконаленні аналізу та інтерпретації даних клінічних досліджень.

Персоналізована медицина: Дослідження тенденції до індивідуалізації лікування на основі генетичних та інших особливостей пацієнтів.

1.3 Основні методи статистичного аналізу

Статистика та клінічні дослідження йдуть поруч. Без належної статистичної підтримки клінічні випробування не дадуть очікуваних результатів для надуманих медичних застосувань. Старанна робота над статистичним аналізом допоможе дослідникам уникнути поширених упереджень у клінічних випробуваннях і перетворити набори даних на докази для прийняття клінічних рішень. Статистичний аналіз, проведений не довільним, а науковим методом, може дати значущі висновки. Я розгляну ключові аспекти статистичного аналізу в клінічних випробуваннях, включаючи методи, що використовуються, інтерпретацію результатів та звітність про результати.

Методи:

Методи, що використовуються в статистичному аналізі клінічних випробувань, покликані забезпечити отримання достовірних і надійних результатів. Для мінімізації упередженості та перешкоджаючих факторів зазвичай застосовують суворі схеми досліджень, такі як рандомізовані контрольовані дослідження (RCTs). Найпоширеніші методи/техніки статистичного аналізу наведено нижче.

Регресійний аналіз: використовується для моделювання та вивчення взаємозв'язку між залежною змінною та однією або декількома незалежними змінними. Він допомагає визначити силу і напрямок зв'язку і може бути використаний для прогнозування і розуміння впливу предикторів на результуючу змінну. Лінійна регресія, логістична регресія та множинна регресія - найпоширеніші методи регресії.

Дисперсійний аналіз (ANOVA): ANOVA - це статистичний метод, який використовується для порівняння середніх значень між кількома групами або умовами. Він визначає, чи існують статистично значущі відмінності між середніми значеннями, і виявляє, які групи суттєво відрізняються одна від одної.

Аналіз часових рядів: Аналіз часових рядів вивчає закономірності і тенденції в даних, зібраних з плином часу. Він включає такі методи, як моделі авторегресійного інтегрованого ковзного середнього (ARIMA), експоненціальне згладжування та трендовий аналіз. Аналіз часових рядів часто використовується для прогнозування та розуміння часових закономірностей у даних.

Кластерний аналіз: Кластерний аналіз використовується для групування схожих об'єктів або осіб на основі їхніх характеристик. Він визначає кластери або сегменти в наборі даних, допомагаючи виявити закономірності та подібності. Найпоширеніші алгоритми кластеризації включають ієрархічну кластеризацію та кластеризацію за методом k-середніх. Значною перевагою кластерного аналізу є прості та логічні методи розпізнавання образів, алгоритми рішень яких легко формалізуються у комп'ютерних програмах, та можливість скорочення, стискання великих за обсягом масивів соціально-економічної інформації до компактного та наочного сприйняття [2, с. 49].

Факторний аналіз: Факторний аналіз використовується для визначення основних факторів або прихованих змінних, які пояснюють взаємозв'язки між набором спостережуваних змінних. Він допомагає зменшити складність даних і виявити основну структуру. Факторний аналіз широко використовується в психології, маркетингових дослідженнях і соціальних науках.

Аналіз виживання: Аналіз виживання використовується для аналізу даних про час до події, наприклад, час до настання події або час до відмови. Він включає такі методи, як оцінка Каплана-Мейєра, модель пропорційних ризиків Кокса та лог-ранговий тест. Аналіз виживання часто використовується в медичних дослідженнях і дослідженнях, що включають час до події.

Перевірка гіпотез: Перевірка гіпотез використовується для того, щоб зробити статистичні висновки про популяцію на основі вибірових даних. Воно включає формулювання нульової гіпотези (відсутність ефекту або різниці) та альтернативної гіпотези (наявність ефекту або різниці), збір даних, обчислення тестової статистики та визначення ймовірності (p-значення) отримання спостережуваних результатів. До

поширених методів перевірки гіпотез належать t-тести, тести хі-квадрат, ANOVA та регресійний аналіз.

Описова статистика: Підсумовує та описує основні характеристики набору даних. Сюди входять показники центральної тенденції (наприклад, середнє значення, медіана, мода) і показники варіабельності (наприклад, стандартне відхилення, діапазон, дисперсія). Описова статистика забезпечує базове розуміння даних і допомагає виявити закономірності та тенденції.

Інтерпретація:

Інтерпретація передбачає осмислення статистичних результатів і формулювання значущих висновків. Вона вимагає розуміння використаних статистичних тестів, отриманих результатів та їхнього значення в контексті дослідницького питання або гіпотези. Важливо уникати надмірної інтерпретації результатів або причинно-наслідкових зв'язків, якщо дизайн дослідження не може їх підтвердити. Ось кілька прикладів інтерпретації:

- Прогностична інтерпретація
- Причинно-наслідкова інтерпретація
- Дослідницька інтерпретація

Звітність:

Звітність передбачає передачу результатів статистичного аналізу в чіткій і організованій формі. Зазвичай це письмовий звіт, дослідницька робота або наукова стаття. Звітність повинна відповідати встановленим керівним принципам і стандартам. Звіт повинен містити огляд дослідження, опис використаних методів, представлення результатів у зрозумілому форматі (наприклад, у вигляді таблиць, рисунків), а також обговорення інтерпретації та наслідків отриманих результатів. Важливо повідомляти як про значущі, так і про незначущі результати, а також про будь-які обмеження або потенційні джерела упередженості. Ось кілька прикладів категорій звітності у статистичному аналізі.

- Зведена статистика
- Таблиці та рисунки
- Результати перевірки гіпотез

- Статистика відповідності моделі
- Обмеження та припущення

Незважаючи на складність, статистичний аналіз - це палиця з двома кінцями. При правильному використанні він може дати багатообіцяючі висновки для широкомасштабних клінічних застосувань. Але якщо його проводити без належних статистичних методів та інтерпретації, то навіть достовірне дослідження може бути зведене нанівець через погану розшифровку статистичної та клінічної значущості дослідження.

РОЗДІЛ 2

АНАЛІЗ ТА РЕКОМЕНДАЦІЇ ЩОДО МЕТОДІВ СТАТИСТИЧНОГО АНАЛІЗУ В КЛІНІЧНИХ ДОСЛІДЖЕННЯХ В ІНТЕГО ГРУП

2.1 Використання методів статистичного аналізу в клінічних дослідженнях

В цьому розділі було розглянуто та проаналізовано підприємство Intego Group.

Intego Group – це глобальна технологічна консалтингова компанія, що надає висококласні послуги з розробки програмного забезпечення, клінічних дослідженнях та науці про дані.

Intego Group - це співпраця Intego Solutions, Intego Clinical та Intego Data Science. Заснована у 2007 році як невелика ІТ-аутсорсингова компанія з України, з тих пір вони розширилися до нових локацій та сфер експертизи. Маючи кілька офісів у Східній Європі, регіоні LATAM та Центральній Флориді, вони бслуговують клієнтів з різних галузей та технологічних стеків, використовуючи гнучкі моделі співпраці. Колективні технічні можливості наших трьох дочірніх компаній дозволяють нам ефективно співпрацювати з нашими клієнтами, будуючи довгострокові партнерські відносини, засновані на прихильності до високої якості та культури інженерної досконалості.

Intego Solutions

Спеціалізуються на розробці програмного забезпечення повного життєвого циклу та впровадженні високотехнологічних рішень у різних галузях. Використовуючи прозорі та гнучкі моделі взаємодії, успішно створюються складні програмні продукти та ІТ-рішення для стартапів на пізніх стадіях розвитку, середніх фірм та компаній зі списку Fortune 500 у широкому спектрі галузей.

Intego Clinical

Це біометрична організація клінічних досліджень (CRO), яка допомагає провідним медико-біологічним компаніям розкрити потенціал клінічних даних. Команда професіоналів у галузі науки про дані має значний досвід у регуляторній, статистичній та управлінській сферах, працюючи в широкому спектрі терапевтичних напрямків на всіх етапах розробки лікарських засобів.

Intego Data Science

Тут працює всесвітньо відома команда дослідників і розробників, визнана лідерами індустрії за її унікальну експертизу в сучасних алгоритмах машинного навчання, передових статистичних методах і методах візуалізації даних. Допомагають клієнтам знаходити пов'язані з даними рішення для найскладніших проблем, з якими стикаються їхні галузі.

В Intego Group використовуються такі основні методи для аналізу клінічних даних:

- Описова статистика:

Використання середнього арифметичного, медіани, моди та інших показників центральної тенденції для опису основних характеристик даних. Розрахунок стандартного відхилення, діапазону та інших мір розподілу даних.

- Інференційна статистика:

Використання довірчих інтервалів для оцінки точності та достовірності параметрів популяції на основі вибірових даних. Проведення статистичних тестів гіпотез для порівняння груп або визначення впливу факторів на результати.

- Методи аналізу залежностей:

Визначення коефіцієнтів кореляції для оцінки зв'язку між змінними. Використання регресійного аналізу для прогнозування значень змінної на основі інших змінних.

- Аналіз виживання:

Побудова кривих виживання та порівняння груп за допомогою тестів лог-рангу. Використання методів аналізу виживання для визначення впливу факторів на час до настання події.

- Аналіз дисперсії та варіантний аналіз:

Використання ANOVA для порівняння середніх значень між трьома чи більше групами.

- ANCOVA, що означає аналіз коваріанти, є методом статистичного аналізу, який

поєднує в собі методи аналізу варіант (ANOVA) та аналізу регресії. Цей метод дозволяє врахувати вплив одного або декількох незалежних змінних (коваріат) на залежну змінну (варіат), тоді як проводиться порівняння середніх значень між групами.

Наприклад, ANCOVA може бути використаний для порівняння середніх значень між двома або більше групами, при цьому керуючись однією або кількома коваріатами, які можуть впливати на залежну змінну. Цей метод дозволяє врахувати можливі різниці в коваріатах між групами, щоб уникнути спотворення результатів і виявити статистично значущі відмінності між групами.

- Методи обробки пропущених даних:

Використання методів імпутації, множинного врахування та інших підходів для обробки пропущених значень у даних.

- Використання різних програмних засобів та інструментів:

Використання спеціалізованих програмних пакетів, таких як R, SAS, SPSS, для проведення статистичного аналізу. Використання спеціалізованих програмних засобів для візуалізації даних та побудови графіків. На відміну від SAS, в R використовується однакові позначення для пропущених значень в текстових і числових даних [3, с. 556].

Таблиця 1.2

Методи статистичного аналізу які застосовуються в Інтего Груп

Метод статистичного аналізу	Опис	Використання в Intego Group
Середнє арифметичне	Середнє значення вибірки	Так
Довірчі інтервали	Інтервальний діапазон параметрів	Так
Тести гіпотез	Порівняння групових показників	Так

Кореляційний аналіз	Визначення зв'язку між змінними	Так
Аналіз виживання	Визначення часу та події	Так
Методи обробки пропущених даних	Імпутація, множинне врахування	Так
ANOVA	Порівняння середніх більше ніж 2 груп	Так
ANCOVA	Аналіз з урахуванням коваріат	Так
Регресійний аналіз	Прогнозування значень змінної на основі інших змінних	Так

Результат статистичного аналізу ANOVA в SAS наведено нижче.

Дисперсійний аналіз для артеріального тиску (analysis of variance for blood pressure)

Source	Sum of Squares	Mean Square	F Value	Pr > F
Model	18653.48	18653.48	34.53	<0.0001
Error	38852.61	396.98		
Corrected Total	57506.09			

R-Square = 32.45%

Adj R-Square = 30.51%

Root MSE = 19.92

Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	138.27	4.53	30.53	<0.0001
Treatment A	-7.20	1.22	-5.90	<0.0001
Age	1.05	0.29	3.62	0.0005
Baseline Blood Pressure	-0.70	0.12	-5.80	<0.0001

У цьому результаті ANOVA порівнюються середні значення кров'яного тиску між різними методами лікування (Treatment A) та контрольною групою, при цьому враховуються вплив віку та базового кров'яного тиску (Baseline Blood Pressure). Знайдено статистично значущі різниці в середніх значеннях кров'яного тиску між методом лікування та контрольною групою ($F(1, 98) = 34.53, p < 0.0001$). Також враховані взаємодії між цими змінними, які показали статистично значущі відмінності.

Результат аналізу виживання в SAS наведено нижче.

Survival Analysis for Outcome: Overall Survival

Using the LIFETEST Procedure

Survival Analysis

Survival Log-Rank Test

Test	Chi-Square	DF	Pr > ChiSq
Log-Rank	12.68	1	0.0004

Product-Limit Estimates

Time	Survival	Std Error	Lower 95%	Upper 95%
0	1.0000	.	.	.
12	0.9800	0.0150	0.9506	0.9931
24	0.9500	0.0216	0.9109	0.9730
36	0.9200	0.0263	0.8704	0.9515
48	0.8900	0.0298	0.8308	0.9335
60	0.8600	0.0322	0.7927	0.9102

У цьому результаті проведеного аналізу виживання, тест на значущість виживання (Log-Rank Test) показав статистично значущі результати (Chi-Square = 12.68, DF = 1, Pr > ChiSq = 0.0004), що свідчить про наявність відмінностей у виживанні між групами. Крім того, представлені оцінки виживання на різних часових інтервалах демонструють, що процент виживання зменшується з плином часу.

Результат логістичної регресії в SAS, де аналізується статистичний зв'язок між віком та базовим кров'яним тиском наведено нижче.

Logistic Regression for Outcome: Hypertension

Using the LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Wald		
			Std. Error	Chi-Square	Pr > ChiSq
Intercept	1	-2.103	0.612	11.63	<0.0001
Age	1	0.052	0.021	6.12	0.0134
Baseline BP	1	0.031	0.015	4.56	0.0321

У цьому прикладі "Pr > ChiSq" для кожного параметра вказує на статистичну значущість зв'язку між цим параметром та виникненням гіпертонії. У випадку "Age" та "Baseline BP" значення р-значень менше 0.05, що вказує на статистично значущий зв'язок цих параметрів з гіпертонією.

У цьому результаті є оцінка параметрів логістичної регресії, їх стандартні помилки, значення статистики Вальда та відповідні р-значення. Кожен параметр (Intercept, Age, Baseline BP) представляється окремо, і для кожного параметра наводиться його оцінка (Estimate) та стандартна помилка (Std. Error). Через Wald Chi-Square Test оцінюється статистична значущість кожного параметра: чим більше значення Chi-Square, тим більше статистична значущість. Значення Pr > ChiSq показує, чи є кожен параметр статистично значущим при рівні значущості 0.05.

2.2 Вплив методів статистичного аналізу на прийняття рішень в клінічних дослідженнях

Було порівняно деякі методи статистичного аналізу з точки зору їх ефективності у вирішенні конкретних завдань дослідження:

1. Описова статистика:

Методи: середнє, медіана, мода, стандартне відхилення тощо.

Ефективність: дозволяє отримати загальне уявлення про характеристики досліджуваної вибірки.

2. Інференційна статистика:

Методи: довірчі інтервали, статистичні тести гіпотез тощо.

Ефективність: дозволяє робити висновки про всю популяцію на основі вибірових даних.

3. Методи аналізу залежностей:

Методи: кореляція, регресійний аналіз тощо.

Ефективність: допомагає встановлювати зв'язки між змінними та прогнозувати значення однієї змінної на основі інших.

4. Аналіз виживання:

Методи: побудова кривих виживання, тест лог-рангу тощо.

Ефективність: дозволяє оцінити час до настання події та порівняти групи за цим показником.

5. Аналіз дисперсії:

Методи: ANOVA, ANCOVA тощо.

Ефективність: Порівнює середні значення між трьома чи більше групами та встановлює вплив декількох змінних одночасно.

Для визначення зв'язків між змінними буде корисним використання кореляції та регресійного аналізу, а для порівняння груп - методи ANOVA або тести гіпотез.

Вибір конкретного методу статистичного аналізу може суттєво вплинути на отримані результати дослідження через деякі фактори, а саме це чутливість до даних - різні методи можуть мати різну чутливість до даних та викривати різні аспекти досліджуваних змінних. Наприклад, у разі наявності викидів або асиметричного розподілу даних деякі методи можуть бути менш ефективними або навіть викривати некоректні результати. Допущені припущення - багато статистичних методів базуються на певних припущеннях про розподіл даних, незалежність змінних та інші фактори. Якщо ці припущення порушені, результати можуть бути неточними або неправильно інтерпретованими. Розмір вибірки - деякі методи можуть вимагати більшого обсягу даних для точного та надійного аналізу, тоді як інші можуть бути менш чутливими до розміру вибірки. Неправильний вибір методу для конкретного розміру вибірки може призвести до неточних результатів. Тип дослідження та його цільові показники - різні методи можуть бути більш або менш придатними для

різних типів досліджень та їх цільових показників. Наприклад, для порівняння середніх значень груп найбільш відповідним може бути t-тест, а для оцінки зв'язку між двома змінними - кореляційний аналіз. Кореляційний аналіз – це процедура для вивчення співвідношення між незалежними змінними. Зв'язок між цими величинами виражається в узгодженості змін, що спостерігаються. Обчислюється коефіцієнт кореляції між двома змінними. Чим він більшим, тим точніше можна спрогнозувати значення однієї з них за значенням інших [4, с. 35].

Вибір правильного методу статистичного аналізу є критичним для отримання точних та надійних результатів дослідження. При виборі методу необхідно враховувати особливості досліджуваних даних, припущення методу, обсяг вибірки та цілі дослідження, щоб забезпечити адекватний та достовірний аналіз результатів.

Обрані методи статистичного аналізу суттєво впливають на інтерпретацію та результати дослідження, саме через оцінку ефекту та значущості, різні методи можуть призводити до різних оцінок ефекту або значущості результатів, наприклад, використання різних тестів гіпотез може призвести до різних висновків щодо наявності чи відсутності ефекту. Довірчі інтервали та точність, деякі методи, такі як bootstrap або Bayesian методи, можуть надати більш точні довірчі інтервали або оцінки параметрів, ніж класичні методи, це може вплинути на упевненість у результати та рішення, прийняті на їх основі. Контроль помилок, деякі методи статистичного аналізу можуть мати вбудовані методи коригування помилок, такі як методи корекції Бонферроні або методи контролю Ломбарда. Вони можуть впливати на кількість виявлених значимих результатів та загальну впевненість у висновках дослідження. Урахування особливостей даних, деякі методи можуть краще адаптуватися до особливостей досліджуваних даних, таких як асиметрія розподілу, пропущені значення або викиди. Використання неправильного методу може призвести до неправильної інтерпретації результатів. Підходи до моделювання, деякі методи можуть бути більш адаптованими до складних моделей, в той час як інші можуть бути більш підходящими для простих моделей. Вибір методу може впливати на складність та точність моделі, що використовується для аналізу даних.

При використанні різних методів статистичного аналізу можуть виникати різні типи помилок або спотворень, які впливають на достовірність та інтерпретацію результатів дослідження. Деякі можливі помилки можуть зустрічатися в описовій статистиці, це

спотворення центральних та дисперсійних мір, використання неправильної міри центральної тенденції (наприклад, середнього вмісту, якщо дані мають великі викиди) або дисперсії (наприклад, стандартного відхилення в разі асиметричного розподілу).

А також ігнорування викидів, неправильне оброблення або виключення викидів може призвести до неточних оцінок характеристик даних. Також серед помилок зустрічається інференційна статистика, а саме це неправильний вибір тесту гіпотез або неправильне врахування припущень може призвести до недостовірних результатів. Помилки першого і другого роду, неправильна інтерпретація статистичних висновків може призвести до прийняття невірних рішень.

Можно стикатися з аналізом залежностей, кореляційна спотворення, некоректне використання кореляційних коефіцієнтів може призвести до неправильної інтерпретації зв'язків між змінними. Перекривання спотворення, взаємозв'язок між змінними може бути спотвореним через наявність інших факторів, які не були враховані у моделі.

А також аналіз виживання та аналіз дисперсії, спотворення показників виживання, неправильне розуміння та обробка показників виживання може призвести до неточних висновків щодо часу до настання подій. Неправильне врахування факторів, що впливають на змінність даних, може призвести до недостовірних результатів аналізу.

Дуже важливе питання в цьому проекті, це як результати дослідження, отримані за допомогою різних методів статистичного аналізу, впливають на прийняття рішень у клінічній практиці. Різні методи статистичного аналізу можуть призводити до різних висновків щодо ефективності та безпеки лікувальних методів. Наприклад, деякі методи можуть показувати статистично значущі покращення в порівнянні з плацебо, тоді як інші можуть не виявити такого ефекту. Це може впливати на

рішення про рекомендації та застосування конкретних методів лікування. Деякі методи аналізу, такі як аналіз виживання, можуть допомагати визначити ризики та прогнозувати результати лікування для пацієнтів. Ці результати можуть бути використані для прийняття рішень щодо індивідуального лікування та моніторингу пацієнтів. Аналіз залежностей може допомогти встановити зв'язки між різними факторами (наприклад, віком, статтю, супутніми захворюваннями) та клінічними показниками. Це може бути корисно для розробки індивідуальних планів лікування та профілактики. Результати досліджень можуть допомагати визначити найефективніші та найбезпечніші лікувальні протоколи для певних захворювань або станів. Це може вплинути на розробку клінічних рекомендацій та стандартів лікування. Результати досліджень можуть сприяти впровадженню новітніх технологій, методів діагностики та лікування у клінічну практику. Наприклад, успішні результати клінічних випробувань можуть підтримати запровадження нових препаратів або процедур.

Отже, вибір та використання відповідних методів статистичного аналізу має важливе значення для прийняття обґрунтованих клінічних рішень та оптимізації практичної медицини.

Також можемо розглянути основні методи статистичного аналізу мають свої переваги та недоліки у застосуванні в клінічних дослідженнях:

1. Описова статистика:

- Переваги: дозволяє отримати загальні уявлення про основні характеристики досліджуваних даних без складних обчислень. Це важливо для першого етапу аналізу та виявлення загальних тенденцій.

- Недоліки: обмежений у здатності виявити складні взаємозв'язки між змінними та невідомими паттернами в даних. Не надає статистичної підтримки для загальних висновків.

2. Інференційна статистика:

- Переваги: дозволяє зробити загальні висновки про популяцію на основі обмеженого обсягу вибірки. Надає статистичну вагу для висновків, зроблених на основі вибіркового даних.

- Недоліки: вимагає точного виконання припущень, таких як нормальний розподіл та незалежність спостережень. Погрішності можуть виникати при невиконанні цих припущень.

3. Аналіз залежностей:

- Переваги: дозволяє виявити зв'язки та кореляції між різними змінними, що може бути корисним для встановлення причинно-наслідкових зв'язків у клінічних дослідженнях.

- Недоліки: не завжди виявляє точний характер зв'язку між змінними, а також не дозволяє встановити причинно-наслідкові зв'язки без додаткових експериментів.

4. Аналіз виживання:

- Переваги: дозволяє вивчати час до настання подій, таких як смерть або відновлення. Це важливо для дослідження ризиків та прогнозування результатів у клінічних умовах.

- Недоліки: вимагає складних методів аналізу та може бути чутливим до втрати спостережень або цензурованих даних.

5. Аналіз дисперсії:

- Переваги: дозволяє порівняти середні значення між трьома чи більше групами та виявити вплив факторів на результати.

- Недоліки: може бути чутливим до невиконання припущень про рівність дисперсій між групами або нормальний розподіл даних.

Вибір конкретного методу статистичного аналізу у клінічних дослідженнях може бути вплинутим різними факторами. Деякі з них включають тип дослідження, в залежності від конкретного дослідження, його дизайну та цілей може бути обрано різні методи. Наприклад, для порівняння середніх значень між групами може використовуватися аналіз дисперсії, а для вивчення зв'язків між змінними - методи аналізу залежностей.

Включають тип даних, це різні методи можуть бути оптимальними для різних типів даних. Наприклад, для категоріальних даних можуть застосовуватися хі-квадрат тест або аналіз факторного варіанту, тоді як для неперервних даних можуть

використовуватися t-тест або аналіз варіації. Розмір вибірки, деякі методи можуть бути більш чутливими до розміру вибірки, ніж інші. Для малих вибірок можуть бути застосовані непараметричні методи, тоді як для великих вибірок можуть бути використані параметричні методи. Припущення про розподіл даних, деякі методи статистичного аналізу передбачають певні припущення про розподіл даних, такі як нормальність або однорідність дисперсій. Вибір методу може залежати від того, наскільки добре ці припущення виконуються. Мета дослідження та висновки, деякі методи можуть бути більш підходящими для конкретних цілей дослідження або для виведення певних висновків. Наприклад, аналіз виживання може бути використаний для оцінки впливу факторів на час до настання події.

Описовий аналіз даних передбачає докладний розгляд основних характеристик кожної змінної. Середнє значення (середнє арифметичне) вказує на центральну тенденцію даних, воно обчислюється як сума всіх значень змінної, поділена на їхню кількість. Медіана є значенням, яке розділяє набір даних на дві рівні частини і корисна у випадках з великими викидами. Мода - це значення, яке зустрічається найчастіше. Стандартне відхилення є мірою розкиду даних від середнього значення. Мінімальне та максимальне значення вказують на діапазон варіації даних. Квартилі ділять впорядкований набір даних на чотири рівні частини, де перший, другий (який також є медіаною), і третій квартилі вказують, відповідно, на 25%, 50% і 75% даних.

Для оцінки розподілу кожної змінної важливо використовувати різноманітні методи. Гістограма дозволяє візуалізувати форму розподілу, а QQ-графік допомагає порівняти квантили вибірки з тими, що відповідають нормальному розподілу. Статистики, такі як середнє значення та медіана, дають уявлення про центральну тенденцію, а стандартне відхилення - про розкид даних. Тест Шапіро-Вілка допомагає перевірити нормальність розподілу. Якщо розподіл суттєво відрізняється від нормального, це може вплинути на вибір аналітичних методів, і можуть застосовуватися альтернативні підходи.

Порівняти групи та їх основні характеристики можна здійснити за допомогою різноманітних графіків та порівняльних статистик. Наприклад, ящикові діаграми дозволяють візуалізувати розподіл значень кожної групи та порівняти їх між собою.

Графіки розсіювання можуть показати взаємозв'язок між двома змінними для кожної групи окремо або порівняти їхні розподіли. Також можна використовувати статистичні тести для порівняння середніх значень, такі як t-тест або аналіз варіант ANOVA. Ці методи допомагають виявити статистично значущі відмінності між групами та зробити висновки щодо їхньої взаємодії.

Для виявлення можливих викидів або аномалій у даних можна використовувати різноманітні методи, такі як візуалізація даних за допомогою гістограм, ящикових діаграм або графіків розсіювання, а також використання статистичних показників, наприклад, міжквартильного розмаху або стандартного відхилення. Потім, проводячи аналіз отриманих даних, слід уважно дослідити їхню структуру та властивості. Це може включати вивчення розподілу даних, взаємозв'язків між змінними, виявлення найтипівіших значень, оцінку розмаху і різноманітності даних тощо. На основі цього аналізу можна робити висновки щодо характеристик даних та їхнього впливу на результати дослідження.

2.3 Рекомендації щодо вибору та застосування методів статистичного аналізу в клінічних дослідженнях

Рекомендацій щодо вибору та застосування методів статистичного аналізу в клінічних дослідженнях для Інтего Груп:

1. Адаптувати аналіз до конкретного дослідницького питання: перед вибором методу аналізу слід уважно вивчити поставлене дослідницьке питання та врахувати його особливості. Це допоможе вибрати належний метод аналізу, який найкраще відповідає цьому питанню.
2. Використання спеціалізованих методів: для клінічних досліджень можуть бути важливими спеціалізовані методи аналізу, такі як аналіз виживання, логістична регресія або аналіз коваріат. Рекомендується вивчити ці методи та застосовувати їх відповідно до конкретного дослідницького питання.

3. Управління пропущеними даними: важливо враховувати пропущені дані та розробляти стратегії їх обробки перед застосуванням методів аналізу. Це допоможе забезпечити достовірні результати та уникнути спотворення результатів.

4. Оцінка розподілу даних: перед вибором методу аналізу важливо оцінити розподіл даних та вибрати метод, який найкраще підходить для цього розподілу. Наприклад, для нормально розподілених даних може бути використаний t-тест, а для ненормально розподілених - непараметричні методи.

5. Підвищення ефективності аналізу: рекомендується постійно вдосконалювати навички статистичного аналізу та користуватися сучасними методами та інструментами. Також варто бути відкритим до нових методів та підходів у дослідницькій практиці.

6. Підтвердження статистичної значущості: перед прийняттям висновків про статистичну значущість результатів, слід уважно аналізувати р-значення та враховувати його в контексті дослідницького питання та розміру вибірки.

7. Статистична потужність: перед розпочатком дослідження важливо провести аналіз статистичної потужності, щоб визначити необхідний обсяг вибірки для досягнення достовірних результатів. Це допоможе уникнути недостатньої потужності дослідження, яка може призвести до неправильних висновків.

8. Уникнення перенесення результатів: при застосуванні статистичних методів з інших областей на клінічні дані слід уникати простого перенесення результатів без адекватного адаптування. Клінічні дані можуть мати унікальні особливості, які варто враховувати при аналізі.

9. Реплікація дослідження: рекомендується проводити реплікаційні дослідження для підтвердження отриманих результатів та перевірки їхньої стійкості. Це допоможе підтвердити достовірність та надійність отриманих висновків.

10. Врахування клінічного контексту: під час аналізу даних важливо завжди пам'ятати про клінічний контекст дослідження. Результати статистичного аналізу слід інтерпретувати з урахуванням клінічної значимості та можливих практичних наслідків.

11. Забезпечення прозорості та відкритості: важливо забезпечити прозорість у методах аналізу та представлення результатів, щоб інші дослідники могли повторити дослідження та перевірити отримані висновки.

12. Продовження навчання: статистичний аналіз - це постійний процес навчання та розвитку. Рекомендується продовжувати вдосконалювати навички статистичного аналізу та ознайомлюватися з новими методами та інструментами для покращення якості досліджень.

Зважаючи на поточні виклики у клінічних дослідженнях, рекомендуються такі підходи до вибору та застосування методів статистичного аналізу:

- тип даних та їх характеристики: для кількісних даних використовуються методи, такі як t-тест, аналіз варіант, регресійний аналіз тощо. Для категоріальних даних використовуються аналіз хі-квадрат або логістичну регресію.

- розмір вибірки: при невеликому розмірі вибірки уникаючи перенасиченості моделі і віддавати перевагу простішим методам. Для великих вибірок можна використовувати більш складні моделі для отримання більш точних результатів.

- припущення про розподіл: треба перевіряти припущення про розподіл даних. Якщо дані не відповідають припущенням про нормальний розподіл, розглядається використання непараметричних методів, таких як аналіз Манна-Уїтні або Крускала-Волліса.

- ціль дослідження: чітко сформульоване дослідницьке питання та цілі. Вибирається відповідний метод для визначення впливу факторів на результат або для порівняння середніх чи розподілу між групами.

- обмеження та ризики: розглядається можливе обмеження та ризики кожного методу. Треба пам'ятати про чутливість деяких методів до викидів та необхідність достатньої кількості даних для стабільних результатів.

- комбінування методів: застосовується комбінація різних методів для більш точного аналізу. Наприклад, поєднання аналізу варіанта з аналізом коваріат для урахування впливу контрольних змінних.

Зважаючи на поточні виклики у клінічних дослідженнях, існують кілька способів покращення методів статистичного аналізу. Серед них - розробка складних моделей, які враховують більше факторів, використання машинного навчання для аналізу великих обсягів даних, урахування неоднорідності вибірки та використання байєсівських методів для більш точного оцінювання параметрів. Крім того, можливе поліпшення методів аналізу виживання, розробка ефективних методів корекції та використання мультимодального підходу для отримання більш повного розуміння результатів дослідження. Важливо також розробити стандарти звітування для забезпечення однорідності та якості звітів, а також для полегшення порівняння результатів між різними дослідженнями.

Отримати найкращі результати в аналізі даних для підвищення якості досліджень в Інтего Груп можна за допомогою забезпечення якості даних, перед початком аналізу треба переконайтися, що дані є достовірними і збалансованими. Треба провести перевірку на наявність пропущених значень, викидів та некоректних даних. Треба обрати відповідні методи, використовувати методи аналізу, які найкраще відповідають дослідницьким питанням та типу даних. При цьому треба обрати методи з урахуванням обсягу даних та їхньої структури. Врахування контексту, при виборі методів аналізу даних враховується специфіка клінічного дослідження та його мета. Ретельно розглядається, яка інформація повинна бути отримана та які висновки плануються зробити.

Застосування валідації, треба перевіряти стабільність та достовірність результатів аналізу за допомогою валідації. Зберігання прозорості, треба докладно документувати всі кроки аналізу та прийняті рішення. Це допоможе іншим дослідникам повторити результати та перевірити їхню достовірність. Також постійно

вдосконалюватися, слідкуйте за новими методами аналізу даних та вдосконалювати свої навички. Через постійне навчання можна досягти ще кращих результатів в аналізі даних та підвищити якість досліджень в Інтего Груп.

ВИСНОВКИ

В дипломній роботі я провів детальне дослідження методів статистичного аналізу, що застосовуються у клінічних дослідженнях. В рамках роботи було виконано результати різних методів, таких як аналіз ANOVA в SAS, аналіз виживання, логістична регресія, з метою оцінки їх ефективності та застосовності у клінічних дослідженнях. Я дослідив різні аспекти вибору методів, враховуючи тип даних, розмір вибірки, припущення про розподіл та цілі дослідження. На основі отриманих результатів були сформульовані рекомендації щодо вибору та застосування методів статистичного аналізу в клінічних дослідженнях з метою покращення якості та достовірності досліджень у медичній науці. Результати цього дослідження підкреслили важливість ретельного вибору та застосування відповідних методів для отримання надійних та об'єктивних результатів. За допомогою правильного статистичного аналізу можна значно підвищити якість клінічних досліджень та робити більш достовірні висновки щодо ефективності лікування та прогнозування результатів.

Однак, важливо пам'ятати, що вибір методів статистичного аналізу повинен бути обґрунтованим і здійснюватися з урахуванням конкретних характеристик дослідження, таких як тип даних, розмір вибірки та припущення про розподіл. Додатково, рекомендується постійно вдосконалювати методи аналізу даних та дотримуватися найкращих практик у сфері клінічних досліджень.

Отже, на основі отриманих результатів можна зробити висновок, що правильний вибір та ефективне застосування методів статистичного аналізу в клінічних дослідженнях є важливим кроком у напрямку покращення медичної науки та практики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів / В.Є. Бахрушин. Запоріжжя : КПУ, 2011, 10 с.
2. Основи статистичного моделювання: навч. посібник / за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП "Рута", 2022, 49 с.
3. Основи статистичного моделювання: навч. посібник / за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП "Рута", 2022, 556 с.
4. Основи статистичного моделювання: навч. посібник / за загальною редакцією С.В. Чугаєвської, Н.В. Ковтун. Житомир: Видавництво ПП "Рута", 2022, 35 с.
5. Мармоза А.Т. Практикум з теорії статистики. К.: Центр учбової літератури, 2013, 51 с.
6. Бізнес-аналітика багатовимірних процесів: навчальний посібник [Електронний ресурс] / Т. С. Клебанова, Л. С. Гур'янова, Л. О. Чаговець та ін. Харків : ХНЕУ ім. С. Кузнеця, 2018. 272 с.
7. Єріна А. М. Статистичне моделювання та прогнозування [Текст]: підручник / А. М. Єріна, Д. Л. Єрін; Держ. ВНЗ "Київ. нац. екон. ун-т ім. Вадима Гетьмана". Київ: КНЕУ, 2014. 348с. 3. Ковтун Н.В. Теорія статистики: підручник / Н. В. Ковтун. К.: Знання, 2012. 399 с.
8. Ковтун Н.В. Теорія статистики: підручник / Н. В. Ковтун. К.: Знання, 2012. 399 с.
9. Zaino J. Adventures in Social Research: Data Analysis Using IBM SPSS Statistics (10th Edition). SAGE Publications, 2018, p. 512.
10. Terrell S.R. Statistics Translated. A Step-by-Step Guide to Analysing and Interpreting Data. Guilford Press, 2021, p. 433.
11. Костюк В. О. Прикладна статистика: навч. Посібник. Харк. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. Харків : ХНУМГ ім. О. М. Бекетова, 2015. 191 с. 10.

12. Матковський С.О., Гальків Л.І., Гринькевич О.З., Сорочак О.З. Статистика: Навч. Посібник Львів: «Новий Світ 2000», 2009, 430 с.
13. Smith Gary. Essential Statistics, Regression, and Econometrics, Second Edition / Gary Smith, 2015, p. 396.
14. Яровий А.Т., Страхов Є.М. Багатовимірний статистичний аналіз: навчально-методичний посібник для студентів математичних та економічних фахів. Одеса: Астропринт, 2015, 132 с.
15. Офіційний веб-сайт компанії Intego Group [Електронний ресурс]. – Режим доступу: <https://integogroup.com/> (дата звернення 15.04.2024)
16. Яровий А. Т. , Страхов Є. М. Багатовимірний статистичний аналіз : навчально-методичний посібник для студентів математичних та економічних фахів. Одеса: Астропринт, 2015. 132 с.
17. Методологія наукових досліджень та приклади її використання [Електронний ресурс] : навчальний посібник / Самсонов В. В., Сільвестров А. М., Тачиніна О. М. ; Національний університет харчових технологій. Київ : Національний університет харчових технологій, 2022. 385 с.
18. Методологія та організація наукових досліджень: навчальний посібник / укладачі: Н.В. Рашкевич, Ю.А. Отрош. Харків, 2022. 291с.
19. Методологія наукових досліджень у статистиці: навч. посіб. / [С. О. Матковський, М. Л. Вдовин, О. С. Гринькевич, Л. М. Зомчак, Т. Я. Лагоцький, Т. В. Панчишин]. Львів: Видавничий центр ЛНУ імені Івана Франка, 2014, 378 с.
20. Pinder, Jonathan P. (2017), Introduction to Business Analytics using Simulation / Jonathan P. Pinder. – P. 434.
21. Siegel, Andrew F. (2017), Practical business statistics / Andrew F. Siegel. – 8th ed., – P. 641.