

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Економічний факультет

Кафедра статистики, інформаційно-аналітичних систем і демографії

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

**ПОРІВНЯЛЬНИЙ СТАТИСТИЧНИЙ АНАЛІЗ МЕТОДІВ
ОЦІНЮВАННЯ ФАКТОРІВ ВИЖИВАНOSTІ ХВОРИХ НА СЕРЦЕВО-
СУДИННІ ЗАХВОРЮВАННЯ**

Студентки II курсу
спеціальності 051 Економіка
за освітньо-науковою програмою
«Економічна аналітика та
статистика»
денної форми навчання
Рудзей Вікторії Олександрівни

(підпис)

Науковий керівник:
к.е.н. доцент
Поплюйко Ярослава Валеріївна

(підпис)

Роботу допущено до захисту на засіданні ЕК рішенням кафедри статистики та демографії, протокол № 10 від «04» травня 2023 року

Завідувач кафедри

д-р екон. наук, професор
Наталія КОВТУН

Київ – 2023

РЕФЕРАТ

Кваліфікаційна магістерська робота містить 51 ст., 16 рис., 15 табл., 25 джерел.

Ключові слова: серцево-судинні захворювання, метод головних компонент, аналіз виживаності, метод Каплана-Мейера, регресія Кокса.

Предмет дослідження: теоретико-методологічні засади та особливості практичного застосування різних методів оцінювання факторів виживаності хворих на серцево-судинні захворювання.

Об'єкт дослідження: госпіталізовані пацієнти, які мають серцево-судинні захворювання.

Мета дослідження: пошук, порівняння та аналіз факторів, що не випадково впливають на ймовірність виживання пацієнтів на базі використання методів аналізу виживаності.

Методи дослідження: метод головних компонент, аналіз виживаності, метод Каплана-Мейера та регресія Кокса.

Наукова новизна, практична значимість дослідження: проведено порівняння та аналіз факторів за допомогою використання методу головних компонент, методу Каплана-Мейера та регресії Кокса, виявлені фактори, що впливають на виживаність пацієнтів хворих на серцево-судинні захворювання.

Практична цінність: виявлено фактори, що впливають на виживаність пацієнтів хворих на серцево-судинні захворювання, адаптовано програмний код для реалізації власних даних. Висновки, отримані в результаті оцінювання, можуть бути корисними у медицині для виявлення ризиків, що негативно впливають на шанси пацієнта на повну реабілітацію, профілактику та корекцію захворювань, а також для вибору оптимального методу лікування.

Основні положення роботи були висвітлені на Міжнародній науковій конференції «Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення» (випуск 75), 6-7 березня 2023 р.

RESUME

Taras Shevchenko National University of Kyiv
Faculty of Economics
Department of Statistics, Information and Analytical Systems
and Demography

Keywords: cardiovascular diseases, method of principal components, survival analysis, Kaplan-Meier method, Cox regression.

Master's qualification work «Comparative statistical analysis of methods of assessing survival factors of cardiovascular diseases patients» consists of 51 pages, 16 figures, 15 tables, 25 sources.

The subject of the study is theoretical and methodological principles and features of the practical application of various methods of assessing the survival factors of patients with cardiovascular diseases. The object of the study is hospitalized patients with cardiovascular diseases.

The purpose of the research is theoretical and practical study of factors affecting survival probability of patients with cardiovascular diseases. The author has compared and analysed factors using the method of principal components, the Kaplan-Meier method and Cox regression, identified factors that affect the survival of patients with cardiovascular diseases.

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ І. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ СТАТИСТИЧНОГО ДОСЛІДЖЕННЯ ВИЖИВАНОСТІ ХВОРИХ НА СЕРЦЕВО-СУДИННІ ЗАХВОРЮВАННЯ	6
1.1. Поширення та основні фактори ризику серцево-судинних захворювань	6
1.2. Статистичний аналіз виживаності	10
1.3. Інформаційне забезпечення статистичного дослідження хворих на серцево-судинні захворювання.....	15
РОЗДІЛ 2. ФАКТОРНИЙ АНАЛІЗ ВИЖИВАНОСТІ ПАЦІЄНТІВ ХВОРИХ НА СЕРЦЕВО-СУДИННІ ЗАХВОРЮВАННЯ.....	18
2.1. Метод головних компонент.....	18
2.2. Формування груп значущих факторів виживаності хворих на серцево судинні захворювання	22
РОЗДІЛ 3. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ВИЖИВАНОСТІ.....	31
3.1. Порівняльний статистичний аналіз виживаності пацієнтів, хворих на серцево-судинні захворювання за стратами.....	31
3.2. Статистичний аналіз факторів, що впливають на виживаність пацієнтів з урахуванням наявності додаткової операції.....	37
3.3. Побудова моделі та порівняльна характеристика впливу факторів виживаності	45
ВИСНОВКИ	50
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	52
ДОДАТКИ.....	55

ВСТУП

На сьогодні серцево-судинні захворювання є основною причиною смерті в усьому світі. За даними Всесвітньої організації охорони здоров'я [1], приблизно 17,9 мільйонів людей щороку помирають від серцево-судинних захворювань, що становить 31% усіх смертей у світі. Попередні оцінки показують, що до 2030 року кількість смертей від серцево-судинних захворювання може збільшитися на 23 мільйони осіб щороку.

Серцево-судинні захворювання також є основним економічним тягарем для систем охорони здоров'я. Прямі витрати на лікування серцево-судинних захворювань, включаючи госпіталізацію, медикаменти та операції, є досить значними.

Перш за все, вивчення серцево-судинних захворювань допомагає зменшити ризик їх виникнення. Знання про фактори ризику, такі як неправильне харчування, недостатня фізична активність, куріння та гіпертонія, дозволяє людям приймати свідомі рішення щодо свого здоров'я. Попередження серцево-судинних захворювань шляхом усунення або зниження цих ризикових факторів може допомогти зберегти життя і покращити якість життя.

Крім того, вивчення серцево-судинних захворювань має велике значення для громадського здоров'я. Попередження та контроль захворювань серця та судин не тільки зберігають життя і покращують якість життя окремих осіб, але і сприяють економічному розвитку країни. Великі витрати на лікування серцево-судинних захворювань можуть стати тягарем для системи охорони здоров'я, але завдяки вивченню цих хвороб можна розробити ефективні стратегії профілактики та лікування, що сприятимуть зменшенню витрат і покращенню економічного становища суспільства.

Нарешті, вивчення серцево-судинних захворювань підкреслює важливість збереження здорового способу життя. Правильне харчування, активний спосіб життя, відмова від шкідливих звичок - все це впливає на здоров'я серця та судин.

Освіченість населення щодо цих питань допомагає підтримувати громадський інтерес до здорового способу життя та відповідальність за власне здоров'я.

Для зменшення рівня серцево-судинних захворювань можна застосовувати два основних підходи: профілактику та діагностику. Профілактика включає різні заходи, призначені для попередження розвитку захворювань у здорових людей. Надання допомоги людям, які мають ризик розвитку серцево-судинних захворювань, а також регулярний моніторинг стану здоров'я, дозволяють виявляти патологію на ранніх стадіях і раніше почати лікування. Також для досягнення позитивних результатів важливо забезпечити доступ до професійної медичної допомоги та заходів по підвищенню рівня освіти та свідомості населення щодо медичних питань.

Статистичні дослідження допоможуть визначити потенційні ризики для розвитку серцево-судинних захворювань та виявити чинники, що впливають на їх розвиток. На основі цих досліджень можна розробляти програми профілактики і діагностики, що забезпечать захист здоров'я населення. Також вони дозволять оцінити ефективність вже існуючих програм профілактики та діагностики.

Метою кваліфікаційної роботи є пошук, порівняння та аналіз факторів, що не випадково впливають на ймовірність виживання у пацієнтів на базі використання методів аналізу виживаності.

Відповідно до поставленої мети визначено *основні завдання* дослідження, спрямовані на її досягнення:

- дослідити інтенсивність поширення серцево-судинних захворювань у світі та фактори ризику, які спричиняють зростання кількості хворих;
- дати статистичну характеристику інформаційній базі дослідження;
- дослідити особливості проведення аналізу виживаності, зокрема метод Каплана-Мейера та регресію Кокса;
- визначити умови та особливості застосування методу головних компонент;
- визначити групи факторів (головні компоненти), що впливають на виживаність пацієнтів, які мають серцево-судинні захворювання;

- визначити фактори, що не випадково впливають на виживаність пацієнтів, хворих на серцево-судинні захворювання;

- побудувати модель впливу факторів на виживаність пацієнтів, хворих на серцево-судинні захворювання.

Аналізу виживаності присвячено безліч праць різних вітчизняних та закордонних вчених. Серед українських вчених необхідно виділити: Ковтун. Н. В., Пальян З.О., Сакало А.В., Чашурін Ф.Н., Щербіна О.В., Яковлев П.Г. Фундаментальну розробку аналізу виживаності здійснили видатні зарубіжні вчені, як Aalen O., Bland J.M., Cox. D., Prentice R.L. тощо.

Об'єктом дослідження є госпіталізовані пацієнти, які мають серцево-судинні захворювання.

Предметом дослідження є теоретико-методологічні засади та особливості практичного застосування різних методів оцінювання факторів виживаності хворих на серцево-судинні захворювання.

Інформаційну базу кваліфікаційної магістерської роботи склали 17287 пацієнтів, які мають серцево-судинні захворювання, були госпіталізовані та були під наглядом у США [2]. Розрахунки проведені в програмі SPSS Statistics 26 та в програмному середовищі SAS.

Апробація результатів дослідження. Основні положення роботи були висвітлені на Міжнародній науковій конференції «Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення» (випуск 75), 6-7 березня 2023 р. на тему «Метод головних компонент: оцінювання ризиків для пацієнтів хворих на серцево-судинні захворювання».

Обсяг та структура роботи. Випускна кваліфікаційна робота магістра складається зі вступу, трьох розділів, висновків, списку використаних джерел та додатків. Перший розділ присвячений огляду проблем та інформаційної бази дослідження. В другому розділі проводиться факторний аналіз для пацієнтів хворих на серцево-судинні захворювання з використанням методу головних компонент. В третьому розділі здійснюється аналіз факторів що впливають на виживаність. Висновки містять основні результати дослідження.

РОЗДІЛ І. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ СТАТИСТИЧНОГО ДОСЛІДЖЕННЯ ВИЖИВАНOSTІ ХВОРИХ НА СЕРЦЕВО-СУДИННІ ЗАХВОРЮВАННЯ

1.1. Поширення та основні фактори ризику серцево-судинних захворювань

Серцево-судинні захворювання є головною причиною смерті у світі. Захворювання включають в себе гострі інфаркти міокарда, хронічні ішемічні захворювання серця, гіпертонію та артеріальну гіпертензію. Вони мають значний вплив на основні демографічні показники населення: смертність, середня тривалість життя, інвалідність, кількість хворих. За останні десятиліття захворювання на серцево-судинні захворювання зазнавали значного зростання, і ця тенденція продовжується. Це поширення зумовлене комбінацією факторів, включаючи зміну способу життя, збільшення вживання нездорової їжі, низьку фізичну активність, куріння, споживання алкоголю, погану якість харчування та зростання захворювань, пов'язаних зі стресом.

Для прогнозування та контролю кількості серцево-судинних захворювань, їх ризиків та факторів, що на них впливають, систематично проводиться статистичний моніторинг та аналіз інформації. Це дозволяє досягти найкращих результатів у медичній діагностиці та лікуванні. Аналіз поширення серцево-судинних захворювань та виживаності пацієнтів допомагає підтвердити факти та прийняти правильні рішення при створенні програм запобігання та контролю захворювань. На основі здобутих даних визначаються ризикові фактори та їх вплив на розвиток захворювань.

Аналіз поширення серцево-судинних захворювань показує, що існує географічна нерівномірність поширення серцево-судинних захворювань. Вони є найпоширеніші в країнах з високим рівнем розвитку, де спостерігаються великі

міські центри, злагоджений спосіб життя та змінений дієтичний режим. Однак, в останні роки спостерігається збільшення поширення серцево-судинних захворювань в країнах з низьким та середнім рівнем розвитку, де зростає кількість несприятливих факторів ризику та змінюється спосіб життя. Зростання на серцево-судинні захворювання в країнах з низьким і середнім рівнем доходу є особливо тривожним, оскільки ці країни мають обмежені ресурси для боротьби з цією проблемою. Крім того, високий показник серцево-судинних захворювань може бути пов'язано з бідністю та низьким рівнем освіти, що може погіршити ситуацію в цих країнах. Наприклад, в середньому приблизно 1600 людей із 100 000 помирають від серцево-судинних захворювань у США та в середньому більше 8000 тисяч людей в Україні (рис. 1.1). Можна спостерігати, що з кожним роком динаміка смертей збільшується.

CENTRAL ILLUSTRATION: Global Burden of Cardiovascular Diseases and Risks

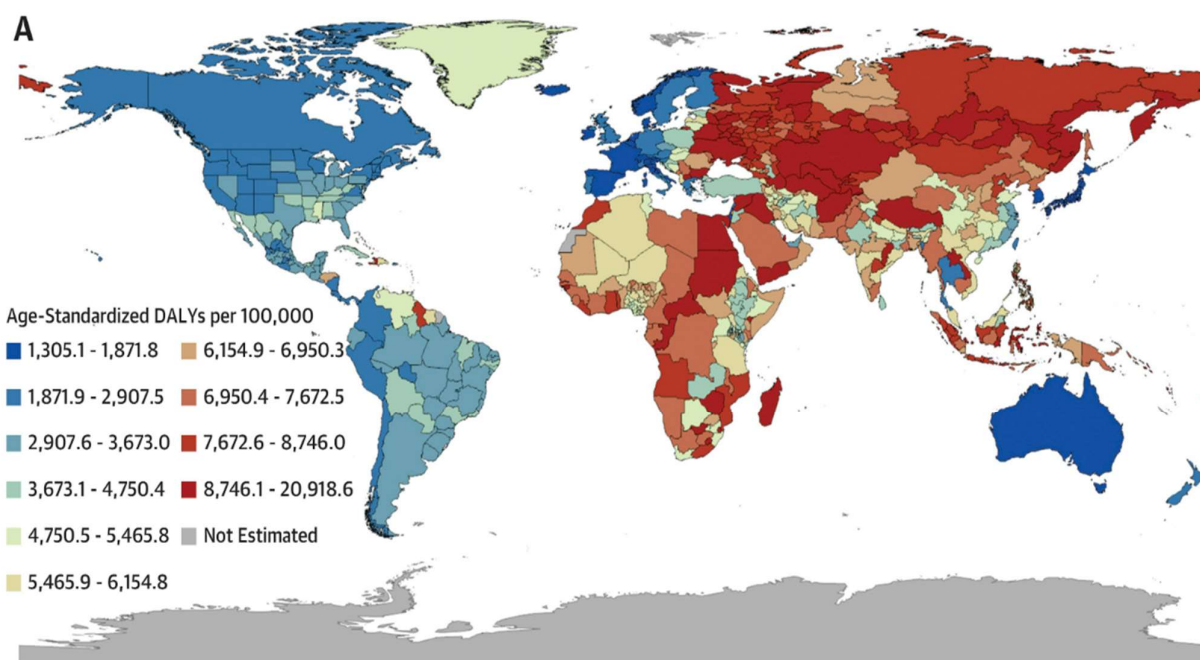


Рисунок 1.1 - Ступінь поширення серцево-судинних захворювань у світі у 2022 році

Джерело: Journal of the American College of Cardiology [3]

Серцево-судинні захворювання включають в себе різні захворювання серця (рис.1.2). Найбільший відсоток становить ішемічна хвороба серця та ішемічний інсульт.

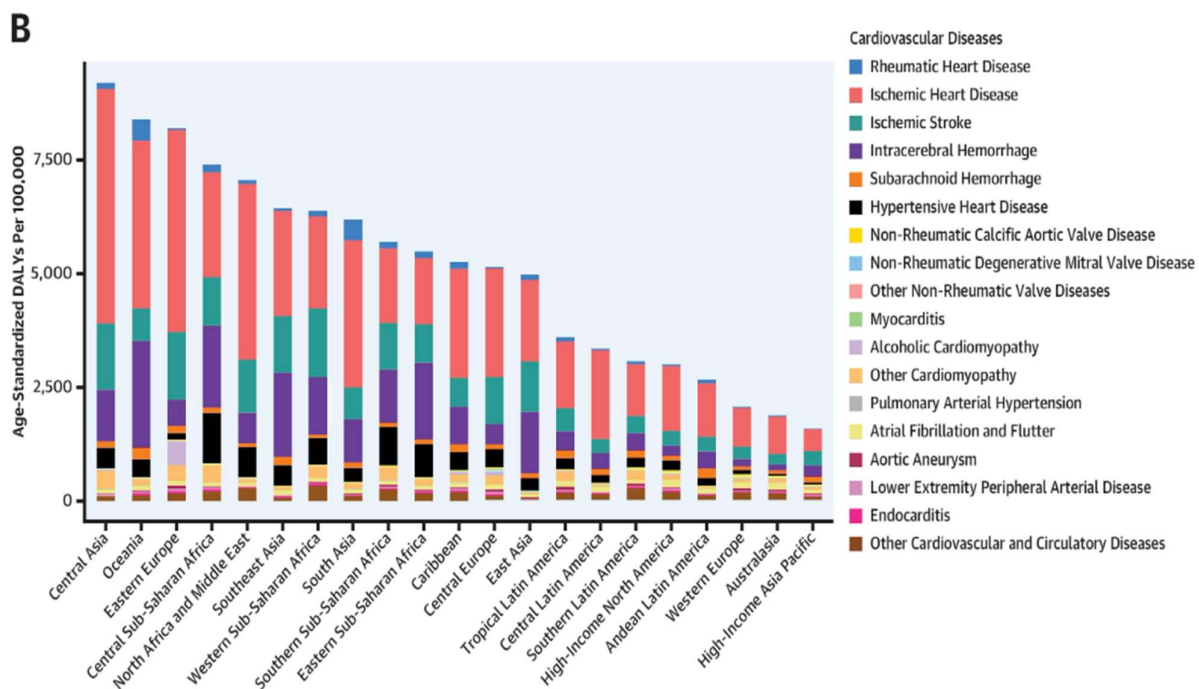


Рисунок 1.2 - Стандартизовані за віком роки життя з поправкою на інвалідність (DALY) на 100 000 у 2022 році для серцево-судинних захворювань за країнами порівняно з теоретичним мінімальним рівнем ризику

Джерело: Journal of the American College of Cardiology [3]

Захворювання серця і судин найчастіше розвиваються при декількох негативних факторах ризику. Основними з них є негативний стиль життя, вік, генетичні ризикові фактори, повнота, артеріальне тиск, стан системи кровообігу, холестерин, діабет, передавання та прийом алкоголю та тютюну. Розглянемо більш детально основні фактори ризику серцево-судинних захворювань (рис.1.3).

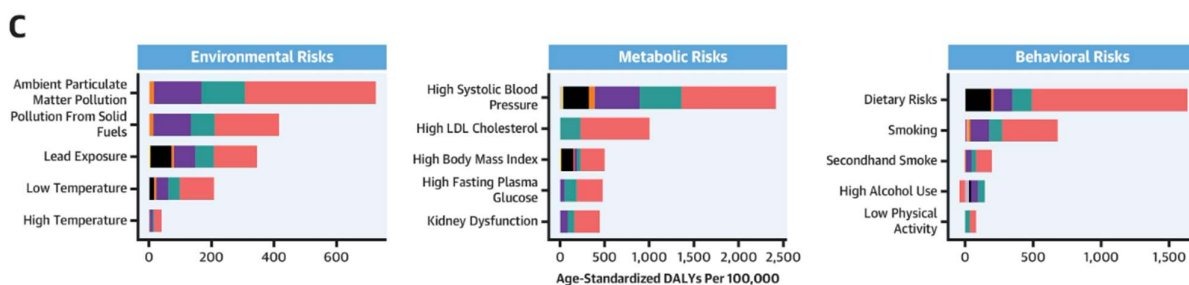


Рисунок 1.3 - Стандартизовані за віком роки життя з поправкою на інвалідність (DALY) на 100 000 у 2022 році для тягара, пов'язаного з вибраними факторами ризику за країнами порівняно з теоретичним мінімальним рівнем ризику

Джерело: Journal of the American College of Cardiology [3]

Відповідно до представленої градації ризиків (рис. 1.3), перша супутня хвороба, на яку варто звернути увагу, - цукровий діабет. Ця хвороба характеризується підвищеним рівнем цукру в крові, що може спричинити ушкодження стінок артерій та сприяти розвитку атеросклерозу. Відновлення нормального рівня цукру в крові і підтримання стійкого контролю цукрового діабету є важливими для зниження ризику серцево-судинних захворювань у пацієнтів.

Гіпертонія, або підвищений артеріальний тиск, є ще однією супутньою хворобою, яка має прямий вплив на серцево-судинну систему. Високий тиск навантажує серце, збільшуючи його робоче навантаження та ризик розвитку серцевих проблем, таких як серцева недостатність, інфаркт міокарда або інсульт. Контроль артеріального тиску за допомогою ліків, змін у способі життя і правильне харчування є важливими факторами у попередженні серцево-судинних ускладнень.

Холестерин - ще один важливий фактор, який впливає на серцево-судинні захворювання. Високий рівень "поганого" холестерину (LDL) та низький рівень "доброго" холестерину (HDL) можуть сприяти утворенню атеросклеротичних бляшок на стінках артерій і збільшенню ризику розвитку серцево-судинних

захворювань, таких як коронарна артеріальна хвороба. Контроль рівня холестерину шляхом здорового харчування, фізичної активності та при необхідності призначення лікарських препаратів є важливими кроками для зниження ризику серцевих проблем.

Ожиріння є ще однією супутньою хворобою, яка може впливати на серцево-судинну систему. Надлишковий жир може ставитися у взаємозв'язку з високим артеріальним тиском, дисліпідемією, цукровим діабетом та запаленням, що збільшує ризик розвитку серцевих захворювань. Збалансована дієта, фізична активність та втрата зайвої ваги можуть допомогти управляти ожирінням та знизити негативний вплив на серцеву систему.

Загалом, супутні хвороби мають значний вплив на розвиток серцево-судинних захворювань. Цукровий діабет, гіпертонія, дисліпідемія, ожиріння та стрес - всі вони взаємодіють і сприяють ризику серцевих проблем.

Саме тому аналіз та оцінювання факторів у хворих на серцево-судинні захворювання є необхідним для виявлення пацієнтів, що мають більший ризик смерті у довготривалій перспективі.

1.2. Статистичний аналіз виживаності

Методи оцінювання виживаності є важливим інструментом в медичній статистиці та епідеміології для дослідження тривалості виживання пацієнтів та інших важливих подій в медичних дослідженнях. Ці методи дозволяють встановити зв'язок між факторами ризику, лікуванням та виживаністю пацієнтів, а також дозволяють оцінити ефективність нових терапевтичних підходів.

Основна мета методів оцінювання виживаності полягає у вивченні тривалості часу до настання певної події, такої як смерть, рецидив захворювання або одужання. Ці методи дозволяють оцінити ймовірність події в залежності від часу та ідентифікувати фактори, що впливають на виживаність.

Основним поняттям в методах оцінювання виживаності є крива виживаності, яка відображає ймовірність виживання пацієнтів протягом певного періоду часу. Ця крива дозволяє спостерігати зміни виживаності протягом часу та порівнювати різні групи пацієнтів або лікувальні підходи.

Важливий аспект методів оцінювання виживаності полягає у врахуванні цензурування даних. Цензурування виникає, коли час до події не може бути повністю спостереженим або вимірним для всіх пацієнтів у дослідженні. Наприклад, деякі пацієнти можуть бути втрачені під час спостереження, або вони можуть мати неповну інформацію про час до події. У таких випадках використовується цензурована інформація, яка враховується при аналізі виживаності [7].

Одним з найпоширеніших та широко використовуваних методів оцінювання виживаності є метод Каплана-Мейєра, який дозволяє будувати криву виживаності на основі цензурованих даних.

Метод Каплана-Мейєра отримав свою назву від статистиків Едварда Каплана та Пола Мейєра, які вперше описали його в 1958 році. Вони розробили цей метод для аналізу виживаності та часу до подій в медичних дослідженнях. Раніше існували інші методи аналізу виживаності, але вони були обмежені, оскільки не брали до уваги цензуровані дані, коли ми не маємо повної інформації про час до події для всіх об'єктів. Метод Каплана-Мейєра був інноваційним, оскільки він враховував цензуровані дані та дозволяв оцінювати функцію виживаності на основі повного набору даних.

Ключовою ідеєю методу Каплана-Мейєра є використання функції виживаності для моделювання ймовірності виживання на різних проміжках часу. Цей метод дозволяє будувати криву виживаності та порівнювати виживаність між різними групами пацієнтів або лікувальними підходами.

Завдяки своїй простоті та ефективності метод Каплана-Мейєра став широко використовуваним у медичних дослідженнях. Він дозволяє аналізувати тривалість виживання та час до подій з урахуванням цензурованих даних, що

робить його корисним інструментом для вивчення різних хвороб, лікувальних підходів та прогнозування результатів.

Метод Каплана-Мейєра використовується для аналізу часу до події, такого як смерть або відновлення, у групі пацієнтів або об'єктів спостереження.

Метод Каплана-Мейєра припускає, що час до події є випадковою величиною, а цензуровані дані враховуються при обчисленні ймовірностей виживання на кожен момент часу. У випадку, коли всі дані є спостереженими, метод Каплана-Мейєра дає точну криву виживаності.

За допомогою методу Каплана-Мейєра можна побудувати криву виживаності, яка показує ймовірність виживання на різних проміжках часу. Ця крива може бути порівняна для різних груп пацієнтів або лікувальних підходів, що дозволяє визначити різницю виживаності між ними [9].

Ймовірність виживання розраховується за формулою:

$$S_t = S_{t-1} * \frac{N_t - E_t}{N_t} \quad (1.1)$$

де S_t – ймовірність виживання в момент t ; S_{t-1} – ймовірність виживання в момент $t-1$; N_t – кількість груп ризику на момент часу t ; E_t – кількість подій (смертей) на момент часу t ; C_t – кількість цензурованих подій на момент часу t .

Також, широкозастосованими є Log rank та Wilcoxon тести.

Log rank та Wilcoxon тести є двома різними методами порівняння кривих виживання у групах. Основна відмінність між ними полягає в тому, як вони враховують цензуровані дані.

Log rank тест є найбільш поширеним і стандартним методом порівняння кривих виживання. Він базується на порівнянні спостережених подій (смертей) у двох або більше групах за певний проміжок часу. Цей тест враховує цензуровані дані (тобто випадки, коли спостереження не завершилися на момент аналізу) і визначає, чи є статистично значущі різниці в ризику виживання між групами.

Wilcoxon тест (або тест Вілкоксона-Бреслоу) також використовується для порівняння кривих виживання, але він зосереджений на рангових порівняннях між групами. Він враховує не тільки час виживання, але й інтенсивність подій (тобто ранги), що стаються протягом цього часу. Wilcoxon тест також може бути використаний для цензурованих даних і дозволяє виявити статистично значущі різниці між групами [10].

Таким чином, метод Каплана-Мейєра є потужним інструментом для аналізу виживаності та часу до подій у медичних дослідженнях. Він враховує цензуровані дані та дозволяє будувати криві виживаності, порівнювати групи пацієнтів та оцінювати ефективність лікування. Цей метод є широко використовуваним інструментом, який допомагає лікарям та дослідникам зрозуміти тривалість виживання та прогнозувати результати в медичних дослідженнях. Використання методу Каплана-Мейєра дозволяє отримати більш об'єктивну оцінку виживаності та встановити статистичну значущість між групами пацієнтів або лікувальними підходами.

Моделі пропорційних інтенсивностей є фундаментальними для аналізу виживання та дослідження впливу факторів на ризик події протягом часу. Ця модель, також відома як модель Кокса, ґрунтується на кількох ключових припущеннях, що дозволяють робити статистично обґрунтовані висновки [11].

Модель Кокса (Cox proportional hazards model) - це статистична модель, яка дозволяє вивчати вплив різних факторів на виживаність, контролюючи інші змінні. Вона базується на припущенні пропорційності ризиків, що означає, що вплив факторів залишається сталим з часом. За допомогою моделі Кокса можна отримати оцінки коефіцієнтів впливу факторів, їх статистичну значущість і інтерпретацію результатів [12].

Модель Кокса може розглядатися в деякому сенсі як непараметрична та може бути записана в наступному вигляді:

$$h_{(t)} = h_0(t) \times \exp(b_1x_1 + b_2x_2 + \dots + b_kx_k) \quad (1.2)$$

де t – час виживання; $h(t)$ – функція ризику визначена набором k факторів (x_1, x_2, \dots, x_k); (b_1, b_2, \dots, b_k) – коефіцієнти детермінації факторів (x_1, x_2, \dots, x_k) відповідно; h_0 – базовий ризик [16].

Головне припущення моделі пропорційних інтенсивностей полягає в пропорційності ефектів факторів на ризик події протягом часу. Це означає, що вплив факторів залишається постійним і не змінюється з часом. Пропорційність виражається через константні співвідношення ризиків (hazard ratios) між порівнюваними групами.

Регресія Кокса є статистичним методом, який дозволяє виміряти відносний ризик виживаності у контексті аналізу виживання. Використовуючи модель пропорційних інтенсивностей Кокса, можна оцінити вплив різних факторів на ризик події та виразити його у вигляді співвідношень ризиків (hazard ratios).

Регресія Кокса базується на припущенні пропорційності ефектів факторів, що означає, що вплив факторів на ризик події залишається сталим протягом часу. За допомогою моделі Кокса, можна оцінити коефіцієнти регресії, які відображають величину та напрямок впливу кожного фактора на ризик події. Результати регресії Кокса можуть бути інтерпретовані як відносний ризик виживаності. Коефіцієнти регресії вказують, наскільки зміна в одиниці фактора впливає на відносний ризик події порівняно з базовим рівнем. Наприклад, якщо коефіцієнт регресії дорівнює 0.5, це означає, що при збільшенні значення фактора на одиницю ризик події зменшується вдвічі порівняно з базовим рівнем.

Модель також припускає лінійність логарифму інтенсивності ризику відносно факторів. Це дозволяє використовувати лінійну регресійну модель для оцінки впливу факторів на ризик події. Такий підхід спрощує аналіз та інтерпретацію результатів [17].

Крім того, модель Кокса враховує цензуровані спостереження, що відбуваються в дослідженні, коли не всі події відбулися до кінця спостереження. Це дозволяє використовувати модель для аналізу неповних даних, які часто зустрічаються в дослідженнях виживання.

Модель пропорційних інтенсивностей Кокса використовує метод часткового правдоподібності для оцінки параметрів. Цей підхід дозволяє оцінювати параметри моделі, не потребуючи повної інформації про розподіл базової інтенсивності. Замість цього, модель фокусується на оцінці співвідношень ризиків між групами, що дозволяє досліджувати вплив факторів на виживання без необхідності визначення точного розподілу базової інтенсивності [18].

Застосування моделі пропорційних інтенсивностей вимагає деяких припущень та обмежень, які варто враховувати. Наприклад, модель передбачає пропорційність ефектів факторів, але якщо це припущення порушується, існують альтернативні моделі, які можуть бути використані.

1.3. Інформаційне забезпечення статистичного дослідження хворих на серцево-судинні захворювання

Інформаційне забезпечення статистичного дослідження хворих на серцево-судинні захворювання полягає в зборі, вивченні, обробці та аналізі даних. Інформація збирається з різних джерел, таких як обстеження хворих, проведення медичних досліджень та показників медичної статистики. Вони можуть включати такі дані, як персональна історія захворювання, показники здоров'я, медикаментозне лікування, дані генетичного тестування, загальний рівень смертності від серцево-судинних захворювань, частота порушень ритму серця, ступінь розвитку серцево-судинних захворювань у певних діапазонах вікових груп та інші. Важливо обрати якісну базу даних, що буде об'єднувати дані з різних джерел та містити потрібну інформацію для проведення статистичних досліджень. Наприклад, можна використовувати централізовані бази даних, які забезпечують доступ до даних з різних регіонів та країн. Крім того, можна використовувати бази даних, які проводять моніторинг захворювань на базі дослідницьких проектів, що посилюють актуальність інформації.

Зібрані дані підлягають обробці та аналізу за допомогою програмних продуктів, таких як Statistica або SPSS, щоб отримати додаткові результати. Результати дослідження можуть бути використані для підтримки прийняття рішень з медичних, економічних та соціальних причин.

Для дослідження було використано базу дослідження, яка включає 17287 пацієнтів, які мають серцево-судинні захворювання та були госпіталізовані. Попередньо, база містила в собі більше записів, з іншим хворобами, але для аналізу було обрано досліджувати лише серцево-судинні захворювання. Інформаційна база містить в собі 85 змінних, 27 з яких фактори впливу на захворювання. Для покращення результатів аналізу дані були попередньо оброблені та усі записи з відсутніми даними були виключені з дослідження.

Розподіл пацієнтів за основними демографічними характеристиками наведений у табл 1.1.

Таблиця 1.1 - Розподіл пацієнтів за демографічними характеристиками

Характеристики	Кількість пацієнтів N=17287	Частка, %
Стать		
Жінки	6957	40,23
Чоловіки	10333	59,77
Вік, років		
17<= x < 40	764	4,42
40<= x < 50	1289	7,46
50<= x < 60	3069	17,75
60<= x < 70	4593	26,57
x>= 70	7572	43,80
Ч-Етнічна приналежність		
Афроамериканці	1901	11,0
Азіати	173	1,0
Європейці	13528	78,26
Іспанці	718	4,15
Корінні Американці	81	0,47
Інше/Невідомо	886	5,13

Джерело: розраховано автором на основі даних [2]

Більшість госпіталізованих пацієнтів це чоловіки – 59,77%, жінки складають 40,23% з усієї вибірки.

Вік хворих коливається в межах від 17 років до 89. Середній вік госпіталізованих становив $65,9 \pm 13,7$ років. Найбільш чисельна вікова група - старше 70 років (43,80%).

В етнічній приналежності переважають пацієнти європейської раси – 78,26%, афроамериканці – 11%, іспанської раси – 4,15%, східних регіонів, менше 1% азіати та корінні американці.

Показник індексу маси тіла є важливим для спостереження, адже люди з категорії передожиріння та ожиріння мають більшу ймовірність розвитку серцево-судинних захворювань, ніж пацієнти з нормальним показником. Розподіл пацієнтів за індексом маси тіла наведений у табл. 1.2.

Таблиця 1.2 - Розподіл пацієнтів за індексом маси тіла

Характеристики	Кількість пацієнтів N=17287	Частка, %
ІМТ, m/h²		
x < 18,5 – Недостатня маса	412	2,24
18,5 ≤ x < =24,9 - Норма	4005	23,17
25 ≤ x < =29,9 - Передожиріння	5597	32,38
x ≥ 30 - Ожиріння	7264	42,02

Джерело: розраховано автором на основі даних [2]

Результати аналізу показали, що найбільше пацієнтів - 7264 (42,02%) відносяться до категорії ожиріння, 32,38% - до категорії передожиріння, та 23,17% пацієнтів мають нормальний показник ІМТ та 2,24% - недостатньої маси.

Отже, за результатами розподілу ознак сукупності можна стверджувати що середній пацієнт це чоловік-європеєць, віком 65 років, основна популяція дослідження це люди віком від 50 років. Також, варто зауважити, що усі пацієнти вже госпіталізовані та знаходяться в тяжкому стані.

РОЗДІЛ 2. ФАКТОРНИЙ АНАЛІЗ ВИЖИВАНOSTІ ПАЦІЄНТІВ ХВОРИХ НА СЕРЦЕВО-СУДИННІ ЗАХВОРЮВАННЯ

2.1. Метод головних компонент

Метод головних компонент (Principal Component Analysis, PCA) є потужним статистичним інструментом для аналізу та визначення структури великого обсягу даних. Він базується на знаходженні головних компонент, які є лінійними комбінаціями початкових змінних і відображають найбільшу дисперсію в даних. Теоретико-методологічні засади методу головних компонент включають у себе розуміння лінійної алгебри, коваріаційної матриці, сингулярного розкладу та властивості головних компонент [4].

К. Пірсон був першим, хто розробив цей метод, після чого він був розвинений, доопрацьований, описаний та обґрунтований в роботах таких вчених, як Г. Хотеллінг, Г. Харман, С. Рао, П. Андрукович, С. А. Айвазян та В.С. Мхитарян.

Однією з основних ідей методу головних компонент є заміна великого набору змінних меншим числом нових змінних, які називаються головними компонентами. Це зменшує розмірність даних і дозволяє зберегти якомога більше інформації. Головні компоненти вибираються таким чином, щоб перша головна компонента виражала найбільшу частку дисперсії в початкових даних, друга - наступну найбільшу частку дисперсії і так далі. Це дозволяє розташувати дані у новому просторі, де осі відповідають головним компонентам, і провести подальший аналіз.

Теоретичні засади методу головних компонент базуються на коваріаційній матриці. Вона вимірює ступінь залежності між різними змінними. Розрахування коваріаційної матриці дозволяє визначити коваріаційні коефіцієнти між усіма парами змінних і визначити ступінь їх взаємозв'язку. Коваріаційна матриця використовується для знаходження сингулярного розкладу, який є

математичною процедурою, яка дозволяє розкласти матрицю на сингулярні значення і вектори. Сингулярний розклад розкриває основну структуру даних і надає можливість знаходження головних компонент.

Одна з важливих властивостей головних компонент - це їх ортогональність. Головні компоненти обираються таким чином, щоб вони були ортогональними між собою, тобто не корелювали між собою. Це дозволяє використовувати їх як незалежні джерела інформації.

Метод головних компонент має широкий спектр застосувань. Він може бути використаний для зменшення розмірності даних, виявлення складових, які найбільше впливають на варіацію в даних, візуалізації інформації, вирішення проблеми колінеарності між змінними та інших аналітичних завдань.

Однак, варто враховувати певні обмеження методу головних компонент. Він передбачає лінійну залежність між змінними і не завжди ефективний для даних з нелінійними залежностями. Також, втрата інформації при зменшенні розмірності даних може бути незначною, але все ж необхідно враховувати цей аспект.

При правильному використанні, метод головних компонент може допомогти виявити ключові фактори, знизити розмірність даних, покращити візуалізацію та спростити аналіз. Він знаходить широке застосування у різних галузях, включаючи фінанси, економіку, соціологію, медицину та багато інших.

Також, важливо розуміти, що метод головних компонент - це статистичний інструмент і не здатний вирішити всі аспекти аналізу даних. Він не заміняє контекстуального розуміння проблеми та експертного досвіду. Крім того, результати PCA можуть бути інтерпретовані по-різному і залежати від обраної кількості головних компонент. Тому важливо виявляти критичне мислення та обговорювати результати з іншими дослідниками або експертами відповідної галузі.

Крім основних теоретичних засад, важливо зазначити кілька ключових практичних аспектів методу головних компонент. Перш за все, перед застосуванням PCA необхідно провести попередній аналіз даних, щоб

переконатися, що вони задовольняють передумови для використання методу, такі як нормальність розподілу змінних, відсутність великих пропусків або викидів даних.

Під час вибору кількості головних компонент необхідно враховувати частку дисперсії, яку вони пояснюють. Зазвичай, вибирають головні компоненти, які включають в себе велику частку дисперсії (наприклад, 80% або 90%), щоб зберегти якомога більше інформації. Однак, важливо зберігати розумний баланс між збереженням інформації та зменшенням розмірності даних.

Також, варто враховувати можливість впливу шуму на результати PCA. Якщо дані містять значну кількість шуму або викидів, вони можуть спотворити результати і вплинути на точність аналізу. Тому, перед застосуванням PCA рекомендується виконати очищення даних або використовувати методи, що відсторонюють викиди.

Одним з важливих аспектів методу головних компонент є його використання для виявлення кореляційних залежностей між змінними. Він дозволяє ідентифікувати складові, які мають високу кореляцію, а також з'ясувати, які змінні найбільше сприяють варіації в даних. Це може бути корисним для виявлення важливих факторів, які впливають на серцево-судинні захворювання, такі як вік, стать, генетичні чинники або ризикові фактори, такі як куріння або неправильне харчування.

Метод головних компонент може бути використаний для визначення складних взаємозв'язків між різними змінними. Наприклад, він може допомогти виявити, як показники крові, такі як рівень холестерину, глюкози та тригліцеридів, взаємодіють між собою та як ці взаємозв'язки можуть впливати на ризик серцевих захворювань. Це може сприяти розумінню патофізіології серцево-судинних захворювань та розробці більш ефективних стратегій профілактики та лікування.

Метод головних компонент може бути корисним інструментом для визначення класифікації або групування пацієнтів за показниками серцево-судинного здоров'я. За допомогою PCA можна виявити схожість або відмінність

між пацієнтами на основі їх профілю змінних. Це може сприяти індивідуалізації лікування та підходу до кожного пацієнта з урахуванням його унікальних характеристик та потреб.

У методі головних компонент також існує можливість використання вагових коефіцієнтів для змінних. Це означає, що можна надати більшу вагу певним змінним, які вважаються більш важливими або інформативними для серцево-судинних захворювань. Наприклад, можна призначити вагові коефіцієнти на основі результатів попередніх досліджень або експертного досвіду, забезпечуючи зосередженість на ключових змінних [8].

Після аналізу факторів, необхідно оцінити їх значення. Фактор є лінійною комбінацією початкових змінних. Значення для f -го фактора можна обчислити за формулою:

$$F_i = b_{i1}X_1 + b_{i2}X_2 + \dots + b_{ik}X_k \quad (2.1)$$

Використання цього методу дає змогу вивчити та пояснити взаємозв'язки між досліджуваними показниками, після чого можна визначити значущі групи факторів для подальшого аналізу.

Метод головних компонент є популярним методом обробки даних, який застосовується для скорочення розмірності вхідних даних з мінімальним втратами корисної інформації. Цей метод зводиться до пошуку додаткової системи прямих або площини у просторі змінних, що максимізують дисперсію вихідних значень. Такі прямі або площини називаються головними компонентами. Головною перевагою методу головних компонент є те, що він єдиний математично обґрунтований метод факторного аналізу.

2.2. Формування груп значущих факторів виживаності хворих на серцево судинні захворювання

Для вибірки пацієнтів, які мають серцево-судинні захворювання було проведено факторний аналіз, а саме використано метод головних компонент задля визначення груп значущих факторів, які надалі будуть використані для аналізу виживаності. Щоб уникнути впливу розмірності та одиниць виміру всі показники були попередньо стандартизовані.

Всі подальші розрахунки було проведено за допомогою відповідної надбудови програми SPSS (Analyze > Dimension Reduction > Factor Analysis).

Важливим кроком у виконанні факторного аналізу є обчислення кореляційної матриці для визначення тісноти зв'язку між обраними 27 змінних за всією сукупністю пацієнтів та в розрізі пацієнтів, які вижили чи померли.

Щоб визначити кількість головних компонент та частку впливу кожного фактору було розраховано таблицю факторизації моделі для усієї вибірки пацієнтів (табл.2.1). Було сформовано лише 10 груп факторів, які пояснюють 71,4% загальної дисперсії.

Таблиця 2.1 - Пояснення загальної дисперсії для усіх пацієнтів

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,913	18,195	18,195	4,913	18,195	18,195
2	3,151	11,670	29,865	3,151	11,670	29,865
3	1,998	7,400	37,265	1,998	7,400	37,265
4	1,721	6,373	43,638	1,721	6,373	43,638
5	1,434	5,309	48,947	1,434	5,309	48,947
6	1,413	5,233	54,181	1,413	5,233	54,181
7	1,346	4,986	59,167	1,346	4,986	59,167
8	1,215	4,499	63,666	1,215	4,499	63,666
9	1,071	3,968	67,634	1,071	3,968	67,634
10	1,030	3,816	71,450	1,030	3,816	71,450

Extraction Method: Principal Component Analysis.

Джерело: розраховано автором на основі даних [2]

Модель обмежує всю кількість досліджуваних змінних до десяти головних компонент, які пояснюють 71,4% загальної дисперсії. За допомогою графіку власних значення можна уточнити результати факторизації моделі і кількості компонентів, які варто розглядати (рис.2.1), в нашому випадку різкий кам'янистий осип починається після 3 компоненти, які загалом пояснюють лише 37,2% загальної дисперсії (табл.2.1).

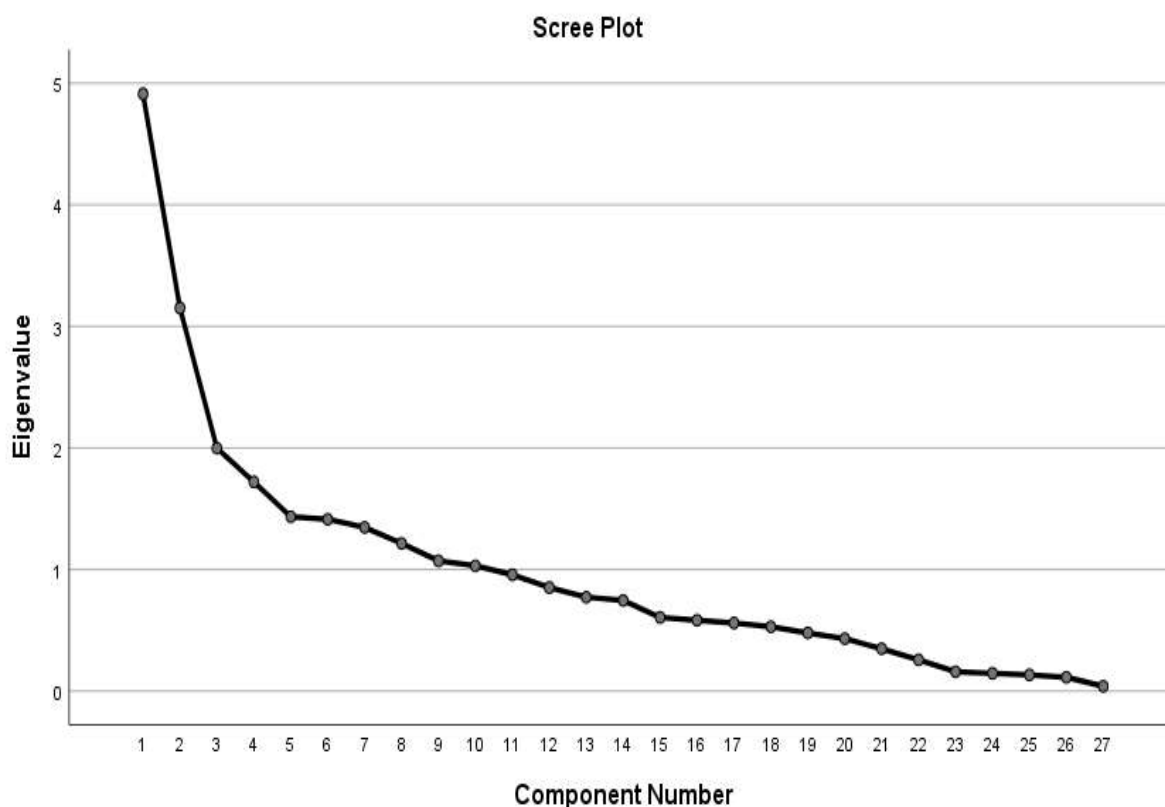


Рисунок 2.1 - Графік власних значень для усіх пацієнтів

Джерело: рохраховано автором на основі даних [2]

Для визначення можливості використання факторного аналізу були використані два критерії: КМО (Кайзера-Мейєра-Олкіна) та критерій сферичності Бартлетта. Значення цих критеріїв підтвердили придатність факторного аналізу для подальшого використання [5]. Результат критерію КМО 0,709, що інтерпретується як прийнятна адекватність. Рівень значущості

критерію сферичності Бартлетта ($p < 0,05$) вказує на те, що дані цілком прийнятні для факторного аналізу (табл.2.2).

Таблиця 2.2 - Критерій КМО та критерій сферичності Бартлетта для усіх пацієнтів

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	,709	
Bartlett's Test of Sphericity	Approx. Chi-Square	223395,893
	df	351
	Sig.	,000

Джерело: розраховано автором на основі даних [2]

Для отримання головних компонент з набору спостережень необхідно проаналізувати перевернуту матрицю компонентів. Метод головних компонент дає змогу отримати матрицю факторних навантажень, де коефіцієнти a_{ij} можуть бути використані для чисельного та формального пояснення коефіцієнтів кореляції. Це дозволяє зробити припущення про наявність фактору, який може бути причиною цих кореляцій. Щоб пояснити відібрані фактори потрібно в кожному рядку повернутої факторної матриці відмітити те факторне навантаження, яке має найбільше абсолютне значення. Факторні навантаження можна розглядати як коефіцієнти кореляції між змінними та факторами [6].

Приклад матриці компонентів представлений нижче (табл 2.3). Повна версія таблиці наведена у додатку А.

Таблиця 2.3 - Матриця компонентів для усіх пацієнтів

Rotated Component Matrix ^a			
	Component		
	1	2	3
h1_mbp_mean	0,905	0,006	0,074
d1_mbp_mean	0,901	-0,022	0,107
h1_diasbp_mean	0,833	0,160	-0,013
h1_sysbp_mean	0,828	-0,151	0,016
d1_diasbp_mean	0,822	0,134	-0,002
d1_sysbp_mean	0,818	-0,193	0,035
map_apache	0,652	0,012	0,070
d1_hearttrate_mean	-0,027	0,929	0,114
h1_hearttrate_mean	0,017	0,902	0,118
heart_rate_apache	-0,018	0,896	0,110
d1_resprate_mean	0,029	0,093	0,875
resprate_apache	0,101	0,062	0,832
h1_resprate_mean	0,092	0,168	0,771

Джерело: рохраховано автором на основі даних [2]

Відповідно до попередніх розрахунків для подальшого аналізу було обрано перші три, найбільш значущих компоненти. До першої компоненти увійшло 7 факторів, які відповідають за життєво-важливі ознаки та пояснюють 18,1% загальної дисперсії. Умовно ми її можемо назвати «основні причинно-наслідкові фактори серцево-судинних захворювань».

Перша компонента містить в собі такі змінні:

- Середній артеріальний тиск пацієнта протягом першої години перебування у відділенні (0,905)
- Середній артеріальний тиск пацієнта протягом перших 24 годин перебування у відділенні (0,901)
- Середній діастолічний артеріальний тиск пацієнта протягом першої години перебування в відділенні (0,833)
- Середній систолічний артеріальний тиск пацієнта протягом першої години перебування у відділенні (0,828)

- Середній діастолічний артеріальний тиск пацієнта протягом перших 24 годин перебування в відділенні (0,822)
- Середній систолічний артеріальний тиск пацієнта протягом перших 24 годин перебування у відділенні (0,818)
- Бал АРАСНЕ III для середнього артеріальний тиск, виміряного протягом перших 24 годин (0,652) .

Другу компоненту відповідно до факторів, які до неї увійшли, можемо назвати «показники пульсу». Вона пояснює 11,6% загальної дисперсії та включає такі змінні:

- Середній показник пульсу пацієнта протягом перших 24 годин перебування у відділенні (0,929)
- Середній показник пульсу пацієнта протягом першої години перебування у відділенні (0,902)
- Бал АРАСНЕ III для середньої частоти пульсу, виміряного протягом перших 24 годин (0,896).

Третя компонента - «показники частоти дихання», які пояснюють 7,4% загальної дисперсії та включає такі змінні:

- Середня частота дихання пацієнта протягом перших 24 годин перебування у відділенні (0,875)
- Бал АРАСНЕ III для середньої частоти дихання, виміряного протягом перших 24 годин (0,832)
- Середня частота дихання пацієнта протягом першої години перебування у відділенні (0,771).

Подальше дослідження було продовжено аналізом групи пацієнтів, які вижили, використовуючи ті ж самі 27 змінних. Факторизація моделі у даній вибірці пацієнтів дала результати, аналогічні до всієї вибірки - 10 головних компонент, які пояснюють 71,1% загальної дисперсії. Результат КМО склав 0.710, що свідчить про прийнятну адекватність. Рівень значущості критерію сферичності Бартлетта становить $p < 0,05$.

За допомогою графіку власних значень (рис.2.2) було прийняте рішення щодо використання, як і в попередньому випадку, трьох компонент, які пояснюють лише 36,6% загальної дисперсії.

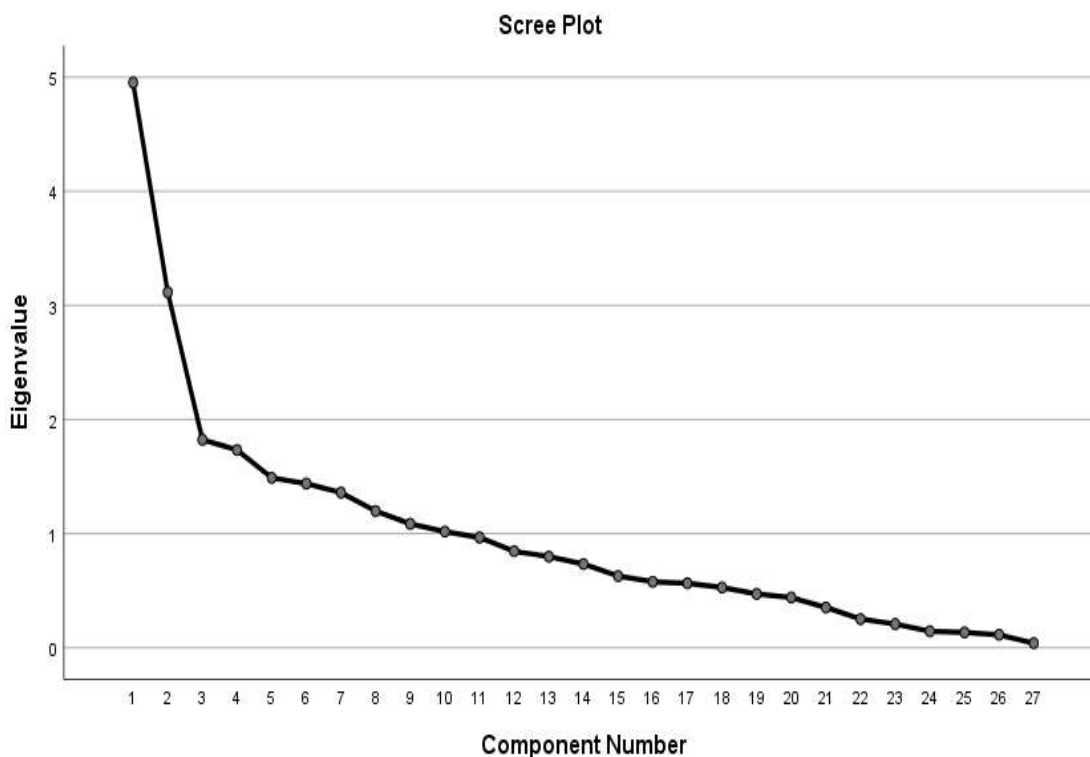


Рисунок 2.2 Графік власних значень для пацієнтів, котрі вижили

Джерело: рохраховано автором на основі даних [2]

Нижче наведена матриця компонентів для пацієнтів котрі вижили (табл 2.4).

Повна версія таблиці наведена у додатку Б.

Аналіз оберненої матриці компонентів для пацієнтів, котрі вижили, показав, що перші 3 компоненти містять в собі аналогічні змінні змінні, що і для всієї вибірки пацієнтів. Перша компонента пояснює 18,3% загальної дисперсії, друга - 11,5%, а третя - 6,7%.

Таблиця 2.4 - Матриця компонентів для пацієнтів, котрі вижили

Rotated Component Matrix ^{a,b}			
	Component		
	1	2	3
h1_mbp_mean	0,906	0,013	0,082
d1_mbp_mean	0,901	-0,012	0,121
h1_diasbp_mean	0,833	0,172	-0,009
h1_sysbp_mean	0,829	-0,149	0,022
d1_diasbp_mean	0,820	0,145	0,002
d1_sysbp_mean	0,820	-0,187	0,043
map_apache	0,658	0,005	0,077
d1_heart_rate_mean	-0,025	0,933	0,107
h1_heart_rate_mean	0,023	0,905	0,114
heart_rate_apache	-0,005	0,900	0,102
d1_resprate_mean	0,037	0,087	0,872
resprate_apache	0,109	0,058	0,831
h1_resprate_mean	0,111	0,161	0,770

Джерело: рохраховано автором на основі даних [2]

Подальше дослідження було продовжено аналогічним аналізом для групи пацієнтів, які померли. Факторизація моделі у цій вибірці пацієнтів показала лише 9 факторів, які пояснюють 67,4% загальної дисперсії. Критерій КМО дав результат 0.689, що свідчить про прийнятну адекватність. Рівень значущості критерію сферичності Бартлетта склав $p < 0,05$. Графік власних значень дозволив встановити, що потрібно враховувати 3 компоненти, які відображають лише 36,5% загальної дисперсії. (див.рис.2.3).

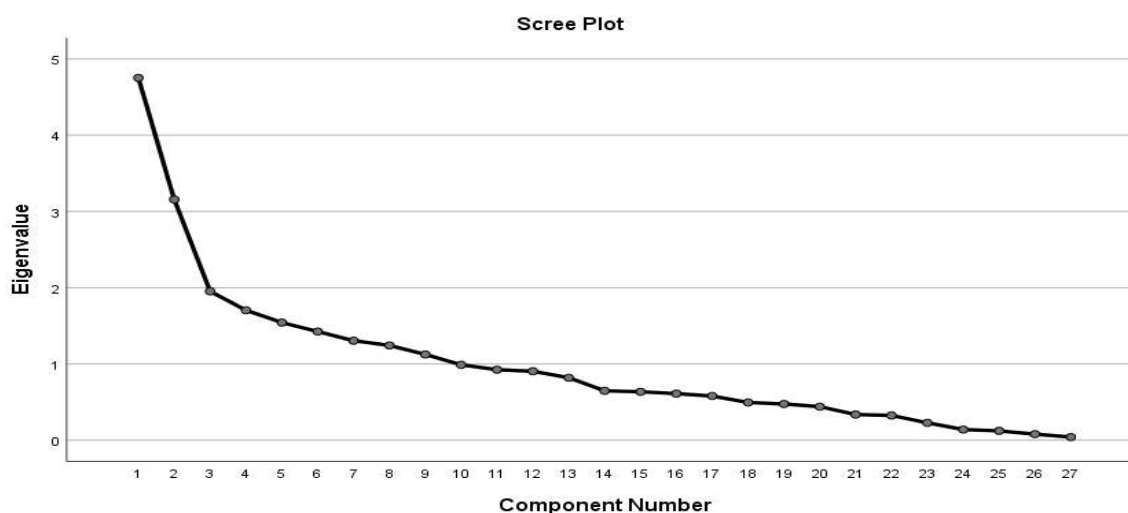


Рисунок 2.3 - Графік власних значень для пацієнтів, котрі померли

Джерело: рохраховано автором на основі даних [2]

Аналіз оберненої матриці компонентів для пацієнтів, які померли, показав ідентичність факторів, що були включені в перші три компоненти, що й для групи усіх пацієнтів та групи пацієнтів, котрі вижили.

Нижче наведена матриця компонентів для пацієнтів котрі померли (табл 2.5). Повна версія таблиці наведена у додатку В.

Таблиця 2.5 - Матриця компонентів для пацієнтів, котрі померли

Rotated Component Matrix ^{a,b}			
	Component		
	1	2	3
d1_mbp_mean	0,877	-0,047	0,035
h1_mbp_mean	0,877	-0,001	0,050
d1_diasbp_mean	0,820	0,071	-0,007
h1_diasbp_mean	0,819	0,097	0,002
h1_sysbp_mean	0,816	-0,118	0,008
d1_sysbp_mean	0,791	-0,162	0,019
map_apache	0,605	0,072	0,000
d1_hearttrate_mean	0,001	0,888	0,122
h1_hearttrate_mean	0,024	0,864	0,093

heart_rate_apache	-0,066	0,842	0,112
d1_resprate_mean	0,009	0,093	0,885
resprate_apache	0,057	0,080	0,819
h1_resprate_mean	0,012	0,125	0,776

Джерело: рохраховано автором на основі даних [2]

Отже, проведене дослідження з використанням методу головних компонент для бази даних пацієнтів, які мають серцево-судинні захворювання, дозволило визначити взаємозв'язки між досліджуваними показниками та обрати 3 групи факторів (головні компоненти), а саме: «основні причинно-наслідкові фактори», «фактори показників пульсу», «фактори частоти дихання». Це дозволило скоротити кількість факторів та визначити найбільш значущі з усіх досліджених. Враховуючи те, що групи значущих факторів містять у собі однаковий набір змінних, можна зробити висновки, що поділ пацієнтів на різні групи в подальшому дослідженні не має сенсу і для подальшого аналізу виживаності варто розглядати всю сукупність пацієнтів загалом.

Визначення значущих факторів допоможе в подальшому аналізі виживаності пацієнтів, і в результаті дозволить більш ефективно проводити лікування серцево-судинних захворювань або навіть запобігти їх появу на ранніх етапах.

РОЗДІЛ 3. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ВИЖИВАНOSTI

3.1. Порівняльний статистичний аналіз виживаності пацієнтів, хворих на серцево-судинні захворювання за втратами

Проведення порівняльного аналізу за допомогою моделі Каплана-Мейера дозволяє оцінити різницю між різними групами. А використання тестів Log rank та Wilcoxon дає можливість проаналізувати та перевірено статистичну значущість відмінностей між кривими виживання.

Хоча обидва тести використовуються для порівняння кривих виживання, вони використовують різні підходи до обробки цензурованих даних та оцінки статистичної значущості.

Проведемо оцінювання імовірності виживаності пацієнтів, хворих на серцево-судинні захворювання, залежно від вікової групи за моделлю Каплана-Мейера (рис. 3.1).

На рис.3.1 зображені криві виживаності за віковими групами, де: 1 – група від 17 до 40 років, 2 – група від 40 до 50 років, 3 – група від 50 до 60 років, 4 – група від 60 до 70 років, 5 – група від 70 років.

Різниця між віковими групами помітна. Можливо зробити висновок, що найменша виживаність у пацієнтів старше 70 років, що очікувано.

Підтвердженням того є результатами Log rank та Wilcoxon тестів p-value складає <0.0001 , що підтверджує суттєвість різниці виживаності у даних групах.

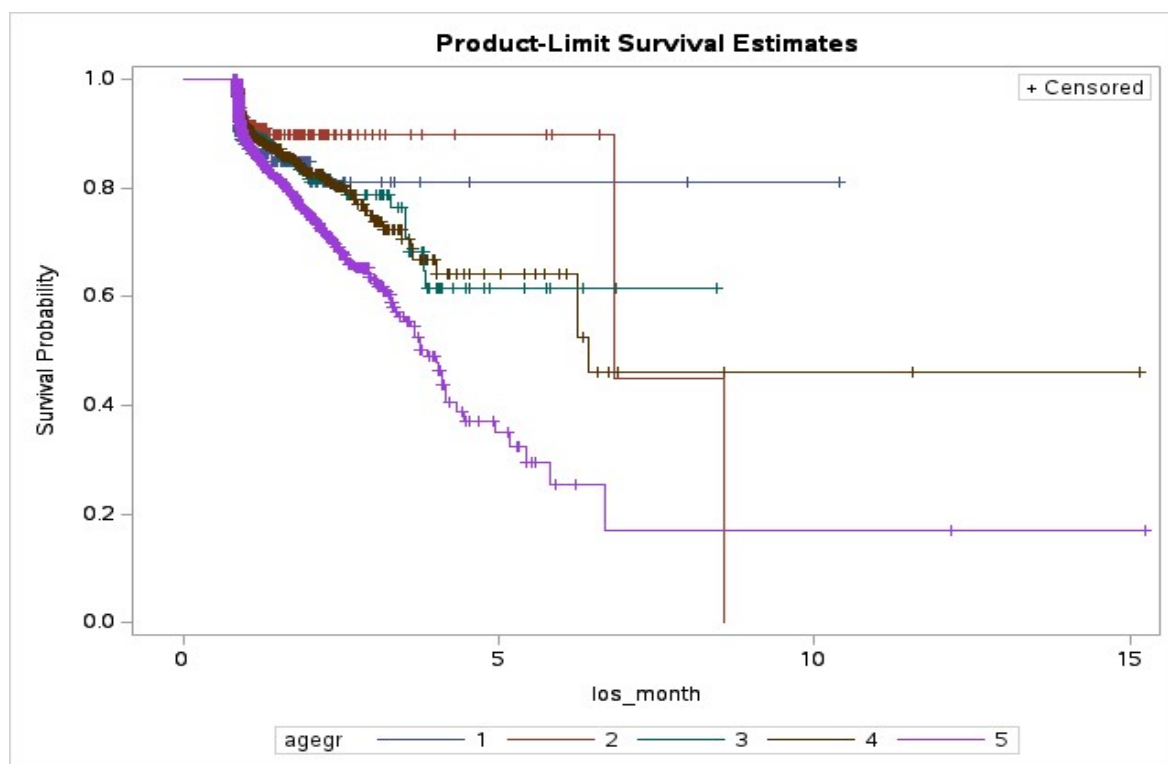


Рисунок 3.1. Оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від вікової групи за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис.3.2 зображені криві виживаності за віковими групами, де: F – група жінок, M – група чоловіків.

Криві Каплана-Мейєра майже збігаються, це означає, що немає статистично значущих різниць у виживаності між різними групами статі у вибірці.

У такому випадку, порівняльний аналіз показує, що ризик виживання та кількість подій (наприклад, смертей) у різних групах практично однакові. Це може мати значення з погляду дослідження, оскільки демонструє, що фактор, який розглядається, не впливає на виживання або що групи є подібними за показником виживання.

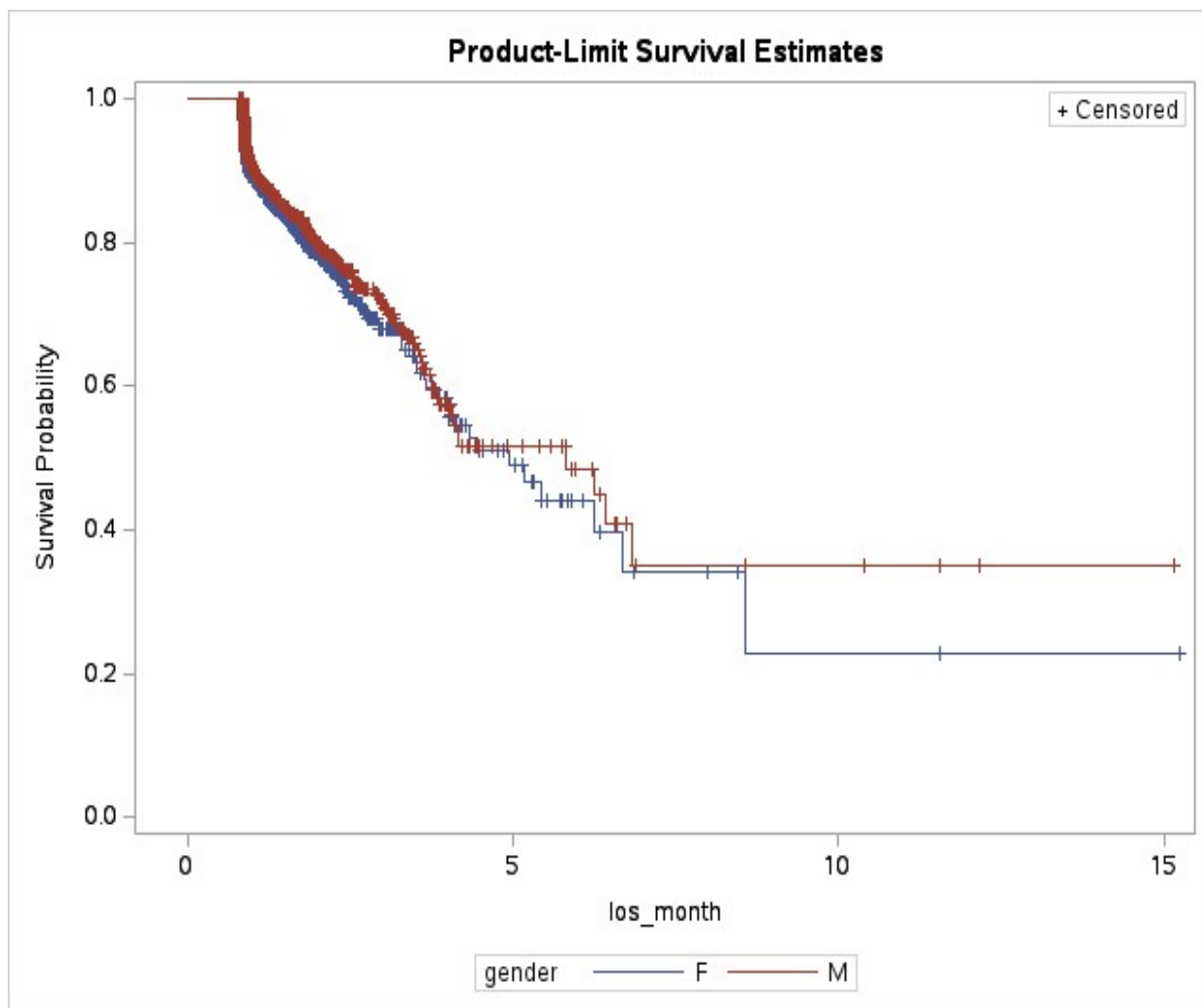


Рисунок 3.2 - Оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від статі за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис.3.3 – 3.4 зображені криві виживаності в залежності від раніше проведеної додаткової операції, як вид лікування, де: 0 – не проведена операція, 1 – проведена операція. Можливо побачити, що різниця у виживаності помітна та пацієнти, які попередньо вже мали додаткову операцію мають вищий відсоток виживаності.

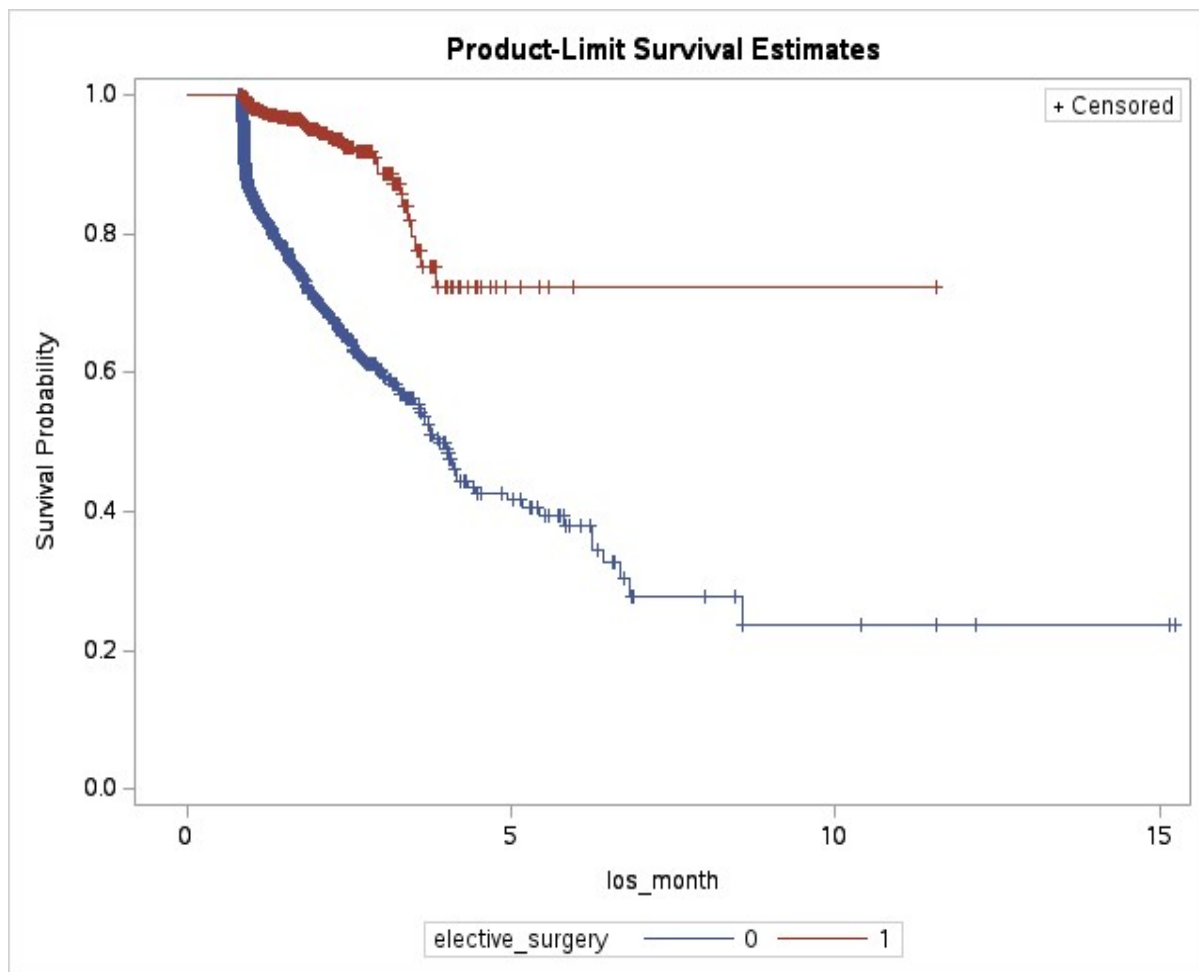


Рисунок 3.3 - Оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від раніше проведеної додаткової операції за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис. 3.4 зображено функція ризику пацієнтів хворих на серцево-судинні захворювання залежно від раніше проведеної додаткової операції за Каплан-Мейером.

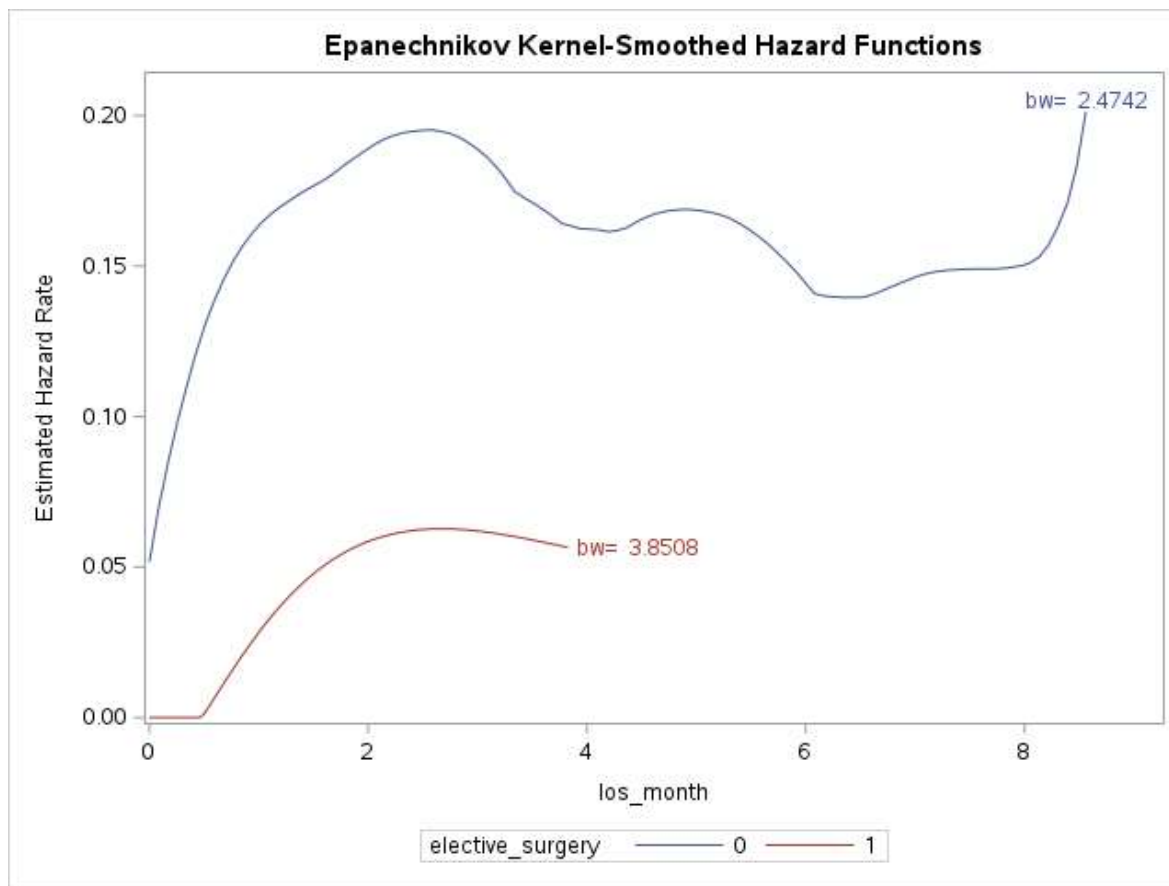


Рисунок 3.4 - Функція ризику пацієнтів хворих на серцево-судинні захворювання залежно від раніше проведеної додаткової операції за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

За результатами Log rank та Wilcoxon тестів виживаності p-value складає <0.0001 , що підтверджує суттєвість різниці виживаності у даних групах та дає підстави для тестування впливу факторів з урахуванням наявності проведення додаткової операції, як тип лікування.

Також, варто перевірити фактор проведеної операції на пропорційність у часі. Для цього використаємо графік LLS (log negative log survival function) - це графік, що відображає логарифм негативного логарифму функції виживання у функції часу. Цей графік може бути корисним для аналізу даних виживання,

оскільки він може допомогти з'ясувати форму розподілу часу виживання та виявити аномальні області в графіку.

Основний спосіб використання графіка LLS полягає в тому, щоб з'ясувати, чи є даний розподіл часу виживання експоненційним або немонотонним.

Фактор проведеної додаткової операції буде змінною у майбутній моделі. На рис. 3.5 видно, що умова пропорційності зберігається.

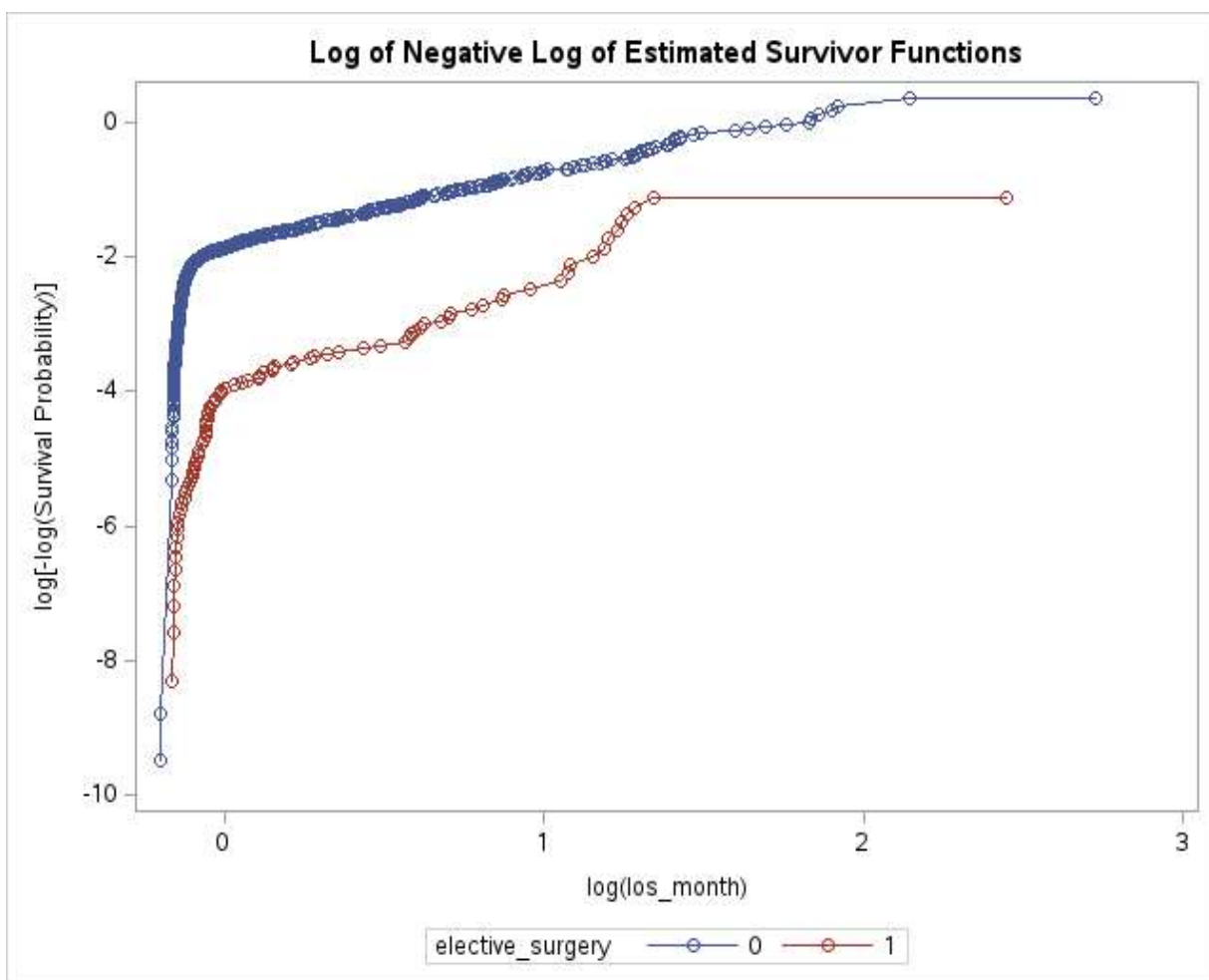


Рисунок 3.5 - LLS криві для фактору «Проведена додаткова операція»

Джерело: розраховано автором на основі даних [2]

Отже, за результатами побудови кривих Каплан-Майєра можемо зробити висновок, що вік та наявність раніше проведеної операції значуще впливають на

виживаність пацієнтів хворих на серцево-судинні захворювання, а стать пацієнтів ніяк не впливає на їх виживаність.

3.2. Статистичний аналіз факторів, що впливають на виживаність пацієнтів з урахуванням наявності додаткової операції

Нижче наведений список факторів, які були протестовані щодо не випадковості впливу на виживаність пацієнтів хворих на серцево-судинні захворювання:

1. Артеріальний тиск.
2. Діастолічний артеріальний тиск.
3. Систолічний артеріальний тиск.
4. Показник пульсу.
5. Частота дихання.
6. Рівень кисню.
7. Рівень глюкози.
8. Температура тіла.
9. Рівень калію.

Для перевірки істотності факторів потрібно: побудувати графік за методом Каплана-Мейера та перевірити за Log rank та Wilcoxon тестами для виживаності стратифікованої з урахуванням наявності проведеної операції.

За результатами перевірки за стратифікованими Log-Rank та Wilcoxon тестами було визначено фактори, що істотно та не випадково впливають на виживаність (табл. 3.1).

Факторами, що мають значний вплив стали: рівень кисню, рівень калію, рівень глюкози, артеріальний тиск, систолічний тиск та діастолічний тиск.

Таблиця 3.1 - Значення p перевірки впливу факторів виживаності пацієнтів хворих на серцево-судинні захворювання

Фактор	Вживаність	
	Log-Rank тест	Wilcoxon тест
Рівень кисню	<.0001	<.0001
Рівень калію	<.0001	<.0001
Рівень глюкози	<.0001	<.0001
Артеріальний тиск	<.0001	<.0001
Систолічний тиск	<.0001	<.0001
Діастолічний тиск	<.0001	<.0001
Температура тіла	0.3331	0.0492
Показник пульсу	0.0074	0.4861
Частота дихання	0.5378	0.0254

p<0.05 свідчить про не випадковість впливу фактору.

Джерело: розраховано автором на основі даних [2]

Далі розглянемо фактори – супутні хвороби, які можуть виникати чи сприяти погіршенню стану, які були протестовані щодо не випадковості впливу на виживаність пацієнтів хворих на серцево-судинні захворювання:

1. Печінкова недостатність
2. Імуносупресія
3. Пухлина з метастазами
4. Цукровий діабет
5. Цироз печінки
6. Лейкемія
7. Лімфома
8. ВІЛ/СНІД

Факторами, що мають значний вплив стали: печінкова недостатність та імуносупресія (рис.3.2).

Таблиця 3.2 - Значення p перевірки впливу факторів (супутніх хвороб) виживаності пацієнтів хворих на серцево-судинні захворювання

Фактор	Вживаність	
	Log-Rank тест	Wilcoxon тест
Печінкова недостатність	0.0096	0.0229
Імуносупресія	0.0281	0.0065
Пухлина з метастазами	0.2497	0.3252
Цукровий діабет	0.3169	0.3784
Цироз печінки	0.5773	0.5526
Лейкемія	0.7266	0.7152
Лімфома	0.7455	0.6832
ВІЛ/СНІД	0.761	0.7542

p < 0.05 свідчить про не випадковість впливу фактору.

Джерело: розраховано автором на основі даних [2]

Нижче наведені побудовані стратифіковані криві виживаності для факторів, що виявилися значущими, також представлений один з факторів, що є незначимим (рис. 3.6-3.10):

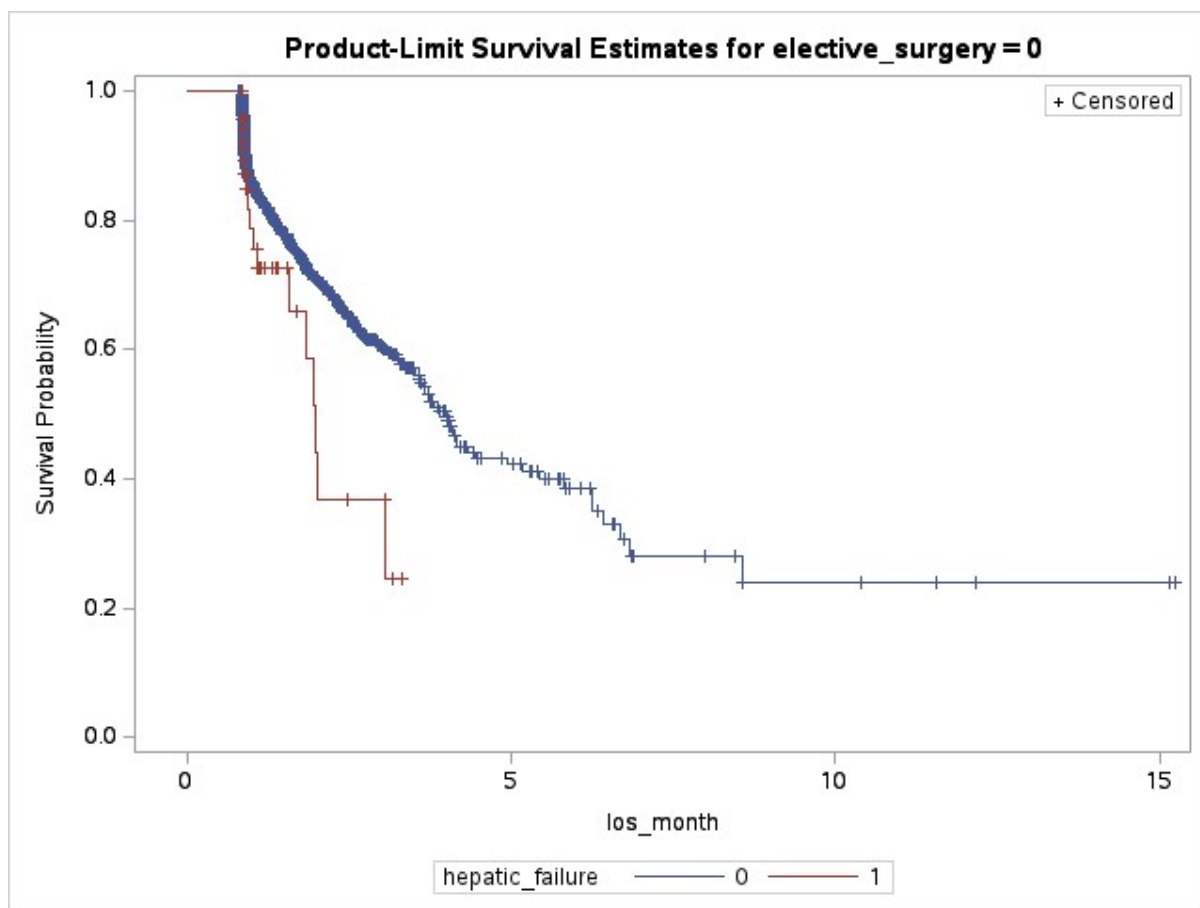


Рисунок 3.6 - Стратифікована оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від печінкової недостатності для тих, хто не мав попередньо проведеної операції за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис. 3.6 представлена функція виживаності для пацієнтів залежно від наявності печінкової недостатності для тих, хто не мав попередньо проведеної операції. З графіку видно, що наявність печінкової недостатності сильно впливає на виживаність пацієнтів. Такі пацієнти мають менші шанси на виживаність особливо в перші 3-4 місяці.

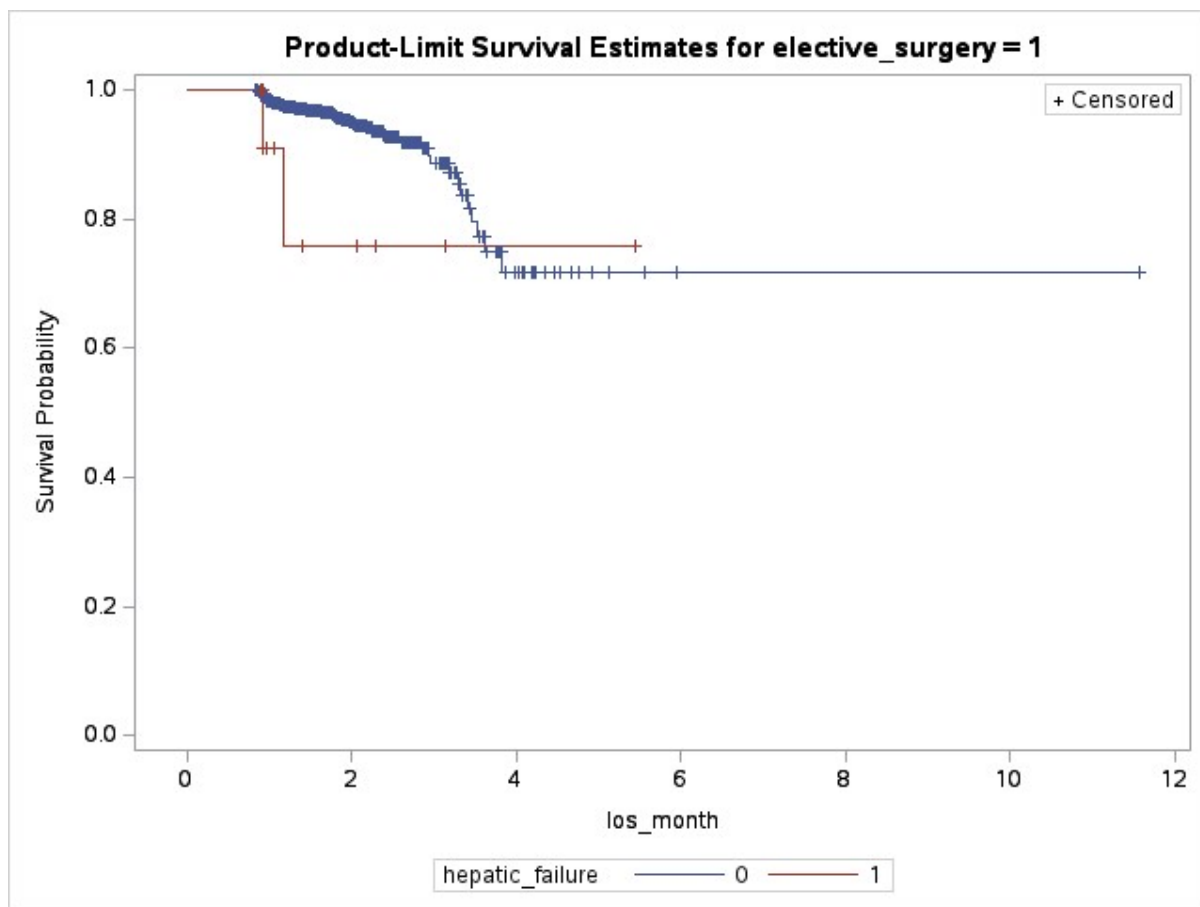


Рисунок 3.7 - Стратифікована оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від печінкової недостатності для тих, хто мав попередньо проведену операцію за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис. 3.7 представлена функція виживаності для пацієнтів залежно від наявності печінкової недостатності для тих, хто мав попередньо проведену операцію. З графіку видно, що наявність печінкової недостатності впливає на виживаність пацієнтів, але все ж умова попередньо проведеної операції дозволяє збільшити шанси на виживаність таких пацієнтів.

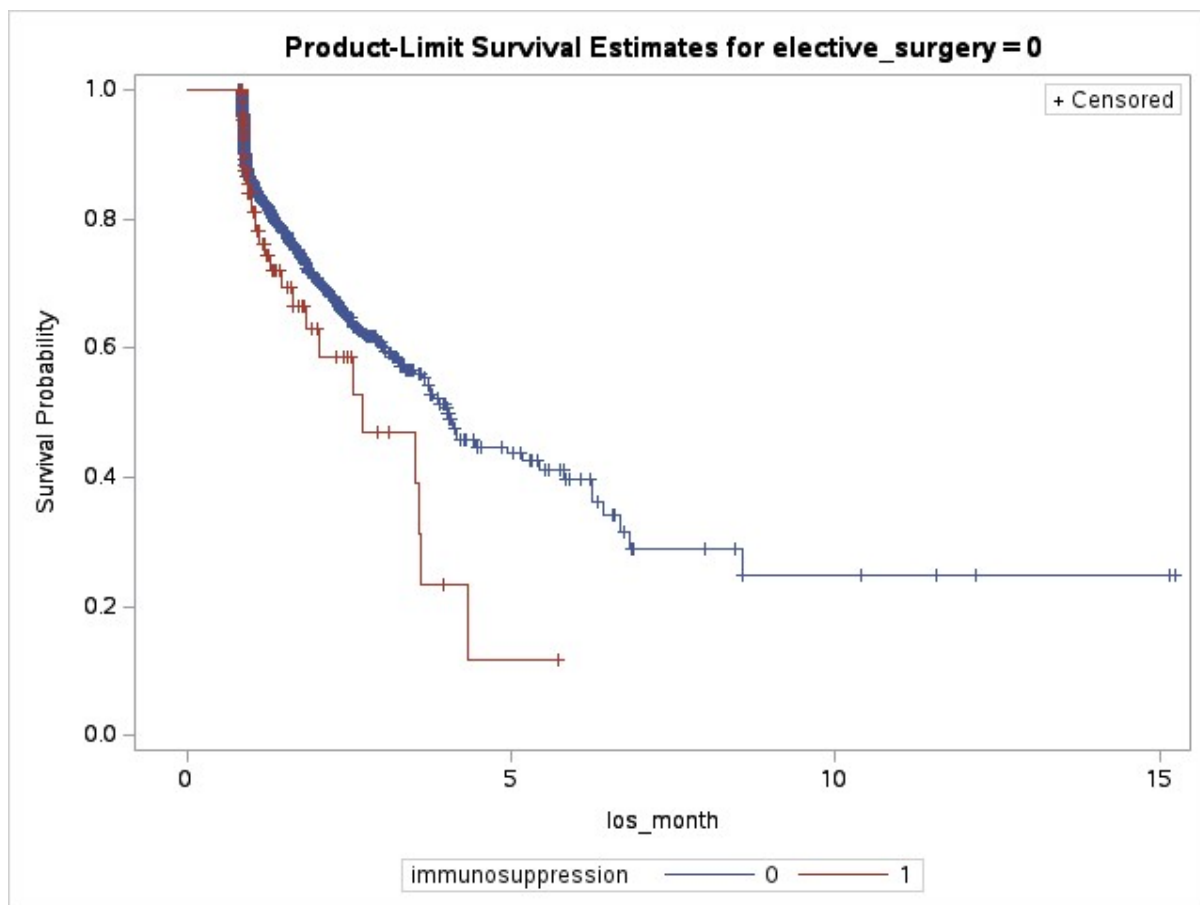


Рисунок 3.8 - Стратифікована оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від імуносупресії для тих, хто не мав попередньо проведеної операції за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис. 3.8 представлена функція виживаності для пацієнтів залежно від наявності імуносупресії для тих, хто не мав попередньо проведеної операції. З графіку видно, що це впливає на виживаність пацієнтів. Такі пацієнти мають менші шанси на виживаність особливо в перші 5 місяці.

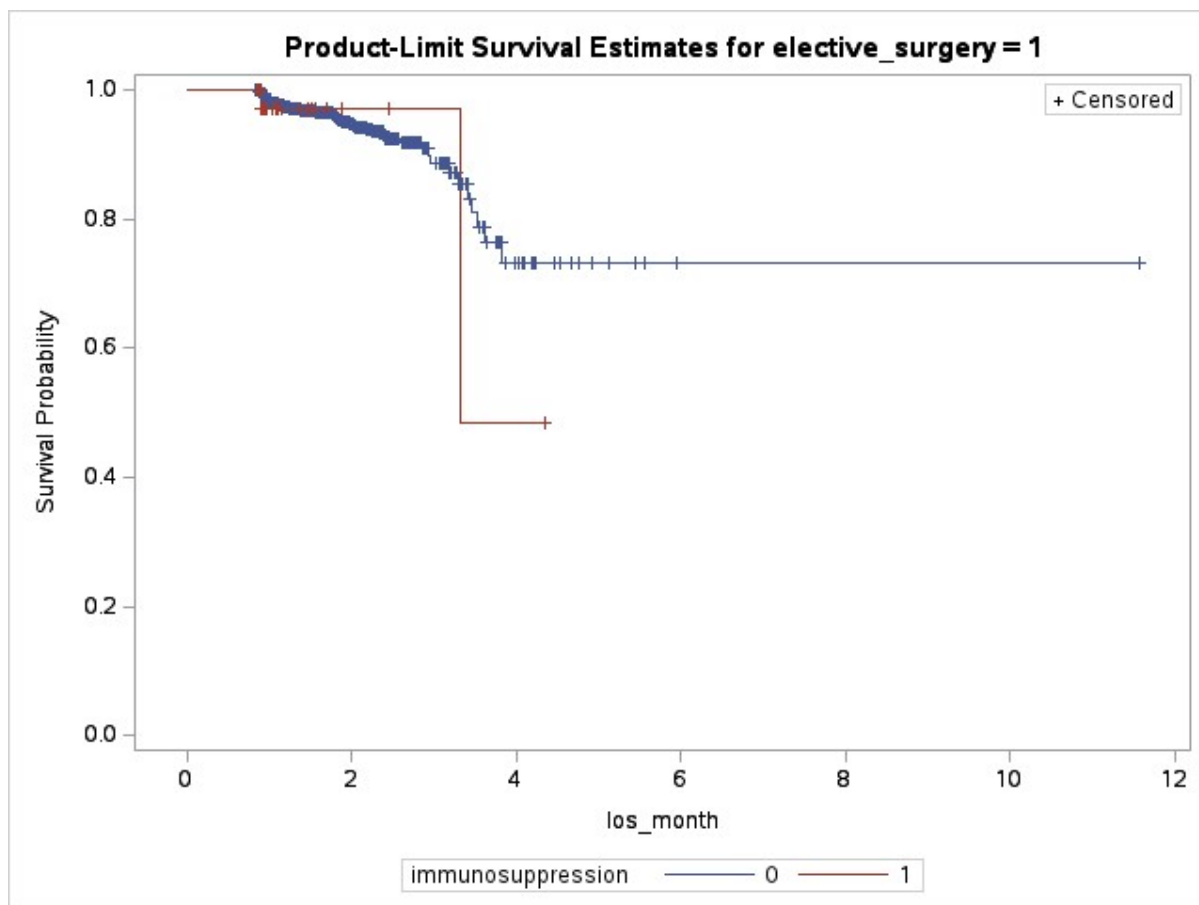


Рисунок 3.9 - Стратифікована оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від імуносупресії для тих, хто мав попередньо проведену операцію за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис. 3.9 представлена функція виживаності для пацієнтів залежно від наявності імуносупресії для тих, хто мав попередньо проведену операцію. Тяжко оцінити результати для цієї категорії пацієнтів, адже представлено всього 36 пацієнтів з усієї вибірки які мали проведену операцію та імуносупресію.

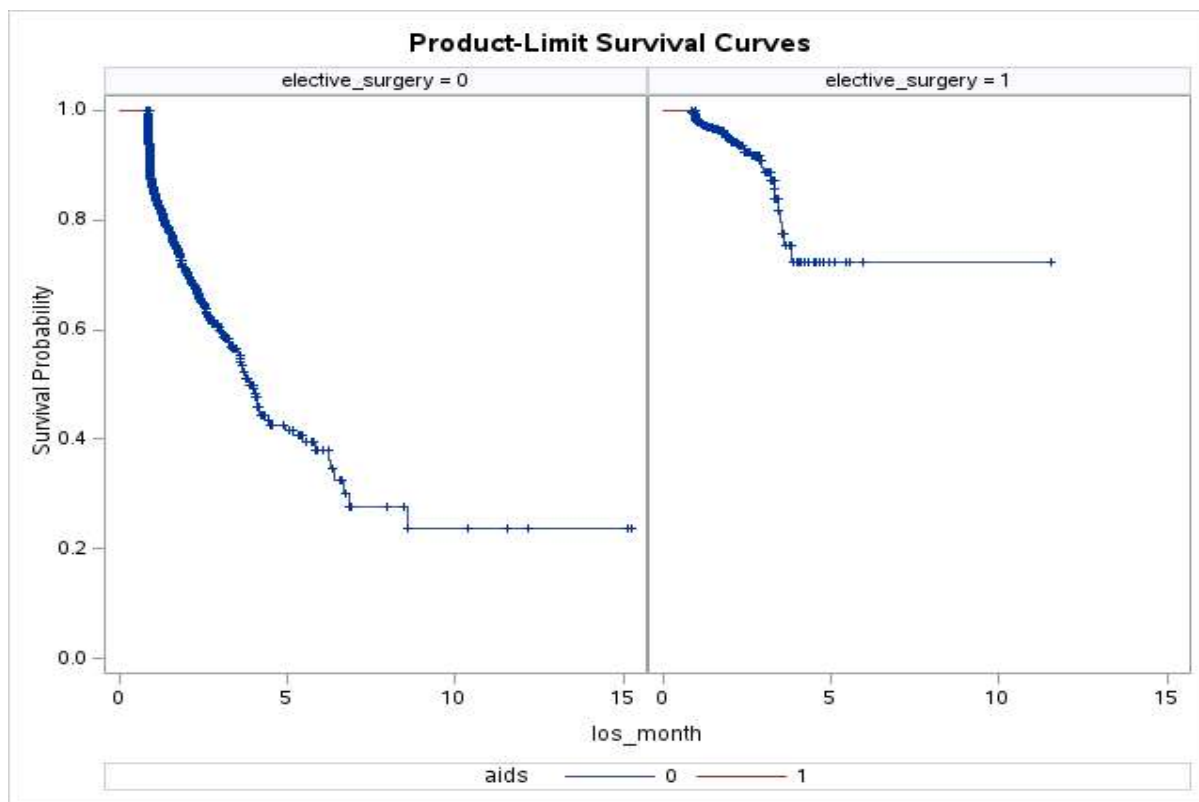


Рисунок 3.10 - Стратифікована оцінка імовірності виживаності пацієнтів хворих на серцево-судинні захворювання залежно від наявності ВІЛ/СНІД за Каплан-Мейером, місяців

Джерело: розраховано автором на основі даних [2]

На рис. 3.10 представлена функція виживаності для пацієнтів залежно від наявності ВІЛ/СНІД. Слід зазначити, що фактор не є значущим через те що вибірка пацієнтів які мають ВІЛ/СНІД відносно дуже мала: менше ніж 3 %.

Результати розрахунків показують нам те, що фактори за наявністю проведення попередньої операції по різному впливають на виживаність пацієнтів хворих на серцево-судинні захворювання.

Отже, такими факторами, що не випадково впливають на виживаність є: печінкова недостатність, імуносупресія (висновок можна зробити лише для популяцію пацієнтів у яких була проведена попередня операція).

3.3. Побудова моделі та порівняльна характеристика впливу факторів виживаності

Список параметрів для багатофакторного регресійного аналізу методом Кокса наведемо нижче:

- Проведення попередньої операції (elective_surgery),
- Печінкова недостатність (hepatic_failure),
- Частота серцебиття (d1_hearttrate_mean)
- Рівень глюкози (d1_glucose_mean)
- Вік (age).

Програмний код процедури наведений у додатку Г.

Маємо біноміальні змінні по стану проведення операції та печінкової недостатності, усі інші – неперервні.

Щоб оцінити та проаналізувати вплив факторів залежно від наявності попередньо проведеної операції до моделі були додані ефекти взаємодії (elective_surgery| hepatic_failure | d1_hearttrate_mean | d1_glucose_mean) [21]. Наявність цих ефектів дозволяє розрахувати відносний ризик залежно від комбінації групи лікування та відповідного фактору.

За результатами перевірки глобального тесту доведено можливість прогнозування виживаності для загальної моделі, на рівні істотності 0,05 (табл 3.3).

Таблиця 3.3 - Тестування глобальної нульової гіпотези

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1131,7868	5	<,0001
Score	1215,0885	5	<,0001
Wald	958,5514	5	<,0001

Джерело: розраховано автором на основі даних [2]

У результаті моделювання усього було оброблено 17287 спостереження, з яких відбулося 1421 подій і 15866 особи було цензуровано, що складає 91.78 % від усієї вибірки.

Фактори до моделі додавалися за допомогою покрокового методу (stepwise selection). Цей метод є одним з популярних підходів до включення факторів у модель і ґрунтується на статистичних критеріях.

Покроковий метод зазвичай має дві фази: початкову фазу додавання факторів та фазу вилучення факторів. У початковій фазі розглядаються всі можливі фактори, а потім використовуються статистичні критерії, для вибору найкращого фактора для включення в модель. У фазі вилучення факторів розглядаються вже включені в модель фактори, і знову застосовуються статистичні критерії для вилучення незначущих факторів [22].

Покроковий метод дозволяє систематично включати або вилучати фактори з моделі залежно від їх статистичної значущості. Це допомагає скоротити модель до найбільш істотних факторів, що знижує ризик перенавчання і поліпшує інтерпретованість результатів.

Розглянемо тест типу 3 (Type 3 Tests), який є одним з методів оцінки статистичної значущості факторів у моделі пропорційних інтенсивностей Кокса (табл. 3.4).

Таблиця 3.4 - Результати Тесту типу 3 покроковим методом

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
elective_surgery	1	299,2011	<,0001
hepatic_failure	1	4,3689	0,0366
d1_heartrate_mean	1	28,1506	<,0001
d1_glucose_mean	1	52,5648	<,0001
d1_heartr*d1_glucose	1	8,3018	0,0040

Джерело: розраховано автором на основі даних [2]

В рамках тесту типу 3 оцінюється вплив окремого фактора, при цьому ураховуючи взаємодію між цим фактором та іншими факторами, які включені в модель. Він дозволяє визначити, чи є статистично значущим вплив окремого фактора на ризик події, при цьому контролюючи вплив інших факторів, які також мають важливість для виживання.

Тест типу 3 базується на методі максимальної правдоподібності і заснований на декомпозиції суми квадратів відхилень (sum of squares, SS). Він враховує унікальний вплив окремого фактора та його взаємодію з іншими факторами, і показує, наскільки зміна фактора впливає на вибірку при фіксованих значеннях інших факторів.

Серед ефектів доданих у модель ефектів взаємодії між частотою серцебиття та рівнем глюкози виявився істотним. Загальновідомим є факт, що, рівень глюкози в організмі може впливати на частоту серцебиття. Глюкоза є основним джерелом енергії для організму, і її рівень може впливати на функціонування різних органів і систем, включаючи серцево-судинну систему.

Зазвичай, при підвищеному рівні глюкози в крові (гіперглікемії), серцебиття може збільшуватись.

З іншого боку, при низькому рівні глюкози в крові (гіпоглікемії), може спостерігатись зменшення частоти серцевих скорочень, цю залежність можна спостерігати за результатами нижче наведеної таблиці (табл. 3.5).

Таблиця 3.5 - Параметри моделі

Analysis of Maximum Likelihood Estimates							
Parameter		Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
elective_surgery	0	1,93068	0,11162	299,201 1	<,0001	6,894	elective_surgery 0
hepatic_failure	0	-0,46040	0,22027	4,3689	0,0366	0,631	hepatic_failure 0
d1_heartrate_mean		0,02003	0,00377	28,1506	<,0001	.	
d1_glucose_mean		0,01320	0,00182	52,5648	<,0001	.	
d1_heartr*d1_glucose		0,0000582	0,000020	8,3018	0,0040	.	d1_heartrate_mean * d1_glucose_mean

Джерело: розраховано автором на основі даних [2]

Також, представлено фактори впливу, які не були включені в модель (рис.3.6), адже вони виявились не значущими. Це означає, що не можна робити висновки про взаємодію факторів за наявності проведення попередньої операції окремо.

Таблиця 3.6 - Параметри моделі, які не були включені

Effect	DF	Score	Pr > ChiSq
		Chi-Square	
elective_*hepatic_fa	1	2,8122	0,0936
d1_heartr*elective_s	1	1,4873	0,2226
d1_glucos*elective_s	1	1,2548	0,2626

Джерело: розраховано автором на основі даних [2]

Інтерпретація усіх інших факторів, які виявились істотними відбуватиметься з поширенням на усю сукупність пацієнтів (табл.3.7).

Таблиця 3.7 - Результати Тесту типу 3 покроковим методом для усієї сукупності

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
elective_surgery	1	299,6129	<,0001
age	1	26,7269	<,0001
hepatic_failure	1	4,4424	0,0351
d1_heartrate_mean	1	52,0239	<,0001
d1_glucose_mean	1	448,7860	<,0001

Джерело: розраховано автором на основі даних [2]

Так само, найбільшу прогностичну силу у моделі має фактор проведення попередньої операції (табл. 3.8).

Таблиця 3.8 - Параметри моделі для усієї сукупності

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
elective_surgery	0	1	1,93134	0,11158	299,6129	<,0001	6,899	elective_surgery 0
age		1	-0,01040	0,00201	26,7269	<,0001	1,010	
hepatic_failure	0	1	-0,46422	0,22025	4,4424	0,0351	0,629	hepatic_failure 0
d1_hearttrate_mean		1	0,01068	0,00148	52,0239	<,0001	1,011	
d1_glucose_mean		1	0,00796	0,0003757	448,7860	<,0001	1.008	

Джерело: розраховано автором на основі даних [2]

З результатів моделювання можна зробити наступний висновок для пацієнтів хворих на серцево-судинні захворювання:

- проведення попередньої операції допомагає збільшити шанси на виживаність пацієнтів в 6,89 разів, ніж у пацієнтів у котрих не була проведена операція:

- спостерігається залежність взаємодії між частотою серцебиття та рівнем глюкози:

- враховуючи фактори взаємодії, які не увійшли в модель, можна зробити висновок, що не можна робити висновки про взаємодію факторів за наявності проведення попередньої операції окремо:

- з кожним роком життя пацієнтів, шанси виживаності зменшуються на 1,04% при незмінності інших факторів.

ВИСНОВКИ

За результатами статистичного дослідження факторів що впливають на виживаність пацієнтів хворих на серцево-судинні захворювання за умовою раніше проведеної операції можна зробити висновки:

1. Загальновідомим є той факт, що, супутні хвороби мають значний вплив на розвиток серцево-судинних захворювань. Цукровий діабет, гіпертонія, дисліпідемія, ожиріння та стрес - всі вони взаємодіють і сприяють ризику серцевих проблем.

2. Статистичні дослідження допоможуть визначити потенційні ризики для розвитку серцево-судинних захворювань та виявити чинники, що впливають на їх розвиток. На основі цих досліджень можна розробляти програми профілактики і діагностики, що забезпечать захист здоров'я населення. Також вони дозволять оцінити ефективність вже існуючих програм профілактики та діагностики.

3. Для проведення аналізу виживаності було взято дані 17287 пацієнтів, які мають серцево-судинні захворювання та були госпіталізовані у США у важкому стані, знаходились під наглядом.

4. Використання методу головних компонент дало нам змогу визначити найбільші групи факторів впливу для пацієнтів котрі вижили чи померли та для усієї вибірки взагалом.

Було виділено 3 групи факторів (головні компоненти), а саме: «основні причинно-наслідкові фактори», «фактори показників пульсу», «фактори частоти дихання», які мають найвищі вагові коефіцієнти.

5. За результатами використання методу Каплана-Мейєра було виявлено, що такими факторами, що не випадково впливають на виживаність є: рівень кисню, рівень калію, рівень глюкози, артеріальний тиск, систолічний тиск та діастолічний тиск, також значущими є фактори печінкової недостатності та імуносупресії.

6. З результатів моделювання можна зробити висновки:

- проведення попередньої операції допомагає збільшити шанси на виживаність пацієнтів в 6,89 разів, ніж у пацієнтів у котрих не була проведена операція:
- спостерігається залежність взаємодії між частотою серцебиття та рівнем глюкози:
- враховуючи фактори взаємодії, які не увійшли в модель, можна зробити висновок, що не можна робити висновки про взаємодію факторів за наявністю проведення попередньої операції окремо:
- з кожним роком життя пацієнтів, шанси виживаності зменшуються на 1,04% при незмінності інших факторів.

Проте, не зважаючи на великий вплив попередньо проведеної операції, не можна робити висновки про взаємодію факторів за її наявністю окремо.

Також було б добре дослідити більше факторів, наприклад, рівень куріння, споживання алкоголю, впровадження здорового харчування, тощо.

Проведене дослідження з використанням методу головних компонент для бази даних пацієнтів, які мають серцево-судинні захворювання, дозволило визначити взаємозв'язки між досліджуваними показниками, які можна пояснити з медичної точки зору.

Аналіз виживаності дозволив виявити фактори, які мають статистично значимий вплив на виживаність пацієнтів, хворих на серцево-судинні захворювання. Використання методів аналізу виживаності забезпечило більш детальне визначення цих факторів та їх значення у забезпеченні ефективного лікування та покращенні прогнозів щодо виживаності пацієнтів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. World Health Organization [Електронний ресурс] // World Health Organization: [сайт]. - Режим доступу: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1
2. Kaggle: Patient Survival Prediction [Електронний ресурс] // Patient Survival Prediction: [сайт]. - Режим доступу: <https://www.kaggle.com/datasets/mitishaagarwal/patient/>
3. Megan Lindstrom, Nicole DeCleene Journal of the American College of Cardiology // Volume 80, Issue 25 Dec 2022
4. Рудзей В. Метод головних компонент: Оцінювання ризиків для пацієнтів хворих на серцево-судинні захворювання [Електронний ресурс] // Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення (випуск 75) [сайт]. - Режим доступу: <http://www.konferenciaonline.org.ua/ua/article/id-1025/>
5. Чугаєвська, С. В., Ковтун, Н. В. *Основи статистичного моделювання*. Вид-во ЖДУ ім. І. Франка // Житомир – «Рута». - 2022
6. Т. С. Клебанова, Л. С. Гур'янова Бізнес-аналітика багатовимірних процесів // Мультимедійний навчальний посібник. - 2020 – Р. 8.4
7. Громова Г.М. Інструменти вимірювання толерантності до невизначеності // Наукові студії із соціальної та політичної психології Вип. 47 (50)
8. Гланц С. Медико-біологіческая статистика. // Пер. с англ., М: Практика 1999; 459 стр.
9. Голубова Г. В. Криві виживаності Каплана – Мейєра: техніка моделювання. *Науковий вісник Національної академії статистики, обліку та аудиту: зб. наук. пр.* 2021. №3-4. С. 15-22. doi: 10.31767/nasoa.3-4-2021.02.
10. Machin D., Cheung Y., Parmar M. Survival Analysis: A Practical Approach. 2nd ed. New York, 2006.

11. Petrie A., Sabin C. Medical Statistics at a Glance. Oxford, 2005.
12. Kleinbaum David G., Klein M. Survival Analysis: A Self-Learning Text. Third Edition. URL: uop.edu.pk/ocontents/survival-analysis-self-learning-book.pdf
13. Brookmeyer, R. and Crowley, J. (1982), "A Confidence Interval for the Median Survival Time," *Biometrics*, 38, 29–41.
14. Collett, D. (1994), *Modeling Survival Data In Medical Research*, London: Chapman and Hall.
15. Cox, D.R. and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman and Hall.
16. Elandt-Johnson, R.C. and Johnson, N.L. (1980), *Survival Models and Data Analysis*, New York: John Wiley & Sons.
17. Royston P., Parmar M.K., Altman D.G. Visualizing of survival in time-to-event studies: a compliment to Kaplan-Meier Plots. // *J. Natl. Cancer Inst.*, 2008, Vol. 100, p. 92-97.
18. Lawless, J.E. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons.
19. Lee, E.T. (1992), *Statistical Methods for Survival Data Analysis*, Second Edition, New York: John Wiley & Sons.
20. *SAS Survival Analysis Techniques for Medical Research*, 2nd Edition. Alan B.Cantor.
21. *SAS Survival Analysis using SAS: A Practical Guide*. Paul D. Allison.
22. Therneau, TM, Grambsch, PM. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer: New York.
23. Grambsch, PM, Therneau, TM. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 81. 515-526.
24. Lin, DY, Wei, LJ, Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*. 80(30). 557-72.
25. Therneau, TM, Grambsch PM, Fleming TR (1990). Martingale-based residuals for survival models. *Biometrika*. 77(1). 147-60.