

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
”МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ)”**

ЖУРНАЛ ПРАКТИКИ

Студента 3 курса

Гординского Дмитрия Михайловича

Институт №8 «Информационные технологии и прикладная математика»

Кафедра №804 «Теория вероятностей и компьютерное моделирование»

Учебная группа М8О-304Б-20

Направление 01.03.04

Прикладная математика

Вид практики Учебная (вычислительная) в Московском Авиационном Институте(НИУ)

Руководитель практики от МАИ Зайцева О.Б.

Гординский Д.М / / 7 июня 2023 г.

Москва, 2023

1. Место и сроки проведения практики

Дата начала практики 9 февраля 2023 г.

Дата окончания практики 7 июня 2023 г.

Наименование предприятия МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ(НИУ)

Название структурного подразделения Кафедра 804

2. Инструктаж по технике безопасности

Платонов Е. Н. / / 7 июня 2023 г.

3. Индивидуальное задание студенту

Выбрать тему дипломной работы, ориентируясь на доступные наборы данных.

4. План выполнения индивидуального задания

1. Изучить теорию по Анализу выживаемости.
2. Ознакомиться с необходимыми библиотеками для работы с данными и их графическим представлением.
3. Ознакомиться с датасетами.
4. Выбрать тему работы и датасет на основании изученной информации.
5. Написать реферат.

Руководитель практики от МАИ: Зайцева О.Б. /

Платонов Е. Н. / 7 июня 2023 г.

5. Отзыв руководителя практики

Задание на практику выполнено в полном объеме. Материалы, изложенные в отчете студента, полностью соответствуют индивидуальному заданию. Рекомендую оценку отлично.

Руководитель

Платонов Е. Н. / _____ */ 7 июня 2023 г.*

Contents

1	Введение	5
2	Краткое описание данных	5
3	Описание полей	5
3.1	О предмете исследования	6
3.1.1	Причины возникновения	6
3.1.2	Факторы риска	7
4	План работ и выводы	7
5	Общая структура работы по выбранной теме	7
5.1	Введение	7
5.2	Какие факторы оказывают наибольшее влияние на выживаемость пациентов с ПБЦ печени?	8
5.3	Методы и подходы	8
5.4	Методы <i>анализа выживаемости</i>	10
5.5	Многомерный анализ	10
5.5.1	Анализ множественной регрессии	11
5.5.2	Анализ логистической регрессии	11
5.5.3	Анализ дискриминантов	12
5.5.4	Многомерный анализ переменных (MANOVA)	12
5.5.5	Факторный анализ	13
5.5.6	Кластерный анализ	13
5.5.7	Каноническая корреляция	13
5.5.8	Моделирование структурных уравнений	13
5.6	Заключение	14
5.7	Методика работы и средства	14
6	Вывод	14
7	Источники	14

1 Введение

Анализ выживаемости — это статистический метод, используемый для изучения времени, которое прошло до наступления определенного события. Этот метод может использоваться для анализа многих видов данных, включая медицинские, социальные и экономические.

В последнее время, анализ выживаемости стал одним из наиболее популярных методов исследования. Это связано с тем, что данный метод позволяет проводить анализы с использованием цензурированных данных. В свою очередь, это позволяет учитывать не только количество произошедших событий, но и время, прошедшее до наступления события.

Одной из наиболее важных областей, в которых используется анализ выживаемости, является медицина, где он используется для анализа выживаемости пациентов после лечения различных заболеваний. Кроме того, анализ выживаемости используется в экономике, социологии, и многих других областях.

В этой работе я хотел бы рассмотреть конкретно сферу медицины.

Выбор данных для анализа — это важный шаг в любом исследовании.

Из-за того, что работа с пациентом подразумевает полный набор данных и характеристик, то в сфере медицинских исследований мы чаще всего можем получить большую репрезентативную выборку с множеством параметров.

Имеено по этой причине я выбрал датасет клиники Майо по первичному билиарному циррозу

2 Краткое описание данных

Эти данные взяты из клинического исследования первичного билиарного цирроза (ПБЦ) печени, проводившегося в клинике Майо в период с 1974 по 1984 год. В общей сложности 424 пациента с ПБЦ, направленные в клинику Майо в течение этого десятилетнего интервала, соответствовали критериям отбора для рандомизированного исследования плацебо. контролируемое исследование препарата D-пеницилламин. Первые 312 случаев в наборе данных участвовали в рандомизированном исследовании и содержат в основном полные данные. Дополнительные 112 пациентов не участвовали в клинических испытаниях, но дали согласие на запись основных измерений и наблюдение за их выживаемостью. Шесть из этих случаев были потеряны для последующего наблюдения вскоре после постановки диагноза, поэтому данные здесь относятся к дополнительным 106 случаям, а также к 312 рандомизированным участникам.

3 Описание полей

```
age:      in years
albumin:  serum albumin (g/dl)
alk.phos: alkaline phosphatase (U/liter)
ascites:  presence of ascites
ast:      aspartate aminotransferase, once called SGOT (U/ml)
bili:     serum bilirunbin (mg/dl)
chol:     serum cholesterol (mg/dl)
copper:   urine copper (ug/day)
edema:    0 no edema, 0.5 untreated or successfully treated
```

1 edema despite diuretic therapy
hepato: presence of hepatomegaly or enlarged liver
id: case number
platelet: platelet count
protime: standardised blood clotting time
sex: m/f
spiders: blood vessel malformations in the skin
stage: histologic stage of disease (needs biopsy)
status: status at endpoint, 0/1/2 for censored, transplant, dead
time: number of days between registration and the earlier of death,
transplantation, or study analysis in July, 1986
trt: 1/2/NA for D-penicillmain, placebo, not randomised
trig: triglycerides (mg/dl)

3.1 О предмете исследования

Цирроз печени — это заболевание, характеризующееся перерождением паренхиматозной ткани печени в фиброзную соединительную ткань. Сопровождается тупой болью в правом подреберье, желтухой, повышением давления в системе воротной вены с характерными для портальной гипертензии кровотечениями (пищеводными, геморроидальными), асцитом и пр. Заболевание носит хронический характер. В диагностике цирроза печени определяющую роль играют данные УЗИ, КИ и МРТ печени, показатели биохимических проб, биопсия печени. Лечение цирроза печени предусматривает строгий отказ от алкоголя, соблюдение диеты, прием гепатопротекторов; в тяжелых случаях — трансплантацию донорской печени.

3.1.1 Причины возникновения

Цирроз характеризуется возникновением в ткани печени соединительнотканых узлов, разрастанием соединительной ткани, формированием «ложных» долек. Цирроз различают по размеру формирующихся узлов на мелкоузловой (множество узелков до 3 мм в диаметре) и крупноузловой (узлы превышают 3 мм в диаметре). Изменения структуры органа в отличие от гепатитов необратимы, таким образом, цирроз печени относится к неизлечимым заболеваниям.

Среди причин развития цирроза печени лидирует злоупотребление алкоголем (от 35,5% до 40,9% пациентов). На втором месте располагается вирусный гепатит С. У мужчин цирроз развивается чаще, чем у женщин, что связано с большим распространением в мужской среде злоупотребления алкоголем.

В подавляющем большинстве случаев причиной развития цирроза печени является злоупотребление алкоголем и вирусные гепатиты В и С, реже - ферментопатии:

- Алкогольная зависимость. Регулярное употребление алкоголя в дозах 80-160 мл этанола ведет к развитию алкогольной болезни печени, которая в свою очередь прогрессирует с возникновением цирроза. Среди лиц, злоупотребляющих алкоголем на протяжении 5-10 лет, циррозом страдает 35%.
- Заболевания гепатобилиарной системы. Хронические гепатиты также зачастую ведут к фиброзу перерождению ткани печени. На первом месте по частоте диагностирования стоят вирусные гепатиты В и С (гепатит С склонен к более деструктивному течению и прогрессирует в цирроз чаще). Также цирроз может

стать результатом хронического аутоиммунного гепатита, склерозирующего холангита, первичного холестатического гепатита, сужения желчных протоков, застоя желчи. Циррозы, развивающиеся вследствие нарушений в циркуляции желчи, называют билиарными. Они подразделяются на первичные и вторичные.

- Метаболические нарушения. Причиной развития цирроза печени может стать обменная патология или недостаточность ферментов: муковисцидоз, галактоземия, гликогеноз, гемохроматоз.

3.1.2 Факторы риска

К факторам риска перерождения печеночной ткани относят:

- гепатолентикулярную дегенерацию (болезнь Вильсона);
- прием гепатотоксичных лекарственных препаратов (метотрексат, изониазид, амиодарон, метил-допа);
- хроническую сердечную недостаточность;
- синдром Бада-Киари;
- операционные вмешательства на кишечнике;
- паразитарные поражения кишечника и печени.
- В 20-30% случаев причину развития цирроза печени установить не удастся, такие циррозы называют криптогенными.

Таким образом, мы понимаем, что данная болезнь может подразумевать глубокое исследование.

4 План работ и выводы

Из-за большого количества параметров в датасете мы можем выбрать разные методы исследования проблемы, в нашем случае — болезни.

Поэтому на основании данных о пациентах, которые болели циррозом печени, может быть Тема *“Какие факторы оказывают наибольшее влияние на выживаемость пациентов с ПБЦ печени”*.

В рамках этой темы можно провести анализ связи между различными факторами, такими как пол, возраст, наличие других заболеваний, уровень печеночных ферментов и т.д., и шансами выживания пациентов с ПБЦ. Это поможет понять, какие факторы могут быть наиболее важными при принятии решений о лечении этой болезни, и какие методы лучше использовать для запуска программ лечения.

5 Общая структура работы по выбранной теме

5.1 Введение

Анализ выживаемости — это метод статистического анализа, который позволяет предсказать вероятность смерти пациентов и определить факторы, которые могут повлиять на выживаемость.

В медицине анализ выживаемости является важной частью медицинских исследований, так как он позволяет определить факторы, которые оказывают наибольшее влияние на выживаемость пациентов.

5.2 Какие факторы оказывают наибольшее влияние на выживаемость пациентов с ПБЦ печени?

ПБЦ печени (пирогенный бактериальный цирроз печени) является одной из форм цирроза печени, которая возникает на фоне инфекции бактериями. В данном случае, исследование влияния различных факторов на выживаемость пациентов с ПБЦ печени имеет большое значение для оптимизации лечения и улучшения прогноза заболевания.

Исходя из опыта и результатов предыдущих исследований, можно выделить несколько факторов, которые оказывают наибольшее влияние на выживаемость пациентов с ПБЦ печени.

- Первый фактор — это стадия заболевания на момент диагностики. Чем раньше была поставлена диагностика и начато лечение, тем больше шансов на выживание у пациента.
- Второй фактор — это возраст пациента. Чем моложе пациент, тем больше шансов на выживание, так как у молодых организмы лучше адаптируются к стрессовым ситуациям.
- Третий фактор — это наличие сопутствующих заболеваний. Наличие других заболеваний может значительно ухудшить прогноз заболевания и увеличить смертность пациентов.
- Четвертый фактор — это наличие осложнений заболевания. Осложнения, такие как гепатическая недостаточность или гепатический кома, могут значительно увеличить смертность пациентов, особенно если не проводится своевременное и адекватное лечение.
- Пятый фактор — это эффективность лечения. Применение современных методов лечения и обслуживания пациентов с ПБЦ печени может значительно увеличить шансы на выживание пациентов.

Для анализа выживаемости необходимо собрать информацию о сроках выживаемости пациентов после диагностики заболевания. Эта информация может быть собрана как проспективно (наблюдение и сбор данных в режиме реального времени), так и ретроспективно (использование архивных данных).

Основным инструментом для анализа выживаемости является кривая выживаемости, которая показывает процент пациентов, которые остаются в живых в зависимости от времени с момента диагностики заболевания.

Также для анализа выживаемости часто используются методы сравнения кривых выживаемости, такие как лог-ранг тест или тест Мантела-Кокса.

5.3 Методы и подходы

Методы и подходы, которые можно использовать для исследования влияния различных факторов на выживаемость пациентов с ПБЦ печени, включают в себя различные методы статистического анализа, методы исследования, проведение клинических испытаний и социальные научные методы. Рассмотрим каждый из них подробнее.

1. Методы статистического анализа

Один из наиболее распространенных методов статистического анализа для определения факторов, влияющих на выживаемость пациентов с ПБЦ печени, — это многомерный анализ. Этот метод анализирует сразу несколько независимых переменных, при этом учитывая взаимодействие между ними. Также широко используется метод ковариации, с помощью которого можно определить, какие факторы имеют наибольший вклад в общее количество изменений.

2. Методы исследования

Для проведения исследования влияния различных факторов на выживаемость пациентов с ПБЦ печени могут быть использованы следующие методы:

- **Наблюдение.** Этот метод предполагает сбор данных о пациентах непосредственно в режиме реального времени. Данные могут быть собраны в рамках мониторинга состояния пациентов в течение всего периода наблюдения.
- **Кросс-секционное исследование.** Этот метод предполагает сравнение данных о пациентах в определенный период времени. Такой подход позволяет получить дополнительную информацию о влиянии различных факторов на выживаемость пациентов.

3. Клинические испытания

Для исследования влияния различных методов лечения и жизненно важных факторов на выживаемость пациентов с ПБЦ печени могут быть проведены клинические испытания. Они могут быть как на протокол, так и на свободном наборе пациентов. Две наиболее популярные методики группа рандомизированных контролируемых клинических испытаний (RCT) и оценка эффективности клиники (CEA).

- RCT — это экспериментальные исследования, в которых пациенты случайным образом распределяются на две группы: группу контроля и группу, которая получает новый метод лечения. Этот метод позволяет определить, какой метод лечения наиболее эффективен для пациентов с ПБЦ печени.
- CEA — это метод, который позволяет проводить оценку эффективности клиники или других медицинских учреждений путем сравнения эффективности различных методов лечения и других медицинских услуг. Он может помочь в определении наиболее эффективных методов лечения для пациентов с ПБЦ печени.

4. Социальные научные методы

Для определения влияния социальных факторов на выживаемость пациентов с ПБЦ печени могут быть использованы социальные научные методы, такие как социологические опросы и мониторинг социальных тенденций. Они позволяют оценить, как социальные факторы, такие как образ жизни, доступ к медицинской помощи и др., влияют на выживаемость пациентов с ПБЦ печени.

В нашем случае мы попытаемся исследовать данное заболевание методами математической статистики.

5.4 Методы *анализа выживаемости*

Анализ выживаемости включает в себя различные методы и подходы для изучения времени до наступления определенного события, такого как смерть пациента. Это может помочь в определении частоты смертности от ПБЦ печени, а также выявлении факторов, которые могут повлиять на выживаемость пациентов.

Для проведения анализа выживаемости могут быть использованы следующие методы:

1. Кривая выживаемости Каплана-Мейера

Кривая выживаемости Каплана-Мейера показывает процент выживших пациентов в зависимости от времени. Этот метод может быть использован для сравнения выживаемости пациентов в группах с разными факторами риска. Также может быть проведено сравнение среднего времени выживаемости между группами.

2. Регрессионный анализ выживаемости

Регрессионный анализ выживаемости является мощным инструментом для изучения факторов, влияющих на выживаемость пациентов с ПБЦ печени. Этот метод позволяет определить факторы, которые могут повлиять на выживаемость пациентов в длинном периоде времени, а также учитывать многие независимые переменные.

3. Многомерный анализ выживаемости

Многомерный анализ выживаемости является более сложным методом, который позволяет изучать взаимодействие многих факторов на выживаемость пациентов. Использование этого подхода может помочь в определении того, как различные факторы коллективно влияют на выживаемость.

4. Анализ времени до события

Дополнением к анализу выживания является анализ времени до события. Это может использоваться для определения того, когда произойдет смерть пациента, а также для изучения влияния различных факторов на возникновение смерти.

Хотелось бы рассмотреть метод *Многомерного анализа выживаемости*, т.к. он с наибольшей вероятностью даст более ощутимый вклад в исследование.

5.5 Многомерный анализ

Многомерный анализ выживаемости — это расширение метода анализа выживаемости, который позволяет исследовать влияние множества факторов на выживаемость пациентов. Он основан на модели Кокса, которая описывает отношение между выживаемостью и независимыми факторами. Многомерный анализ выживаемости позволяет оценить влияние нескольких факторов на выживаемость пациентов, учитывая взаимодействия между ними. Это позволяет получить более точные и надежные результаты при исследовании выживаемости пациентов с циррозом печени.

Прежде чем следовать многомерной методике анализа данных, необходимо собрать соответствующие данные о главном факторе, который мы будем изучать. Данные могут быть метрическими или неметрическими. Следовательно, наш анализ зависит от используемых нами данных. Более того, когда мы находим некоторые данные, вам нужно следить за пропущенными данными во время анализа. Необходимо знать, являются ли данные существенными, или мы сможем найти результат без этих данных.

5.5.1 Анализ множественной регрессии

Эта техника используется чаще всего. В этой многомерной методике анализа данных мы рассмотрим, как две или более независимых метрических переменных соотносятся с одним зависимым метрическим вариантом. Аналитики также используют множественную регрессию для прогнозирования переменной.

Вычислительные аспекты

Общая вычислительная задача, которую требуется решать при анализе методом множественной регрессии, состоит в подгонке прямой линии к некоторому набору точек.

- Метод наименьших квадратов
- Уравнение регрессии
- Однозначный прогноз и частная корреляция
- Предсказанные значения и остатки
- Остаточная дисперсия и коэффициент детерминации R-квадрат
- Интерпретация коэффициента множественной корреляции R

5.5.2 Анализ логистической регрессии

Логистическая регрессия — это статистический метод для анализа набора данных, в котором есть одна или несколько независимых переменных, которые определяют результат. Результат измеряется с помощью дихотомической переменной (в которой есть только два возможных результата). Он используется для прогнозирования двоичного результата (1/0, Да / Нет, Истина / Ложь) с учетом набора независимых переменных.

Также можно рассматривать логистическую регрессию как особый случай линейной регрессии, когда исходная переменная является категориальной, где мы используем логарифм шансов в качестве зависимой переменной. Проще говоря, он предсказывает вероятность возникновения события путем подгонки данных к логит-функции.

В некоторых случаях зависимые переменные могут иметь более двух результатов, например, в браке / не замужем / в разводе, такие сценарии классифицируются как полиномиальная логистическая регрессия. Хотя они работают одинаково, чтобы предсказать результат. Несколько знакомых примеров логистической регрессии:

- Фильтр спама в электронной почте: спам / нет спама
- Обнаружение мошенничества: транзакция является мошеннической, да / нет
- Опухоль: доброкачественная / злокачественная

Другое название для этого — “выбор моделей”. Эта техника помогает предсказать событие. Например, мы можем найти, какой выбор сделает клиент, когда у него есть разные варианты. Чтобы показать классификацию наблюдений, мы создаем таблицу непредвиденных обстоятельств.

5.5.3 Анализ дискриминантов

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие совокупности (группы). Например, некий исследователь в области образования может захотеть исследовать, какие переменные относят выпускника средней школы к одной из трех категорий: (1) поступающий в колледж, (2) поступающий в профессиональную школу или (3) отказывающийся от дальнейшего образования или профессиональной подготовки. Для этой цели исследователь может собрать данные о различных переменных, связанных с учащимися школы. После выпуска большинство учащихся естественно должно попасть в одну из названных категорий. Затем можно использовать Дискриминантный анализ для определения того, какие переменные дают наилучшее предсказание выбора учащимися дальнейшего пути.

Медик может регистрировать различные переменные, относящиеся к состоянию больного, чтобы выяснить, какие переменные лучше предсказывают, что пациент, вероятно, выздоровел полностью (группа 1), частично (группа 2) или совсем не выздоровел (группа 3). Биолог может записать различные характеристики сходных типов (групп) цветов, чтобы затем провести анализ дискриминантной функции, наилучшим образом разделяющей типы или группы.

В этой методике независимые переменные должны быть метрическими и включать высокий уровень нормальности.

5.5.4 Многомерный анализ переменных (MANOVA)

ANOVA («Дисперсионный анализ») используется для определения наличия или отсутствия статистически значимой разницы между средними значениями трех или более независимых групп. Двумя наиболее распространенными типами дисперсионного анализа являются однофакторный дисперсионный анализ и двухфакторный дисперсионный анализ.

Однофакторный дисперсионный анализ: используется для определения того, как один фактор влияет на переменную отклика.

MANOVA («Многомерный дисперсионный анализ») идентичен ANOVA, за исключением того, что в нем используются две или более переменных отклика. Подобно дисперсионному анализу, он также может быть односторонним или двусторонним.

Дисперсионный анализ также может быть трехфакторным, четырехфакторным и т. д., но они менее распространены.

Односторонний пример MANOVA: Мы хотим знать, как уровень образования (т. е. средняя школа, степень младшего специалиста, степень бакалавра, степень магистра и т. д.) влияет как на годовой доход, так и на сумму долга по студенческому кредиту. В этом случае у нас есть один фактор (уровень образования) и две переменные отклика (годовой доход и задолженность по студенческому кредиту), поэтому нам нужно провести односторонний MANOVA.

Этот метод будет анализировать связь между двумя или более зависимыми метрическими переменными и независимыми переменными. Необходимо определить среднее значение вектора для различных групп. Метрикой является зависимая переменная, а категорической — независимая переменная.

5.5.5 Факторный анализ

Главными целями факторного анализа являются: (1) сокращение числа переменных (редукция данных) и (2) определение структуры взаимосвязей между переменными, т.е. классификация переменных. Поэтому факторный анализ используется или как метод сокращения данных или как метод классификации.

Если у нас есть много переменных для разработки плана исследования, мы можем уменьшить некоторые переменные в различных меньших факторах. В этой методике мы не найдем зависимых переменных. Размер выборки в этом методе должен быть более 50 наблюдений, и в каждом пяти переменных должно быть, по крайней мере, три наблюдения.

5.5.6 Кластерный анализ

Кластерный анализ помогает в подгруппировке объектов или отдельных лиц из больших данных по принципу “like attract like”... С помощью характерного анализа можно упростить объекты на различные наборы и группы. Кластерный анализ помогает в сегментации рынка. Можно выбрать один из трех видов методов кластеризации. Есть:

- Иерархический .
- Неиерархический
- Комбинация обоих методов
- Многомерное масштабирование (MDS)

Этот метод помогает трансформировать суждения потребителя в расстояния в многомерном пространстве. Мы можем распознать размеры продукта и обнаружить его сравнительные оценки, когда у нас нет факторов для сравнения. Мы можем интерпретировать измерения, позволяя респонденту идентифицировать их, а затем исследователь может проанализировать данные, полученные в результате идентификации респондента.

5.5.7 Каноническая корреляция

Этот метод связывает различные зависимые и независимые переменные. Это мощный метод и включает в себя независимые метрические переменные, такие как уровень использования, уровень удовлетворенности и продажи. Можно также использовать не метрические переменные. Среди многих методов многомерного анализа данных, этот метод имеет меньшие ограничения.

5.5.8 Моделирование структурных уравнений

В этой методике необходимо изучить взаимосвязь между различными наборами переменных. Эта методика включает в себя дополнительные методы, такие как подтверждающий факторный анализ, анализ латентных переменных и LISREL. Мы можем использовать эту методику для разработки суммарных шкал и оценки масштабируемых атрибутов.

5.6 Заключение

Многомерный анализ выживаемости является мощным инструментом, который может быть использован для изучения выживаемости пациентов с циррозом печени. Этот метод позволяет определять факторы, влияющие на выживаемость, а также учитывает взаимодействия между ними. Исследование на основе многомерного анализа выживаемости может помочь в определении факторов, которые важны для прогнозирования выживаемости пациентов с циррозом печени и в анализе эффективности новых методов лечения.

5.7 Методика работы и средства

В зависимости от конкретного случая, могут использоваться различные средства анализа данных. Один из самых распространенных инструментов — это программный пакет R, который позволяет проводить многомерный анализ с использованием разных методов.

Чтобы начать работу, нужно импортировать датасет в программу, которую вы будете использовать. Стоит убедиться в том, что датасет имеет правильный формат и соответствующую структуру, иначе анализ может получиться не полным. Обычно данные представлены в виде таблицы, где каждая строка соответствует конкретному наблюдению, а каждый столбец представляет собой конкретный атрибут.

В случае с заданным примером исследования, мы должны проанализировать датасет, чтобы понять, какие атрибуты в нем существенны. Возможно, нам нужно будет выполнить предварительную обработку данных, такую как удаление выбросов, обрезание или заполнение пропущенных значений. После этого мы можем начать проводить многомерный анализ.

Важным шагом процесса является выбор метода анализа данных. Так, для выявления взаимосвязи между факторами и выживаемостью пациентов, можно использовать метод главных компонент (PCA), факторный анализ (ФА) или кластерный анализ. Выбор метода зависит от структуры данных и конечного результата, который нужно получить.

После проведения анализа, мы можем получить графические отчеты, таблицы и другие результаты. Они позволяют визуализировать структуру и связи между переменными в нашем исследовании.

Для анализа датасета, я бы использовал язык программирования Python в среде разработки Jupyter Notebook и некоторые библиотеки для работы с данными, такие как pandas, matplotlib и seaborn.

Либо я бы использовал язык R из-за возможности более просто и красиво работать с графиками, а также из-за обилия пакетов для прикладной статистики.

6 Вывод

Чем данная работа может быть полезна?

После выявления факторов, можно использовать эти данные для повышения качества медицинской помощи и объяснения пациентам потенциального исхода лечения. Возможно, в дальнейшем, будут разработаны более точные алгоритмы предсказания выживаемости пациентов, больных циррозом печени.

7 Источники

- https://scikit-survival.readthedocs.io/en/stable/user_guide/00-introduction.html

- <https://lifelines.readthedocs.io/en/latest/index.html>
- <https://paperswithcode.com/task/survival-analysis>
- <https://www-eio.upc.edu/~pau/cms/rdata/doc/survival/pbc.html>
- <http://statsoft.ru/home/textbook/modules/stsurvan.html#general>
- <https://datascience.eu/mathematics-statistics/multivariate-regression-models/>
- <http://statsoft.ru/home/textbook/modules/stmulreg.html>
- <https://machinelearningmastery.ru/logistic-regression-for-dummies-a-detailed-explanation-9597ff76edf46/>
- <http://statsoft.ru/home/textbook/modules/stdiscan.html>
- <https://www.codecamp.ru/blog/differences-between-anova-ancova-manova-mancova/>
- <https://www.ibm.com/docs/ru/spss-statistics/beta?topic=statistics-multivariate-analysis-variance-manova>
- <http://statsoft.ru/home/textbook/modules/stfacan.html>