

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

**ЖУРНАЛ ПРАКТИКИ**

Студента 2 курса

Гординского Дмитрия Михайловича

Институт №8 «Информационные технологии и прикладная математика»

Кафедра №804 «Теория вероятностей и компьютерное моделирование»

Учебная группа М8О-204Б-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 /

/ 11 июля 2022 г.

**Москва, 2022**

## **1. Место и сроки проведения практики**

Дата начала практики 29 июня 2022 г.

Дата окончания практики 11 июня 2022 г.

Наименование предприятия МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ(НИИУ)

Название структурного подразделения Кафедра 804

## **2. Инструктаж по технике безопасности**

\_\_\_\_\_/\_\_\_\_\_/ 29 июня 2022 г.

## **3. Индивидуальное задание студенту**

1. Разобраться с теорией.
2. Привести пример решения задачи.
3. Написать отчет.

## **4. План выполнения индивидуального задания**

1. Изучить теорию по Моделям выживаемости.
2. Ознакомиться с необходимыми библиотеками для работы с данными и их графическим представлением.
3. Решить задачу по анализу данных с применением методов анализа выживаемости.

Руководитель практики от МАИ: \_\_\_\_\_/\_\_\_\_\_/

\_\_\_\_\_/\_\_\_\_\_/ 29 июня 2022 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 \_\_\_\_\_/\_\_\_\_\_/ 11 июля 2022 г.

# Отчет студента

## Содержание

<b>1</b>	<b>Что такое “Анализ выживаемости”?</b>	<b>3</b>
1.1	Основные понятия . . . . .	3
1.1.1	Функция выживания (Survival function) . . . . .	3
1.1.2	Функция риска (Hazard function) . . . . .	3
1.1.3	Цензурирование (censoring) . . . . .	4
1.1.4	Медиана ожидаемого времени жизни (median number of survival days)	4
1.1.5	Доверительный интервал (confidence interval) . . . . .	4
1.1.6	Усечение (truncation) . . . . .	4
1.1.7	Оценка Каплана — Мейера, оценка Нельсона — Аалена . . . . .	4

# 1 Что такое “Анализ выживаемости”?

*Анализ выживаемости* — набор статистических моделей, благодаря которым можно оценить вероятность наступления того или иного события. Анализ занимается моделированием процессов наступления *интересующих* нас (критических) событий для элементов той или иной совокупности (изначально — «смерти» для элементов совокупности живых существ).

*Интересным* событием может быть что угодно. Это может быть фактическая смерть, рождение, выход на пенсию и т. д.

Название “*survival analysis*” взято из медицины, т.к. цель анализа заключается в изучении продолжительности жизни пациента после приема препарата или других факторов влияния на здоровье.

## 1.1 Основные понятия

### 1.1.1 Функция выживания (Survival function)

Пусть  $T$  — неотрицательная случайная величина, представляющая собой время ожидания до наступления некоторого события. Для простоты будем использовать терминологию анализа выживаемости, называя исследуемое событие «смертью», а время ожидания — временем «выживания»

*Функция выживания* сопоставляет некоторому числу  $t$  вероятность того, что случайная величина  $T$  примет значение, не меньшее  $t$ . Иначе говоря, это вероятность того, что некоторое состояние «проживет» как минимум  $t$  единиц времени:

$$S(t) = \mathbb{P}\{T > t\} = 1 - \mathbb{P}\{T \leq t\}$$

Например, если мы хотим знать, какова вероятность того, что безработный индивид не сможет найти работу в течение полугода после начала поиска, то достаточно рассмотреть функцию выживания для  $t = 6$  месяцев.

### 1.1.2 Функция риска (Hazard function)

*Функцию риска* можно охарактеризовать как вероятность того, что событие произойдет за бесконечно малый интервал времени при условии, что оно не произошло к моменту времени  $t$ .

$$h(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + dt | T \geq t)}{dt}$$

Числитель этого выражения — условная вероятность того, что событие произойдет в интервале  $(t, t + dt)$ , если оно не произошло ранее, а знаменатель — ширина интервала. Разделив одно на другое, получаем интенсивность осуществления события в единицу времени. Устремляя ширину интервала к нулю и переходя к пределу, получаем *мгновенную интенсивность осуществления события*.

Т. к. вышесвязанные функции связаны друг с другом, можно показать, что:

$$S(t) = \exp\left(-\int_0^t h(x)dx\right)$$

Интеграл в фигурных скобках в этом уравнении называют *кумулятивным риском* и обозначают как:

$$H(t) = \int_0^t h(x)dx$$

Можно рассматривать  $H(t)$  как сумму всех рисков при переходе от момента времени 0 к  $t$ .

### 1.1.3 Цензурирование (censoring)

*Цензурирование* — вид неполноты информации, при котором наблюдения не содержат точной длительности изучаемого состояния. Различают цензурирование справа, слева и интервальное:

1. Цензурировано справа — о наблюдаемом состоянии известно лишь, что оно продлилось не менее определенного времени.
2. Цензурировано слева — о состоянии известно лишь, что оно продлилось не более определенного времени.
3. На интервале — известны только границы длительности.

### 1.1.4 Медиана ожидаемого времени жизни (median number of survival days)

Это точка на временной оси, в которой кумулятивная функция выживания равна 0,5.

Другими словами, *медиана* — время, выраженное в месяцах или годах, когда ожидается, что половина пациентов будет жива. Это означает, что шанс выжить после этого времени составляет 50 процентов.

### 1.1.5 Доверительный интервал (confidence interval)

Доверительный интервал — интервал, который покрывает неизвестный параметр с заданной надёжностью. Вероятность, с которой в условиях данного эксперимента полученные экспериментальные данные можно считать надёжными (достоверными), называют доверительной вероятностью или надёжностью. Величина доверительной вероятности определяется характером производимых измерений. Мы будем считать доверительную вероятность равной 95 %.

### 1.1.6 Усечение (truncation)

*Усечением*, или урезанием, называется вид неполноты информации, при котором какая-то область возможных значений длительности оказывается недостаточно представленной в выборке: состояния, длительность которых слишком велика или, наоборот, слишком мала, просто не включаются в анализируемые данные. В нашей задаче мы будем называть их (removed) — пациенты, которые больше не являются частью нашего эксперимента. Если человек умирает или подвергается цензуре, то он попадает в эту категорию.

### 1.1.7 Оценка Каплана — Мейера, оценка Нельсона — Аалена

При отсутствии цензурирования и усечения для оценивания закона распределения вероятностей может использоваться эмпирическая функция распределения, из которой легко получить оценки для других характеристик случайной величины: survival function etc. Но в нашем случае это невозможно, т. к. мы имеем дело с неполнотой данных. Эту проблему решают непараметрические методы оценки.

#### **Оценка Каплана — Мейера**

*Оценка Каплана-Мейера* — это непараметрическая статистика, используемая для оценки функции выживания на основе данных о жизни. В медицинских исследованиях он часто используется для измерения доли пациентов, живущих в течение определенного времени после лечения или постановки диагноза. Например: подсчет

количества времени, которое прожил конкретный пациент после того, как у него был диагностирован рак или началось его лечение.

$$\hat{S}(t) = \prod_{t_j \leq t} \frac{n_j - d_j}{n_j}$$

где

$\hat{S}(t)$  = Вероятность того, что испытуемый жив в момент времени  $t$

$n_j$  = Количество испытуемых, оставшихся в живых непосредственно перед моментом времени  $t_j$

$d_j$  = Количество событий в момент времени  $t_j$

Можем переписать формулу выше так:

$$S(t_j) = S(t_{j-1}) \left(1 - \frac{d_j}{n_j}\right)$$

где

$S(t_j)$  = Вероятность того, что испытуемый жив в момент времени  $t_j$

$n_j$  = Количество испытуемых, оставшихся в живых непосредственно перед моментом времени  $t_j$

$d_j$  = Количество событий в момент времени  $t_j$

$S(0) = 1$

$t_0 = 0$

#### **Оценка Нельсона — Аалена**

Мы можем визуализировать совокупную информацию о выживании, используя функцию риска *Нельсона-Аалена*  $h(t)$ . Функция риска  $h(t)$  дает нам вероятность того, что субъект, находящийся под наблюдением в момент времени  $t$ , имеет интересующее событие (смерть) в это время. Чтобы получить информацию о функции опасности, мы не можем преобразовать оценку Каплана-Мейера. Для этого существует соответствующая непараметрическая оценка кумулятивной функции опасности:

$$\hat{H}(t) = \sum_{t_j \leq t} \frac{d_j}{n_j}$$

где

$\hat{H}(t)$  = Кумулятивная вероятность опасности

$n_j$  = Количество испытуемых, оставшихся в живых непосредственно перед моментом времени  $t_j$

$d_j$  = Количество событий в момент времени  $t_j$