



**С. В. Малов  
И. Ю. Малова  
А. В. Процветкина**

**БАЗОВЫЕ МОДЕЛИ  
БИОСТАТИСТИКИ  
АНАЛИЗ РЕЗУЛЬТАТОВ  
КОГОРТНЫХ ИССЛЕДОВАНИЙ**

**Санкт-Петербург  
2022**

МИНОБРНАУКИ РОССИИ

---

Санкт-Петербургский государственный электротехнический  
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

---

С. В. МАЛОВ    И. Ю. МАЛОВА    А. В. ПРОЦВЕТКИНА

# БАЗОВЫЕ МОДЕЛИ БИОСТАТИСТИКИ АНАЛИЗ РЕЗУЛЬТАТОВ КОГОРТНЫХ ИССЛЕДОВАНИЙ

Учебное пособие



Санкт-Петербург  
Издательство СПбГЭТУ «ЛЭТИ»  
2022

УДК 519.2(07)

ББК В 172я7

М19

**Малов С. В., Малова И. Ю., Процветкина А. В.**

М19 Базовые модели биостатистики: анализ результатов когортных исследований: учеб. пособие. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2022. 72 с.

ISBN 978-5-7629-2966-0

Рассмотрено использование методов математической статистики для анализа результатов когортных исследований.

Предназначено для поддержки дисциплины «Биостатистика» на ФКТИ СПбГЭТУ «ЛЭТИ», будет интересно студентам и аспирантам, обучающимся по программам подготовки специалистов в области информационных технологий.

УДК 519.2(07)

ББК В 172я7

Рецензенты: кафедра биоинформатики и математической биологии СПбАУ РАН им. Ж. И. Алфёрова; д-р техн. наук, проф. Л. В. Уткин (СПбПУ им. Петра Великого).

Утверждено

редакционно-издательским советом университета

в качестве учебного пособия

553817

ISBN 978-5-7629-2966-0



© СПбГЭТУ «ЛЭТИ», 2022

## Оглавление

1. ПОСТАНОВКА СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ . . . .	3
1.1. Различные подходы к организации статистического эксперимента . . . . .	4
1.2. Постановка когортного исследования . . . . .	8
2. МОДЕЛИ СО СЛУЧАЙНЫМИ ПАРАМЕТРАМИ И СМЕШАННЫЕ МОДЕЛИ . . . . .	10
2.1. Регрессионная модель для зависимых наблюдений . . . . .	11
2.2. Модели линейной регрессии со случайными параметрами . . . .	18
2.3. Смешанные модели . . . . .	19
3. МОДЕЛИ ДИСПЕРСИОННОГО АНАЛИЗА СО СЛУЧАЙНЫМИ ФАКТОРАМИ . . . . .	25
3.1. Однофакторный дисперсионный анализ . . . . .	26
3.2. Двухфакторный дисперсионный анализ со случайными факторами . . . . .	31
3.3. Двухфакторный дисперсионный анализ со случайным и неслучайным эффектами . . . . .	35
4. ОБОБЩЕННЫЕ СМЕШАННЫЕ МОДЕЛИ . . . . .	38
4.1. Введение в обобщенные смешанные модели . . . . .	38
4.2. Обобщенные смешанные модели с категориальным случайным фактором . . . . .	43
5. АНАЛИЗ ЛОНГИТУДНЫХ ДАННЫХ . . . . .	46
5.1. Разведочный анализ имеющихся данных . . . . .	49
5.2. Анализ лонгитюдных данных в предположении нормальности	51
5.3. Анализ лонгитюдных данных в общем случае . . . . .	59
6. АНАЛИЗ ЛОНГИТУДНЫХ ДАННЫХ В ПАКЕТЕ R . . . . .	62
Список литературы . . . . .	70

В настоящем издании упор делается на когортный план эксперимента, в основе которого лежит наблюдение за пациентом в течение определенного времени. Рассматриваются модели анализа коррелированных данных. Для анализа данных используются смешанные и обобщенные смешанные модели регрессионного анализа, а также семипараметрические обобщенные модели. В учебном пособии также рассматриваются некоторые компьютерные реализации методов анализа лонгитюдных данных в пакете **R**. Предполагается, что читатель знаком с основными понятиями теории вероятностей и математической статистики, линейной алгебры, математического анализа, теории меры и интеграла.

## 1. ПОСТАНОВКА СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Первым шагом в постановке статистического исследования является выдвижение основной гипотезы, подлежащей проверке. Необходимо, соответственно, также четко определить независимые и зависимые переменные, взаимосвязь между которыми предстоит изучить в рамках исследования. Независимую переменную обычно называют «воздействием» (*exposure*, англ.), и обычно она отражает фактор, которому индивид оказался подвергнут, например уровень облучения в результате аварии на предприятии. Зависимую переменную именуют *исходом* (*outcome*, англ.), и она является собой результат эксперимента, например возникновение осложнений после операции.

Далее следует определить, какой тип исследования наиболее точно подойдет для проверки поставленной гипотезы. Принято выделять 2 подхода к организации исследований: *активный, или экспериментальный* (*experimental, interventional*, англ.) и *пассивный, или наблюдательный* (*observational*, англ.).

Основная цель настоящего издания – рассмотреть методы статистического анализа динамики развития заболеваний. Для изучения изменений исследуемых характеристик и сравнения их распределений наилучшим образом подходит когортный дизайн статистического эксперимента, которому будет уделено особое внимание.

## 1.1. Различные подходы к организации статистического эксперимента

В рамках активного подхода исследователь подвергает объект изучения некоторому «вмешательству», после чего наблюдает за ним с целью выявления результатов. Под «вмешательством» понимают применение некоторого нового вида лечения или препарата. К данному классу относят обычно эксперименты на животных и клинические испытания.

*Клиническим испытанием* называют исследование выбранной группы людей на предмет будущего влияния и эффективности «вмешательства». В любом клиническом испытании формируются *контрольная* и *экспериментальная* группы. В ходе исследования только к экспериментальной группе применяется «вмешательство». Контрольная группа должна быть максимально похожа на экспериментальную, чтобы различия в результатах эксперимента могли быть разумно обоснованы применением «вмешательства». Для исключения субъективности (исследователя или участника) в выборе группы распределение в контрольную и экспериментальную группы обычно осуществляют случайно (*рандомизация*). Также при проведении испытания желательно, чтобы участники не были информированы о том, к какой из групп они принадлежат, в целях исключения влияния психологического фактора. Такой подход называется слепым (*blind*, англ.). Чтобы исключить предвзятость исследователя, применяется двойной слепой подход (*double-blind*, англ.) – ни участник, ни взаимодействующий с ним исследователь не информированы о принадлежности участника к группе. В силу этических соображений данный вид исследований жестко ограничен в используемых методах.

Ввиду этических ограничений и затрат на проведение экспериментальных исследований большинство исследований являются наблюдательными. Основные типы наблюдательных исследований: *разовый скрининг* (*cross-sectional*, англ.), *когортное исследование* (*cohort*, англ.) и *случай-контроль* (*case-control*, англ.).

В плане разового скрининга данные о «воздействии» и исходе для группы людей получают одновременно. Данный подход обычно направлен на исследование распространенности того или иного признака в выбранной популяции в текущий момент времени вне зависимости от продолжительности обладания этим признаком. Например, исследование распространенности диабета в рамках разового скрининга будет включать в себя здоровых людей, а также как больных диабетом уже десятки лет, так и

недавно диагностированных больных. Обычно данный дизайн подразумевает сравнение распространенности признака в подгруппах с различным уровнем «воздействия».

В когортном исследовании группа индивидов, обладающих общими характеристиками (*когорта*), наблюдается в течение определенного периода времени с целью изучения динамики изменений изучаемой характеристики, сопутствующих факторов возникновения или развития заболевания (*факторов риска*) и их влияния на исход (обычно возникновение или переход болезни в критическую стадию). Факторы риска могут быть постоянными (например, возраст на момент начала исследования) или зависящими от времени (например, уровень антител к исследуемому заболеванию). Довольно часто целью когортного исследования является выявление причины наступления исхода, поэтому до начала исследования формулируется возможная причина («воздействие»). На основе уровней «воздействия» внутри когорты выделяются группы, в которых по окончании исследования сравнивается распределение исхода (заболеваемость). Если выделяется две подгруппы, обычно одну называют *подверженной риску* (*exposed group*, англ.), а другую, соответственно, *не подверженной риску* (*unexposed group*, англ.). Например, в когорте работников предприятия можно выделить группы курящих и некурящих сотрудников. Конечно же, в рамках исследования можно формулировать и более одной потенциальной причины. В отличие от клинических испытаний рандомизация обычно не практикуется при проведении когортных исследований.

Обычно в исследованиях типа случай-контроль сначала формируется *экспериментальная* группа людей, например обладающих заболеванием, – «случаи». Затем набирается *контрольная* группа полностью здоровых людей. Исследователь сравнивает результаты «воздействия» между двумя группами: контрольная группа несет информацию о распределении изучаемого признака у здоровых людей той же популяции. Если удастся найти значимые различия распределений между экспериментальной и контрольной группами, то говорят, что болезнь *ассоциирована* с «воздействием». При постановке экспериментов данного типа необходимо аккуратно подходить к формированию контрольной группы, чтобы, по возможности, исключить ее отличия от экспериментальной по всем доступным признакам, за исключением «воздействия».

Особенности организации каждого типа исследований всегда ограничивают исследователя в получаемых результатах. Не каждое исследование позволяет выявить причинно-следственные связи между «воздействием» и