Положительные отзывы на книгу «Анализ поведенческих данных на R и Python»

«В отличие от некоторых книг по науке о данных, в которых авторы стремятся научить своих читателей новым техническим приемам, цель Флорана – иная и более глубокая. Он стремится научить нас мудрости, ориентированной на данные: как строить подробное и тонкое понимание данных, содержащих следы человеческого поведения».

– *Стив Вендель*,
руководитель отдела бихевиористики, Morningstar

«Книга “Анализ поведенческих данных” поможет вам разбираться в данных, даже если вы не можете проводить контролируемые эксперименты».

– *Колин Макфарланд*,
директор платформы экспериментирования, Netflix

«Мы переполнены данными, и эта книга является давно востребованным ресурсом, который направляет практиков в том, как использовать эти данные для строительства достоверных причинно-следственных моделей, которые предсказывают и объясняют поведения в реальном мире».

– *Дэвид Льюис*, президент научно-иссследовательского
института BEworks в компании BEworks

«Для всех, кто хочет применять бихевиористику в качестве проводника в принятии деловых решений, эта книга представляет собой ценное подробное введение в принципы эффективного использования причинно-следственных диаграмм во время экспериментирования и в поведенческом анализе».

– *Мэтт Райт*, директор по бихевиористике, WiderFunnel

«Часть того, что делает бихевиористику работоспособной, заключена в бесшовном сочетании количественных и качественных выводов в поддержку нашего понимания причин, почему люди делают то, что они делают. Эта книга поможет любому человеку, обладающему несколькими базовыми навыками работы с данными, принимать осмысленное участие в этом процессе бихевиористики».

– *Мэтт Уоллерт*,
руководитель отдела бихевиористики в frog,
автор книги «Начало в конце: как создавать продукты,
которые создают изменения»

Содержание

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Скачивание исходного кода примеров

Скачать файлы с дополнительной информацией для книг издательства «ДМК Пресс» можно на сайте www.dmkpress.com на странице с описанием соответствующей книги.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и O'Reilly очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Предисловие

Статистика является предметом удивительно многих применений и инструментом удивительно немногих эффективных практиков.

– Брэдли Эфрон и Р. Дж. Тибширани, «Введение в бутстрап» (1993)

Добро пожаловать в «Анализ поведенческих данных на R и Python»! Высказывание о том, что мы живем в век данных, уже стало банальным. Инженеры теперь регулярно используют сенсорные данные на машинах и турбинах, чтобы предсказывать время, когда они выйдут из строя, и проводят превентивное техническое обслуживание. Аналогичным образом маркетологи используют массивы данных, начиная с вашей демографической информации и заканчивая вашими прошлыми покупками, чтобы определять вид объявления, которое вам следует показывать, и время его показа. Как говорится, «данные – это новая нефть», а алгоритмы – это новый двигатель внутреннего сгорания, двигающий нашу экономику вперед.

В большинстве книг по аналитике, машинному обучению и науке о данных авторы неявно предполагают, что задачи, которые пытаются решать инженеры и маркетологи, могут решаться с помощью одних и тех же подходов и инструментов. Разумеется, переменные имеют разные имена, и необходимо приобретать некоторые знания, относящиеся к конкретной области, но кластеризация *k*-средних – это кластеризация *k*-средних, независимо от того, кластеризуете вы данные о турбинах или сообщения в социальных сетях. Принимая на вооружение инструменты машинного обучения в таком ключе, компании нередко могли точно предсказывать поведения, но ценой более глубокого и богатого понимания того, что на самом деле происходит. Это привело к критике моделей науки о данных как «черных ящиков».

Вместо того чтобы стремиться к точным, но непрозрачным предсказаниям, эта книга стремится ответить на вопрос «Что движет поведением?». Если мы решим отправить электронное письмо потенциальным клиентам, то купят ли они подписку на нашу службу в результате отправки этого электронного письма? И какие группы клиентов должны получать это электронное письмо? Склонны ли пожилые клиенты покупать разные товары, потому что они старше? Как влияет опыт клиентов на лояльность и удержание клиентов? Изменив нашу точку зрения с предсказания поведений на их объяснение и измерение причин, мы сможем снять проклятие «корреляция не есть каузация», которое мешало поколениям аналитиков быть уверенными в результатах своих моделей.

Этот сдвиг не будет связан с введением новых аналитических инструментов: мы будем использовать только два инструмента анализа данных: старую добрую линейную регрессию и ее логистического собрата. Указанные две

модели по своей сути читаются намного легче, чем другие типы моделей. Определенно, это нередко происходит ценой более низкой предсказательной точности (т. е. они допускают все больше и больше ошибок в предсказании), но здесь для нашей цели измерения взаимосвязей между переменными это не имеет значения.

Вместо этого мы потратим много времени на то, чтобы научиться разбираться в данных. В своей роли специалиста, проводящего собеседование по науке о данных, я повидал немало кандидатов, которые были способны использовать сложные алгоритмы машинного обучения, но не развили в себе сильное чувство данных: у них мало интуиции относительно того, что, собственно, происходит в их данных, кроме того что им говорят их алгоритмы.

Я твердо убежден, что вы можете развить эту интуицию и попутно повысить ценность и результаты ваших аналитических проектов – нередко значительно, – приняв следующие меры:

- бихевиористский менталитет, который взирает на данные не как на самоцель, а как на линзу для изучения психологии и поведений людей;
- инструментарий причинно-следственной (каузальной) аналитики, который позволяет нам уверенно утверждать, что одна вещь обусловливает другую, и определять силу этой взаимосвязи.

Хотя каждая из них может приносить большие выгоды сама по себе, я считаю, что они являются естественными дополнениями, которые лучше всего использовать вместе. Учитывая, что словосочетание «бихевиористский менталитет с использованием инструментария причинно-следственной аналитики» трудно выговорить, вместо него я будут называть его причинно-поведенческим подходом, или каркасом. Указанный каркас имеет дополнительную выгоду: он в равной степени применим к экспериментальным и историческим данным, используя при этом различия между ними. Это контрастирует с традиционной аналитикой, которая манипулирует ими с помощью совершенно других инструментов (например, ANOVA и T-тест для экспериментальных данных), и наукой о данных, которая не трактует экспериментальные данные отлично от исторических данных.

Для кого эта книга предназначена

Если вы анализируете данные в бизнесе на R или Python, то эта книга для вас. Я использую слово «бизнес» в широком смысле для обозначения любой коммерческой, некоммерческой или правительственной организации, где важны правильные идеи и практические выводы, которые движут действиями.

С точки зрения математики и статистики, не имеет значения, кем вы являетесь: деловым аналитиком, строящим ежемесячные прогнозы, исследователем опыта пользователей (UX), изучающим поведения на основе кликабельности, или исследователем данных, строящим модели машинного обучения. У этой книги есть одно фундаментальное условие: вы должны быть хотя бы немного знакомы с линейной и логистической регрессией. Если вы понимаете регрессию, то вы сможете проследить за аргументами этой кни-

ги и извлечь из нее большую пользу. С другой стороны, я убежден, что даже опытные исследователи данных с докторскими степенями в области статистики или компьютерных наук найдут этот материал новым и полезным, при условии что они еще не являются специалистами в области поведенческой или причинно-следственной аналитики.

С точки зрения подготовленности в качестве программиста, вы должны уметь читать и писать исходный код на R или Python, в идеале на том и другом. Я не буду показывать вам, как определять функцию или как манипулировать структурами данных, такими как кадры данных в pandas. Уже есть отличные книги, которые справляются с этим лучше, чем я, например «Python для анализа данных» Уэса Маккинни (*Python for Data Analysis*, Wes McKinney, O'Reilly)[1] и «R для науки о данных» Гарретта Гролемунда и Хэдли Уикхэма (*R for Data Science*, Garrett Grolemund and Hadley Wickham, O'Reilly)[2]. Если вы читали какую-либо из этих книг, посещали вводные занятия или использовали хотя бы один из двух языков на работе, то здесь вы будете подготовлены к излагаемому материалу. Точно так же я обычно не буду представлять и обсуждать исходный код, используемый для создания многочисленных рисунков в книге, хотя он будет размещен в репозитории книги на GitHub[3].

Для кого эта книга не предназначена

Если вы работаете в академических кругах или в области, которая требует от вас соблюдения академических норм (например, фармацевтические испытания), то эта книга все еще может представлять для вас интерес, но рецепты, которые я описываю, могут вызывать у вас проблемы с вашим консультантом/редактором/менеджером.

Эта книга не является обзором традиционных методов анализа поведенческих данных, таких как T-тест или ANOVA. Мне еще не приходилось сталкиваться с ситуацией, когда регрессия была менее эффективной, чем эти методы для предоставления ответа на деловой вопрос, поэтому я намеренно ограничиваю эту книгу линейной и логистической регрессией. Если вы хотите изучать другие методы, то вам придется поискать в другом месте (например, в книге «Практическое машинное обучение с помощью Scikit-Learn, Keras и TensorFlow» Орельена Жерона (*Hands-On Machine Learning with Scikit-Learn*, Keras, and TensorFlow, Aurélien Géron, O'Reilly)[4] в отношении алгоритмов машинного обучения).

Понимание и изменение поведений в прикладных условиях требует как анализа данных, так и качественных навыков. В этой книге основное внимание уделяется первому, в первую очередь по соображениям пространства. В дополнение к этому уже есть отличные книги, которые охватывают послед-

---

[1] См. https://www.oreilly.com/library/view/python-for-data/9781491957653/.

[2] См. https://www.oreilly.com/library/view/r-for-data/9781491910382/.

[3] См. https://oreil.ly/BehavioralDataAnalysisCh8.

[4] См. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/.

нее, такие как «Толчок в верном направлении: совершенствование решений о здоровье, богатстве и счастье» Ричарда Талера и Касса Санштейна (*Nudge: Improving Decisions About Health, Wealth, and Happiness*, Richard Thaler and Cass Sunstein, Penguin) и «Дизайн для изменения поведения: применение психологии и поведенческой экономики» Стивена Венделя (*Designing for Behavior Change: Applying Psychology and Behavioral Economics*, Stephen Wendel, O'Reilly)[1]. Тем не менее я дам введение в концепции бихевиористики, чтобы вы могли применять инструменты из этой книги, даже если вы – новичок в данной области.

Наконец, если вы – абсолютный новичок в анализе данных на R или Python, то эта книга не для вас. Я рекомендую начать с нескольких отличных введений, таких как те, которые упомянуты в этом разделе.

Исходный код на R и Python

Почему именно R и Python? Почему бы не выбрать один язык из перечисленных? Дебаты по теме «R против Python» все еще оживленны и продолжаются в интернете. Этот вопрос, по моему скромному мнению, в сущности, тоже не имеет значения. Реальность такова, что вам придется применять любой язык, который используется в вашей организации, и точка. Однажды я работал в медицинской компании, где по техническим и нормативным причинам доминирующим языком был SAS. Я регулярно использовал R и Python для своих собственных аналитических расчетов, но так как я не мог избежать работы с унаследованным исходным кодом SAS, в течение первого месяца работы я заставил себя усвоить SAS настолько, насколько мне было нужно. Если вы не проведете всю свою карьеру в компании, в которой не используется R или Python, то вы, скорее всего, в любом случае подхватите некоторые основы и того, и другого, так что с таким же успехом вы могли бы постичь двуязычие. Я еще не встречал никого, кто заявил бы, что «обучение чтению исходного кода на [другом языке] было пустой тратой моего времени».

Если исходить из допущения, что вам повезло работать в организации, в которой используется и то, и другое, с каким языком вам следует работать? Я думаю, что это на самом деле зависит от вашего контекста и задач, которые вам приходится выполнять. Например, я лично предпочитаю выполнять разведывательный анализ данных (EDA) на R, но нахожу, что Python намного проще использовать для создания веб-страниц. Советую выбирать, исходя из специфики вашей работы и опираясь на актуальную информацию: оба языка постоянно совершенствуются, и то, что было верно для предыдущей версии R или Python, может оказаться неверным для текущей версии. Например, Python становится гораздо более дружественной средой для EDA, чем когда-либо. Лучше потратить свою энергию на изучение обоих языков, чем на изучение форумов, посвященных выбору лучшего из двух.

---

[1] См. https://www.oreilly.com/library/view/designing-for-behavior/9781492056027/.

Среды исходного кода

В начале каждой главы я буду называть пакеты R и Python, которые необходимо загружать специально для каждой отдельной главы. В дополнение к этому я также буду использовать несколько стандартных пакетов по всей книге; во избежание повторов они называются только здесь (они уже включены во все скрипты в репозитории на GitHub). Вы всегда должны начинать свой исходный код с них, а также с нескольких параметрических настроек:

```
## R
library(tidyverse)
library(boot)       #Требуется для бутстрап-симуляций
library(rstudioapi) #Для загрузки данных из локальной папки
library(ggpubr)     #Для генерирования мультиграфиков

# Задание начального значения случайного числа
# будет обеспечивать воспроизводимость случайных чисел
set.seed(1234)
# Я лично нахожу используемую по умолчанию научную числовую нотацию
# (т. е. с экспонентами) менее удобной для чтения в распечатках, поэтому я ее отменяю
options(scipen=10)

## Python
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
from statsmodels.formula.api import ols
import matplotlib.pyplot as plt # Для графики
import seaborn as sns           # Для графики
```

Условные обозначения в исходном коде

Я использую R в RStudio. R 4.0 был запущен, когда я писал эту книгу, и я принял эту версию за основу, чтобы сделать книгу как можно более актуальной.

Исходный код R пишется шрифтом, специально предназначенным для исходного кода, с комментарием, указывающим используемый язык, вот так:

```
## R
> x <- 3
> x
[1] 3
```

Я использую Python в среде интерактивной разработки Spyder дистрибутива Anaconda. Обсуждение темы «Python 2.0 против 3.0», надеюсь, уже позади (по меньшей мере, в отношении нового исходного кода; унаследованный исходный код – это уже другая история), и я буду использовать Python 3.7. Условные обозначения, принятые для исходного кода Python, несколько похожи на условные обозначения для R:

```
## Python
In [1]: x = 3
In [2]: x
Out[2]: 3
```

Мы часто будем смотреть на результаты регрессий. Они бывают довольно многословными, с большим объемом диагностики, которая не имеет отношения к аргументам этой книги. Вы не должны пренебрегать ими в реальной жизни, но данный вопрос лучше освещен в других книгах. Поэтому я буду сокращать результат следующим образом:

```
## R
> model1 <- lm(icecream_sales ~ temps, data=stand_dat)
> summary(model1)

...
Coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4519.055    454.566  -9.941   <2e-16 ***
temps        1145.320      7.826 146.348   <2e-16 ***
...
```

```
## Python
model1 = ols("icecream_sales ~ temps", data=stand_data_df)
print(model1.fit().summary())

...
                 coef   std err         t     P>|t|      [0.025         0.975]
------------------------------------------------------------------------------
Intercept  -4519.0554   454.566    -9.941     0.000    -5410.439     -3627.672
Temps       1145.3197     7.826   146.348     0.000     1129.973      1160.666
...
```

Программирование в функциональном стиле

Один из шагов перехода от начинающего программиста к программисту среднего уровня состоит в том, чтобы перестать писать скрипты, в которых ваш исходный код представляет собой просто длинную последовательность инструкций, и вместо этого структурировать свой исходный код в функции. В этой книге мы будем писать и многократно использовать функции в разных главах, наподобие приведенных ниже, для строительства бутстраповских[1] интервалов уверенности[2]:

---

[1] Термин «бутстрап» (bootstrap) дословно означает «вытягивание себя за шнурки ботинок». Неплохой аналогией является история барона Мюнхгаузена, который вытянул себя вместе с лошадью из болота за волосы. – *Прим. перев.*

[2] Указанный термин (confidence interval), обозначающий вычисляемый из наблюдаемых данных диапазон, ограниченный нижним и верхним пределами, переведен в книге именно как интервал уверенности, поскольку речь идет об уверенности (confidence) исследователя в своих данных, а не о доверии к ним (trust), а это, как говорят в Одессе, две большие разницы. – *Прим. перев.*

```
## R
boot_CI_fun <- function(dat, metric_fun, B=20, conf.level=0.9){

  boot_vec <- sapply(1:B, function(x){
    cat("итерация бутстрапа ", x, "\n")
    metric_fun(slice_sample(dat, n = nrow(dat), replace = TRUE))})
  boot_vec <- sort(boot_vec, decreasing = FALSE)
  offset = round(B * (1 - conf.level) / 2)
  CI <- c(boot_vec[offset], boot_vec[B+1-offset])
  return(CI)
}
```

```
## Python
def boot_CI_fun(dat_df, metric_fun, B = 20, conf_level = 9/10):

  coeff_boot = []
  # Вычислить коэффициент, представляющий интерес для симуляции
  for b in range(B):
      print("Номер итерации " + str(b) + "\n")
      boot_df = dat_df.groupby("rep_ID").sample(n=1200, replace=True)
      coeff = metric_fun(boot_df)
      coeff_boot.append(coeff)

  # Извлечь интервал уверенности
  coeff_boot.sort()
  offset = round(B * (1 - conf_level) / 2)
  CI = [coeff_boot[offset], coeff_boot[-(offset+1)]]

  return CI
```

Функции также имеют добавочное преимущество в лимитировании остатков непонимания: даже если вы не понимаете, как работают приведенные выше функции, вы все равно можете считать само собой разумеющимся, что они возвращают интервалы уверенности, и следовать остальным рассуждениям, откладывая более глубокое погружение в их исходный код на потом.

Использование примеров исходного кода

Дополнительные материалы (примеры исходного кода и т. д.) доступны для скачивания по адресу https://oreil.ly/BehavioralDataAnalysis.

Адаптированный вариант примеров в виде электронного архива вы можете скачать со страницы книги на веб-сайте https://dmkpress.com/.

Навигация по книге

Стержневая интуитивная мысль книги состоит в том, что эффективный анализ данных основывается на постоянном взаимодействии между тремя компонентами:

- фактическими поведениями в реальном мире и связанными с ними психологическими явлениями, такими как намерения, мысли и эмоции;
- причинно-следственной аналитикой и в особенности причинно-следственными диаграммами;
- данными.

Книга разделена на пять частей:

часть I «Понимание поведений».
Эта часть закладывает основу для причинно-поведенческого каркаса и взаимосвязей между поведениями, причинно-следственным рассуждением и данными;

часть II «Причинно-следственные диаграммы и распутывание».
В этой части вводится понятие спутывания и объясняется, каким образом причинно-следственные диаграммы позволяют нам распутывать наши аналитические расчеты на данных;

часть III «Устойчивый анализ данных».
Здесь мы занимаемся разведкой инструментов для работы с пропущенными данными и знакомим с бутстраповскими симуляциями, поскольку в остальной части книги мы будем широко опираться на бутстраповские интервалы уверенности.
Данные, которые малы по объему, неполные или имеют неправильную форму (например, с несколькими пиками или выбросами), не являются новой проблемой, но она бывает особенно острой с поведенческими данными;

часть IV «Дизайн и анализ экспериментов».
В этой части мы обсудим вопросы дизайна и анализа экспериментов;

часть V «Расширенные инструменты анализа поведенческих данных».
Наконец, мы сводим все вместе, чтобы разведать модерацию, опосредование и инструментальные переменные.

Различные части книги в некоторой степени основаны друг на друге, и поэтому я рекомендую читать их по порядку, по меньшей мере при вашем первом подходе к книге.

Условные обозначения в книге

В книге используются следующие типографические условные обозначения.

Курсивный шрифт
Обозначает новые термины, URL-адреса, адреса электронной почты, имена файлов и расширения файлов.

`Моноширинный шрифт`
Используется для листингов программ, а также внутри абзацев для ссылки на элементы программ, такие как переменные или имена функций, базы данных, типы данных, переменные среды, инструкции и ключевые слова.

`Жирный моноширинный шрифт`
Показывает команды либо другой текст, который должен быть набран пользователем.

`Моноширинный шрифт курсивом`
Показывает текст, который должен быть заменен значениями, передаваемыми пользователем, либо значениями, определяемыми по контексту.


Этот элемент обозначает общее замечание.


Данный элемент обозначает предупреждение или предостережение.

Благодарности

Авторы часто благодарят своих супругов за терпение и называют особенно проницательных рецензентов. Мне посчастливилось иметь и то, и другое в одном человеке. Я не думаю, что кто-либо другой осмелился бы или сумел бы так много раз отправлять меня обратно за «чертежную доску», и по этой причине данная книга стала намного лучше. Поэтому моя первая благодарность – моему партнеру по жизни и менталитету.

Несколько моих коллег и соратников – ученых-бихевиористов – были достаточно великодушны, чтобы посвятить свое время чтению и комментариям к более раннему черновику. Данная книга стала от этого только лучше. Спасибо (в обратном алфавитном порядке) Джин Утке, Джессике Якубоуски, Чинмайе Гупте и Федре Дайфе!

Особая благодарность Бетани Винкель за ее помощь в написании.

Теперь я съеживаюсь при воспоминании о том, насколько грубыми и запутанными были самые первые наброски. Мой редактор по разработке и технические рецензенты терпеливо подталкивали меня к тому, где эта книга сейчас находится, делясь своим богатым опытом и знаниями. Спасибо вам, Гэри О'Брайен, и спасибо вам, Сюань Инь, Шеннон Уайт, Джейсон Стэнли, Мэтт Лемей и Андреас Кальтенбруннер.

Об авторе

Флоран Бюиссон – поведенческий экономист с 10-летним опытом работы в бизнесе, аналитике и бихевиористике. Еще недавно он основал и в течение четырех лет возглавлял научную группу по бихевиористике в страховой компании Allstate.

Ранее он работал во французской консалтинговой фирме по стратегиям, где использовал экономику и анализ данных для ответа на сложные вопросы эконометрии, например для построения индекса, измеряющего стабильность сельскохозяйственной политики в развивающихся странах от имени Продовольственной и сельскохозяйственной организации ООН. Он также работал в области специализированной медицинской аналитики, анализируя поведение пациентов с тяжелыми заболеваниями.

Флоран публикует научные статьи в таких журналах, как рецензируемый журнал Journal of Real Estate Research, посвященный исследованиям в сфере недвижимости. Он имеет степень магистра эконометрии, а также степень доктора философии в области поведенческой экономики в Университете Сорбонны в Париже.

Об иллюстрации на обложке (колофон)

На обложке книги «Анализ поведенческих данных на R и Python» изображена южноамериканская гремучая змея (Crotalus durissus). Этот вид очень ядовитой гадюки обитает в районах по всей Южной Америке, за исключением высоких Анд и крайнего юга. Его также можно найти на нескольких Карибских островах.

Эти гремучие змеи изменчивы по внешнему виду, как правило, с бледным подбрюшьем и более темно-коричневыми ромбовидными формами или полосами на спине, выделяющимися на более бледном фоне. Они питаются как грызунами, так и ящерицами. Взрослые особи могут вырастать до 6 футов в длину, а в неволе жить до 20 лет. Они размножаются сезонно, и самки рожают до 14 живых детенышей одновременно.

По оценкам, в Северной и Южной Америке от укуса этой змеи умирают около 400 человек в год, а укус южноамериканской гремучей змеи, как известно, особенно смертоносен. Ее яд содержит четыре основных токсина: кротоксин, конвульсин, гироксин и кротамин, – которые змея использует для захвата и переваривания своей добычи.

Гремучие змеи часто используют свой загадочный камуфляж в качестве первой защиты и остаются неподвижными при приближении более крупного животного; в результате этой стратегии иногда случаются укусы людей, потому что они подходят слишком близко к змее или даже наступают на нее. Еще одна защита является источником их общепринятого названия: уникальная предупреждающая функция «погремушек» на их хвостах. Они состоят из кератиновых чешуек с несколькими рыхлыми слоями, и когда змея использует набор уникальных мышц хвоста для вибрации своего хвоста, сухие слои ударяются друг о друга и издают характерный звук. Каждый раз, когда змея сбрасывает кожу, добавляется набор погремушек, что делает число сегментов одним из потенциальных индикаторов (наряду с размером и длиной) возраста змеи.

Южноамериканская гремучая змея занесена Международным союзом по охране природы IUCN в список животных, вызывающих наименьшее беспокойство. Многие животные на обложках издательства O'Reilly находятся под угрозой исчезновения, и все они важны для мира.

Цветная иллюстрация на обложке выполнена Карен Монтгомери на основе черно-белой гравюры из Малой энциклопедии Мейерса.

Часть I

ПОНИМАНИЕ ПОВЕДЕНИЙ

В этой первой части книги дается объяснение причины, почему анализ поведенческих данных требует нового подхода.

В главе 1 будет описан этот новый подход – причинно-поведенческий каркас анализа данных. Мы рассмотрим конкретный пример, показывающий, как даже самые простые аналитические расчеты на данных бывают сорваны присутствием спутывающего фактора. Решение этой проблемы в лучшем случае осложнено, а в худшем – невозможно при использовании традиционных подходов, но новый каркас обеспечивает простой процесс.

В главе 2 будет продолжено изучение особенностей поведенческих данных, обеспечивая при этом осторожное введение в бихевиористику и процесс обеспечения того, чтобы наши данные адекватно отражали соответствующие реально существующие поведения.