

Норман Р. Дрейпер

Прикладной регрессионный анализ

Книга 1

**Москва
«Книга по Требованию»**

УДК 51
ББК 22.1
Н83

Н83 **Норман Р. Дрейпер**
Прикладной регрессионный анализ. Книга 1 / Норман Р. Дрейпер – М.: Книга по Требованию, 2021. – 364 с.

ISBN 978-5-458-25571-4

Работа американских ученых посвящена регрессионному анализу, применяемому во всех отраслях народного хозяйства и научных исследованиях. Второе издание (1-е изд. перевода 1973 г. вышло в одной книге) значительно переработано и дополнено новыми алгоритмами и сравнением их достоинств. Книга 1 содержит классическое описание модели линейной регрессии, включая описание алгоритмов для ЭВМ. В книге 2 приводится описание модели, нелинейной по параметрам регрессии, обширная библиография и приложения. Для специалистов статистиков, экономистов, социологов, научных работников.

ISBN 978-5-458-25571-4

© Издание на русском языке, оформление
«YOYO Media», 2021
© Издание на русском языке, оцифровка,
«Книга по Требованию», 2021

Эта книга является репринтом оригинала, который мы создали специально для Вас, используя запатентованные технологии производства репринтных книг и печати по требованию.

Сначала мы отсканировали каждую страницу оригинала этой редкой книги на профессиональном оборудовании. Затем с помощью специально разработанных программ мы произвели очистку изображения от пятен, клякс, перегибов и попытались отбелить и выровнять каждую страницу книги. К сожалению, некоторые страницы нельзя вернуть в изначальное состояние, и если их было трудно читать в оригинале, то даже при цифровой реставрации их невозможно улучшить.

Разумеется, автоматизированная программная обработка репринтных книг – не самое лучшее решение для восстановления текста в его первоизданном виде, однако, наша цель – вернуть читателю точную копию книги, которой может быть несколько веков.

Поэтому мы предупреждаем о возможных погрешностях восстановленного репринтного издания. В издании могут отсутствовать одна или несколько страниц текста, могут встретиться невыводимые пятна и кляксы, надписи на полях или подчеркивания в тексте, нечитаемые фрагменты текста или загибы страниц. Покупать или не покупать подобные издания – решать Вам, мы же делаем все возможное, чтобы редкие и ценные книги, еще недавно утраченные и несправедливо забытые, вновь стали доступными для всех читателей.

известна и подробно изложена. Тем не менее существовал отчетливо выраженный временной лаг: практические приложения резко отставали.

В 20-е годы сложилось новое направление в экономике — эконометрия. Она взяла на вооружение регрессионные методы и весьма способствовала их распространению. Другой толчок произошел в связи с развитием способов измерения психических свойств личности, имевших большое значение не только для психологии, но и для тесно связанных с нею педагогики, социологии и отчасти медицины. Лишь вторая мировая война и особенно послевоенное время привели к широчайшему внедрению регрессии во все области научных исследований, экономического анализа и промышленного производства. Возник процесс, результаты которого имеют уже экологическое значение. События такого масштаба не могут проходить ни с того ни с сего. В данном случае решающую роль сыграла вычислительная техника. Появление в 50-е годы массового производства ЭВМ привело к регрессионному буму.

Классический регрессионный анализ опирается на некоторую систему постулатов в основном статистического характера. Эти постулаты гласят, что регрессия представляет собой линейную комбинацию некоторых линейно независимых базисных функций от факторов с неизвестными коэффициентами (параметрами). Факторы являются детерминированными. То же справедливо и для параметров. Что же касается откликов (измеряемых зависимых переменных), то считается что это равноточные (с одинаковой дисперсией) некоррелированные случайные величины. Кроме того, предполагается, что это нормально-распределенные случайные величины. И наконец, принимается, что все переменные измеряются в непрерывных шкалах. Такая основа позволяла благополучно довести до числа процесс получения оценок регрессионных коэффициентов и осуществить проверки основных статистических гипотез об уравнении регрессии, его коэффициентах и прогнозируемых значениях отклика.

Заметим кстати, что постулат о равноточности и некоррелированности откликов не является слишком жестким. Если отклики не равноточны и коррелированы, то вычисления коэффициентов регрессии практически не усложняются. Саму процедуру в этом случае называют взвешенным методом наименьших квадратов. По существу, это означает, что указанный постулат можно заменить на более общий, когда предполагается, что априори с точностью до сомножителя известна дисперсионная матрица измеряемых откликов.

Без всяких на то оснований мы до сих пор считали, что набор независимых переменных (факторов) задан однозначно, что все существенные переменные в модели присутствуют и что никаких альтернативных способов выбора факторов нет. Все это, конечно, не так. Выбор переменных, тесно связанный с выбором модели объекта, представляет собой одну из извечных и наисложнейших проблем. Никаких стандартных рецептов здесь нет. Основной аппарат — преобразования. Причем преобразования могут быть содержательными и формальными. Понятно, что содержательные всегда лучше, но не всегда

доступнее. Поэтому чаще встречается такой случай, когда сначала находятся удачные формальные преобразования, а затем выискивается их интерпретация, подбирается физический смысл. Удача на этом пути — всегда событие в науке. Для поиска формальных преобразований разработано множество статистических моделей. Прежде всего — это модели факторного и дискриминантного анализов. Они опираются на линейные преобразования факторного пространства, которые позволяют находить такие новые координаты, что обеспечивается выполнение того или иного условия оптимальности. Разновидностей и модификаций методов подобного рода столь много, что даже перечислить их нет никакой возможности. Постепенно выяснилось, что ЭВМ допускает отказ от жесткой модели объекта исследования и подбор в ходе обработки данных некоторой «наилучшей» модели. После публикаций М. Эфраимсона, известного американского статистика, разработчика первых машинных алгоритмов, предназначенных для решения регрессионных задач, в конце 50-х годов такой новый подход был взят на вооружение, и уже к середине 60-х годов появился целый набор методов, опирающихся на идею последовательного построения подходящих моделей.

Обычная процедура регрессионного анализа исходит из предпосылки, что все нужные данные для построения модели уже собраны. Но ведь на самом деле данные всегда в процессе сбора, их всегда мало. Можно ли обрабатывать данные, которые еще не собраны до конца? Конечно, можно. Для этого разработан целый веер приемов, в основе которых лежит идея последовательного анализа, принадлежащая А. Вальду. Последовательный анализ предполагает, что к некоторому исходному массиву данных добавляется по одной строке и после каждого такого добавления оценки уточняются в свете новой информации. Иногда такой частый пересчет не оправдан и он осуществляется блоками, шагами, после нескольких новых строк. Но можно вообразить и такой вариант, когда одновременно с добавлением новых строк могут вычеркиваться старые, причем не обязательно, чтобы их числа совпадали. Здесь мы оказываемся в рамках моделей стохастической аппроксимации, которые, применительно к нашему случаю, называют еще и текущим регрессионным анализом. В наше время он находит применение в алгоритмах управления некоторыми производственными процессами.

Приемы классического регрессионного анализа в основном обсуждаются в первой книге монографии. Здесь детально рассматриваются исходные предпосылки, процедуры отыскания оценок параметров, свойства этих оценок. Значительное внимание уделяется статистическим аспектам регрессионного анализа, включая проверку гипотез относительно параметров и линейных функций от них. Обосновывается процедура проверки адекватности регрессионной модели.

Процедуры выбора «наилучшей» регрессии из множества возможных сосредоточены в гл. 6, с которой начинается вторая книга.

Исходные предпосылки классического регрессионного анализа выполняются далеко не всегда. Как обнаружить нарушение этих предпосылок? В каких случаях и какие нарушения можно считать допу-

стимыми? Что делать, если нарушения признаются недопустимыми? Эти вопросы давно занимают специалистов по математической статистике.

Мощным средством обнаружения некоторых отклонений от исходных предпосылок регрессионного анализа является анализ остатков, представляющих собой разности между экспериментальными и расчетными значениями откликов. Исследованию остатков посвящена гл. 3 данной книги. Но мало просто обнаружить, что предпосылки нарушены. Нужна конкретная программа действий в указанных условиях. В силу сказанного со всей остротой возникла потребность пересмотра, смягчения основных постулатов регрессионного анализа. Это привело к появлению целого набора новых статистических методов, являющихся продолжением, развитием методов классического регрессионного анализа.

Начнем с пересмотра постулатов относительно базисных функций от факторов и относительно самих факторов. Еще в 20-е годы Р. Фишер разработал дисперсионный анализ. Этот прием, сыгравший огромную роль в развитии планирования эксперимента, породил массу частных моделей и соответствующих методов обработки данных. понадобилось около 30 лет, чтобы началась консолидация процедур регрессионного и дисперсионного анализа. Стало ясно, что основная особенность задач дисперсионного анализа, если их трактовать в терминах регрессий, состоит не столько в том, что факторы здесь измеряются в дискретных шкалах, сколько в том, что соответствующие базисные функции от факторов оказываются линейно зависимыми. А это в свою очередь приводит к тому, что матрица системы нормальных уравнений вырождена и задача отыскания оценок параметров не имеет единственного решения. Усилиями К. Точера, С. Рао и других исследователей был найден прием, позволяющий свести любую задачу дисперсионного анализа к задаче регрессионного анализа, но с вырожденной матрицей системы нормальных уравнений. Для решения этой системы предлагается использовать так называемые обобщенные обратные матрицы Мура—Пенроуза.

Одновременно шел и «встречный» процесс: дисперсионный анализ начал широко применяться при изучении результатов регрессионного анализа. Это направление отчетливо прослеживается вплоть до наших дней.

От модели дисперсионного анализа оставался один шаг до смешанной модели, в которой представлены как регрессионные, так и дисперсионные переменные. Такая модель стала называться моделью ковариационного анализа. Ее введение тоже связано с именем Р. Фишера. В итоге удалось объединить в рамках одной формальной процедуры регрессионного анализа три типа моделей. Подобное объединение создает удобство при программировании и вычислениях на ЭВМ.

Дисперсионный анализ весьма обстоятельно описан в гл. 9 работы Н. Дрейпера и Г. Смита. Причем авторы уделили большое внимание сопоставлению регрессионного и дисперсионного анализа. В книге рассмотрены разные приемы элиминирования вырожденности исходной системы нормальных уравнений.

В классической регрессии факторы предполагаются детерминированными. Это означает, что в условиях реального эксперимента мы должны знать о них все, знать с бесконечной точностью. Реально ли это? Конечно, нет.

Отказ от детерминированности независимых переменных ведет к новой модели — модели корреляционного анализа. В одном частном случае, для парной корреляции, такая модель играет выдающуюся роль в статистическом анализе. Проявляется это и при исследовании регрессионных моделей. Но все попытки существенного обобщения этой модели на многомерный случай наталкиваются пока на серьезные препятствия. Главный камень преткновения здесь — требования к многомерным функциям распределения, которые не известно ни как обеспечить, ни как проверить.

Трудности многомерного корреляционного анализа привели в 30-е годы к созданию компромиссной модели — модели конфлюэнтного анализа, предложенной Р. Фришем. В этой модели допускается, что при нормально-распределенном отклике факторы тоже могут иметь некоторый разброс значений, тоже нормально-распределенный и усеченный. Причем никаких многомерных условий не налагается. В такой ситуации удастся построить процедуру обработки данных, сводящую дело к многократному решению регрессионной задачи.

Теперь коснемся постулата о параметрах моделей. Модели со случайными параметрами рассматриваются в современном дисперсионном анализе, их именуют иногда моделями со случайными компонентами. Отказ от детерминированности параметров регрессионных моделей приводит к более серьезным последствиям, поскольку при этом затрагиваются статистические основы регрессионного анализа. Тем не менее такие модели имеют право на жизнь. Можно себе представить, что иногда существует информация о параметрах регрессионной модели, позволяющая задать некоторое априорное распределение этих величин, рассматриваемых как случайные. Тогда в качестве оценок параметров можно использовать их условные математические ожидания, если только имеют место наблюдаемые значения откликов. Когда условные распределения параметров используются для получения оценок, говорят о байесовском регрессионном анализе, поскольку условные (апостериорные) распределения и ожидания вычисляются по обобщенной формуле Байеса.

И наконец, обсудим постулаты, относящиеся к отклику регрессионной модели.

Регрессионные модели нередко применяются для описания процессов, развивающихся во времени. Заметим, что отклики при этом могут измеряться дискретно, а в определенные моменты времени непрерывно, на некотором временном интервале. В таком случае от рассмотрения случайных величин откликов придется перейти к анализу случайных последовательностей и случайных процессов. А в более общей ситуации, когда процесс развивается и во времени, и в пространстве, — может быть даже и к анализу случайных полей. Это приводит к серьезным осложнениям. Одна из распространенных про-

стейших моделей такого рода — модель авторегрессии. Она предполагает, что отклик зависит не только от ряда изучаемых входных переменных (факторов), но и от времени. Если последнюю зависимость удастся выявить, то задача сводится к стандартной, но для преобразованного отклика. Если же нет — требуются специальные, более сложные приемы.

В обычной регрессионной модели предполагается, что неизвестные параметры сосредоточены в зависимости математического ожидания от факторов. Что же касается дисперсий и ковариаций измеряемых откликов, то считается, что они известны с точностью до сомножителя, отождествляемого часто с дисперсией ошибки эксперимента. В реальных задачах информация о дисперсиях и ковариациях откликов отнюдь не столь полна. В этой связи представляет интерес обобщенная регрессионная модель, допускающая зависимость дисперсий измерений от факторов. В эту модель может входить несколько неизвестных параметров. Это обобщение называют *F*-моделью. Разработана итерационная процедура «ИРДЖИНА» для поочередного оценивания параметров, входящих в выражение для математического ожидания отклика и в выражение для дисперсий измерений. *F*-модель имеет ряд преимуществ перед классической регрессией. К ней сводится, в частности, модель конфлюэнтного анализа.

Пока нормальный закон считался само собой разумеющимся, особых проблем не возникало. Тем более, что он опирался на авторитет центральной предельной теоремы теории вероятностей. Но когда под воздействием практики от этой догмы пришлось отказаться, стало ясно, что мы существенно зависим от априорной информации о законе распределения отклика. Ее уровень в разных задачах может быть совершенно различным, да и распорядиться ею можно по-разному. Когда мы заранее знаем, каков закон распределения, можно построить процедуру обработки данных, использующую эту информацию. Метод такого рода был разработан также Р. Фишером. Он называется методом максимума правдоподобия. Ясно, что стандартный классический вариант регрессионного анализа — частный случай этого метода. Хотя с вычислительной точки зрения возникающая процедура гораздо менее приятна, чем классическая, ничего страшного она не привносит. Вполне понятно, что учет надежной информации о фактическом законе распределения, скажем логнормальном вместо нормального, улучшит оценки, а в качестве платы за улучшение придется дольше считать по более сложной программе. Но в практике столь высокий уровень априорной информации встречается крайне редко. А что делать, если нам неизвестно истинное распределение?

В математической статистике давно была высказана мысль о том, что возможно получение некоторой полезной информации и в том случае, когда мы не можем или не хотим воспользоваться информацией о законе распределения изучаемой случайной величины. Пока мы верили в нормальность, эта идея не получала признания. Действительно, если нормальность на самом деле выполняется, то такие «непараметрические» процедуры будут существенно менее эффектив-

ными, чем процедуры классической теории. Они годились лишь для каких-то исключительных ситуаций. Когда же нормальность превратилась не более чем в частный случай, пусть распространенный, положение резко изменилось. Выяснилось, что когда отсутствует достаточно обоснованная информация о функции, описывающей регрессию и известной с точностью до параметров, можно построить такую регрессионную процедуру, которая по своей эффективности приближается к классической, а в ряде случаев она оказывается практически единственно возможной. Так появился еще один конкурент классической регрессии — непараметрический регрессионный анализ.

К нашему распределению, какое бы оно ни было, часто примешиваются чужеродные элементы, даже в малых количествах существенно ухудшающие ситуацию. Опыт показывает, что в больших массивах данных появление засорений практически неизбежно. Долгое время разрабатывались методы выявления подозрительных наблюдений, которые называют «дикими» или сорными. Отбрасывание таких наблюдений существенно улучшало положение. Однако, чтобы их выявить, надо снова знать закон распределения. В 1950 г. Дж. Бокс, занимаясь дисперсионным анализом, пришел к мысли о том, что можно не выявлять и не отбрасывать дикие наблюдения, а строить такие процедуры оценки, которые были бы нечувствительны к наличию в выборке засоряющих наблюдений. Он назвал такие процедуры робастными, или устойчивыми. С тех пор теория робастного оценивания вообще и для регрессии в частности быстро развивается. Выведены специальные формулы для робастных оценивателей. Исследованы ранее предложенные методы отыскания параметров регрессии. Выяснилось, что повышенной устойчивостью обладают оценки параметров, полученные по методу минимизации суммы модулей ошибок и максимального модуля ошибки (чебышевский метод оценивания). Новые веяния, относящиеся к робастному оцениванию, кратко описаны в гл. 6 монографии Н. Дрейпера и Г. Смита.

Робастные алгоритмы в известном смысле можно рассматривать как промежуточные компромиссные между параметрическими методами стандартной теории и непараметрическими подходами: они используют некоторую информацию о законах распределения, хотя и «распоряжаются» ею иначе.

По мере того как накапливался опыт работы с регрессиями, все больше и больше обнаруживались их «коварные» свойства. Выяснилось, например, что даже при соблюдении всех исходных постулатов МНК-оценки параметров, несмотря на все их оптимальные свойства, нередко с большой ошибкой оценивают параметры модели. И это вовсе не обязательно связано с плохим выбором условий эксперимента. Часто виновата сама регрессионная модель, ее структура. Если регрессия выражается в виде линейной комбинации экспонент или полиномом высокой степени, то столбцы матрицы X могут оказаться почти линейно зависимыми. Это явление, называемое мультиколлинеарностью, приводит к плохой обусловленности матрицы системы нормальных уравнений и к неустойчивости оценок параметров.

Плохо обусловленные задачи оценивания регрессии составили целое направление в регрессионном анализе. Они породили специальные, тонкие методы поиска оценок параметров. Практика показала, что повышения устойчивости оценок параметров можно добиться, если отказаться от требования их несмещенности, строго соблюдаемого в обычной регрессии. Так появилась гребневая, или ридж-регрессия. Гребневая регрессия достаточно подробно описана в гл. 6 данной книги.

До сих пор речь шла о регрессиях с одномерным откликом. Однако реальные объекты, для описания которых привлекается регрессионный анализ, нередко имеют несколько откликов. В связи с этим представляет интерес многомерная (многооткликовая) регрессия. Появились такие модификации многомерной регрессии, как псевдонезависимая регрессионная модель, модель в виде системы одновременных (синхронных) уравнений.

В первом случае речь идет о ряде стохастически связанных между собой одномерных регрессионных уравнений. Во втором предполагается, что между разными откликами системы существуют линейные связи. Одновременные уравнения находят широкое распространение в эконометрии.

Регрессионные модели, построенные на базе полиномов, носят, как правило, формальный характер. Их используют для описания изучаемых объектов, относительно которых нет достаточно четких количественных представлений. Однако исследователей чаще интересуют содержательные, физические модели, отражающие механизм, сущность явлений. Если разработаны теоретические основы исследуемого явления, то может быть заранее известна структура модели. В этом случае экспериментальные данные служат лишь для определения отдельных параметров. Выбор же типа модели объекта — традиционный удел всякого специалиста. Это вообще один из центральных вопросов науки.

Содержательные, физические модели, как правило, нелинейны по параметрам. Методология их создания составляет один из интенсивно развивающихся разделов математической статистики — нелинейный регрессионный анализ. Нелинейный регрессионный анализ привнес в статистику целый клубок трудно решаемых проблем. Эти проблемы связаны не только с нелинейным характером зависимости откликов от параметров. Как правило, физические модели являются многомерными, отклики нередко связаны между собой. К тому же и сама регрессионная зависимость, связывающая отклики с факторами, выражается неявно. Она обычно представляет собой решение системы алгебраических или дифференциальных уравнений, которое чаще всего не может быть представлено аналитически. В результате появляется проблема точечного оценивания параметров нелинейных моделей. Она намного сложнее, чем в случае линейной параметризации. Оценки — чаще всего смещенные, причем степень смещения оценить нелегко.

Задача оценивания параметров нелинейных моделей, как правило, имеет не одно, а множество решений. А иногда решение вообще от-

существует. Неустойчивость оценок резко обостряется. Одним словом, здесь мы сталкиваемся со всеми атрибутами некорректно поставленных задач.

Однако недостаточно просто найти точечные оценки параметров, не менее важно их охарактеризовать статистически, определить их дисперсии и ковариации. В условиях нелинейной параметризации это — нелегкая задача. Линеаризация нелинейных по параметрам зависимостей может приводить к резко искаженным величинам дисперсий и ковариаций оценок параметров. Выходом из этой ситуации может быть использование асимптотических разложений функции отклика, в которых участвуют производные от функции отклика по параметрам более высоких степеней. Заметим, что алгоритм вычисления слагаемых таких разложений является очень трудоемким.

Весьма проблематичными становятся процедуры статистического анализа нелинейной регрессии. Даже если известно, что отклики подчиняются нормальному распределению, то что можно сказать про распределение оценок параметров? Как проверять гипотезы об адекватности модели, о значимости параметров? На эти вопросы пока нет исчерпывающих ответов.

Проблеме оценивания параметров нелинейных моделей посвящена гл. 10 монографии Н. Дрейпера и Г. Смита. Авторам удалось четко проследить сходство и различие между линейным и нелинейным оцениванием.

Регрессионный анализ — методологическая основа теории планирования эксперимента. Многие критерии оптимальности эксперимента заимствованы из соотношений, характеризующих свойства оценок параметров. Планирование эксперимента для линейно параметризованных моделей превратилось в хорошо разработанный, обширный раздел математической статистики. В настоящее время интенсивно развивается планирование эксперимента в случае нелинейной параметризации. В обсуждаемой монографии эти вопросы затронуты вскользь. Интересующийся читатель должен обратиться к специальной литературе (некоторые дополнительные ссылки на литературу приведены в примечаниях переводчиков к гл. 10).

Сейчас наступил новый этап развития вычислительной техники. Появились мини- и микроЭВМ, персональные компьютеры. Повышение быстродействия, увеличение памяти и удешевление ЭВМ, а также значительный прогресс в сервисных устройствах вызвали к жизни новые подходы к анализу данных, основанные на применении вычислительной техники. Это прежде всего относится к имитационному моделированию, предложенному Т. Нейлором. Не менее важное значение имеет концепция анализа данных, вытекающая из работ Дж. Тьюки. Большие надежды возлагают на разработанный и предложенный в 1979 г. Б. Эфроном метод «бутстреп». Все эти методы в совокупности с известными методами многомерной статистической классификации данных обогатили методологию регрессионного анализа. С другой стороны, сама регрессионная модель выступает теперь в качестве цементирующего начала, связывающего эти методы в нечто целостное.

Монография Н. Дрейпера и Г. Смита не охватывает все аспекты регрессионного анализа, что сделать, пожалуй, и невозможно. Важно другое: она дает фундаментальные представления о регрессии — как линейной, так и нелинейной. Опираясь на них, можно при желании углубить и расширить свои знания по регрессионному анализу,² обратившись к другим источникам.

Предлагаемая книга может служить путеводителем по соответствующей литературе. К обширной библиографии, завершающей книгу и охватывающей период приблизительно до 1980 г., добавлен список литературы, где в основном приведены работы на русском языке.

Мы надеемся, что книга Н. Дрейпера и Г. Смита представит интерес для советского читателя и поможет статистикам, экономистам, социологам, научным работникам овладеть приемами и методами регрессионного анализа.

Ю. АДЛЕР,
В. ГОРСКИЙ

● ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

В 1962 г. к нам обратились представители химического отдела Американского общества контроля качества (A. S. Q. C.) с предложением подготовить краткий курс регрессионного анализа. В этой связи мы составили ряд конспектов по темам, которые, по нашему мнению, важны для практиков, применяющих регрессионный анализ. Эти конспекты были хорошо приняты, в дальнейшем в них было внесено много дополнений и изменений. Данная книга является результатом этой работы.

Мы попытались объединить в книге ряд методов, развитых для регрессионных задач и распространенных в настоящее время. Так как мы сделали акцент на практическом применении регрессионного анализа, то теоретические результаты во многих случаях приводятся без доказательств. Хотя обучение регрессионному анализу проводится без использования вычислительных машин или со сравнительно примитивной техникой, тем не менее работы по применению регрессии выполняются теперь исключительно с помощью быстродействующих вычислительных машин. Поэтому, хотя данной книгой можно пользоваться вообще без всяких вычислительных средств (или, быть может, только с настольной вычислительной машиной), мы специально употребили в нескольких местах машинные распечатки. Все десятичные знаки в этих данных вряд ли нужны на практике, но числа писались так, будто они получены на обычной вычислительной машине. Мы составили также различные упражнения: некоторые из них можно легко решить «вручную», для решения других, более сложных, была бы полезна, хотя и не абсолютно необходима, вычислительная машина.

Эта книга представляет собой стандартный основной курс множественной линейной регрессии, но она включает также материал, который либо совсем не появлялся в учебниках, либо если и появлялся, то был труднодоступен для понимания. Например, в гл. 3 обсуждается исследование остатков; в гл. 6 рассматриваются процедуры отбора факторов в регрессионных программах разных типов; в гл. 8 обсуждается планирование больших регрессионных исследований; гл. 10 дает основное введение в теорию нелинейного оценивания.

Гл. 1 и 3 вместе представляют собой курс по подбору уравнения прямой линии вообще без использования алгебры матриц. Если же прибавить часть гл. 2, то вдобавок можно получить введение в идеи матричного представления регрессионных задач. Односеместровый