

Д. Хейс



ПРИЧИННЫЙ
АНАЛИЗ
В СТАТИСТИЧЕСКИХ
ИССЛЕДОВАНИЯХ

Д. Хейс

CAUSAL ANALYSIS

David R. Heise



518

ПРИЧИННЫЙ АНАЛИЗ В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Перевод с английского Ю. Н. ГАВРИЛЬЦА,
Л. М. КУТИКОВА и М. А. РОДИОНОВА

Предисловие
Т. В. РЯБУШКИНА и Ю. Н. ГАВРИЛЬЦА



A Wiley-Interscience publication

JOHN WILEY & SONS, NEW YORK · LONDON · SYDNEY · TORONTO

Москва «Финансы и статистика» 1981

ВЫШЛИ ИЗ ПЕЧАТИ

1. Ли Ц., Джадж Д., Зельнер А. Оценивание параметров марковских моделей по агрегированным временным рядам.
2. Райфа Г., Шлейфер Р. Прикладная теория статистических решений.
3. Клейнен Дж. Статистические методы в имитационном моделировании. Вып. 1.
4. Клейнен Дж. Статистические методы в имитационном моделировании. Вып. 2.
5. Бард Й. Нелинейное оценивание параметров.
6. Болч Б. У., Хуань К. Д. Многомерные статистические методы для экономики.
7. Иберла К. Факторный анализ.
8. Зельнер А. Байесовские методы в эконометрии.

ГОТОВИТСЯ К ПЕЧАТИ

Пуарье Д. Эконометрия структурных изменений.

Редколлегия: А. Г. Аганбегян,
Ю. П. Адлер, Ю. Н. Благовещенский,
А. Я. Боярский, Н. К. Дружинин,
Э. Б. Ершов, Т. В. Рябушкин,
Е. М. Четыркин

X 10805*-046 45-81(C) 0702000000
010(01)-81

Второй индекс 10803.

© John Wiley & Sons, Inc., 1975

© Перевод на русский язык, предисловие, предметный и именной указатель, «Финансы и статистика», 1981

● ПРЕДИСЛОВИЕ К РУССКОМУ ПЕРЕВОДУ

Предлагаемый читателю перевод работы профессора социологии университета Северной Каролины (США) Д. Хейса является введением в «модное» направление в эконометрии и социологии — «причинный анализ». Это направление имеет целью дать такое описание системы взаимозависимых переменных, при котором можно указать переменные, являющиеся «причинами», переменные, являющиеся «следствиями», и прогнозировать вторые по первым. В этом узкоспециальном смысле и будет употребляться дальше выражение «причинный анализ», который понимается не тождественно этому термину в его общепризнанном значении.

Известно, что одной корреляции или даже зависимости между двумя признаками недостаточно для утверждения о наличии причинного отношения между ними. Поэтому нужны какие-то дополнительные основания уверенности исследователя в том, что один из признаков (одно из событий, состояний) является причиной другого. Особенно это важно в прикладных исследованиях, когда прогнозируются возможные состояния или когда управляют некоторым процессом с целью получения нужных результатов. И в том и в другом случае знание действительных причин совершенно необходимо. Прогнозирование становится научно обоснованным и более точным, если оно опирается на реальные причины; для управления же причинные связи являются совершенно необходимыми, поскольку управлять на основе только «корреляционных» зависимостей практически невозможно.

Поэтому строгое и операциональное определение причинно-следственных отношений, их формализация в соответствии с содержательной концепцией делают научный анализ явлений более глубоким и эффективным, а качественные рассуждения о социально-экономических явлениях благодаря этому подкрепляются расчетами.

Чаще всего под «причинным анализом» (не в широком философском смысле) понимают определенную систему представлений, формализованную в вид линейных моделей взаимозависимостей между переменными (рекурсивные системы, структурные уравнения и т. п.). Однако имеются и другие подходы, когда отношения между переменными могут быть нелинейными и сугубо вероятностными.

Данная работа посвящена частному разделу причинного анализа — линейным системам — и не охватывает всех проблем данного направления, но тем не менее она выгодно выделяется среди многих других. Причины этому следующие.

Во-первых, книга написана в стиле учебников по «программированному обучению», когда обучающийся усваивает материал

в строго продуманной («спрограммированной») последовательности, с разбором примеров, с самостоятельным решением большого числа задач и примеров по всем разделам «курса».

Во-вторых, в тексте совершенно не дается формальных доказательств математических результатов и положений; в то же время изложение и анализ этих результатов являются основным содержанием книги. Благодаря этому читатель, даже недостаточно знакомый с математикой, усваивает материал, не пропуская отдельных страниц книги, а читая ее целиком. В этом смысле книга является как бы справочником с чрезвычайно обильными комментариями.

В-третьих, и это главное, автор старается все математические результаты и представления причинного анализа излагать в графическом виде — на рисунках, схемах и т. п. Это ему вполне удастся, причем без потери строгости изложения. В результате сами положения даются очень наглядно, что существенно облегчает их усвоение, а использование формул становится легкодоступным делом даже для людей с гуманитарным образованием.

Книга состоит из шести глав, библиографических указателей, упражнений — к каждой главе и ответов на упражнения — в конце книги.

В первой главе автор излагает свою концепцию причинного анализа с целью дать некоторую методологическую основу для дальнейших утверждений и схем. Подчеркивается, что причинные отношения между событиями зависят от некоторых физических, т. е. материальных, механизмов, действующих в изучаемой реальности. Он подробно анализирует понятие причинной зависимости, указывает его отличие от других форм связи. Событие C будет причиной события E , если в присутствии некоторого «оператора» оно происходит до события E и с необходимостью его вызывает. При этом событие E может наступать и без наступления события C .

Исходным понятием автор считает событие как «однородный поток, подверженный увеличению или ослаблению», описываемый некоторой переменной. При этом «любое интересующее нас событие — это смесь событий более низкого уровня». Нам представляется, что легче было бы говорить просто о переменных модели и о том, что одни переменные являются «причинами», а другие — «следствиями». Интерпретировать линейные причинные схемы в терминах потоков, как и предполагать обязательно линейными всякие потоковые схемы, совершенно не обязательно. Не все явления описываются на языке потоков, поэтому излишнее стремление следовать «потоковой» интерпретации может только затруднить понимание.

Надо сказать, что философская концепция автора по причинному анализу, хотя и является стихийно-материалистической, не выглядит достаточно четкой. Он понимает под «оператором» нечто материальное, без чего не будет действовать механизм причин — следствий. Так, твердое тело, по его словам, — это оператор, превращающий силу в ускорение. На наш взгляд, автор,

желая подчеркнуть, что не любое событие одного рода может быть причиной события другого рода, в некотором смысле усложняет проблему. Аргументы типа того, что «нельзя сдвинуть куб воздуха» или «нельзя поднять тень», не слишком убедительны и только затемняют идею того, что в каждом конкретном случае исследователь ограничивается определенным уровнем рассмотрения с целью познания *относительной*, а не абсолютной истины. Аналогично за рассуждениями об «операторах», «компонентах операторов», «полях операторов», «градиентах полей» и т. п. стоит простая идея о всеобщей взаимосвязи одних явлений через другие с третьими и о возможности сколь угодно дробной дифференциации явлений в процессе познания взаимодействий.

Во второй главе автор знакомит читателя с основной для книги формальной конструкцией — графическим изображением причинных зависимостей между переменными с помощью так называемых потоковых графов. Эти графы, являющиеся наглядным выражением систем структурных уравнений, осваиваются нематематиками достаточно легко и образуют, по справедливому замечанию автора, мост между вербальными теориями и более абстрактными математическими построениями.

Весьма подробно объясняется смысл потоковых графов, на которых могут иметься цепочки связей, циклы, петли обратных связей, вызываемые опосредованным влиянием переменной на самое себя через опущенные промежуточные переменные. При этом факт отсутствия стрелки между парой переменных также представляет собой некоторую информацию, т. е. означает отсутствие одной из возможных непосредственных связей. Поэтому при построении причинных графов в сомнительных случаях надо включать стрелки, а не опускать их.

В третьей главе изложены основные и простейшие понятия математической статистики: распределения, среднее, дисперсия, ковариация, корреляция, регрессия. Указывается, что регрессионные уравнения и структурные уравнения — это не одно и то же, но объяснения и тем более доказательств тех или иных положений нет. Приводятся формулы для расчета простейших коэффициентов и характеристик связей.

Если диаграммы второй главы не предполагают наличия стохастичности в изучаемых системах, то «путевой анализ» четвертой и пятой глав является собой более реалистичные модели и методы для изучения социально-экономической действительности. В основе применения развиваемых далее схем, как подчеркивает автор, лежит предположение, что все многочисленные наблюдения (элементы выборки) имеют одну и ту же причинную структуру связи между переменными, а сама организация наблюдений позволяет фиксировать совместно и причины, и следствия. При этом возникают два типа задач. Если известна причинная структура системы стохастических переменных, то можно пытаться находить статистические характеристики «выходных» переменных («следствий») по статистическим характеристикам «входных» переменных

(«причинам»). Зная же статистические характеристики тех и других переменных, можно попытаться найти причинную структуру, которая преобразует данным образом «причины» и «следствия». Четвертая глава имеет дело с первой задачей, пятая — со второй.

Сам путевой анализ как статистический метод возник еще в 1910—1920 гг. благодаря работам С. Райта. Он позволяет определять (для данной структуры) ковариации, корреляции и дисперсии в виде произведений и сумм соответствующих коэффициентов на диаграмме. Правила и условия четвертой главы несколько отличаются от традиционного подхода Райта и других исследователей. Основное внимание уделено простоте используемых процедур, их наглядности, а также распространению традиционных процедур на системы с обратными связями.

Показав эффективность процедур путевого анализа, автор подчеркивает роль внешней среды, определяющей значение статистических параметров, а также относительность предположений о причинах и следствиях на языке переменных. Он указывает на наличие определенного рода внешних факторов, называемых пропускными механизмами, которые организуют изучаемую совокупность и часто сужают область значений наблюдаемых переменных. В результате одни и те же переменные могут, например, быть коррелированными или некоррелированными в зависимости от границ, определяющих эти области. Рассматриваются различные варианты связи между дисперсиями и корреляциями входных и выходных переменных. Особенно подчеркивается роль положительных и отрицательных обратных связей. Показывается, как изменением только дисперсий переменных, влияющих на петли обратной связи, можно делать корреляцию между переменными петли и положительной, и отрицательной, и нулевой.

Для того чтобы можно было эффективно использовать путевой анализ в задачах анализа, прогноза и управления, необходимо количественно оценить структурные параметры причинных систем. Если в технических системах эти параметры являются известными и даже задаваемыми самим исследователем-проектировщиком, то в социальных системах они исследователю обычно неизвестны. Поэтому параметры социальных систем должны быть определены из эмпирических наблюдений — апостериорно. Пятая глава излагает процедуры вычисления структурных параметров, логической базой которых является метод наименьших квадратов, поскольку при достаточных естественных предположениях он дает наилучшие оценки коэффициентов линейной регрессии.

Если причинная структура известна, то по правилам путевого анализа можно выписать соотношения, связывающие статистические характеристики переменных со структурными параметрами системы. Из этих соотношений можно попытаться определить значения параметров. Однако ясно, что если число уравнений будет меньше числа неизвестных, то значения параметров не определяются однозначно. Такие системы называются недоидентифицируемыми и порождают саму проблему идентификации.

Для решения задачи идентификации необходимо сделать ряд априорных предположений о системе, важнейшим из которых является предположение о некоррелированности возмущений на переменные с предшествующими (по причинному пути) переменными. Наиболее изучены так называемые рекурсивные системы, для оценки параметров которых хорошо работает обычный регрессионный анализ. Рекурсивные системы автор определяет как системы с линейными связями, без циклов и с отсутствием корреляций между «предшествующими» переменными и возмущением зависимой переменной. Указываются процедуры определения параметров для рекурсивных систем. На простейших примерах показывается, что при коррелированности предшествующих переменных с возмущениями последующих переменных, а также при наличии циклов идентификация обычными процедурами становится невозможной. Правда, в зависимости от структуры некоторые параметры петель иногда могут быть определены.

Дополнительно вводимая инструментальная переменная обладает тем свойством, что ее коэффициент связи с основной переменной может быть определен, только если их отношение рекурсивно.

Общий метод для оценивания коэффициентов нерекурсивных систем (предложенный эконометриком Г. Тейлом в 1950 г.), излагаемый в данной книге под названием «двухэтапного (двухшагового) метода наименьших квадратов», включает использование инструментальных переменных на первом этапе множественного регрессионного анализа с целью построения новой системы переменных, свободной от недостатков коррелированности возмущений.

Шестая глава посвящена изучению с позиций причинного анализа простейших динамических систем с запаздыванием. По-прежнему изучение причинных взаимозависимостей проводится в терминах потоков. Поскольку требования однородности процесса и одновременно «замера» причин и следствий весьма редко удовлетворяются, автор рассматривает причинные взаимосвязи между переменными, относящимися к разным моментам времени, в соответствии с лагами, разделяющими во времени причины и результаты. Подчеркивается важность фактора времени, в течение которого осуществляется преобразование всех входных сигналов в выходные. Приводятся расчетные формулы для вычисления значений параметров связи системы по статистическим характеристикам (дисперсиям, ковариациям); кратко рассматривается схема определения параметров системы, когда наблюдения за входными параметрами ведутся бесконечно долго.

Как уже отмечалось, большое место в книге занимают примеры, а также упражнения и ответы на них. Их важная роль определяется не только тем, что они иллюстрируют возможности техники причинного анализа, но и тем, что с их помощью можно овладеть этой техникой и развить в себе модельное мышление. Приводится и анализируется большое количество моделей самых различных явлений. Эти модели иллюстрируют и сильные, и слабые стороны причинного анализа, иногда кажутся удачными,

иногда — наивными, но все равно они достаточно интересны и не слишком узки по тематике (так, например, встречаются примеры из области анализа социальной структуры, криминологии, экономики и др.).

Здесь, как нам кажется, следует сделать существенную оговорку. Конечно, прежде всего эти примеры предназначены для обучения читателя приемам и технике причинного анализа, но они иллюстрируют и интерпретацию автором некоторых социально-экономических явлений, серьезно беспокоящих прогрессивную общественность Запада. Именно серьезностью таких проблем, как безработица, наркомания, преступность, обусловлен подбор тем для примеров, обсуждаемых в книге. Советскому читателю хорошо известны факты, свидетельствующие о беспрецедентном росте наркомании в США или о взрывоопасной обстановке в негритянских гетто Америки. Руководящие политические деятели дают широковещательные обещания снизить уровень хронической безработицы до «безобидных» четырех-пяти процентов, в то время как этот уровень уже достиг восьми процентов. Однако указанные проблемы вызваны самой природой капитализма и в силу этого неразрешимы. Именно поэтому в книге встречаются и упрощенная интерпретация динамики социально-профессиональной структуры, причин преступности или военно-политического равновесия; здесь и признание в качестве «нормы» пятипроцентного уровня безработицы и т. п. Как часто бывает, характер применения формальных методов очень сильно зависит от тех целей, которые ставят перед собой исследователи-прикладники, поэтому приводимые здесь модели далеко не всегда могут считаться достаточно адекватными. Советский читатель без труда сумеет отделить технику причинного анализа от субъективного методологического подхода автора к анализу общественных явлений.

Несмотря на все эти оговорки, книга Д. Хейса имеет прикладную и педагогическую ценность, так как позволяет специалистам с гуманитарным образованием овладеть современными эффективными методами анализа систем взаимозависимых переменных. Да и читателям, имеющим техническое образование, данная книга будет полезной и интересной, потому что содержит оригинальное изложение сравнительно нового направления — причинного анализа.

Для знакомства с вышедшими у нас работами по причинному анализу можно рекомендовать следующую литературу:

Бунге М. Причинность. М., Изд. иностр. лит., 1962.

Закон. Необходимость. Вероятность. Сб. статей. М., Прогресс, 1967.

Математика в социологии. М., Мир, 1977.

Моделирование социально-экономических процессов и социальное планирование. М., Наука, 1979.

Член-корреспондент АН СССР
Профессор

Т. В. Рябушкин
Ю. Н. Гаврилец

• ПРЕДИСЛОВИЕ

Эта книга посвящена изучению линейных систем и представляет собой попытку подать информацию по этому вопросу в довольно элементарной форме. Работая над книгой в 70-х годах, я имел возможность привлечь основные идеи из работ философов, инженеров и методологов общественных наук. Здесь добавлены некоторые новшества, но главная цель состоит не столько в том, чтобы «поднять целину», сколько в том, чтобы возделывать ту почву, которая уже вспахана, и, таким образом, исподволь расширять знание и понимание этого поля и его плодов.

Книга адресована довольно широкой аудитории исследователей-практиков в области общественных наук, студентов и интересующихся неспециалистов. Для того чтобы донести идеи до этой обширной аудитории и не скомпрометировать их, я целиком положился на педагогические возможности причинных схем. Эти схемы позволяют отчетливо увидеть и понять даже некоторые из более сложных вопросов методологии общественных наук, а сопровождающие их правила обращения позволяют почти каждому многократно проводить математический анализ с точностью алгебраиста. Педагогическое преимущество, даваемое причинными схемами, может наводить на мысль, что они являются «грубым и лежащим на поверхности» подходом к анализу систем. Однако это не так. Причинные схемы со своим множеством правил в них относятся к математике. Такими схемами ежедневно пользуются ученые-практики и инженеры.

Преподаватель, просматривающий эту книгу, может поинтересоваться тем, какое место она занимает в курсе обучения. Статистика в книге обсуждается подробно, но акцент здесь делается на статистическом описании социальных систем, а не на обычном статистическом выводе по выборкам для совокупностей. В примерах и упражнениях анализируются самостоятельные темы, но специальные примеры моделей охватывают большую часть вопросов социологии и отражают много точек зрения. Тем не менее как руководство по построению моделей эта книга может быть использована в курсах методологии и статистики, где главное внимание обращено на использование данных для проверки и разработки теорий, и может служить дополнительным источником в курсах по построению теорий.

В течение последнего десятилетия многочисленные коллеги и студенты помогли мне подготовиться к написанию этой книги.

Особенно значительна роль как учителей четверых из них — это Эдгар Боргатта, Джордж Борштедт, Артур Голдбергер и Денис Уиллиген. Джеймс А. Дэвис, Дункан Мак-Рэй младший и Роналд Бурт дали ценную критику первоначальных вариантов некоторых глав. То же сделало множество анонимных читателей-студентов, в том числе один шутник (впоследствии оказалось, что это Д. Ковэн), предложивший озаглавить книгу «Исследования по динамике стрелок». Для применения при изучении курса в Чепел Хилле рукопись была напечатана на машинке Г. Риппи.

Я особенно ценю личную поддержку Эльзы Льюис и Стефена Хейса, жизнерадостность и остроумие которых неоднократно возвращали меня из области абстракций, пока я писал эту книгу.

Дэвид Р. Хейс

Отделение социологии

Университет Северной Каролины, Чепел Хилл

Май 1975 г.

• ПРОЛОГ

Абу-Хамид* говорит:

По нашему мнению связь между тем, что обычно считается причиной, и тем, что считается следствием, не есть необходимая связь; о каждом из них нельзя сказать, что это есть то, а то есть это. В равной мере утверждение или отрицание одного не включает в себя утверждения или отрицания другого, потому существование и несуществование одного не вытекает с необходимостью из существования и несуществования другого. Так, например, утоление жажды не заключает в себе питья, насыщение — еды, горение — соприкосновения с огнем, свет — восхода солнца, смерть — обезглавливания, выздоровление — принятия лекарств, очищение желудка — принятия слабительного средства. Точно так же дело обстоит со всеми прочими связями (между явлениями), наблюдаемыми во врачебной науке, науке о звездах, в искусствах и ремеслах. Ибо связь между этими явлениями основана на предопределении всевышнего, вызывающего эти явления в последовательном порядке; но эта связь отнюдь не необходима сама по себе и может быть разорвана: всевышний обладает властью создавать сытость без еды, смерть без обезглавливания, продление жизни — после обезглавливания, и так обстоит дело со всеми прочими связями (между явлениями)...

Я говорю:

Что касается отрицания существования действующих причин, наблюдаемых в чувственно воспринимаемых вещах, то это софистика; тот, кто говорит так, либо говорит одно, а думает другое, либо же увлекается софистическими лжеумствованиями, осаждающими его во время обсуждения этих вопросов. Ибо тот, кто отрицает это, не может уже признавать, что всякое действие необходимо имеет свою причину. А достаточны ли эти причины сами по себе для того, чтобы вызвать исходящие от них действия, или же для того, чтобы их действие было законченным, требуется (дополнительно) некая внешняя причина — все равно, будет ли она (действовать) отдельно (от них) или нет — этот вопрос не самоочевиден, он требует глубокого рассмотрения и кропотливого исследования. Если мутакаллимы сомневаются в наличии дей-

* В оригинале употребляются имена Газали (Абу-Хамид) и Аверроэс (Ибн Рушд), под которыми они известны на Западе. — *Примеч. пер.*

ствующих причин, кои воспринимаются как обуславливающие друг друга, сомневаются потому, что имеются действия, причины которых не воспринимаются, то для этого нет никаких оснований. Те явления, причины которых не воспринимаются, остаются пока неизвестными и должны быть исследованы именно потому, что их причины не воспринимаются...

Ибн Рушд (1126—1198 гг.) *

* Цитируется по русскому переводу книги «Опровержение опровержения» в сборнике «Избранные произведения мыслителей стран Ближнего и Среднего Востока IX—XIV вв.», Издательство социально-экономической литературы, М., 1961, с. 506 и 508 (перевод А. И. Рубина и А. В. Сагадеева с бейрутского издания 1930 г.). — *Примеч. пер.*

Понятие причинности применяется всегда, когда осуществление одного события оказывается достаточным основанием для ожидания того, что произойдет другое. Причинное мышление устанавливает связь с деятельностью, потому что осуществление события подразумевает какой-нибудь вид изменения. Однако процедуры причинного анализа обычно сосредоточивают внимание на конфигурациях событий, — во времени или в единственный момент, — а не на изменениях как таковых. Причинная обусловленность порождает модели событий, а изучение моделей может обеспечить понимание причинных отношений, которые их порождали.

Причинное мышление регулярно используют в обыденной жизни, особенно когда объектами манипулируют или когда они переходят из одного состояния в другое. Возможно, именно поэтому манипулирование объектами кажется столь важным при введении причинных зависимостей и поэтому представляется, что причинные объяснения дают не только ощущение понимания, но также ощущение возможности управления. Причинные объяснения можно абстрагировать от манипуляций объектами или, по крайней мере, рассматривать на уровне, при котором манипуляции являются лишь чисто гипотетическими; например, можно было бы сказать, что поле тяготения солнца является причиной некоторых особенностей в движении планет, даже если нет возможности экспериментальной проверки. С помощью такого обобщенного представления иногда можно неэкспериментально исследовать причинность путем использования построенных на событиях моделей.

Возможность причинного анализа и вывода без манипулирования объектами имеет решающее значение в общественных науках, в которых так много политических, практических и этических проблем, суживающих возможности выполнения классических экспериментов. Таким образом, наша первая цель состоит в таком достаточно общем определении причинности, чтобы понятие было применимым, даже если связанное с манипулированием объектами управление недостижимо.

ПРИЧИННОЕ УПОРЯДОЧЕНИЕ

События являются исходным пунктом причинного анализа. Событие — это осуществление в некотором объекте определенного состояния или конфигурации состояний. Простые изменения

расположения или физической ориентации объекта составляют элементарные события. Можно также определять события на более абстрактном языке и сигнализировать о событиях изменениями наблюдаемых характеристик, изменениями скорости деятельности или предрасполагающими изменениями (т. е. изменениями возможности других событий).

Идея причинности становится уместной, если события упорядочены и устроены определенным образом. Прежде всего высказывание, что одно событие является причиной другого, требует, чтобы первое событие — назовем его C — вызывало надежду на осуществление второго события, E . Подобной надежды не существует в обратном направлении. Если по некоторой причине известно только, что случилось E , не особого основания верить, что C случится или даже с необходимостью произошло ранее. Эти особенности можно суммировать в виде таблицы, как показано в 1.1.

1.1.

		E	
		происходит	не происходит
C	происходит	да	нет
	не происходит	да	да

«Да» означает, что «эта комбинация может существовать», а «нет» означает — «эта комбинация не может существовать». Модель соответствует утверждению, что « C является причиной E », которое означает более определенно, что « C не может произойти без того, что произойдет E ».

Таблицу в 1.1 можно прочесть следующим образом: осуществления C подразумевают осуществления E , но осуществления E не означают осуществлений C ; или же осуществления C достаточны, но не необходимы для осуществлений E .

Представление о том, что C предполагает E , является определяющим для причинного мышления, но самого по себе этого не достаточно для того, чтобы ухватить сущность причинности. Импликация просто связывает возможности различных совместных осуществлений событий без каких бы то ни было ограничений на их расположение во времени или на физическую структуру; например, не принимая в расчет упорядоченность во времени и структуру, можно, в качестве альтернативы, интерпретировать модель в 1.1 как «осуществления C вытекают (развертываются) из осуществлений E ».

Рассмотрение временного упорядочения делает ясным направление, в котором должна читаться логическая импликация. Следствия не происходят до своих причин. Причина всегда предшествует своему следствию в том смысле, что причинное событие всегда начинается до того, как начинается следствие. Путем объединения логического и временного критериев мы получаем более

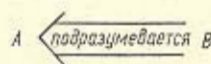
адекватную концепцию причинности. Существование причинной связи подозревают, если осуществление одного события влечет более позднее осуществление другого.

Добавление условия временной очередности отличает причинные связи от «связей развертывания», при которых одно событие предполагает другое, но второе происходит до первого. Оба эти типы связей иллюстрируются в 1.2.

1.2*. Причинная связь — это связь, при которой осуществление первого события является достаточным условием для осуществления более позднего события.



Связь развертывания — это связь, при которой осуществление первого события необходимо для осуществления последующего события.



Причинность не требует отсутствия E , когда отсутствует C . Может быть, что осуществления C в более ранний момент нет, но E так или иначе тем не менее происходит. По существу, это допускает ситуацию, когда E может вызываться событиями, отличными от C . Так, хотя и верно, что «следствие не происходит до своей причины», надо осознавать, что следствие может вызываться одной из причин до осуществления другой его причины. Следовательно, для точности, первая схема в 1.2 должна получить следующую интерпретацию: C является причиной E , только если после осуществления C неизменно обнаруживают осуществление E , даже если бы E ранее не было заметно.

Сочетание логического условия и условия временного упорядочения дает улучшенную формулировку, касающуюся причинности, чем каждое из двух условий по отдельности, но результат все еще неадекватен. С одной стороны, эта формулировка дала бы нам возможность полагать, что любое событие вызывает любое другое событие при непроверяемом предположении, что второе событие неизменно следует за первым «где-то во вселенной». Нам надо ясно осознать, что причинность действует при физических ограничениях. Кроме того, причинные связи не имеют постоянной силы даже в ограниченных областях, если не выполняются некоторые минимальные условия, и это тоже следует понять.

* Большая стрелка указывает направление логической импликации —
Примеч. пер.

Причинная обусловленность зависит от чрезвычайного изменения реальности в отдельный момент, такого, что одно событие превращается в другое. Требуемая для причинности специальная структура очевидна в обычно приводимом примере — спичка, которая производит взрыв. Горение спичек обычно не вызывает взрывов. Взрывы бывают результатом, только если зажженная спичка попадает в помещение, где присутствует легковоспламеняющийся газ, или же в случае соприкосновения с должным образом упакованным взрывчатым веществом. Обстоятельства, наделенные структурой, также подразумеваются даже в самых общих физических примерах причинности; например, сила вызывает движение тела, только если тело обладает достаточной жесткостью (нельзя сдвинуть куб воздуха) и только если оно обладает массой (у тени есть форма, но ее нельзя поднять). В применении к элементарной механике это представление о причинности, зависящей от чрезвычайных и преходящих обстоятельств, кажется понятным лишь посвященным, потому что требуемые физические условия столь вездесущи, устойчивы и общеизвестны, что мы считаем их само собой разумеющимися. Однако зависимость причинной обусловленности от предварительно существующих структур важна при изучении жизни общества, в которой требуемые структуры часто отсутствуют. Даже довольно общее социологическое утверждение вроде «внешние угрозы вызывают групповую сплоченность» применимо только к некоторым типам людей (например, к детям) с некоторым минимумом групповой принадлежности и структуры. Без требуемых условий угрозы могут вызывать другие реакции, например панику, или вообще отсутствие реакции.

Оператором называется вещественное устройство или наделенный структурой процесс, которые обеспечивают выполнение причинной связи. Причинное превращение событий не может иметь места без соответствующего оператора, а оператор должен существовать до превращения, которому он способствует. Причинная связь не обязательно зависит только от одного оператора. И биологические системы, и человеческие производственные системы обычно дают классы по существу эквивалентных устройств, которые осуществляют одни и те же связи. К тому же часто причинные связи могут осуществляться операторами с различным внутренним устройством (например, автомобиль работает независимо от того, поршневой или роторный двигатель он имеет). Итак, причинная связь эволюционно зависит от существования класса операторов.

Оператор состоит из организованных компонент. Сами компоненты являются операторами в том отношении, что они являются устройствами, которые служат для превращения одного события в другое; например, у автомобиля есть и двигатель, и передача, и каждая из этих частей сама является оператором для преобра-

зования событий одного рода в события другого. Кроме того, так как каждая компонента оператора является оператором, ее тоже можно разложить на подкомпоненты. Теоретически процесс разложения можно продолжать неограниченно. Эта разложимость операторов соответствует разложимости причинных связей; иными словами, превращение события *C* в событие *E* можно разложить на элементы множества промежуточных причинных связей, которые обнаруживаются в операторе и осуществляются компонентами оператора. Эти связи в свою очередь можно разложить на еще более дифференцированное множество составляющих причинных связей. В принципе этот процесс анализа может продолжаться неограниченно. Таким образом, причины и следствия связаны непрерывными цепями промежуточных событий.

Аналитическое рассечение операторов и причинных связей является важным аспектом понимания. Мы разделяем новые операторы и причинные связи на компоненты и связи до тех пор, пока не почувствуем, что знаем, как происходит превращение одного события в другое — как одно событие «производит» другое. В обыденной жизни мы, как правило, продолжаем процесс анализа до тех пор, пока не достигнем уровня известных, обычно принятых связей. Даже научный анализ редко продолжается за пределы, за которыми дальнейшее разделение оператора потребовало бы его трактовки как статистической совокупности мельчайших устройств, осуществляющих связи, которые являются объектом другой дисциплины. Например, исследователь теоретически может определить, какая сила, приложенная к стальному шарiku, вызывает направленное движение, спустившись на молекулярный уровень и анализируя воздействия силы на отдельные молекулы и их влияние друг на друга в рамках ограничений, налагаемых структурой подразумеваемого вещества. Движение шарика выглядит тогда как совокупный результат, определенный в терминах миллионов параллельных и подобных событий на более низком уровне. Это, однако, гораздо подробнее, чем обычно хочется, и такой анализ на микроуровне не представляет для нас большого интереса. Точно так же социологическое объяснение явлений обычно останавливается, не достигая уровня нейробиологического анализа индивидуумов, психологическая интерпретация обычно останавливается, не достигая уровня биохимического анализа отдельных клеток, а биохимические исследования редко доводятся до уровня структурного анализа атомов.

Одного множества компонент недостаточно для создания оператора. Оператор появляется только тогда, когда компоненты соотностающим образом связаны друг с другом. Фактически можно представлять себе оператор как своеобразную и мимолетную конфигурацию (своего рода событие), производимую существованием множества частей совместно с организующим связующим процессом. Операторы могут получаться эволюционным путем или конструироваться, как показано в 1.3.

1.3. Части оператора O можно рассматривать как множество объектов S , которого достаточно для создания O , когда действует соответствующий организующий (или связующий) процесс L . Таким образом, множество частей влечет образование оператора в присутствии соответствующего организующего процесса.

$S \text{ влечет } O, \text{ при данном } L$

Одна интерпретация этой формулировки заключается в том, что наличие частей является причиной образования оператора, если с неизбежностью происходит организующий процесс. Эта формулировка дает «конструктивный» подход к образованию оператора. Когда организующие процессы заранее заданы (скажем, в виде генетических механизмов или квалифицированного, чем-то побуждаемого к деятельности мастера), то для производства отдельных устройств (структур) достаточно собирать необходимых материалов.

Как альтернатива результатом оказывается «эволюционный» подход в том случае, если множество компонент рассматривается как заранее существовавшее,

а организующий процесс происходит случайно, $L \text{ влечет } O, \text{ при данном } S \text{ т. е.}$

при наличии совокупности компонент случайно образующиеся связи порождают устройства, которые могут служить в качестве операторов для новой причинной связи.

Оператор образуется только из совместимых компонент. Компонента реагирует только на события особых видов, а если компонента должна работать и делать вклад в действие большего оператора, она должна получать от компонент, которые предшествуют ей в системе, входную информацию надлежащего вида. Часто условие совместимости выполняется довольно простым образом — причинное событие для одной компоненты совпадает с результирующим событием другой. Однако совместимости можно также достичь в случае, когда одна компонента будет реагирующей на конфигурацию исходов от других компонент; например, может потребоваться, чтобы исход первой компоненты «задавал» вторую прежде, чем она сможет породить событие, которое послужит входной информацией для третьей (поршень в автомобильном двигателе не может работать, если в цилиндре нет горючей смеси и искра не воспламеняет горючее). Компонента может быть чувствительной к пороговому числу параллельных событий, произведенных предыдущими компонентами (например, нейрон срабатывает только тогда, когда накопление возбуждения от другого нейрона достигает критического уровня). Исследование совместимости компонент включает определение реакции одного объекта на события, порождаемые другим (или сочетанием других). Это — вид независимого анализа, который должен проводиться заново в каждой новой области исследования.

Часто операторы можно расклассифицировать с точки зрения таксономий или типологий, уменьшая разнородность трактовкой некоторых операторов как эквивалентных, хотя и различающихся в деталях. Есть множество методов для создания этих классификаций: внутренний структурный анализ; сравнения моделей роста

или построения; оценка параллелизма и дополнительности в распределении; разбиение на кластеры с точки зрения следствий, порождаемых различными операторами; группировка с точки зрения функционального подобия согласно подобному расположению в больших системах. Идеальная же цель состоит в определении всеобъемлющей классификации, которая соответствует всем основам одновременно, позволяя таким образом легко переводить информацию одного типа в остальные значения. Подобные исследования независимы в том смысле, что их также надо проделывать заново для каждого класса изучаемых явлений.

Поля. Следствие — это осуществление отдельного состояния или конфигурации состояний в операторе. Поэтому следствие имеет определенное местоположение в пространстве и времени. Любой второй оператор, который должен связываться с первым для образования оператора более высокого порядка, должен быть совместимым; иными словами, он должен реагировать на следствие, производимое первым. К тому же для того чтобы находиться в области предшествующего события, он должен быть соответствующим образом согласован в пространстве и времени. В механических устройствах это, по существу, требование смежности. Для того чтобы события передавались, две компоненты должны быть «соприкасающимися». Однако для охвата других явлений нужна более общая формулировка.

Ради упрощения события могут пониматься как поля, берущие начало в операторе, который их производит, но распространяющиеся в пространстве и времени от своего начального пространственно-временного местоположения. Вообще, чем больше расстояние от начала, тем ниже интенсивность поля и его способность производить действие в совместимом операторе. Таким образом, одна компонента согласуется с другой, только если она находится в действующем поле события другой. Построение оператора требует как минимум такого расположения компонент, что некоторые из них существуют в полях событий других.

Между двумя событиями может существовать причинная связь только в том случае, когда некоторый оператор осуществляет превращение. Теперь мы видим, что оператор, если он должен производить следствие, должен быть совместимым и согласованным с причинным событием. Следовательно, мы можем сказать, что первое событие является причиной второго, только если первое согласуется с некоторым оператором, который на него реагирует и обладает способностью вызвать второе.

Таким образом, событие определенного типа само по себе никогда не является универсальной причиной другого. У него есть возможность быть причиной лишь в том случае, если оно подходящим образом располагается относительно реагирующего оператора. Событие, которое раньше не порождало следствия, может начать делать это, если в его поле доставляется подходящий оператор. Событие, которое действовало как причина в прошлом,

перестало бы действовать так, если бы все относящееся к делу операторы из его поля удалялись.

Общий принцип, рассматриваемый здесь, является жизненно важным в обыденной жизни — суп не разогревается, если кастрюля не помещена на огонь, а голод не утоляется, если пища не проглатывается. Принцип мотивирует конструктивные действия, при которых два или больше оператора совместно вводятся для создания большего оператора со специальной функцией, и дает средства управления нежелательными событиями (например, путем изоляции некоторых людей от средств осуществления их идей или страстей).

Поле события может заполнять пространство в относительно простом физическом смысле. Например, поле магнитного события или события «тепловое излучение» имеет однородную интенсивность по всем направлениям, но градиент интенсивности быстро падает с увеличением расстояния, и реагирующая структура должна была бы находиться вблизи от начала события для того, чтобы событие на нее влияло. Конечно, близость относительна. Поля механических событий имеют такой быстрый пространственный градиент, что компоненты должны «соприкасаться» для того, чтобы влиять друг на друга. Электромагнитные события имеют меньший градиент, таким образом допуская более ослабленный взгляд на близость.

Однако поля событий, и физических, и социальных, могут также направляться, распространяться и сохраняться множеством средств; например, голос человека обычно дает простое круглое поле для коммуникативных событий, но с помощью специальных устройств поле может направляться (бычьим рогом или мегафоном), распространяться (с помощью телефона или посредством передачи через плотное вещество) или сохраняться (магнитофоном). Когда специальные средства или устройства искажают поля событий, идея близости по отношению к одиночному месту в обычных пространстве и времени больше не помогает нам понять, как влияют операторы друг на друга. Мы должны либо действовать в специфических пространствах *ad hoc**, в которых принцип смежности сохраняет свое значение, либо учитывать искажение поля события в обычном пространстве (например, постулированием воспроизведений первоначального события в различных точках пространства и времени для создания подходящего поля в соответствии с обычными принципами расстояния).

Вместе с идеей искаженных полей событий была введена возможность того, что событие в одном месте вызывает следствие в некоем другом отдаленном месте. Это противоречит традиционному канону причинной философии: запрету действия на расстоянии. Конечно же, жесткое требование «отсутствия действия на расстоянии» слишком ограничительно, и подход с полем разумен при видоизменении его в требование «уменьшения возможности

для действия с увеличением расстояния», объясняющее хорошо известные явления вроде магнетизма и гравитации. Может показаться, что искаженные поля открывают возможность отсутствия каких бы то ни было ограничений. Любые два отдаленных события могут быть причинно связанными простым постулированием того, что у первого есть искаженное поле, достигающее оператора для второго события. Требуется добавить лишь, что для распространения полей необходимы операторы — усилители для увеличения их амплитуды, устройства памяти для хранения до более поздних моментов. Следовательно, искажение поля само должно быть причинно объяснимым.

Вообще поле подвергается причинному анализу более мелких операторов, действующих совокупно; например, мы можем говорить о развитии авторитаризма в одном государстве как о событии с полем, влияющим на другие государства. Однако такое поле действует через посредство отдельных личностей (операторы по праву), и при желании мы могли бы исследовать его статистическую механику путем обзора действий индивидуумов. Так как поле порождается событиями на уровне с большой дробностью, его можно распространить от одной точки до другой, только если имеется структурно-причинная основа для распространения на более низком уровне анализа. В частности, искажения полей-событий не произвольны. Они должны иметь материальную основу и быть причинно интерпретируемыми с точки зрения более мелких структур. (В настоящее время для электромагнитных и гравитационных полей структурная основа не постулирована. Это означает, что могут быть существенные возражения против идеи причинного истолкования полей или что искажения таких полей до сих пор не вызывали беспокойства, достаточного для того, чтобы побудить к выбору другого структурного уровня.)

ОПРЕДЕЛЕНИЕ ПРИЧИННОСТИ

В начале этой главы говорилось, что причинность подразумевается, когда «осуществление одного события является достаточным основанием для того, чтобы ожидать, что произойдет другое». Для того чтобы направлять и ограничивать применение принципа причинности при построении теории и при планировании исследования, требуется более точное утверждение.

1.1. Событие *S* является причиной другого события *E*, если и только если

(а) существует оператор, который порождает *E*, который реагирует на *S* и устроен так, что связь между *S* и *E* можно разложить в последовательность совместимых компонент с перекрывающимися полями событий;

(б) осуществления события *S* согласованы с наличием такого оператора — он существует в поле события *S*;

(в) когда условия (а) и (б) выполняются, оператор изолирован от полей событий, отличающихся от *S*, и ни *S*, ни *E* не при-

* Для данной цели (лат.). — Примеч. пер.

существуют заранее, тогда осуществления *C* неизменно начинаются до начала осуществления *E*;

(г) когда условия (а) и (б) выполняются, *C* влечет *E*; иными словами, на протяжении некоторого временного промежутка осуществления *C* всегда сопровождаются осуществлениями *E*, хотя *E* может присутствовать без *C* или же оба события могут отсутствовать.

Условие (а) отражает тот факт, что до того, как создается возможность существования определенной причинной связи, должны быть налицо обстоятельства, наделенные очень развитой структурой. Условие (б) подчеркивает, что события до того, как они смогут иметь следствия, должны согласовываться с такими обстоятельствами. Вместе эти условия определяют физическую основу причинности. Причины связаны со следствиями устройствами (структурами), которые можно точно описать, с определенными местоположениями в пространстве и времени.

Временная направленность определяется в условии (в). Условие (г) формулирует требование логической импликации, направленной от причины к следствию. Относительно последних двух критериев в широком масштабе разработаны современные методы социального исследования. Временной очередностью манипулируют в формальных экспериментах, а сложный статистический анализ служит для проверки логических зависимостей в ситуациях, запутанных многими процессами.

ПРИЧИННЫЙ ВЫВОД

Предшествующее обсуждение было сосредоточено на установлении смысла причинности, с указанием условий, которые характеризуют тип связи между событиями, называемый причинным. Теперь точка зрения изменяется. Определение 1.1 дает условия, которые должны выполняться для того, чтобы события были причинно связанными. Следовательно, эти условия могут служить критериями при решении вопроса о существовании причинной связи между двумя типами события.

Причинный вывод начинается с предположения, что любое событие могло бы быть причиной любого другого события. Затем мы принимаем исключать связи, которые невозможны или неправдоподобны при конкретных обстоятельствах. Этот «исключающий» подход диктуется той предпосылкой, что детерминистические связи охватывают физический и социальный мир и что они могут существовать, даже если люди о них не подозревают. Если бы мы должны были строить наши модели детерминистических отношений с точки зрения тех связей, которые, как мы знаем, существуют, модели могли бы быть опасно неполными; иными словами, они могли бы игнорировать важные процессы и служить основанием для ошибочных выводов. Вместо этого модели разрабатываются путем исключения связей, которые, как мы уверены, не существуют, и сохранения тех, относительно существо-

вания которых мы не уверены, равно как и тех, о которых известно, что они действующие.

1.2. Одно событие не является непосредственной причиной другого, если нет никакого действующего оператора, который мог бы служить основанием этой связи.

Иногда можно с некоторой уверенностью установить тип оператора, который требовался бы для того, чтобы служить основанием данной причинной связи, и сделать вывод, что такого оператора нет; например, на этой основе мы можем заключить, что процент смертности в технологическом обществе не влияет на уровень загрязненности воздуха. Этот принцип обычно применяется в социальных исследованиях, часто без подробного теоретического обсуждения, причем предполагается, что отсутствие особого оператора общезвестно. Поэтому нужно подчеркнуть главную погрешность подхода. Этот принцип требует вывода, что для данной связи вообще нет оператора, а не просто, что отсутствует очевидный оператор. Таким образом, мы всегда должны рассматривать возможность того, что различные виды оператора могли бы служить основой связи, и отбрасывать эти возможности одну за другой, используя всю имеющуюся информацию о предмете обсуждения.

Иногда делается вывод с одним или двумя возможными исключениями, что для данной связи не существует никакого оператора. В таком случае все еще может оказаться возможным достичь интересующего уровня причинного вывода путем тщательной разработки анализа, так чтобы можно было сделать вывод, что между интересующими событиями не существует никакой непосредственной связи, допуская в то же время опосредованное действие через другие события. Например, предположим, что мы делаем вывод об отсутствии оператора, который дал бы возможность обычным колебаниям производительности компании влиять на семейную адаптацию рабочего, за исключением, может быть, психологических механизмов, затрагивающих мораль. При явном включении морали в анализ мы можем сделать вывод, что изменения производительности компании не являются непосредственной причиной изменений семейной адаптации рабочего, хотя они могут быть причиной изменений морали, которые в свою очередь могут вызвать изменения семейной адаптации. Это утверждение существенно сильнее, чем простое высказывание о том, что производительность может или не может влиять на семейную адаптацию рабочего. Возрастающая определенность при этом подходе является общим побуждением для разработки теорий за пределы нескольких событий, представляющих основной интерес.

Оператор для данной связи может существовать, но в определенной ситуации он может оказаться несоединенным или бездействовать по другой причине и может трактоваться как практически отсутствующий; например, радиосообщение об иностранном вторжении могло бы быть причиной паники в 6 часов вечера, когда включены миллионы радиоприемников и телевизоров, но

этого не случилось бы в 3 часа ночи, когда большинство приемников бездействует. Точно так же административный аппарат, который обычно преобразует некоторую информацию на входе в соответствующую информацию на выходе, может перестать действовать, если дезорганизуется бедствием. Таким образом, иногда могут находиться или создаваться особые ситуации, в которых возможны некоторые причинные выводы, даже если те же самые выводы невозможны при более общих обстоятельствах.

1.3. Событие не может быть причиной другого события, если первое событие не согласуется с существующими операторами.

Этот принцип является решающим при планировании классических экспериментов, в которых оператор устраняется от влияния всех событий, кроме одного, с тем чтобы можно было исследовать определенную причинную связь. Это может включать удаление или фиксацию требуемого множества промежуточных операторов для уменьшения влияния событий — идея изоляции; наложение специального оператора, который отклоняет или поглощает деятельность, — идея экранирования; порождение противоположных событий, погашающих те события, которые нежелательны, — идея гомеостатического управления.

Принцип 1.3 может также использоваться при неэкспериментальном исследовании для того, чтобы делать обстоятельные выводы об отсутствии определенных связей; например, культурные события в одном сообществе не имеют последствий в другом, если между ними нет никакого социального или материального взаимодействия. Точно так же исторические события не влияют на индивидуумов, которые физически, социально и информационно устранены от них (например, изолированные заключенные, больные и монахи). События не имеют определенных последствий, если в конкретной ситуации нет требуемых операторов; например, уличные преступления в гетто обычно остаются безнаказанными, если по чистой случайности вблизи не окажется блюстителя закона, и при этих обстоятельствах преступное поведение не вызывает юридических санкций. Преднамеренная тайная деятельность не возбуждает определенных операторы и успешная секретность исключает причинные зависимости между событиями, которые в противном случае могли бы быть им свойственны.

Кроме того, люди иногда создают полную окружающую обстановку (например, церкви и учреждения), в которой они укрываются от влияния расстраивающих их событий, так что некоторые обычные детерминации могут считаться при этих обстоятельствах отсутствующими.

1.4. Причинами события не являются другие события, которые происходят после него.

При экспериментах генерируют отдельное событие для определения того, влияет ли оно на последующие события, и при этом не должны интересоваться, могли ли воздействовать другие события. При неэкспериментальном исследовании выбираются события, которые произошли раньше и влияние которых остается

постоянным (например, половая или расовая классификация индивидуумов), после чего можно оценивать влияние этих событий, не беспокоясь о том, что на них могли воздействовать более поздние события. Действительно, такие ранние события могут быть довольно интересными по аналитическим соображениям, даже если они не особенно интересны теоретически, потому что могут служить инструментом для анализа сложных моделей взаимной причинной обусловленности между другими событиями (эти методы обсуждаются в гл. 5).

Исключение причинной связи с точки зрения временного упорядочения является предметом исследования истории при определенных обстоятельствах и не предотвращает возможности существования причинности при других обстоятельствах; например, раз человек окончил колледж, никакая прибавка богатства, побуждений или знания не будет причиной того, что человек опять станет выпускником колледжа, — это уже было сделано, — но некоторая комбинация данных факторов может стать причиной такого результата для другого человека без диплома.

Иногда может показаться, что принцип временного упорядочения нарушается в том, что более поздние события определяют более ранние случаи; например, вследствие того, что человек должен иметь образование для того, чтобы стать врачом, может показаться, что более позднее состояние «быть врачом» есть причина образования. Однако это связь развертывания. «Быть врачом» подразумевает образование, но не является его причиной. Самое большее, образование вызывается стремлением быть врачом, которое существует до или в течение процесса образования.

1.5. Если событие A происходит без последующего осуществления события B, то A не является причиной B при данных обстоятельствах.

Этот принцип причинного вывода основан на логической импликации, заключенной в причинности. С точки зрения строгой логики единственный пример A без B служит достаточным основанием для заключения, что A — не причина B. Фактически тем не менее установившаяся практика состоит в переходе к статистической точке зрения и принятии в расчет нескольких случаев «A без B», если A обычно влечет B. Одна причина заключается в том, что осуществления событий не могут всегда наблюдаться без неопределенности. Одно событие может маскировать другое, и несколько случаев «A без B» могут обуславливаться ошибками наблюдений. Другая причина становится ясной из выражения «при данных обстоятельствах» в утверждении 1.5. Если A происходит без B (и эвристически предполагается, что B не замаскировано), это означает одно из следующих утверждений:

1) в особой окружающей обстановке нет оператора для преобразования A в B;

2) такой оператор может существовать, но во время осуществления A его не было в поле события A;

3) такой оператор существовал в поле A , но был во время осуществления A разобраным, неустановленным или бездействовал по другой причине.

Обычно, когда мы заключаем, что A не является причиной B , мы хотим думать, что (1) верно, так что имеется некоторая обобщаемость для вывода, по крайней мере, пока физическая обстановка остается одной и той же. Однако если верно именно (2) или (3), то A может быть причиной B при той же самой обстановке в другое время. Следовательно, нельзя сделать вывод, что A не является причиной B , однако в отдельный момент можно заключить, что произошел случай отрицания.

Психологические эксперименты по обучению показывают, что и животные, и люди трактуют случайную причинную связь как достаточно реальную для того, чтобы на нее полагаться. Эта психологическая предрасположенность является умеренной в том отношении, что она исключает меньше причинных связей, чем позволяет логика, и, как говорилось выше, такая склонность к умеренности — в интересах обобщения за пределы временных обстоятельств в момент наблюдения. В действительности эта терпимость к случайному отсутствию причинности означает, что организмы вообще выбирают скорее статистическую, чем строго логическую ориентацию на отношение импликации к причинности. Психологические эксперименты по обучению также показывают, что люди (и животные) сведущи в обнаружении указаний, которые сигнализируют, если оператор действующий. Такая проникаемость соответствует методам уточнения и детализации в науке, согласно которым мы формулируем точные специальные условия, при которых имеет силу причинная связь, и этим путем совершенствуем статистические выводы в пользу более точных утверждений детерминизма. Таким образом, статистическая ориентация в сочетании с продолжающимися усилиями в определении точных условий существования причинной связи служит стратегически в качестве канала, доставляющего сведения о причинных связях, которые трудно наблюдать, а также понимание их и даже тех связей, которые фактически являются только возможностью.

Может показаться, что статистическая точка зрения подрывает прочный фундамент науки — критическое наблюдение, при котором отсутствие следствия доказывает отсутствие связи. Критические наблюдения являются обоснованными только в том случае, если мы сосредоточиваемся на отдельном операторе, о котором известно, что он находится в рабочем состоянии и должным образом согласован с событием A . Только тогда единственный случай « A без B » приводит к обобщенному выводу, что оператор не служит основанием отношения « A является причиной B ». Следовательно, критическое наблюдение является средством определения того, служит ли основой для отдельной связи *точно определенный* оператор, а не того, служит ли основой для отдельной связи некий точно не указанный оператор.

Заметная часть развития в области методологии общественных наук была направлена на создание усложненных статистических методов распознавания соответствия между событиями даже тогда, когда оно является редким или маскируется осуществлением событий, не относящихся к делу. Вследствие умеренности статистического подхода (в смысле возможности отвергнуть относительно немного гипотетических связей) специальные усилия были направлены на разработку методов установления отдельных обстоятельств, при которых видимая корреляция между событиями исчезает, таким образом, показывая, что события в действительности не являются причинно связанными вообще. Эта техника является главным в этой книге, особенно в гл. 3, 4 и 5.

ПОТОКИ

При причинном анализе способности анализировать можно добиться путем добавления дополнительных утверждений о природе реальности и (или) путем анализирования только тех частей реальности, которые удовлетворяют некоторым ограничениям. Возможно несколько различных линий развития, и одна, представленная здесь, была выбрана потому, что основные идеи повсеместно встречаются в современных социальных исследованиях, потому, что аналитические принципы широко разработаны математиками-прикладниками и статистиками и потому, что с этой точки зрения можно объяснить, по крайней мере приближенно, очень много явлений.

С этого момента мы будем рассматривать события как однородные потоки, подверженные увеличению и ослаблению. Согласно этой интерпретации любое интересующее нас событие — это смесь событий более низкого уровня, происходящих неоднократно с данной интенсивностью, пока условия остаются неизменными. Оператор, создающий поток, — это устройство, которое снова и снова порождает одни и те же события и действует непрерывно (по крайней мере, в совокупности) вследствие постоянного притока стимулирующих событий. Не все события могут на самом деле характеризоваться на языке потоков, но идея оказывается легко приспособляемой, и даже разрушительные и созидательные события часто могут характеризоваться как потоки в рамках большего устройства — интенсивность взрывов на поле сражения или интенсивность сборок на заводе. Кроме того, многие социальные и культурные переменные, которые кажутся чисто категориальными, могут пониматься в терминах процессов, поддающихся потоковой интерпретации; например, половая принадлежность людей может рассматриваться «процессуально» как длящееся исполнение некоторых видов ролей во взаимодействии. В этом смысле половая принадлежность является пожизненным потоком событий, закрепленным в биологии и увеличенным ранней социализацией.

Принципы причинности применяются к однородным потокам; иными словами, если A и B понимаются теперь как два различных вида потока, то A является причиной B , если A (т. е. вид протяженного во времени процесса, происходящего с данной интенсивностью) влечет B (т. е. точно указанное количество процесса другого вида), если бы потока A добились одновременно с созданием потока B или до этого и если поток A согласуется с действующим анализируемым оператором, порождающим B . Аналогичным образом применяются принципы причинного вывода. A не является причиной B , если не имеется действующего оператора, если A оказывается несогласованным с любым таким оператором и если уровень процесса B был задан до того, как начался процесс A , или если изменения в процессе A не раскрываются как изменения в процессе B , даже если другими факторами управляют.

С этого момента анализ ограничивается *однородными* потоками, в противном случае приведенные выше принципы могут не иметь силы. Поток однороден, если он вызывает один и тот же результат безотносительно к тому, как он образуется. Таким образом, когда мы имеем дело с однородными потоками, нам не требуется рассматривать, как развивался причинный поток исторически. Для того чтобы провести причинный анализ, нам нужно знать только значение потока и связи между потоком и последующими операторами. Следует заметить, что являющееся однородным потоком при одном анализе, может не быть им при другом; например, продукция винокуренного завода, который иногда разбавляет свой продукт, является однородной по отношению к отправке оптовым фирмам (проблемы транспортировки и хранения одни и те же безотносительно к составу продукта). Этот поток, однако, не однороден по времени в отношении степени опьяняющего воздействия на потребителей.

Аддитивность

Один образ действий, которым анализ потоков отличается от обычного причинного анализа, можно пояснить рассмотрением примера, в котором A и B являются различными причинами E . Если бы A , B и E были обычными событиями, то наличие A , или B , или же обоих событий влекло бы осуществление E . В частности, наличие обоих производило бы то же самое E , что либо A , либо B по отдельности. С другой стороны, предположим, что A , B и E являются потоками. Теперь A или B поодиночке влекут E , но наличие A и B вместе не влечет того же самого E . Скорее, вызывается другой поток E' . Этот новый поток является увеличенным вариантом E . При потоках действие многих причин производит накопление следствий.

То, что называют потоком E , в действительности является множеством различных потоков $E(1)$, $E(2)$, $E(3)$ и т. д., упорядоченным таким образом, что когда два вызываются одновременно, ре-

зультатом является третий поток «более высокого уровня» из того же самого множества. В частности, предполагается, что E имеет неограниченное число потоковых уровней, что уровни эти упорядочены количественно и что на самом деле они ассоциируются с обычными числами таким образом, что алгебра чисел определяет, как накапливаются значения. Предположим, например, что A производит значение E , которое мы идентифицируем как $E(2)$, а B производит значение, называемое $E(3)$. Теперь предполагается, что значения E были определены так, что когда имеются оба, A и B , вызывается значение $E(5)$. Сложный результат является просто суммой отдельных следствий. Таким образом, мы имеем «правило суммирования» для причинного анализа сложения потоков.

Пропорциональность следствия

Дальнейшее осложняющее обстоятельство обнаружится, когда мы вспомним, что причины тоже являются потоками и что значения причинного потока, вызывающего данное следствие, можно определить точнее; например, точное описание связей в приведенном выше примере могло бы иметь следующий вид: $A(4)$ является причиной $E(2)$, $B(9)$ является причиной $E(3)$, а следствие $A(4)$ и $B(9)$ вместе определяется правилом суммирования $E(2) + E(3) = E(5)$. Теперь для различных значений A и B мы должны были бы сделать дополнительные утверждения о связи с E ; например, $A(2)$ является причиной $E(1)$, $B(6)$ является причиной $E(2)$ и т. д. Ясно, что это может стать громоздким при множестве значений пары A и B . Нужен другой упрощающий принцип, сравнимый с правилом суммирования, который позволил бы экономно выразить отношения импликации между всеми значениями двух потоков.

Связи между значениями потоков, направленные от причины к следствию, описываются посредством правила умножения, которое определяет значение следствия как пропорциональное значению причины. Например, в приведенном выше примере точно указывалось, что $A(4)$ является причиной $E(2)$, а $A(2)$ является причиной $E(1)$. Эти детальные описания теперь порождаются, а отношение между A и E описывается для всех значений формулировкой « $A(i)$ является причиной $E(i/2)$ », или просто $E = A/2$. Аналогично точные указания того, что $B(6)$ является причиной $E(2)$, а $B(9)$ является причиной $E(3)$, обобщаются до утверждения, что $B(j)$ является причиной $E(j/3)$, или просто $E = B/3$. Таким образом, основной идеей, заключенной в правиле умножения, является идея, что значение следствия может устанавливаться по значению причины следующим процессом пересчета: задайте численное значение причины, умножьте это значение на постоянную (вроде $1/2$ или $1/3$) и результатом явится соответствующее численное значение следствия. Чем больше значение причинного по-

тока, тем больше значение потока следствия. Действительно, между ними существует строгая пропорциональность.

Фактически полезность принципа можно увеличить путем включения одного незначительного усложнения. Мы допускаем, что пересчет только на множители всегда может «отклоняться» на определенную величину. Следовательно, мы можем оказаться вынужденными после выполнения умножения добавить другое постоянное число; например, если у нас есть отношение вроде « $C(4)$ является причиной $E(2)$, а $C(8)$ является причиной $E(3)$ », то соответствующей формулой пересчета будет « $C(k)$ является причиной $E(1 + k/4)$, или $E = C/4 + 1$ ». Вообще формулой для перехода от некоторого численного значения причинного потока к численному значению потока следствия является $E = a + bC$, где b является константой, используемой в качестве коэффициента, а a является числом, используемым для подгонки результатов на постоянную величину.

Когда для следствия имеется больше одной причины, правило суммирования и принцип пропорциональности определяют исходы совместно. Однако опять допускается, что пересчет всегда может «отклоняться» на постоянную величину. Предположим, например, что поток E вызывается парой потоков — A и B . Тогда полным уравнением для предсказания значения E при данной информации о значениях A и B могло бы быть уравнение вроде $E = 4 + A/2 + B/4$. Если бы E имел три причины вместо двух, формула могла бы быть суммой постоянной и трех членов, относящихся к потокам. Если бы имелись четыре причины, она могла бы быть суммой постоянной и четырех других членов и т. д.

Отрицательные коэффициенты

Уравнения вида $E = a + bC$ дают гибкий и общий способ описания влияния потока C на другой поток E . Действительно, такие уравнения учитывают возможности, которые не были рассмотрены, так как один из двух параметров в уравнении (т. е. либо a , либо b) мог бы иметь отрицательное значение. Фактически знак минус, приписываемый корректирующей константе a , не сильно усложняет вопрос. Он просто означает, что преобразование причинного потока в поток следствия должно корректироваться вычитанием, а не добавлением постоянной. Однако отрицательное значение для коэффициента b указывает на особый род причинной связи между двумя потоками. Это иллюстрируется примером в 1.4.

1.4. Предположим, что причинная связь между двумя потоками описывается уравнением

$$E = 10 - C/2.$$

В этом случае поправочной постоянной является $(+10)$, а постоянной-коэффициентом является $(-1/2)$. Это уравнение описывает особый род связи между потоками, как показывают следующие примеры:

если C есть	2,	то E есть	9
»	4	»	8
»	6	»	7
»	8	»	6
»	10	»	5

Таким образом, чем выше уровень C , тем ниже уровень E . Зависимость потока E от потока C обратная.

Отрицательные коэффициенты представляют обратные, или отрицательные, связи. Чем выше значение причинного потока, тем ниже значение потока-следствия.

Отрицательная связь, по-видимому, требует, чтобы причинный поток служил препятствием деятельности, лежащей в основе потока-следствия, или сдерживал ее. Кроме того, кажется, что отрицательные связи могут существовать только тогда, когда поток-следствие имеет в основе непрерывно длящийся уровень деятельности, независимый от сдерживающего потока. В противном случае нет ничего, что надо было бы сдерживать. Эти представления действительно применимы в определенных случаях отрицательных связей. Предположим, например, что C является популяцией лис на острове, а E является популяцией кроликов (такие величины являются потоками, потому что они представляют агрегированные явления жизни лис и кроликов). Здесь имеется отрицательная связь — чем больше лис, тем меньше кроликов; она существует потому, что лисы препятствуют жизни кроликов. Кроме того, связь может существовать только в том случае, если популяция кроликов поддерживается посредством процессов воспроизводства и пропитания на достаточно высоком уровне. Если бы не было достаточного запаса кроликов, лисы съели бы их всех, и связь между лисами и кроликами на острове перестала бы существовать.

С другой стороны, понятие потоков иногда можно развить с тем, чтобы включить как отрицательные, так и положительные уровни, и в этом случае отрицательные связи могли бы существовать даже тогда, когда поток-следствие не подкреплён другими факторами. Хорошим примером является ситуация, когда два взаимно препятствующих потока объединяются в один поток — назовем его Z . При преобладании одной компоненты-потока уровни Z положительны, при преобладании другой — отрицательны. Формирование установки, по-видимому, является такой переменной. Эмоциональные реакции на стимул могут включать ощущения удовольствия и ощущения неудовольствия, и эти два рода реакции стремятся взаимно препятствовать друг другу. Установка является чистой ответной реакцией — положительной, если преобладают приятные ощущения, отрицательной, если преобладают неприятные. Раз определен такой двухполюсный поток, возможно определить неограниченные отрицательные связи: чем выше

поток C , тем ниже поток E , причем связь продолжается даже в область отрицательных значений E .

В следующих главах будет показано, что потоки часто измеряются произвольными шкалами с нулевой точкой, расположенной на среднем уровне, найденном во множестве наблюдений, — процедура, которая обычно дает отрицательные значения для некоторых наблюдаемых случаев. Эта процедура представляет собой, однако, просто аналитическое удобство, которое исключает постоянные корректировки из причинных уравнений (смотрите обсуждение, сопровождающее 2.12). Вычисленный нуль не является непременно истинным нулем, который свидетельствует об отсутствии потока, а отрицательные значения не обязательно являются настоящими отрицательными, которые означают противодействие положительному потоку.

Терминологические соглашения

В этом месте желательно привести терминологию в соответствии с другой литературой по причинному анализу. В дальнейшем специальный тип потока будет называться «переменной», если только мы не хотим подчеркнуть, что он состоит из событий более низкого уровня. Уравнение, которое описывает причинную связь между двумя переменными, например $E = a + bC$, называется «структурным уравнением». Постоянная корректировки a — просто «постоянная». Множитель b — «коэффициент» или «структурный коэффициент». Постоянная и коэффициент вместе — это «параметры» уравнения.

ЛИНЕЙНОСТЬ

Внимание было направлено на потоки событий, происходящих непрерывно на микроуровне. Вдобавок требовалось, чтобы поток мог измеряться таким образом, что значения, обусловленные различными причинами, аддитивно накапливались, а значения причины были пропорционально связаны со значениями следствия (после того как, возможно, сделана постоянная аддитивная поправка). Вместе эти ограничения, на которых концентрируется внимание, определяют границы особой области причинного анализа — анализа *линейных* связей. На самом деле линейные причинные отношения предполагают просто, что события можно оценить с точки зрения величин, что следствия, обусловленные различными причинами, объединяются аддитивно и что уровни следствия пропорциональны уровням причины после учета постоянной поправки.

Концентрация внимания на линейных связях, принятая в этой книге, оставляет незатронутыми много возможных тем в причинном анализе, потому что не все причинные явления могут описываться на таком ограниченном языке. Рассмотрение только линейных связей оправдывается тем, что они были тщательно изу-

чены, что для обращения с ними есть сильные аналитические методы и что многие интересные системы работают почти линейно, пока условия работы остаются довольно устойчивыми. Кроме того, концентрация внимания на линейных связях позволяет развить в причинном анализе на предварительном уровне много тонких и полезных идей.

Линейная формулировка причинной связи допускает удивительный и полезный перевод с языка взаимосвязанных состояний, который мы использовали до сих пор, на язык взаимосвязанных «изменений состояний», которого полезно придерживаться при обсуждениях. Основные идеи, затрагиваемые при таком переводе, иллюстрируются в 1.5.

1.5. Предположим, что линейная причинная зависимость E от C описывается уравнением

$$E = \frac{C}{2}.$$

Это означает, что уровни E , вызываемые уровнями C , представляют собой точно половину численного значения соответствующего уровня C ; например, $C(6)$ является причиной $E(3)$. Теперь вопрос состоит в том, что случилось бы, если бы мы внесли в уровень C изменение, скажем, на 4 единицы. При продолжении рассмотрения примера это дало бы уровень C , равный $C(6 + 4) = C(10)$, и, используя приведенное выше уравнение, мы получаем в качестве нового уровня E следующее:

$$E = \frac{1}{2}(10) = 5.$$

Новый уровень E на две единицы выше, чем первоначальный; таким образом, изменение C на четыре единицы вызывает изменение E на две единицы.

Можно было бы подумать, что величина изменения E зависит от начальных значений C и E до внесения изменения в C . Однако это не так при линейной связи, что иллюстрируется в таблице.

Если C составляет	E составляет	C увеличивается на 4 единицы до	то E принимает значение	которое представляет собой изменение на
2	1,0	6	3,0	2 единицы
3	1,5	7	3,5	2 единицы
4	2,0	8	4,0	2 единицы
5	2,5	9	4,5	2 единицы

Другими словами, изменение C на четыре единицы всегда вызывает изменение E в линейном соотношении на две единицы. Можно было бы представить дополнительные примеры того, что возрастание C на две единицы всегда вызывает возрастание E на одну единицу, возрастание C на шесть единиц вызывает возрастание E на три единицы и т. д. Кроме того, если мы уменьшаем C на две единицы, E уменьшится на одну единицу, т. е. изменение C на минус две единицы вызывает изменение E на минус одну единицу. Аналогично, изменение C на минус четыре дает изменение E на минус две, изменение C на минус шесть дает изменение E на минус три и т. д.

Эти примеры иллюстрируют тот факт, что уравнение, которое описывает причинную связь между уровнями двух потоков, описывает также связь между

изменениями потоков. Уточним это. Пусть Δ (дельта) представляет численное значение изменения уровней. Тогда на основании соотношения $E = C/2$, которое связывает уровни C и E , мы можем высказать догадку, что

$$\Delta E = \Delta C/2.$$

Предположим теперь, что у нас есть поток F , который определяется потоком A и потоком B следующим образом:

$$F = A/4 + B/3.$$

Читатели могут проверить различные численные значения для того, чтобы убедиться, что здесь формулу, которая точно описывает связи между уровнями, тоже можно перевести прямо в формулу для изменений, а именно:

$$\Delta F = \Delta A/4 + \Delta B/3.$$

Моменты, иллюстрируемые в 1.5, можно сформулировать в виде общего принципа.

1.6. Если причинные связи между потоками линейны по форме, структурные уравнения, которые описывают связи между значениями переменных, можно перевести прямо в форму, которая описывает связи между «изменениями значений». Коэффициенты одни и те же в обеих формах, но постоянная из уравнения для изменений вычеркивается.

Этот принцип обычно не применим, когда связи между потоками отличаются от линейных.

Если нам заранее дано только уравнение для изменений, мы не можем полностью выписать структурное уравнение, потому что не знаем значения постоянной. С другой стороны, уравнение для изменений дает информацию, касающуюся структурного уравнения. Численные значения всех коэффициентов те же самые, и вообще если мы знаем, что изменение потока A является причиной изменения потока B , ясно, что структурное уравнение для B должно включать член для A . Этот момент часто используется при переводе словесных теорий, которые выражают связи между изменениями, в математические и графические формулировки, которые касаются связей между состояниями (см. гл. 2).

Возможность переводить формулировки линейных связей через состояния в формулировку линейных связей через изменения и наоборот особенно полезна при словесных рассуждениях, потому что обычно легче и привычнее говорить об изменениях, являющихся причинами изменений, чем о состояниях, являющихся причинами состояний. (К тому же, конечно, язык изменений исключает надобность постоянно иметь дело с поправочными постоянными.) Раз причинные связи системы точно описаны на языке изменений, можно, используя процедуры гл. 2, выдвинуть структурную формулировку с тем, чтобы дать основу для опознания численных значений коэффициентов (гл. 5). Наконец, раз есть численные значения для коэффициентов, интерпретации опять можно выразить на языке изменений, производящих изменения.

Истолкование структурных коэффициентов

Коэффициент-множитель в линейном структурном уравнении вроде $E = a + bC$ — это величина, имеющая решающее значение. Действительно, последующие главы в этой книге посвящаются главным образом различным видам анализа, затрагивающим такие структурные коэффициенты. Их важность заключается в том, что они могут теоретически интерпретироваться с нескольких различных точек зрения.

Во-первых, как показано в гл. 5, возможно оценить значение структурного коэффициента по эмпирическим наблюдениям переменных. Если значение структурного коэффициента — нуль, то это говорит о том, что никакой действующий оператор не связывает две переменные. Заключение об отсутствии причинности, когда коэффициент — нуль, является преобладающим способом обеспечения выполнения принципа причинного вывода 1.5.

Во-вторых, структурный коэффициент обеспечивает свидетельство силы существующей причинной связи между переменными. Предположим, например, что основной связи « C — причина E » служат два различных оператора и что $E = C/4$ описывает последствия одного, а $E = C/2$ — другого. Причинная связь, опирающаяся на второй оператор, сильнее, потому что при некотором изменении C он вызывает большее изменение E (например, если C изменяется на четыре единицы, то E изменяется на две единицы при использовании второго оператора, но только на одну единицу при использовании первого). Разница в силе здесь отражается в структурных коэффициентах: $1/2$ больше, чем $1/4$. Вообще если несколько различных видов оператора служат основой некоторой причинной связи, самый сильный ассоциируется с наибольшим структурным коэффициентом (не обращая внимания на знаки плюс и минус). Сила связи часто является важным вопросом при планировании систем для достижения определенных целей. Такая информация может также быть полезной при определении того, как различные результаты могут порождаться системами с одной и той же схемой связей, но с различными компонентами.

В-третьих, знак коэффициента показывает, является ли соответствующий оператор основой для положительной связи между потоками (возрастания вызывают возрастания) или для отрицательной, обратной связи (возрастания вызывают убывания). Такая информация становится интересной, когда имеют дело с взаимосвязанными операторами, в которых изменения, внесенные в одну переменную, проходят через сеть связей, иногда вызывая увеличение других переменных, а иногда — уменьшение. Методы прослеживания последствий в таких случаях описаны в гл. 2.

Наконец, структурные коэффициенты всегда связаны с некоторым оператором. Коэффициент может рассматриваться как в высшей степени сжатое описание оператора, которое особое значение придает его результатам в связывании потоков, а не его

внутренней структуре и механизмам. Время от времени обнаруживают, что операторы, служащие основой для связи между двумя потоками, можно систематически описать с точки зрения некоторого свойства, величина которого соответствует структурному коэффициенту каждого; например, твердые тела являются операторами, которые превращают силы F , приложенные к телу, в ускорение тела A ; масса тела M определяет соответствующий коэффициент согласно элементарному закону физики: $A = (1/M)F$. Возможно, подобные законы есть и в общественных науках.

ПРИЧИННЫЕ СИСТЕМЫ

Немногие интересные явления, особенно в общественных науках, зависят только от одной причины и следствия. Явления, изучаемые общественными науками, обычно включают много различных видов события, определяемых некоторым количеством различных факторов, причем каждый влияет на некоторое количество других факторов. Сети причинных отношений, в которых многие различные переменные связаны друг с другом, называются системами.

Превращение причинных связей в системы, или сети причинной обусловленности, делает причинность трудной для изучения. В то же время именно это свойство делает причинность более интересной для нас. Трудности появляются потому, что сложные данные и анализ нуждаются в исследовании причинного отношения, введенного в сеть других причинных отношений. Дополнительная же выгода проистекает из того, что причинные сети вызывают большое разнообразие явлений — колебание, рост, распад, управление и усиление.

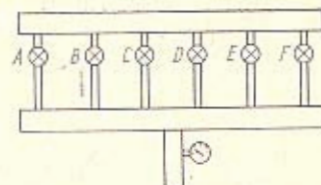
Избитая фраза гласит, что система больше суммы ее частей. Эта фраза вводит в заблуждение постольку, поскольку она предполагает, что система не поддается пониманию и рациональному анализу, но она верна, когда означает, что импликация систем причинных связей представляют собой гораздо большее, чем мы могли бы предполагать, глядя на каждое причинное отношение в отдельности. Для того чтобы охватить сложность реальных явлений, причинный анализ обязательно должен быть развит в системный анализ.

Множественные причины

Данное следствие может быть вызвано любой из нескольких причин. Такая множественная причинная обусловленность вносит в анализ осложнения, особенно когда мы заинтересованы в том, чтобы сделать вывод о линейных причинных связях между переменными. Основная проблема состоит в том, что когда имеются множественные причины, то следствие может составляться многими различными способами. Таким образом, становится трудно

определить соответствие между следствием и другой переменной, которая может быть причиной. Проблема иллюстрируется в 1.6.

1.6.



Эта система состоит из шести труб-ответвлений, отходящих от источника высокого давления. Каждая труба снабжается водопроводным краном неизвестной пропускной способности, и все трубы входят в одну выпускную трубу. Пропускная способность измеряется прибором, отсчитывающим число литров в минуту. Не изображенный на рисунке исследователь может управлять водопроводным краном A , а пять обезьян управляют водопроводными кранами B , C , D , E и F ; обезьяны обучены поворачивать соответствующие водопроводные краны в одном из двух направлений. Ученый хочет определить воздействие крана A на общий расход жидкости. Очевидная вещь, которую он должен сделать, — это открыть кран A для того, чтобы увидеть, насколько увеличится поток. Он делает это, но в тот же самый момент одна обезьяна прикрывает кран B , другая обезьяна немного приоткрывает кран C , обезьяна, управляющая краном D , спит, четвертая обезьяна до конца открывает кран E , а пятая обезьяна полностью закрывает кран F . Чистый результат состоит в отсутствии изменения или даже в уменьшении потока. Конечно, регулирование крана ученым оказывало причинное влияние на полный поток. При открывании крана он имел достаточное основание ожидать увеличения потока. Однако вследствие всех других событий это не обязательно происходило. Следствие особой причины может маскироваться следствиями других причин.

Множественная причинная обусловленность усложняет причинный анализ, потому что она создает ситуацию, в которой значение следствия определяется не только той причиной, на которой сосредоточено внимание, но также и другими. При попытке исследовать связь между конкретной причиной и следствием все другие причины действуют в качестве мешающих факторов, которые заводят анализ в тупик и делают его бесполезным. Для того чтобы исследовать интересующее отношение, мешающими факторами надо как-то управлять. Имеются три стратегии.

Один подход, связанный главным образом с классическим экспериментированием в физических науках, предполагает изоляцию интересующего отношения. Любое возмущение зависит от причинного отношения, поэтому возмущения происходят вследствие существования некоторого оператора, согласованного с неким событием-причиной. Следовательно, если оператор для возмущения можно отсоединить, сделав недействующим, или изолировать от своих событий-стимулов, то эта причинно-следственная связь больше не окажет своего влияния. Кроме того, если причинные связи для всех возмущающих факторов разрываются, связь между остающейся причиной и следствием можно анализировать, не беспокоясь о возмущениях. Применение этого подхода

в примере 1.6 потребовало бы устранения обезьян, приваривания кранов в закрытом положении или герметической закупорки их. Использование этой стратегии не всегда так просто, как это может показаться. Разрушение каждого мыслимого оператора или управление всеми нежелательными событиями требует значительного искусства и знаний. Эта стратегия все-таки может оказаться применимой только в задачах, уже содержащих достаточно научных знаний, которые являются основанием планирования эксперимента.

Второй подход к проблеме возмущений состоит в создании типа пассивного контроля над мешающими факторами путем подробного их наблюдения с целью определения, когда тот или другой создает нежелательное следствие, так чтобы его можно было не принимать в расчет или в противном случае регулировать. В своей простейшей форме это означало бы ограничение анализа до периодов, в которые мешающие факторы не создают возмущений. (При применении к проблеме, обсуждаемой в 1.6, эта стратегия привела бы ученого к испытанию крана А только в то время, когда все обезьяны спят или же отвлечены.)

В-третьих, к проблеме мешающих факторов можно было бы подходить с точки зрения статистики. Допускают, что любое одиночное наблюдение безнадежно запутывается мешающими факторами, но если наблюдается достаточно отдельных случаев, в которых предполагаемая причина действует на определенных уровнях, то должна существовать возможность определить, имеет ли данная причина предполагаемое следствие *в среднем*. При усреднении по многим случаям следствия мешающих факторов будут погашать друг друга, что оказывается многообещающим. В применении к проблеме в 1.6 эта стратегия требовала бы, чтобы ученый открыл и закрыл кран А 200 раз или около того и добился бы на выходе среднего результата. При рассмотрении в среднем он допустил бы, что случайные события одного испытания погашают случайные события другого.

Основным требованием здесь является то, что мешающие факторы не должны быть связанными с интересующим нас причинным событием. Если это неверно, — если одно мешающее событие всегда имеет тенденцию осуществляться вместе с интересующим нас причинным событием, — то следствие этого возмущения при усреднении не исчезало бы. Кроме того, всегда необходимо получать наблюдения в присутствии и при отсутствии предполагаемой причины или с предполагаемой причиной, заданной на различных уровнях, для того чтобы выяснить, вызывает ли предполагаемая причина какое-нибудь различие.

Идея статистического усреднения развивалась в двух различных направлениях. Наиболее точной процедурой (но часто невыполнимой в социальных ситуациях) является статистический эксперимент, в котором отсутствие согласования между мешающими факторами и интересующей предполагаемой причиной гарантируется случайным отнесением подлежащих наблюдению ситуаций

в два различных множества, с последующим обеспечением гипотетического причинного события только для одного из них. Тогда, так как предполагаемая причина не согласована с любым из возможных мешающих факторов, ее следствие может определяться путем нахождения средней величины результата в присутствии предполагаемой причины и сравнения ее со средней величиной, найденной, когда предполагаемая причина отсутствует.

При втором подходе наблюдения мешающих факторов используются для определения групп результатов измерений, в которых каждое возмущение имеет постоянное значение. Тогда значение переменной-следствия в каждой группе подправляется арифметически до величины, которую она имела бы, если бы не было возмущений, после чего можно исследовать интересующую связь для того, чтобы увидеть, имеется ли соответствие между гипотетической причиной и регулируемым следствием, служащее признаком причинности. При действительных применениях подход с регулировкой на самом деле равнозначен исследованию нескольких причинных отношений одновременно. Процедура поправки на воздействия второй причины при исследовании первой требует проведения регулировки для первой причины с тем, чтобы определить, на сколько регулировать для второй причины. Сначала это звучит безнадежно закрученным и сложным. Фактически же существуют процедуры для того, чтобы произвести это разделение при данных многочисленных наблюдениях всех относящихся к делу переменных (принципы обсуждаются в гл. 3 и 5).

Множественные следствия

Причинная переменная может быть определяющим фактором нескольких различных следствий. Одним важным результатом такого моделирования является ложная корреляция, при которой следствия имеют тенденцию возрастать и убывать совместно ввиду их обоюдной зависимости от одной причины. Ложную корреляцию можно рассматривать как проблему, когда мы хотим проанализировать соответствие между переменными. С другой стороны, ложные корреляции являются основой вызывающего существенный интерес явления, свойственного системам.

Когда большое количество подобных следствий зависит от одной или нескольких причин, можно создать новую переменную более высокого уровня. Значение причины влияет на значения всех следствий, таким образом создавая согласованность их величин. Тогда, если причина начинает носить характер повторяющегося изменения, она дает начало модели повторяющихся и согласованных следствий, которые можно в совокупности считать другим потоком более высокого уровня. Этот новый поток может служить в качестве элемента при причинном анализе на другом уровне абстракции или может включаться в качестве упрощаю-

щего средства в первоначальный анализ; например, первоначальная причина может пониматься как определяющая сложный поток при простом игнорировании отдельных следствий.

Кроме того, мы можем рассматривать элементарные следствия как зависящие от переменной более высокого уровня, которая сама представляется как зависящая от причины. Этой искусственной концептуализации может оказаться достаточно при некоторых видах анализа, если большинство элементарных следствий имеют параллельные связи с другими переменными в интересующей системе и если составное значение большого количества элементарных следствий почти независимо от единственного значения какого-нибудь одного из них. Такие «причинные приближения» нередки в научном и обыденном мышлении, а идея часто встречается в общественных науках. Когда отдельные действия имеют одни и те же причины, согласованность действия может считаться отдельным явлением, и это явление даже может «объяснять» отдельные действия, если это полезно для анализа.

Взаимная причинная обусловленность

Взаимная причинность существует всегда, когда две переменные связаны друг с другом двумя причинными преобразованиями — одним в каждом направлении. Такое двойное отношение является петлей обратной связи. В петле обратной связи изменение *A* является причиной изменения *B*, а это изменение в свою очередь направляется обратно, для того чтобы вызвать другое изменение *A*, которое затем вызывает другое изменение *B* и т. д. Таким образом, начальное изменение *A* вызывает целый ряд последующих изменений и *A*, и *B* (приращение на каждом цикле), и вследствие принципа аддитивности изменения каждой переменной накапливаются. Петли являются существенным элементом в трех важных явлениях, свойственных системам (см. гл. 6 о динамике петель).

Усиление. Если изменение *A* вызывает аналогичное изменение *B* (например, возрастание вызывает возрастание), а изменение *B* вызывает аналогичное изменение *A*, механизм равносильно усиливающей системе, которая берет малые изменения и кумулятивно накапливает их в большие. Простой пример этого дает нам связь между сбытом и рекламой. Усиление рекламирования вызывает увеличение продаж, больший объем продаж ведет к большему рекламированию, добавочное рекламирование влечет еще больший объем продаж и т. д.

Управление. Предположим, что изменение *A* вызывает аналогичное изменение *B*, но изменение *B* вызывает обратное изменение *A*. Тогда, если значение *A* повышается, *B* повышается, но вследствие того, что *B* повышается, *A* теперь понижается. Понижение *A* является причиной понижения *B*, но вследствие того, что

B понижается, *A* теперь повышается. Это устройство находится в центре многих управляющих механизмов, включая социальное управление. Классическим примером в технике является терморегулятор. Когда температура в помещении слишком высока, источник тепла прикрывается или отключается; это позволяет температуре понизиться, а затем нагревание усиливается; таким образом, колебание температуры в комнате остается в ограниченном диапазоне — оно управляемо. Механизмы социального управления могут работать подобным же образом; например, рост числа преступлений определенного вида ведет к более бдительному надзору за соблюдением закона, уменьшающему сферу распространения преступлений, что в свою очередь уменьшает бдительность; поэтому теоретически доля преступлений остается в ограниченном диапазоне.

Неустойчивость. Если приращения изменения с каждым циклом становятся все больше и больше из-за петли обратной связи и продолжают накапливаться с каждым циклом, в конце концов значения переменных могут достичь уровней столь крайних, что операторы, служащие основой процессов, будут подвергаться опасности; например, микрофон, помещенный вблизи громкоговорителя, образует неустойчивую петлю обратной связи и создает вой возрастающей громкости, что в конце концов может испортить узлы системы. Связи между численностью населения и уровнем технологии могут составлять пример из области общественных наук (в предположении, что непосредственный или опосредованный рост численности населения вызывает повышение уровня технологии, а более высокая технология порождает большую численность населения). По крайней мере в современный период, рост и численности населения, и уровня технологии носит характер взрывов, хотя мы, по-видимому, все еще не убеждены, что основополагающие социальные структуры подвергаются опасности. Эти примеры относятся к неустойчивому усилению, при котором изменения значения переменной неослабно продолжают в одном направлении до поломки. Петли управления тоже могут быть неустойчивыми. В этом случае величины переменных колеблются вверх и вниз, доходя до все больших и больших крайностей, пока не произойдет разрушение одного или обоих причинных отношений.

Не все петли неустойчивы. В последующих же главах нас интересуют только системы, в которых петли устойчивы.

ИСТОЧНИКИ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Причинность является одним из традиционных вопросов в философии науки, и по этому вопросу имеются многочисленные философские очерки. Последней и наиболее заметной из них является работа Mario Bunge. *Causality: The Place of the Causal Principle in Modern Science* (Cambridge, Mass.; Harvard University Press, 1959).

Среди авторов, пишущих о теории познания, есть такие, которые подчеркивают, по существу, объективную природу науки: например, Karl R. Popper. *The Logic of Scientific Discovery* (New York, Harper & Row, 1968; first published 1934) и Carl G. Hempel. *Fundamentals of Concept Formation in Empirical Science* (Chicago, University of Chicago Press, 1952). Другие подчеркивают релятивизм и субъективность основ научного знания, как Thomas S. Kuhn*. *The Structure of Scientific Revolutions* (Chicago, University of Chicago Press; 2nd ed. enlarged, 1970), Norwood R. Hanson. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science* (London, Cambridge University Press, 1958) и Roger Poole. *Towards Deep Subjectivity* (New York, Harper & Row, 1972). Некоторые ученые-социологи недавно выказали интерес к эпистемологии и психологии познания, что отражено, например, в Jean Piaget. *Psychology and Epistemology: Towards a Theory of Knowledge* (New York, Viking Press, 1971, Translated by Arnold Rosin); Donald T. Campbell. *Natural Selection as an Epistemological Model*, Chapter 3 в Raoul Naroll and R. Cohen (Eds.). *A Handbook of Method in Cultural Anthropology* (Garden City, N. Y., Doubleday-Natural History Press, 1970); Oswald Werner and Joann Fenton. *Methods and Theory in Ethnoscience or Ethnoepistemology*, Chapter 29 в Naroll and Cohen, *ibid.*; Edward E. Jones et al. *Attribution: Perceiving the Causes of Behavior* (Morristown, N. J., General Learning Press, 1972) и Harold H. Kelley. *The Processes of Causal Attribution*, *American Psychologist*, 28 (1973), 107—128.

Исследователи в области общественных наук, в частности, касались методологии причинного вывода в ситуациях, в которых классическое экспериментирование невозможно. Двумя плодотворными работами являются Hubert M. Blalock, Jr., *Causal Inferences in Non-experimental Research* (Chapel Hill, N. C., University of North Carolina Press, 1961) и Donald T. Campbell and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research* (Chicago, Rand McNally, 1963).

Современное развитие общей теории систем сильно обобщило и оживило причинный анализ введением идеи операторов и приданием особого значения сетям причинности. Попытки дать беглый очерк общей сферы системного анализа и кое-что из того, что в ней подразумевается, сделаны в W. R. Ashby**. *Introduction to Cybernetics* (New York, Wiley, 1963) и Ludwig von Bertalanffy. *General System Theory: Foundations, Development, Applications* (New York, George Braziller, 1968). Хорошо написанное введение в методы имеется в Van Court Hare, Jr., *Systems Analysis: A Diagnostic Approach* (New York, Harcourt, Brace & World, 1967).

* Имеется русский перевод: Кун Т. Структура научных революций, 2-е изд. М., Прогресс, 1977. — *Примеч. пер.*

** Имеется русский перевод: Эшби У. Р. Введение в кибернетику. М. — Л., ИЛ, 1959. — *Примеч. пер.*

Лекции, интересные для ученых, работающих в области общественных наук, собраны в Walter Buckley, Ed. *Modern Systems Research for the Behavioral Scientist: A Sourcebook* (Chicago, Aldine, 1968). Его книга Walter Buckley. *Sociology and Modern Systems Theory* (Englewood Cliffs, N. J., Prentice-Hall, 1967) сосредоточивает внимание на значении современного системного анализа для социальной теории.

Анализ потоков можно формализовать в значительной степени в рамках традиционной математики, что иллюстрируется в следующих главах этой книги. Однако строгая формальная логика, пригодная для анализа структур операторов, была разработана математиками только недавно в виде алгебры категорий. Категориальная алгебра дает средства анализа структуры общих систем J. A. Goguen. *Mathematical Representation of Hierarchically Organized Systems*, pp. 112—128 в E. O. Atlinger, Ed. *Global Systems Dynamics* (New York, Wiley, 1970) и I. Băianu. *Organismic Supercategories and Qualitative Dynamics of Systems*, *Bulletin of Mathematical Biophysics*, 33 (1971), 339—354. Она применялась также более специальным образом к социальным структурам: Francois Lorrain and Harrison White. *Structural Equivalence of Individuals in Social Networks*, *Journal of Mathematical Sociology*, 1 (1971), 49—80. Изучение категорий требует математической подготовки, но полезным введением для людей с требуемой подготовкой является Saunders Mac Lane and G. Birkhoff. *Algebra* (New York, Macmillan, 1967). Более углубленную трактовку дает Mac Lane. *Categories for the Working Mathematician* (New York, Springer-Verlag, 1971).

Принципы построения словесных теорий периодически возбуждали интерес ученых в области общественных наук. Некоторые трактовки этого предмета имеются в Arthur L. Stinchcombe. *Constructing Social Theories* (New York, Harcourt, Brace & World, 1968), Hubert M. Blalock, Jr. *Theory Construction: From Verbal to Mathematical Formulations* (Englewood Cliffs, N. J., Prentice-Hall, 1969), Jerald Hage, *Techniques and Problems of Theory Construction in Sociology* (New York, Wiley, 1972) и Abraham Kaplan. *The Conduct of Inquiry* (San Francisco, Chandler, 1964). Пример построения такой теории для науки с точки зрения систем дается в William T. Powers. *Behavior: The Control of Perception* (Chicago, Aldine, 1973).

УПРАЖНЕНИЯ

1. Предположим, что почти все потребители героина в американском обществе до того, как они приняли более сильный наркотик, были знакомы с марихуаной. Следует ли отсюда, что употребление марихуаны является причиной употребления героина? (Употребление какого наркотика влечет знакомство с другим? Что предшествует чему?)

2. Статус индивидуума можно представлять себе как интенсивность привилегированных действий, которые производит этот индивидуум, или, в качестве альтернативы, как интенсивность действий подчинения, которые присутствие

этого индивидуума вызывает у других. Кто является оператором, наделяющим статусом, при каждой из этих интерпретаций статуса? Проиллюстрируйте, как различные точки зрения могут привести к различным тактикам увеличения статуса индивидуума.

3. Мобильность между поколениями часто исследуется путем рассмотрения образования и социально-профессионального статуса отца вместе с образованием и социально-профессиональным статусом его сына. Какие причинные связи между этими четырьмя переменными можно исключить путем использования правил причинного вывода?

4. Эмпирические исследования мобильности между поколениями в современной Америке показали, что соответствие между социально-профессиональными статусами отцов и сыновей ничтожно для сыновей с одним и тем же уровнем образования. Другими словами, анение того, что отец имеет престижную профессию, не позволяет нам предсказать многое о статусе профессии сына относительно профессий других сыновей с тем же самым образованием. Как эта эмпирическая информация видоизменяет модель, разработанную в упражнении 3?

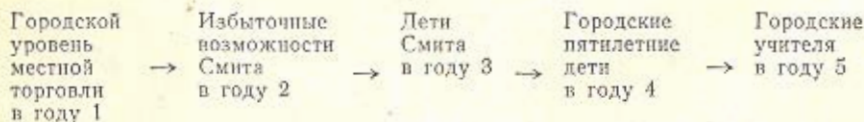
5. Каким образом можно было бы рассматривать реактивно-истребительную авиацию страны А как поток? Разумно ли предположить, что такой поток линейно связан со своими причинами? Предположим, что воинские части оснащены самолетами страны А и самолетами нейтральной страны. Укажите какие-нибудь обстоятельства, при которых они могут считаться однородными, и обстоятельства, при которых эти части должны были бы трактоваться как неоднородные.

6. Если безработица снижается приблизительно до 5%, правительство, вероятно, должно выбрать такую политику, как будто никакой безработицы нет вовсе, как будто бы цифра 5% в действительности является нулем. Если безработица падает ниже 5%, правительство может даже начать ликвидацию рабочих мест, действуя так, как если бы наблюдалась отрицательная безработица какого-либо вида. По-видимому, это так потому, что 5%-ная безработица является нормативным уровнем, и норма определяет практическую нулевую точку для этой переменной. Какая разница между нормативно определенной нулевой точкой и истинным нулем?

7. Предположим, что увеличение доходов людей делает их большими сторонниками существующего строя, а повышение их образованности делает их большими противниками существующего строя. Предположим также, что ориентация на существующий строй может измеряться как одна переменная установка с положительными и отрицательными значениями, соответствующими «за» и «против». Опишите вероятные положения индивидуумов со следующими видами несогласованности статусов: высокий уровень образования, но низкий доход; слабая образованность в сочетании с чрезмерным доходом. Чего следует ожидать от индивидуума с согласующимися статусами, а именно с доходом, должным образом соответствующим уровню образования? Имеется ли какое-нибудь основное различие в отношении бедного необразованного индивидуума и богатого образованного индивидуума?

8. Во время местного экономического бума мистер и миссис Смит приходят к заключению, что их уровень материальной удовлетворенности выше и свободного времени гораздо больше, чем они когда-либо надеялись достигнуть. Одним из следствий этого является то, что они прекращают пользоваться противозачаточными средствами, пытаются завести ребенка и даже с этой целью обращаются за медицинской помощью. В результате миссис Смит в следующем году рождает двойню. Примерно через пять лет растущая семья Смита является одним из факторов, которые вынуждают директора городской школы увеличить количество учителей для первого класса.

(а) Покажите, как во всех городских семьях происходят параллельные события, путем использования множественных цепей стрелок, подобных следующей:



(б) Используйте «причинное приближение» со второй и третьей переменными в цепи (уровни жизни и уровни рождаемости) для получения упрощенного представления на макросоциологическом уровне. Каковы преимущества и недостатки этого представления?

9. Опишите промежуточные процессы, посредством которых избыточные возможности семьи, или уровень жизни, могли бы определять коэффициент воспроизводства. В частности, рассмотрите, что могла бы сделать семья с неожиданными экономическими излишками для увеличения рождаемости и что семья могла бы сделать для сдерживания рождаемости, когда ее уровень жизни ниже желаемого. При предположении, что все эти действия действительно образуют один оператор, как он отличается от физического оператора вроде автомобиля, который непрерывно реагирует на увеличение расхода горючего возросшей скоростью? В каком смысле он подобен?

10. Рассмотрим двухрасовое общество с людьми P одной расы и людьми R другой. Люди P , по существу, контролируют городское недвижимое имущество, торговлю и законодательный совет, посредством чего они в состоянии сдерживать размер совокупности R . Степень власти людей P зависит, однако, непосредственно от размера совокупности P , и, следовательно, сдерживание R также зависит от размера совокупности P в данный момент. Кроме того, случается, что люди P пугаются людей R до такой степени, что простое присутствие R в окрестности терроризирует людей P . Действительно, занятие каждым R местожительства в городе, как правило, вызывает то, что несколько P уезжают. Таким образом, увеличение числа людей R вызывает пропорциональное убывание людей P .

(а) Используйте стрелки (как в упражнении 8) для того, чтобы показать причинные связи в этой системе, буквами s и d пометьте эти стрелки.

(б) Представьте себе s и d как структурные коэффициенты для двух связей. Каковы знаки этих чисел? Образуют ли две связи вместе систему управления?

(в) Предположим, что абсолютные значения s и d больше, чем 1,0, и что внезапная миграция производит существенное увеличение людей R . Что будет конечным следствием миграции R в городской состав?

Достоинством качественных утверждений общественных дисциплин является то, что они широко понятны. Их недостаток — в ограниченных возможностях логических выкладок. Поэтому весьма желательна разработка математизированных формулировок, которые обеспечивали бы возможность выведения неочевидных следствий данной теории. Кроме того, математические утверждения обычно пригодны для сопоставления теории с реальностью и проверки ее адекватности с помощью статистических процедур.

Ученые-социологи часто более искусно владеют словом, нежели математическим аппаратом, и для них формулы означают скорее бессмыслицу, чем возможность увеличения логической мощи. Надеюсь уменьшить математическую отчужденность этого типа, в данной работе мы делаем акцент на конструировании теоретических моделей, основанных на специальном виде математических формулировок — потоковых графах, представляющих уравнения в наглядной форме. Следуя небольшому числу поддающихся интерпретации правил, можно получать математические выводы, используя лишь процедуры вычерчивания и изучения этих диаграмм. Потоковые графы представляют собой мост между качественными теориями и более абстрактными представлениями других теорий, использующих язык уравнений.

ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ ПРИЧИННЫХ ОТНОШЕНИЙ

Построение потокового графа, представляющего теоретическую модель, начинается с введения обозначений всех переменных, подлежащих рассмотрению.

II.1. Каждая переменная представляется на диаграмме символом или аббревиатурой — начальными буквами словосочетания.

Если бы социально-экономический статус был важной теоретической переменной в исследуемой системе, то диаграмма могла бы начинаться с краткого обозначения SES («социально-экономический статус») или, например, просто X. Краткие обозначения не имеют математического смысла, и их использование вызвано тем, что более длинные обозначения ухудшают графическую картину, делая диаграмму слишком громоздкой.

По условию буквы V, W, X, Y и Z служат для обозначения переменных. Иногда одна и та же буква используется несколько раз, но с различными нижними индексами (например, X_1 , X_2). В случае малоразмерных систем можно обойтись и без индексов, поскольку их употребление придает модели более технический вид, чем это есть на самом деле; но когда имеют дело с взаимоотношениями многих переменных, применение индексов действительно упрощает обозначения.

При употреблении абстрактных символов — X, Y, Z — рекомендуется сопроводить диаграмму экспликацией — указателем того, что обозначает каждый символ; это делает диаграмму вполне интерпретируемой и без дополнительного обращения к справкам.

Фактический процесс конструирования потокового графа предполагает рассмотрение одной переменной в некоторый момент времени и исследование ее причинных отношений с другими переменными системы. Процедура построения диаграммы, очерченная далее, применяется только для непосредственных причинных отношений. Причинное отношение называется непосредственным, если изменение значений одной из переменных ведет к изменению значений другой переменной — без необходимого изменения специально указываемой третьей переменной между ними. Например, если теория утверждает, что X имеет влияние на Z, но только при воздействии X на Y, то отношение между X и Z будет косвенным, хотя и существует непосредственное причинное влияние X на Y, и Y на Z.

II.2. Сплошная (предпочтительно прямая) линия между символами обозначает непосредственную причинную связь для данной пары переменных. Стрелка указывает направление причинности — в сторону зависимой переменной.

2.1. $X \rightarrow Y$ означает: «X является причиной Y»

Вообще говоря, желательно помещать причину слева или сверху, а следствие — справа или внизу. Это соображение не имеет математического обоснования, и не стоит следовать ему неукоснительно. Однако оно помогает аккуратному построению диаграммы и усиливает ее интерпретируемость тем, что совокупность входов расположена в левой или верхней части диаграммы, а совокупность выходов — в нижней или правой части.

II.3. Каждый причинный путь обозначается единственным определяющим символом.

Путевой символ допускает две интерпретации. Во-первых, он ясно выражает, что оператор должен устанавливать определенное причинно-следственное отношение, и дает способы спецификации этого оператора при обсуждении. Во-вторых, символ означает количественное описание оператора — каким образом этот оператор преобразует изменение одной переменной в изменение другой.

В качестве количественной характеристики символ может означать характеристическую таблицу (как в случае усилителя

высокого класса), сложную формулу или единственное число. В стадии построения диаграмм теории не содержится ограничений на сложность описания. Однако в данной работе мы будем иметь дело только с линейными преобразованиями, которые могут быть выражены единственным числом, и в этом случае путевой символ представляет собой структурный коэффициент частного причинного отношения.

Как и в случае переменных, для обозначений структурных коэффициентов принимаются некоторые условия. Обычно используются стандартные символы a, c, d, e, f, g, p и q (буква b резервируется для других целей). В системах с большим числом отношений использование различных букв становится малоприменимым, и поэтому необходимо добавлять индексы. При индексировании структурному коэффициенту всегда приписываются два индекса: первый индекс характеризует зависимую или выходную переменную данного отношения, второй — независимую или входную переменную. Например, если X — причина Y и буква a обозначает структурный коэффициент, то индексация приводит к виду a_{yx} . Если же сами переменные имеют индексы, то структурному коэффициенту приписываются только индексы этих переменных. Например, если X_1 — причина X_2 , а коэффициент обозначается буквой d , то окончательный вид будет d_{21} .

2.2. $X \xrightarrow{a} Y$ означает, что X есть причина Y , а символ a характеризует тип операции, которая преобразует изменения X в изменения Y . В данной книге предполагается, что преобразование линейно. В этом случае a оказывается также структурным коэффициентом, описывающим это самое линейное преобразование.

a_{yx}
 $X \xrightarrow{a_{yx}} Y$ означает то же самое, что и выше, но здесь структурный коэффициент определен индексом более конкретно.

a_{21}
 $Z_1 \xrightarrow{a_{21}} Z_2$ означает, что Z_1 есть причина Z_2 , и a_{21} снова обозначает структурный коэффициент, который описывает, как величина Z_1 преобразуется в величину Z_2 .

Если численное значение структурного коэффициента неизвестно, но знак его известен, все равно на поясняющей части диаграммы может быть дана полезная информация. Например, если теория указывает положительную связь (возрастание X приводит к росту Y), диаграмма будет содержать запись вида $a_{yx}(+)$. Если из теоретических соображений связь должна быть отрицательной, так что возрастание Y приводит к убыванию Z , то записывается $a_{zy}(-)$. Когда же количественное значение коэффициента полностью известно (например, $a_{yx} = 0,53$), то над стрелкой соответствующего графа вместо символа пишется это число.

СЛОЖНОСТИ В ДИАГРАММАХ

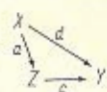
Графическое изображение системы получают постепенно, рассматривая пары переменных и используя указанные выше принципы снова и снова, включая в диаграмму все больше перемен-

ных и стрелок, пока не будут разобраны и отображены все важные стороны теоретической модели. Поточный граф, построенный таким упрощенным способом, в конечном виде представляет систему как целое. Однако при построении поточного графа могут возникать определенные затруднения.

Ветви

Система была бы малоинтересной, если бы она включала только отношения между переменными изолированных пар и не содержала отношений, связывающих разные пары. Поэтому на практике некоторые переменные будут соотноситься не только с какой-то одной переменной, но и с несколькими другими. В данном случае множество стрелок будет направлено на эту переменную или (и) выходить из нее. Когда несколько разных стрелок входят в переменную, это означает, что зависимая переменная имеет много причин. Когда же несколько стрелок выходят из данной переменной, это значит, что данная переменная является источником многих последствий.

Задача вычерчивания становится более трудной, когда необходимо изобразить несколько причин и следствий. Например, если у X две зависимые переменные — Y и Z , то как их надо изобразить на диаграмме? По условию они обе должны быть помещены справа или снизу от X , но должен ли быть Y сверху и слева от Z или наоборот? Разобраться в этом нет никакой возможности, пока не будут изучены все отношения Y и Z . Так, позже может выясниться, что Z имеет причинное влияние на Y , и тогда Z может быть расположен сверху или слева от Y , как показано на схеме 2.3.

2.3.  означает, что X — непосредственная причина как для Z , так и для

Y ; кроме того, Z — непосредственная причина Y .

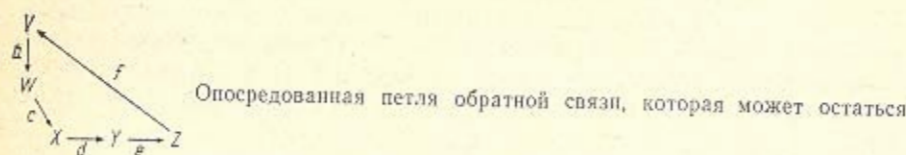
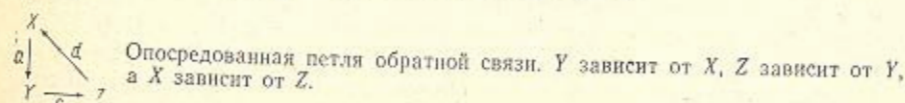
При первом вычерчивании очень редко удастся расположить все точки, как это требуется, и с минимальным количеством искривлений линий и пересечений стрелок. Однако поскольку слишком много для последующих выводов зависит от визуальной наглядности диаграммы, категорически необходимо делать ее максимально ясной и прозрачной. Обычно это приводит к необходимости перерисовывания диаграммы один или два раза с тем, чтобы расположить точки так, что все стрелки будут вести в основном в одну сторону, а пересечения и искривления линий либо совсем отсутствуют, либо имеются в минимальном количестве.

Всякий раз, когда некоторая переменная не просто пассивно зависит от своих причин, но в свою очередь влияет на одну или несколько из них, появляется причинная петля. Если пара переменных образует петлю, переменные связываются двумя стрелками, указывающими противоположные направления. Каждая из них метится своим собственным символом. Причинные петли мешают построению диаграммы с соблюдением условий «сверху вниз» и «слева направо», и, как ни стараться, все равно стрелки будут идти снизу вверх и справа налево.

Иногда петли обратной связи оказываются опосредованными и даже не могут быть точно отражены в качественной теории, хотя они и подразумеваются при описании свойств других отношений. Некоторый намек на петлю появляется, когда стрелки в окончательной диаграмме идут в направлении, противоречащем названному выше условиям. Ситуация, в которой это может произойти, отражена примером 2.4.

2.4. Примеры петель обратных связей.

$X \xrightarrow{a} Y$ означает, что X — причина Y , а Y — причина X .

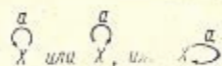


неотмеченной в качественной теории, становится явной, как только диаграмма будет нарисована.

Простые петли

Переменная может входить в петлю обратной связи, даже если в эту петлю никакая другая переменная не входит. По условию диаграммы стрелка будет выходить из этой переменной и входить в нее же. Направление стрелки несущественно.

2.5. Простая петля. Все три изображения эквивалентны:



Символом a обозначены операция или множество операций, которые преобразуют начальное изменение X в дополнительное последующее изменение X же.

То, что взаимная причинность между двумя переменными выражается парой стрелок, тогда как односторонняя причинность представлена одной стрелкой, подчеркивает важное свойство потоков графов. *Отсутствие стрелки выражает определенное количество информации: возможное причинное влияние не присутствует в данной системе.* Например, отсутствие на диаграмме стрелки от переменной Y к переменной X указывает на то, что Y не имеет прямого причинного влияния на X , и изменение в Y либо вообще не воздействует на X , либо, самое большее, косвенное влияние опосредуется изменениями других специфических переменных. Отсутствие причинных операторов — главное свойство системы — делает ее структурной и характерной. Проблеме невычерчивания стрелки между переменными необходимо уделять столько же внимания, сколько проблеме ее изображения.

Диаграмма с небольшим количеством звеньев, вообще говоря, более интересна, чем схема с причинными стрелками, связывающими почти все друг с другом и в обоих направлениях. Более простые теоретические схемы предпочтительнее и обычно легче соотносятся с эмпирическими исследованиями. В самом деле, структуры, в которых все переменные соединены стрелками в обоих направлениях, вообще дают мало возможностей для проверки каких-либо гипотез. Тем не менее простота, достигаемая предположениями об отсутствии причинных взаимодействий, когда они в действительности могут иметь место, гораздо хуже усложнения схемы. Слишком сильное упрощение теоретической конструкции может привести к ошибочной интерпретации и ложным выводам, полученным на основе исследуемых данных. Поэтому вводим следующее правило.

II.4. Всякий раз, когда есть сомнение по поводу наличия причинного взаимодействия, лучше добавить стрелку, чем ее опустить.

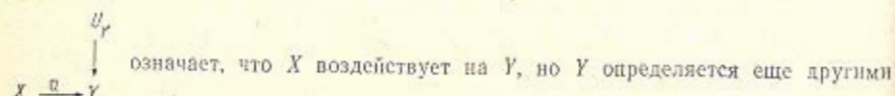
Никакой граф, отражающий существо содержательной теории, не может считаться действительно окончательным до тех пор, пока не будет обоснована допустимость каждого пропуска стрелки. Возможная стрелка между парой переменных может быть отброшена лишь в том случае, когда существуют определенные серьезные аргументы в пользу предположений об отсутствии влияния. Такие аргументы определяются принципами причинного вывода, установленными в гл. I.

Требование систематического рассмотрения каждого отбрасывания возможного взаимодействия способно привести в замешательство исследователя больших систем. Помочь в этом может составление «таблицы отношений» с переменными, выписанными вдоль строк и столбцов. Клетки этой таблицы соответствуют всем возможным причинным отношениям в системе, и последовательный перебор всех клеток таблицы представляет собой систематическую процедуру, гарантирующую раздельные рассмотрения каждой возможной пары отношений.

В большинстве социальных теорий широко употребляются фразы типа «вообще говоря», «обычно», «в среднем», «при прочих равных», «частично объясняет» и т. п. Причина этого в том, что реальные социальные системы чрезвычайно редко проявляют себя изолированно от других систем. Социальные системы функционируют внутри среды, состоящей из других систем, и разные системы оказывают возмущающие влияния друг на друга. Эти возмущения обычно не имеют общего, раз и навсегда определенного характера. Иногда они заставляют переменную изменяться в одном направлении, иногда — в другом; иногда они вносят большие изменения, иногда — малые. При построении и использовании диаграмм возмущения рассматриваются в терминах их чистого влияния на системную переменную, т. е. интерпретируются совокупно; как если бы они были единственной внешней входной переменной с неизвестным значением.

11.5. Возмущения представлены отдельным символом рядом с возмущаемой переменной; стрелка ведет от возмущения к переменной.

2.6. Изображение на диаграмме внешних возмущений.



неопределенными факторами, представленными совокупно через U_Y .

Большинство социальных переменных подвержено возмущениям. Однако возмущающие члены не изображают для входных переменных, потому что их значения не зависят ни от каких переменных внутри системы. Они целиком определяются внешними силами, и было бы излишне представлять на диаграмме внешние источники и для них.

Возмущающие члены обычно обозначаются буквой U с нижним индексом. Индекс определяется той переменной, на которую действует данное возмущение, как это показано в примере 2.6, где возмущающий член переменной Y обозначен через U_Y .

Для идентификации стрелки, идущей от возмущений к зависимым переменным, можно воспользоваться одним из двух соглашений. С одной стороны, на гипотетический возмущающий фактор можно смотреть как на некоторую переменную со своими специфическими внутренними свойствами, измеряемую в некоторой специальной шкале. В этом случае она сравнима с другими входными переменными системы, и стрелка должна ставиться так, чтобы указывать, как изменения в возмущающей «переменной» преобразуются в изменения зависимой переменной. Стрелка отмечается буквами a , c , p , q и т. п. с двумя нижними индексами: первый совпадает с символом зависимой переменной, второй — с

буквой U , указывая, какой именно фактор рассматривается. Например, стрелка от U к Y помечается символом a_{UY} .

С другой стороны, возмущающий фактор может быть выражен в той же самой шкале, что и переменная, на которую он воздействует. В этом случае изменение в возмущающей «переменной» всегда точно совпадает с ее влиянием на зависимую переменную. Если все источники возмущения зависимой переменной приводят к ее изменению на одну единицу, то изменение самой гипотетической переменной принимается равным тоже единице. Поэтому вообще нет необходимости помечать возмущающую стрелку, так как она всегда обозначает одно и то же преобразование: зависимая переменная изменяется на ту же величину, что и возмущающая. Это — тождественное преобразование со структурным коэффициентом, равным 1, и по условию отсутствие метки обозначает именно это. В силу того что при употреблении второго условия элиминируются дополнительные символы, мы будем его придерживаться, пока не возникнут какие-либо новые соображения в пользу другого условия.

Несмотря на то что почти все социальные переменные, вероятно, подвержены возмущающим влияниям, в этой книге не во всех потоковых графах вводятся символы для возмущений. Обычно это делается с целью упрощения анализа и обсуждения. Однако если наличие возмущений влияет на общие заключения, они вводятся в диаграммы обязательно.

Нелинейные отношения

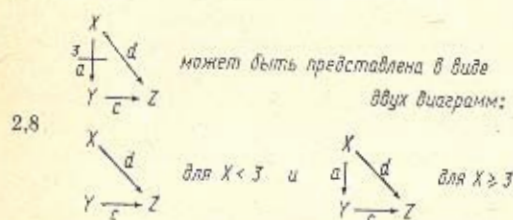
Процедуры составления диаграмм предполагают выполнение условий линейности связей. В частности, предполагается, что обозначаемые стрелками операции никоим образом не зависят от значения входной переменной или какой-либо другой переменной системы. Могут быть приведены примеры, когда это не выполняется: расходы на развлечения могут быть определены как постоянная доля дохода, но только после того, как доход достигнет определенного уровня, который позволит удовлетворить первоочередные жизненные потребности.

Для указания на диаграмме ситуации, когда преобразование имеет данный вид лишь при превышении входной переменной некоторого порога, вводится специальное правило. Обычно рисуется стрелка, помеченная символом преобразования; однако чтобы отразить указанное ограничение, стрелка перечеркивается короткой поперечной чертой, которая в свою очередь помечается указанием, с какого значения входной переменной действует причинный оператор. Например, диаграмма в 2.7 имеет следующий смысл: X определяет Y посредством преобразования a . Однако преобразование имеет силу, только когда X достигает по величине значения 5 и более. По условию это понимается так, что если X принимает значения меньше 5, то изменения в X не оказывают влияния на Y .

2.7. Пример ограничителя. $X \xrightarrow{\frac{a}{s}} Y$

С помощью комбинаций ограничителей могут быть выражены сложные типы преобразований. Здесь эти возможности не рассматриваются, но их можно найти в литературе, перечень которой дан в конце этой главы.

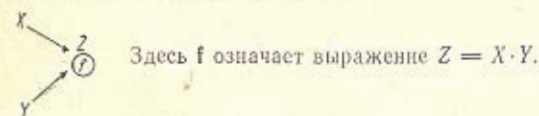
Ограничители на диаграмме указывают на нелинейность и не дают возможности пользоваться простыми и изящными методами анализа графов, обсуждаемого в этой книге. Единственный ограничитель, однако, может быть легко исключен путем превращения диаграммы в две другие — одна применяется ниже граничного значения, другая — выше, как это показано на 2.8.



Иногда форма связи между двумя переменными зависит от значения некоторой третьей переменной системы. Например, как доход, так и образование, вероятно, детерминируют потребление «высокой культуры» (книг, классической музыки, посещений театра и т. д.). Однако сила этого отношения, возможно, растет с повышением уровня образования. Таким образом, доход и образование *взаимодействуют* в их детерминации потребления высокой культуры.

Для изображения такого взаимодействия можно использовать другое условие построения диаграмм. В конце стрелок от взаимодействующих переменных помещается кружок, отмечающий зависимую переменную извне и содержащий внутри символ (обычно f или g), выражающий специальное преобразование для превращения значений входных переменных в значение зависимой переменной.

2.9. Пример, иллюстрирующий графическое представление взаимодействия.



Взаимодействия тесно связаны с ограничителями; подобно последним они затрудняют прямое использование анализа потоков

графов. Иногда, как и в случае ограничителей, диаграмму можно переделать в несколько отдельных диаграмм, в каждой из которых эффект взаимодействия сведен к нулю или уменьшен. Так, например, диаграмма 2.9 может быть переделана в две различные диаграммы, соответствующие высшему и низшему значениям Y . В другом случае можно рассматривать одну из переменных как параметр, который должен фиксироваться заранее для каждого конкретного случая.

2.10. Альтернативный метод изображения диаграммы мультипликативного взаимодействия схемы 2.9: $X \xrightarrow{Y} Z$.

Иногда взаимодействия могут быть исключены из диаграммы путем изменения шкал, в которых измерены переменные. Отношение между переменными в 2.9 удовлетворяло бы нашим условиям, если бы X , Y и Z измерялись в логарифмах их действительных значений, так как логарифмическое преобразование превращает произведение в сумму. Другие более общие приемы представления взаимодействия на потоковых графах требуют привлечения математики более высокого уровня, чем используется в этой книге. Интересующийся читатель может познакомиться с ними по ссылкам этой главы.

Анализ потоковых графов

Линейные потоковые графы можно «читать» так, что будет легко определять чистый эффект влияния одной переменной на другую, даже если эти переменные связаны опосредованно, а связи содержат большое количество причинных цепей и петель обратной связи. Этот анализ позволит нам проследить за всеми последствиями входов: он укажет различные типы всевозможных результатов действий системы; он покажет, как будут коррелировать между собой переменные на основе эмпирических данных, и даст основу для последующих причинных выводов; он, наконец, будет способствовать пониманию того, как необходимо видоизменить систему, чтобы нужным образом изменились ее выходные характеристики.

Описанные в этой главе принципы применимы только к линейным системам определенного типа. Во-первых, они применяются только к системам, являющимся стационарными в течение всего рассматриваемого периода. Всякое изменение в X ведет к ожиданию определенной величины изменения в Y , независимо от того, когда эти изменения в X возникли. Во-вторых, главное внимание в этой главе уделяется конечным эффектам от изменений входных величин, а не промежуточной динамике, благодаря которой переменные достигают своего конечного состояния. С одной стороны, это означает, что мы ограничиваемся рассмотрением *устойчивых* систем, которые достигают установившегося состояния, но не рассматриваем такие, истинная природа которых ведет

к бесконечным колебаниям или неограниченному росту. С другой стороны, наше предположение требует, чтобы все переменные во время наблюдения были стабилизированы в неизменяющихся или статических значениях. (Динамика систем кратко затрагивается в гл. 6.)

В дополнение к сказанному необходимо отметить, что ограничение нелинейности отношений требует измерения переменных в интервальной шкале, так что выполняются свойства аддитивности и пропорциональности. В действительности же не требуется, чтобы шкала была действительно интервальной, достаточно ее некоторого приближения; например, то, в чем измеряется произведение масс, не является, строго говоря, интервальной шкалой, но это приемлемая аппроксимация для изучения многих обыкновенных объектов. Техника измерения в социальных науках достаточно развита, чтобы обеспечить подходящую аппроксимацию интервального измерения для анализа многих каузальных систем.

Потоковый граф в соединении с небольшим количеством правил позволяет выразить значения зависимых переменных в терминах значений их входных переменных.

СОСТАВЛЕНИЕ УРАВНЕНИЙ

II. 6. Значение переменной, определяемой одним-единственным входом, равно значению входа, умноженному на структурный коэффициент.

2.11. $X \xrightarrow{a} Y$ означает: $Y = a \cdot X$.

Уравнение 2.11 может показаться недостаточно общим, так как обычно даже простые уравнения содержат некоторую константу, уточняющую значение фактора, например $Y = a \cdot X + 5$. В нашем случае выражения, полученные на основании графа, не содержат констант. Это означает, что каждая переменная должна быть измерена в шкале, приведенной к такому виду, когда эти константы отсутствуют.

Если необходимо, поправочные константы могут быть сохранены в графическом анализе путем незначительной модификации диаграмм, как это показано в 2.12 (написание уравнения требует применения правила II. 7).

2.12. $X \xrightarrow{a} Y$ означает $Y = a \cdot X + 5$.

Не надо путать константы с возмущающими членами. Константы суть входные величины специального типа, остающиеся неизменными, что бы ни происходило, в то время как возмущения

суть такой тип входных величин, который символизирует переменные, влияющие со стороны.

Ясно, что включение констант усложняет диаграммы, и легко показать, что это не является необходимостью, когда имеют дело с линейными системами. Обращаясь к примеру 2.12, мы произведем новое измерение переменной Y' , вычитая 5 из каждого первичного измерения Y : $Y' = Y - 5$, так что $Y = Y' + 5$.

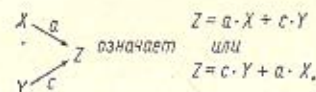
Теперь, подставляя это в уравнение 2.12, получим:

$$Y' + 5 = a \cdot X + 5 \text{ или } Y' = a \cdot X.$$

Таким образом, несколько измененный способ измерения Y приводит к уравнению без константы, а этому более простому уравнению соответствует и более простой граф, подобный графу в 2.11.

II. 7. Значение переменной, определяемой одним или несколькими входными величинами, равно сумме входных значений, умноженных соответственно на их структурные коэффициенты. Порядок суммирования не имеет никакого значения.

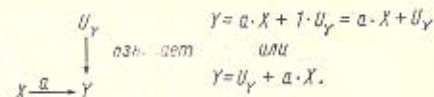
2.13.



Правило суммирования применяется независимо от числа входных величин. Однако оно применимо лишь к непосредственным входам; например, в 2.13 в суммирование не включены переменные, которые влияют на Z косвенно, скажем, через их действие на Y . Такие косвенные эффекты анализируются с помощью правил, приводимых далее.

Так как большинство диаграмм в социальных науках содержит возмущающие члены в качестве одного из входов для всякой зависимой переменной, то полезно рассмотреть пример применения правила суммирования в этом случае.

2.14.



Таким образом, правило суммирования применяется к возмущающим членам точно так же, как и к любым другим входам. Более того, как это видно из 2.12, правило суммирования применяется даже, если один из входов равен константе.

Представление структурных уравнений

Предыдущие правила позволяют нам выписывать структурные уравнения соответствующих теорий. Эти правила указывают, как качественные теоретические положения превращаются в выраже-

ния потоковых графов и как потоковый граф определяет математические уравнения. Формулировки с помощью структурных уравнений играют особенно важную роль в современных средствах анализа сложных систем.

Для каждой зависимой переменной системы структурное уравнение получается по правилу II.7, как только потоковый граф будет нарисован. Пути, отходящие от переменной при написании уравнения для этой переменной, не учитываются, но каждая входящая стрелка указывает член, который должен быть учтен. Переменная с петлей обратной связи рассматривается точно так, как и любая другая переменная: уравнение для нее содержит члены для каждой из ее непосредственных входов даже в том случае, когда эта входная переменная содержится в некоторой другой петле обратной связи. Поскольку большая часть зависимых переменных подвержена влиянию возмущений, большинство уравнений содержит также возмущающие члены.

Для того чтобы применить эту процедуру к уже разбиравшемуся примеру, рассмотрим систему, приведенную на первом графе в 2.43 (вначале). Имеются три зависимые переменные, т. е. три переменные со стрелками, идущими к ним. Математическая формализация включает три уравнения:

$$X = aW + eY + cZ + U_X,$$

$$Y = dZ + U_Y,$$

$$Z = fX + U_Z.$$

Эти одновременные уравнения являются статической моделью в виде структурных уравнений системы, представленной на графе. В более специальной литературе зависимые переменные (X, Y, Z предыдущего примера) часто называются *эндогенными* переменными. Переменные, значения которых определяются внешними для системы числами (в нашем примере — только W), называются *экзогенными*. В данной книге используется термин «входы» вместо «экзогенные переменные».

Правила редукции

II.8. Если одна переменная определяет вторую переменную, а другая — определяет третью, то значение третьей переменной может быть выражено как значение первой переменной, умноженное на произведение структурных коэффициентов вдоль цепи. Этот же принцип применяется, когда цепь имеет более чем два звена.

$$2.15. X \xrightarrow{a} Y \xrightarrow{c} Z \text{ означает: } Z = a \cdot c \cdot X.$$

Это «правило цепи» определяет опосредованное влияние исходной переменной с помощью исключения из рассмотрения пе-

ременных, расположенных между исходной переменной и интересующим нас выходом.

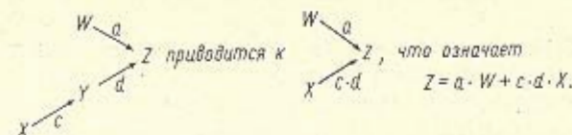
Выражение, полученное применением правила цепи к 2.15, может быть переписано как $Z = (a \cdot c)X$, а это выражение должно соответствовать диаграмме $X \xrightarrow{a \cdot c} Z$. Действительно, что касается переменных X и Z , то получившаяся новая диаграмма совершенно эквивалентна первоначальной, и правило цепи имеет смысл приведения графа к более простому виду.

$$2.16. X \xrightarrow{a} Y \xrightarrow{c} Z \text{ приводится к } X \xrightarrow{a \cdot c} Z.$$

Этот принцип применим также и в случае любого количества звеньев в данной цепи.

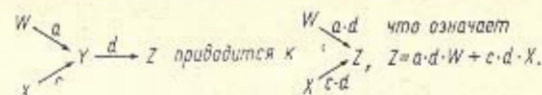
II.9. Чтобы выразить значение зависимой переменной через многочисленные непосредственные и косвенные входы, сначала получают отдельные влияния вдоль каждой цепи по правилу II.6 или II.8. Затем суммируют все влияния в соответствии с правилом II.7.

2.17.



Специальный и весьма важный случай возникает, когда на выходную переменную действует несколько косвенных входных источников через одну и ту же промежуточную переменную. Цепь от каждого источника определяется автономно, как если бы других источников не было.

2.18.



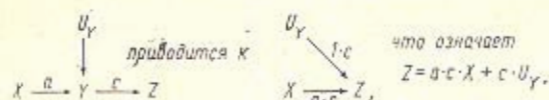
Это правило также предпочтительнее применять, когда имеется несколько выходных переменных, а не одна. Мы просто применяем это правило повторно, чтобы получить свое выражение для каждой выходной переменной.

2.19.



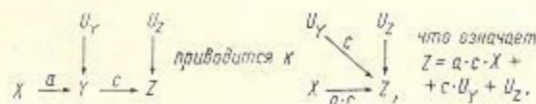
Другой интересный случай возникает, когда промежуточная переменная в цепи подвержена возмущениям. Возмущающий член интерпретируется как еще один источник, и правило II.9 применимо.

2.20.



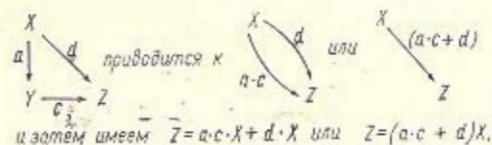
Требуется совсем незначительное усилие, чтобы включить и возмущающий член для Z.

2.21.

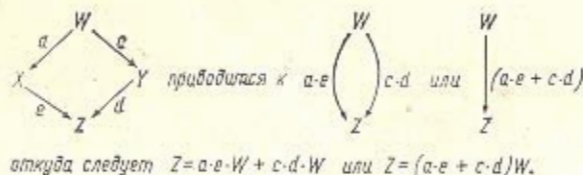


II. 10. Если источник и зависимая переменная связаны многими путями, то результирующая связь между ними равна сумме влияний вдоль каждого отдельного пути. Влияние вдоль пути получается по правилам II.6 и II.8. Порядок суммирования не имеет значения.

2.22.



2.23.

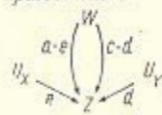


Может возникнуть вопрос относительно процедуры, когда промежуточные переменные имеют дополнительные входы. В этом случае сначала мы разделяем влияния, обусловленные разными входами, используя правило II.9, а затем применяем правило II.10.

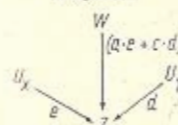
2.24.



с помощью правила II.9
приводится к



с помощью правила II.10
получаем



Выражение для определения величины Z может быть теперь получено с помощью правила II.7.

Дополнительная интерпретация правил

Некоторые из указанных выше правил имеют более интересные приложения, чем это может показаться на первый взгляд. Так, например, правило цепи (II.8) означает, что если все коэффициенты вдоль цепи положительны, то возрастание источника ведет к возрастанию конечной переменной, а убывание источника ведет к убыванию конечной переменной. С другой стороны, единственный отрицательный знак где-либо вдоль цепи приводит к противоположному правилу направления изменения от входа к выходу. В общем случае, если в цепи отсутствуют отрицательные коэффициенты или число их четно, то изменение в источнике ведет к изменению зависимой переменной того же знака. Если в цепи появляется нечетное число отрицательных коэффициентов, изменения в значении входной переменной приводят к изменениям конечной переменной в противоположных направлениях.

В социальных системах абсолютные значения структурных коэффициентов часто оказываются дробными (т. е. их абсолютные значения находятся между нулем и единицей), когда все переменные измерены в сопоставимых шкалах. Перемножаясь между собой, дроби дают дробь, по абсолютной величине меньшую каждого из сомножителей. Вследствие этого влияние всей цепи имеет тенденцию быть меньше, чем любое прямое влияние на этой цепи. Чем длиннее цепь, тем меньше чистый эффект влияния, потому что произведение дробей становится все меньше — с увеличением их числа. Это означает, что длинные причинные цепи в анализе социальных систем не могут иметь большого практического значения, а теории, рассматривающие конечное влияние в терминах длинных последовательностей зацепляющих звеньев, могут объяснить только совсем малую долю вариаций значений выхода.

Правило П.10 показывает, что воздействия одной переменной на другую складываются по всем цепям, которые связывают обе эти переменные. Особый интерес имеет ситуация, когда один из всего множества путей является прямым и содержит только одну стрелку от источника к зависимой переменной; другие пути (или путь) являются опосредованными и включают цепи, проходящие через несколько других переменных системы, входы которых имеют непосредственное влияние через данный прямой путь и косвенные влияния через другие пути. Величины этих двух типов влияний можно сравнивать. Если мы должны манипулировать источником, чтобы добиться нужного влияния на зависимую переменную, то необходимо учитывать как прямые, так и косвенные влияния.

Потоковые графы с петлями

П.11. Цепочка стрелок, отходящая от некоторой переменной и возвращающаяся в конце концов туда же, называется петлей при условии, что вдоль всего пути ни разу не происходит изменения направления стрелок на обратное и что весь путь не проходит ни через одну переменную более одного раза, за исключением переменной, принятой за начало, через которую, считается, путь проходит дважды.

Типичные примеры петель приведены в 2.25. Заметим, что всякий раз изображается и входная переменная, не входящая в петлю. Вход должен иметься у петли по необходимости, а петля без единого входа представляется нереалистической. Обратим внимание также, что петли могут быть «спрятаны» в контексте большей диаграммы, как это показано в нижней части 2.25.

2.25.



Для идентификации петель здесь принимается простое условие — последовательно перечислять переменные, через которые проходит петля. Начальная переменная второй раз не называется, хотя должно быть понятно, что она является также конечной

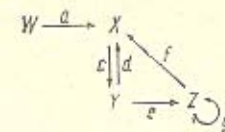
переменной. Такая идентификация не является единственно возможной, так как любая переменная в цепочке петли может быть принята в качестве начальной. Так, в нижнем примере 2.25 любое из следующих обозначений относится к одной и той же петле: (WXY) , (XYW) , (YWX) .

Условия вычерчивания стрелок (слева направо и сверху вниз) помогают нахождению петель; для примера заметим, как стрелка e указывает на наличие петли в нижней схеме 2.25. Однако иногда при вычерчивании диаграмм весьма затруднительно строго следовать принятым соглашениям, поэтому стрелки, идущие в «неправильном» направлении, в лучшем случае играют роль ориентира.

Цепь (VWX) на нижнем графе 2.25 не является петлей, потому что стрелка g имеет направление, обратное пути. Заметим также, что цепочка типа $(WXYW)$ не должна рассматриваться как петля, потому что путь пересекает W более одного раза (в предположении, что он кончается в переменной X); поэтому данная цепь проходит через две переменные по два раза.

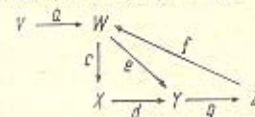
Многие системы содержат по несколько петель.

2.26. Система с тремя петлями (XY) , (XYZ) , (Z) .



Эта система указывает гнездо петель, т. е. множество петель, цепочки которых частично пересекаются. Петлю гнезда необходимо идентифицировать как отдельную, если только эта петля отличается от других по крайней мере единственной стрелкой. На схеме 2.27 изображен другой пример группы петель.

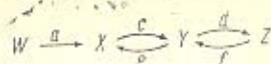
2.27. Система с двумя петлями: $(WXYZ)$, (WYX) .



В этом примере имеются две петли, но только одна стрелка неправильного направления. Это еще раз подчеркивает, что принцип направленности есть лишь вспомогательное средство обнаружения петель. В частности, принцип «неправильного направления» может не иметь успеха при выявлении существования гнезда петель, имеющих один и тот же обратный путь.

Петли, являющиеся лишь смежными друг с другом, не порождают гнезда.

2.28. Система с двумя петлями: (XY) , (YZ) .



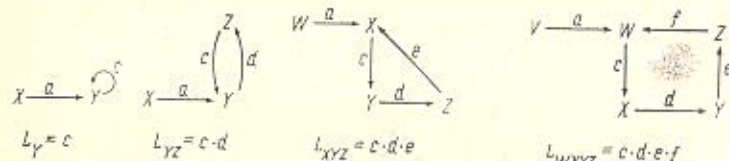
Путь $(XYZY)$ не является петлей, потому что дважды проходит через промежуточную переменную Y .

Выявление всех различных петель — важнейший шаг в анализе систем с обратными связями, потому что не существует простого правила, гарантирующего нахождение петель, и способность к этому должна быть развита до уровня мастерства.

II.12. «Эффект обратной связи» петли, обозначаемый буквой L , равен произведению коэффициентов вдоль цепи петли.

Диаграмма верхней части примера 2.25 воспроизводится в 2.29, поясняя сказанное. Множества символов, обычно служащих идентификаторами петель, теперь используются как индексы, отмечающие отдельный обратный эффект.

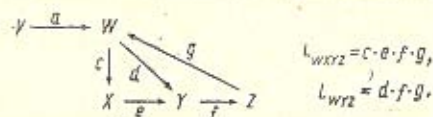
2.29.



Выбор начальной точки не имеет значения, например, для третьей диаграммы.

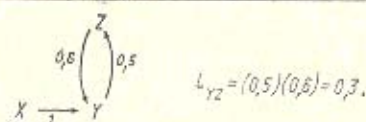
На следующей схеме приводится более сложный пример.

2.30.



Эффект обратной связи, или «обратный эффект», допускает содержательную интерпретацию. Он указывает, на сколько единиц изменится переменная петли в течение одного цикла изменений вдоль всей петли, прошедшего после начального изменения значения на одну единицу. Рассмотрим, например, ситуацию на схеме 2.31.

2.31.



Предположим, что значением X управляют так, чтобы вызвать начальное изменение Y на одну единицу. Это приведет к изменению Z на 0,5. Изменение в Z приводит к новому изменению в Y на $(0,5 \cdot 0,6) = 0,3$ единицы. В результате после одного цикла Y возрастет на величину обратного эффекта. Конечно, эти 0,3 единицы изменения должны пройти вдоль петли, вызвав еще раз новое возрастание в Y (на 0,09 единицы), а оно в свою очередь снова пройдет через петлю и т. д. Однако обратный эффект — приращение, возникшее за один цикл после начального единичного изменения, — является наиболее важной количественной характеристикой. Он дает достаточную информацию для измерения чистого эффекта всех циклов, следующих за начальным изменением (используя приводимое далее правило II.16.)

II.13. Цепочка стрелок представляет собой «открытый путь», если вдоль нее нет ни одного обращения направления и если стрелки данного пути не направляются ни на одну переменную более одного раза»

Простой пример открытого пути дается в 2.16.

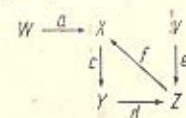
Открытые пути противоположны петлям, являющимся «замкнутыми путями», не имеющими реального начала или конца. (Выбор «начальной точки» в петле произволен и необходим для идентификации петли; любая переменная петли будет одинаково хорошо служить в качестве начальной точки.)

Обозначения, которые идентифицируют открытые пути, подобны тем, которые используются для идентификации петель. Например, (XYZ) относится к открытому пути с началом в X , идущему в направлении к Y и заканчивающемуся в Z . Конечно, эта цепочка не возвращается к начальной точке в открытом пути. Кроме того, последовательность символов задается однозначно: (XYZ) не то же самое, что (YZX) . Указателями петли и открытого пути всегда служат нижние индексы — наборы различных символов, и слишком мала вероятность перепутать их на практике.

Открытые пути, по существу, не представляют собой ничего сложного, а правила II.6—II.10 относятся исключительно к ним. Например, диаграмма в 2.22 содержит два открытых пути между X и Z : (XZ) и (XYZ) ; однако они представляют более общий интерес, потому что могут быть идентифицированы либо как пути, проходящие через петлю, либо как пути внутри нее.

Это показано в 2.32.

2.32. У этой системы один открытый путь между W и Z : $(WXYZ)$. Между Y и Z два открытых пути: (YZ) и (ZXY) .



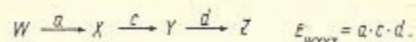
Между парой переменных может не проходить вообще ни одного открытого пути (например, W и V в 2.32), либо ровно один

открытый путь, либо несколько их. Открытые пути, связывающие Y и Z на 2.32, противоположны по направлению. Это существенная черта открытых путей, которые связывают переменные в одной и той же петле.

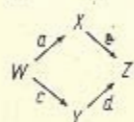
П. 14. «Эффект открытого пути», обозначаемый через E , равен произведению коэффициентов вдоль всей цепи открытого пути.

На схеме 2.33 приведено несколько примеров эффекта открытого пути. Наборы символов, используемых для идентификации открытых путей, употребляются как индексы для идентификации эффекта открытого пути.

2.33.

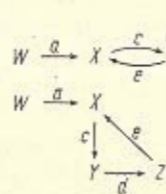


$$E_{\text{work}} = q \cdot \mathcal{E} \cdot d$$



$$\bar{c}_{\text{max}} = a - e,$$

$$E_{\text{max}} = c \cdot d.$$



$$E_{\text{max}} = a \cdot c \cdot d.$$

$$E_{wx} = a; E_{xr} = c; E_{rz} = d;$$

$$E_{xz} = e; E_{wxyz} = a - c$$

$$E_{WXYZ} = a \cdot c \cdot d; \quad E_{XYZ} = c \cdot d;$$

$$E_{XYZ} = d \cdot e; \quad E_{YXZ} = e \cdot c.$$

В первых трех примерах приведены выражения для небольшой отобранной части эффектов открытых путей; в последнем примере указаны все эффекты открытого пути.

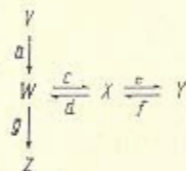
Соприкасающиеся пути

П.15. Две петли «соприкасаются», если их идентификаторы содержат один или несколько общих символов. Петля касается открытого пути, если какой-либо символ из идентификатора открытого пути (за исключением самого первого) присутствует также в идентификаторе петли.

Идентификаторы петель и открытых путей представляют собой список переменных, идущих вдоль соответствующих путей. Например, в 2.32 ($WXYZ$) — идентификатор открытого пути от W к Z , а (XYZ) — идентификатор петли. Эти пути касаются друг друга, поскольку их идентификаторы имеют по три общих символа. С другой стороны, можно заметить, что E_{WXYZ} — эффект открытого пути, а L_{XYZ} — обратный эффект петли. У этих эффектов по три одинаковых символа, поэтому они относятся к соприкасающимся путям.

В 2.34 приведен другой важный пример. Открытый путь (WZ) не касается ни петли (WX), ни петли (XY), потому что у него ни одного общего символа (*не считая первого*) с петлями.

2.34.



Петля (WX) касается открытого пути (VWZ) , но не (WZ) . Петли (WX) и (XY) соприкасаются.

Концепция «касания» позволяет нам определить релевантную обратную связь следующим образом.

Пусть переменная X будет входной переменной для Y , т. е. X влияет на Y непосредственно или опосредованно, а Y не воздействует на X ни-прямо, ни косвенно. Очевидно, от X к Y должен быть один или несколько открытых путей. Если в этой системе присутствуют также петли, то релевантная обратная связь для $X-Y$ -отношения есть множество всех петель, касающихся хотя бы одного из открытых путей от X к Y , плюс дополнительные петли, касающиеся петель первого множества, плюс новые петли, касающиеся предыдущих, и т. д.

В качестве иллюстрации рассмотрим снова систему в 2.34. Переменная W представляет собой естественный вход для Z , но ни одна из петель системы не является релевантной обратной связью для данного отношения, потому что ни одна из них не касается открытого пути (WZ) . V , однако, также является нормальным входом для Z , но в этом случае обе петли оказываются релевантными обратными связями: петля (WX) — потому, что она касается открытого пути (VWZ) , и петля (XY) — потому, что она касается петли (WX) .

Редукция петель

Основные принципы анализа графа системы с петлями были предложены для технических приложений Самуэлем Мэсоном в 1951 г. Основное правило имеет следующий вид.

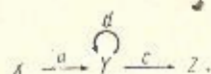
11.16. Полный эффект влияния T некоторой входной переменной на произвольную зависимую переменную может быть рассчитан следующим образом: пусть $E, E', E'' \dots$ суть отдельные эффекты открытых путей входа на зависимую переменную; $L, L', L'', \dots$ — обратные эффекты для всех отдельных петель, обеспечивающих релевантные обратные связи. Тогда

$$T = \left[\frac{(E + E' + E'' + \dots) \cdot (1 - L) \cdot (1 - L') \cdot (1 - L'') \dots}{(1 - L) \cdot (1 - L') \cdot (1 - L'') \dots} \right]^*$$

где []* есть специальный оператор, обозначающий, что сначала и в числителе, и в знаменателе производится перемножение, затем

вычеркиваются члены, равные произведениям эффектов касающихся путей, и только потом производится деление. Простой пример приводится в 2.35.

2.35. Найдем полное влияние X на Z в следующей диаграмме:



Как легко видеть, между X и Z имеются только один открытый путь с эффектом $E_{XYZ} = a \cdot c$ и только одна петля со своим обратным эффектом $L_Y = d$. Поэтому

$$T_{ZX} = \left[\frac{E_{XYZ} \cdot (1 - L_Y)}{1 - L_Y} \right]^* = \left[\frac{E_{XYZ} - E_{XYZ} \cdot L_Y}{1 - L_Y} \right]^*.$$

Открытый путь от X к Z и петля (Y) касаются друг друга, поэтому оператор $[]^*$ требует уничтожения члена $E_{XYZ} \cdot L_Y$. В результате получим

$$T_{ZX} = \frac{E_{XYZ}}{1 - L_Y} = \frac{a \cdot c}{1 - d}.$$

На 2.35 показана самая простая из возможных систем с петлями. Тем не менее в этом случае могут быть хорошо проиллюстрированы некоторые из принципов, затрагиваемых правилом II.16. Положим, переменной X придадут такое значение, что Y возрастает ровно от нуля до 1 сразу вслед за изменением в X . В предположении, что X сохраняется на своем новом значении (что мы всегда можем иметь в статическом случае), Y должно было бы сохранять значение, равное 1. Однако теперь по петле передается собственное изменение в Y . Оно умножается на обратный эффект петли L_Y и это увеличение в свою очередь будет добавлено к Y . Так как Y уже достиг значения 1, то его новое значение будет равно $1 + L_Y$. Кроме того, это второе приращение зависит косвенно от X , и пока X сохраняет постоянное значение, значение Y остается равным $1 + L_Y$. С этой точки зрения вторичное изменение Y на L_Y единиц также пройдет дальше по петле. Умноженное на обратный эффект петли оно дает Y приращение L_Y^2 .

В результате значение Y станет равным $1 + L_Y + L_Y^2$ и будет сохраняться на этом уровне, пока X остается неизменным. Очевидно, что третье изменение в Y также передается по петле и вызовет новое возрастание L_Y^3 переменной Y . Этот процесс продолжается неограниченно, и конечное значение Y должно быть равно следующей сумме:

$$1 + L_Y + L_Y^2 + L_Y^3 + L_Y^4 + \dots$$

Это суммирование продолжается до бесконечности. Тем не менее можно показать математически, что Y сходится к конечному окончательному значению, если только L_Y есть дробь, т. е. если

$-1 < L_Y < +1$ (если L_Y не является правильной дробью, сумма ряда расходится, и петля будет неустойчивой). Кроме того, значение Y быстро достигает своего конечного значения — всего за несколько циклов. Это является следствием того, что каждый из последующих членов ряда имеет все более высокую степень и последующие приращения становятся пренебрежимо малыми. Например, если $L_Y = \frac{1}{2}$, то $L_Y^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$. Значение такой суммы может быть определено совершенно точно. В частности, можно показать математически, что следующая формула будет справедливой (когда L_Y есть дробь):

$$\frac{1}{1 - L_Y} = 1 + L_Y + L_Y^2 + L_Y^3 + \dots$$

Если L_Y — положительно, то $1 - L_Y$ есть число, меньшее единицы, и $\frac{1}{1 - L_Y}$ будет числом, превышающим единицу. В этом случае эффекты открытых путей, касающихся петли, будут увеличены или усилены. Если L_Y — отрицательно, то $1 - L_Y$ — больше единицы, а $\frac{1}{1 - L_Y}$ меньше единицы. В этом случае влияние открытого пути, проходящего через петлю, уменьшается, ослабляется, регулируется или сдерживается.

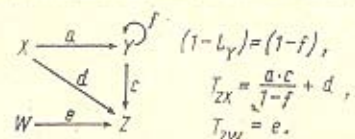
Как было мимоходом замечено выше, о свойствах стабильности системы с единственной петлей можно заключить по значениям эффекта обратной связи. Если абсолютная величина обратного эффекта больше единицы или равна ей (т. е. величина обратного эффекта ≤ -1 или $\geq +1$), система будет неустойчивой. При любом заданном входе значения переменных неограниченно возрастают в том или ином направлении. Если величина обратного эффекта находится между -1 и $+1$, то система с единственной петлей будет устойчивой. Система со многими петлями может быть неустойчивой, даже если каждая индивидуальная петля в отдельности порождала бы устойчивость. И, наоборот, возможен случай, когда отдельные петли индивидуально неустойчивы, а система в целом находится в равновесии, потому что эти петли нейтрализуют друг друга. Чистое воздействие множества петель на устойчивость системы может быть оценено, но в этом случае требуется техника более сложного математического анализа, которая в явном виде применяется в анализе динамических систем (см. примечания к гл. 6).

Релевантные петли, которые фактически не касаются открытого пути, не оказывают полного воздействия на этот путь. В общем, их влияния ослабляются их удаленностью от основной части пути, и для отражения этого факта¹ требуется поправка числителя коэффициента передачи. Однако эта поправка может быть получена традиционно, когда применяется правило II.16.

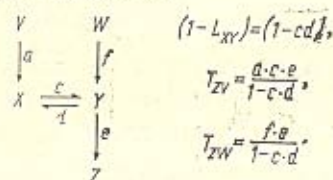
¹ На эту функцию перемножаемых членов, получаемых по правилу Мэсона, обратил мое внимание Дж. А. Дэвис.

Правило II.16 сформулировано так, что его можно использовать даже в случае чрезвычайно запутанных сетей. Однако если на диаграмме имеется только одна-единственная петля, то ее влияние сказывается просто в поправке всех эффектов открытых путей, ее касающихся, на множитель $1/(1-L)$. На схеме 2.35 представлен один пример такого рода. Другой пример дается в 2.36.

2.36.

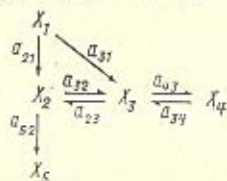


Лишь один открытый путь (XYZ) касается петли. Поэтому выше только один эффект открытого пути подправлен.



Следующий более сложный пример демонстрирует все ключевые аспекты правил II.15 и II.16. Он также показывает, как пользуются численными индексами, когда имеют дело с большими системами.

2.37. Найдем полный эффект (влияние) X_1 на X_5 .



От X_1 к X_5 имеются два открытых пути:

$$E_{125} = a_{21} \cdot a_{52},$$

$$E_{1325} = a_{31} \cdot a_{23} \cdot a_{52}.$$

В системе две петли:

$$L_{23} = a_{32} \cdot a_{23},$$

$$L_{34} = a_{43} \cdot a_{34}.$$

Поэтому полный эффект от X_1 к X_5 равен:

$$T_{51} = \left[\frac{(E_{125} + E_{1325}) \cdot (1 - L_{23}) \cdot (1 - L_{34})}{(1 - L_{23})(1 - L_{34})} \right]^*.$$

Обычно сначала раскрывают скобки нижнего выражения, потому что оно появляется и в числителе, и в знаменателе. В результате перемножения скобок получим

$$1 - L_{23} - L_{34} + L_{23}L_{34}.$$

Два эффекта в последнем члене относятся к касающимся петлям, поэтому данный член отбрасывается:

$$[1 - L_{23} - L_{34} + L_{23}L_{34}]^* = 1 - L_{23} - L_{34}.$$

С помощью этого соотношения можно переписать формулу для T_{51} :

$$T_{51} = \frac{[(E_{125} + E_{1325})(1 - L_{23} - L_{34})]^*}{1 - L_{23} - L_{34}}.$$

Произведение в верхней части этого выражения имеет вид

$$E_{125} + E_{1325} - E_{125}L_{23} - E_{1325}L_{23} - E_{125}L_{34} - E_{1325}L_{34}.$$

Подчеркнутые члены содержат сомножители — эффекты, относящиеся к касающимся путям, поэтому их опускаем и получаем выражение

$$T_{51} = \frac{E_{125} + E_{1325} - E_{125}L_{34} - E_{1325}L_{34}}{1 - L_{23} - L_{34}} = \frac{E_{125}(1 - L_{34}) + E_{1325}}{1 - L_{23} - L_{34}}.$$

Подставляя теперь вместо каждого эффекта их более конкретное значение, получим:

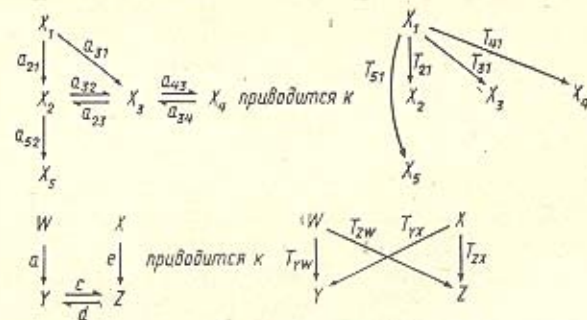
$$T_{51} = \frac{a_{21}a_{52}(1 - a_{43}a_{34}) + a_{31}a_{23}a_{52}}{1 - a_{32}a_{23} - a_{43}a_{34}}.$$

Эти примеры показывают, как проводится анализ петель в системах без возмущений. Ради простоты на этих диаграммах возмущения не показывались, но при их наличии применяются те же самые процедуры и без всяких модификаций.

Приведенная форма системы

II.17. Как только будут получены все полные эффекты, соотносящие каждую зависимую переменную со всеми входными переменными, можно построить приведенную форму диаграммы для того, чтобы показать связь каждой зависимой переменной непосредственно с входными и отбросить промежуточные переменные.

2.38.



Как показывают эти примеры, в диаграммах приведенной формы петли отсутствуют.

Диаграммы приведенной формы подчиняются обычным правилам потоковых графов. Так, начертив эти приведенные формы и применяя соответствующие правила, легко выразить значение любой зависимой переменной в виде взвешенной суммы входов системы, даже если система включала сложные петли обратной связи. Коэффициенты диаграмм приведенной формы (T -коэффициенты) состоят из выражений для полного эффекта от каждого входа к каждой зависимой переменной, определяемых по правилу П. 16.

Имеется одно замечательное свойство анализа потоковых графов, заключающееся в том, что влияние (эффект) единственного входа на интересующую нас переменную может быть определено без получения всей приведенной формы¹. Но алгебраически это не может быть обобщено.

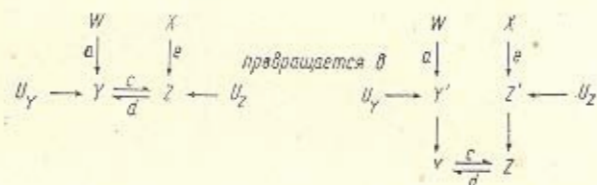
Частичное приведение

То обстоятельство, что приведенные формы не содержат петель, играет в причинном анализе чрезвычайно важную роль. Часто, однако, бывает весьма желательным упростить только часть графа, исключив некоторые петли и сохранив остальные без изменения. Это может быть сделано введением дополнительных гипотетических переменных, соответствующих переменным в петлях, и превращением в них петель.

П. 18. Петли могут быть элиминированы из графа превращением их в гипотетические «входные» переменные, лежащие между данной петлей и остальной частью графа.

Если, например, Y и Z суть переменные, связанные петлей, то вводятся новые переменные Y' и Z' . Гипотетическая переменная Y' введена как промежуточная переменная между входными переменными Y (вне петли) и самим Y . Аналогично Z' становится промежуточной переменной между внешними входами Z и самим Z . В 2.39 дается пример этого.

2.39.

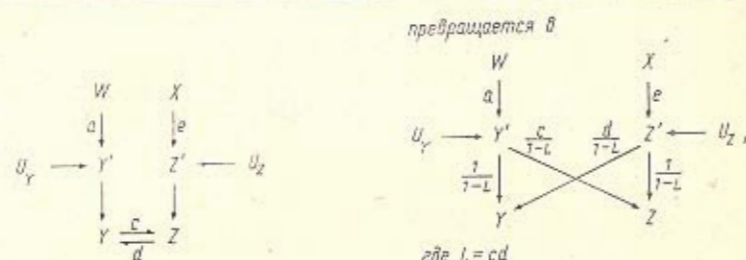


Заметим, что стрелки от гипотетических входных переменных символами не метаются. Это означает, что коэффициенты при этих стрелках равны единице. Так, скажем, влияние эффекта на Y от любого первоначального входа Y на измененном графе будет точно таким же, как и на исходном графе.

¹ Это соображение принадлежит Д. Энтвайсу.

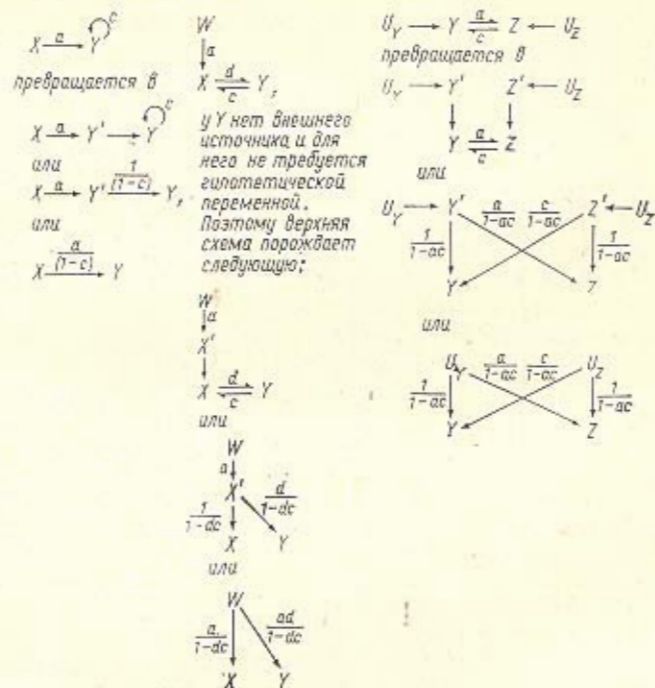
Теперь, используя обычные процедуры приведения, определенные правилом П. 17 применительно к новому графу, включающему гипотетические переменные и петлю, можно выразить в приведенной форме значения первоначальных переменных из петли. Соответствующий пример приведен в 2.40.

2.40.



Если у переменных петли только один внешний источник, гипотетические переменные становятся излишними и могут быть исключены, как это показано в 2.41.

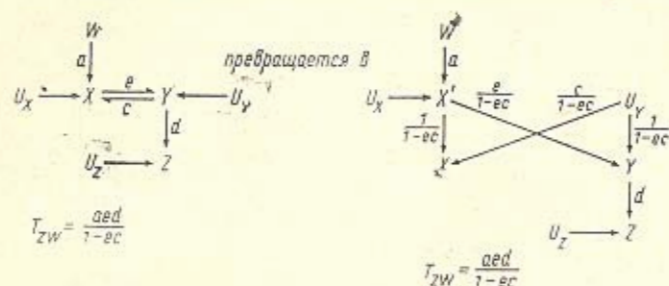
2.41.



Исключение петли, таким образом, сохраняет правильные отношения между переменными петли и всеми первоначальными

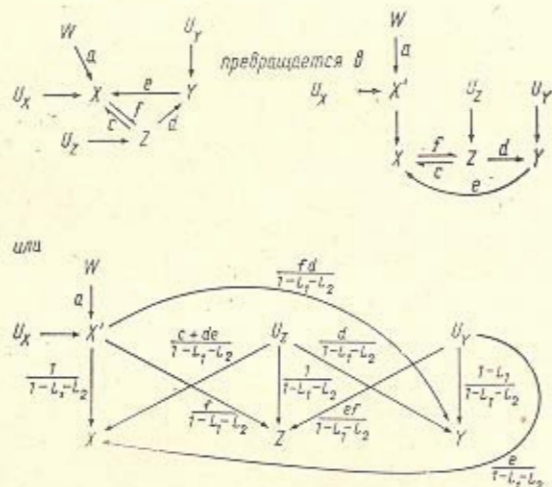
входами. Более того, приведенные петли сохраняют должные отношения между переменными «выше» и «ниже», как это показано в 2.42.

2.42.



Гнезда петель также могут быть частично редуцированы, как это показано в 2.43.

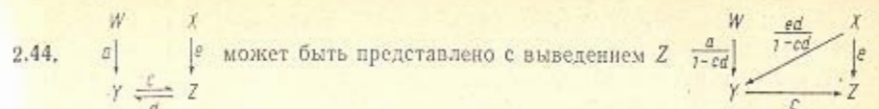
2.43.



где $L_1 = L_{XZ} = f \cdot c$ и $L_2 = L_{XZY} = f \cdot d \cdot e$.

Выведение является другой формой частичного приведения. Оно пригодится нам в гл. 5 при изучении некоторой процедуры, называемой оцениванием двухшагового метода наименьших квадратов.

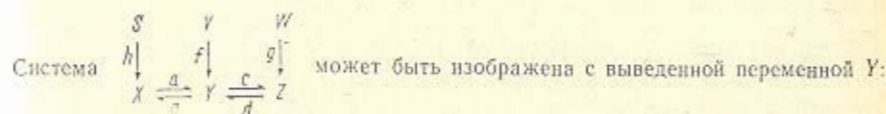
II. 19. Выбранные переменные могут быть «выведены» из петли исключением любых стрелок обратной связи, идущих от них, и представлением всех других зависимых переменных системы в приведенной форме.



Зависимость Y от Z скрыта в измененной диаграмме, но все полные эффекты от источников представлены правильно:

$$T_{YW} = \frac{a}{1 - cd}, \quad T_{YX} = \frac{ed}{1 - cd}, \quad T_{ZW} = \frac{ac}{1 - cd}.$$

$$T_{ZX} = e + \frac{edc}{1 - dc} = \frac{e - edc + edc}{1 - dc} = \frac{e}{1 - dc}.$$



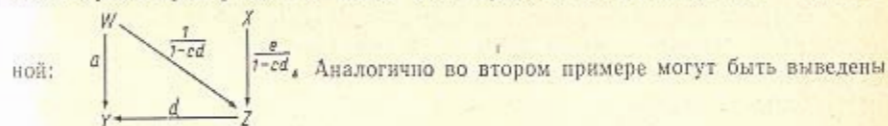
Здесь $\Delta = 1 - ae - cd$. И снова полные эффекты

в этой измененной диаграмме представлены совершенно правильно:

$$T_{YS} = \frac{h(1 - cd)}{\Delta} \cdot a + \frac{hac}{\Delta} \cdot d = \frac{ha}{\Delta} (1 - cd + cd) = \frac{ha}{\Delta}.$$

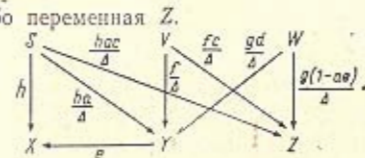
Выведение переменных дает еще один путь редукции петель; оно сохраняет без изменений больше отношений первичных переменных петель, чем частичное приведение по правилу II. 18. Редукция с помощью вывода, однако, не симметрична, и всегда возможен другой способ редукции, как показано в 2.45.

2.45. Первый пример из 2.44 может быть представлен с выводением Y-переменной:

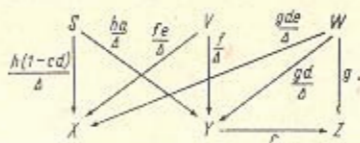


либо переменная X, либо переменная Z.

Выводим X:



Выводим Z:



Выведение дает возможность разглядеть отношения переменных петли к внешним источникам (не из петли). Первое выведение в 2.44 показывает, что Z изменится на ϵ единиц, когда X изменится на одну единицу, но изменение X порождает дальнейший опосредованный эффект, вследствие изменения Y через петлю. Выведение делит эффект влияния внешнего источника на прямую компоненту и косвенную компоненту, вызываемую динамичностью петли. Оно показывает, что эти две различные части эффекта суть просто части одной суммы.

Более общие схемы анализа потоковых графов

Существуют дополнительные правила анализа графов и, в частности, имеются правила обращения графов. Грубо говоря, это означает изменение направления стрелок в соответствии с переходом от записи уравнения в виде $Y = a \cdot X$ к записи в виде $X = Y/a$. Правила обращения графов здесь не рассматриваются, но более сложные, чем эта, работы, включенные в список литературы, содержат правила для обращения графов.

Девятнадцать правил, введенных выше, позволяют нам изображать графически и анализировать системы самых различных структур, в том числе с петлями и гнездами петель. Необходимо, однако, отметить, что сложные большие системы требуют таких запутанных графов, что практически удобнее превратить их во множество структурных уравнений и для анализа пользоваться соответствующими стандартными математическими средствами.

ПРИКЛАДНЫЕ ИССЛЕДОВАНИЯ

Основная цель теоретизирования состоит в организации информации таким образом, чтобы можно было получать из нее нетривиальные выводы. Утверждения теории на языке парных отношений обычно этого не обеспечивают, поэтому их необходимо так логически преобразовать, чтобы можно было сделать выводы, содержащие новые знания. Предыдущие правила дают средства извлечения более глубоких пластов теории, выраженной в форме диаграммы.

Критические случаи

Концепция причинности предполагает, что наблюдаемое событие должно иметь определенную конфигурацию значений входов, как только зафиксированы значения входов, а причинные процессы длятся достаточно долго, чтобы себя проявить. Более того, причинная теория предполагает определенную последовательность, в соответствии с которой эта конфигурация должна развиваться: этот процесс упорядочивания должен начаться с входной переменной и передаваться вдоль стрелок графа к разным зависимым переменным. То, что может наблюдаться после того,

как процесс достигнет конечного состояния, может служить в качестве критического факта теоретического анализа.

Предположим, рассматривается простая теоретическая конструкция в виде цепи $X \rightarrow Y \rightarrow Z$, в которой все отношения положительны. Предположим далее, что может возникнуть случай, когда X резко примет очень высокое значение, в то время как Y и Z остаются на низком уровне. Теоретически такая ситуация не может быть окончательной, поскольку Y непосредственно зависит от X и должен увеличить свое значение, а Z , зависящий в свою очередь от Y , также должен возрасти. Если эти гипотезы будут подтверждены последующими исследованиями, теория будет иметь определенную степень истинности.

Такой подход используется многими теоретиками для подтверждения их теорий; например, Меггерс¹ использовала эту технику, чтобы обосновать правдоподобие тезиса о том, что сложность культуры непосредственно зависит от возделывания земель, занимаемых обществом. Общины, подчиненные племени майя, изгнанные из центра государства на окраины в джунгли, характеризуются крайне низким значением показателя возделывания почвы, тогда как зависимая переменная сложности культуры в начальной стадии сохраняет высокое значение. Соответственно если эта теория верна, то крайне низкий уровень возделывания почв, напротив, должен был бы вызвать снижение сложности культуры. Посредством реконструкции истории Меггерс нашла, что в действительности дело обстояло именно так. Подчиненные племена за несколько столетий превратились в «примитивные» по культуре.

Объяснение

Выражения редуцированной формы определяют значения зависимых переменных через значения входных переменных системы. В линейных системах такие алгебраические выражения могут быть инверсированы; этим подтверждается, что отдельное значение зависимой переменной содержит в себе предопределенное значение входной переменной. Положим, например, что уровень преступности (R) возрастает и известно, что он зависит от трех входов: W , X и Y . Положим, что W и X в течение недавнего времени не показали заметного изменения, но что Y недавно принял новое значение, которое сейчас сохраняется как константа. Было бы тогда естественно, что изменение в Y объясняет рост преступности или что этот рост вызван возмущениями, здесь не рассматриваемыми. Если мы считаем, что возмущения минимальны, выражение приведенной формы (T_{RY}), соотносящее R к Y , может представлять интерес, отвечая на вопрос, достигла ли волна преступности своего пика или будет продолжать расти? Если

¹ Meggers Betty J. Environmental Limitation on the Development of Culture. — American Anthropologist, 1954, № 56, p. 801—824.

настоящий уровень Y , умноженный на выражение приведенной формы, дает значение, близкое к настоящей степени преступности, необходимо считать, что ее подъем должен скоро закончиться. Если он дает большее значение, можно ожидать дальнейшего роста преступности.

Польза такого анализа для объяснения заключается в том, что мы можем идти дальше простейшей интерпретации и выявлять отдаленные, косвенные причины даже в сложных системах. При большом количестве путей и пересекающихся петель становится почти невозможно дать удовлетворительное качественное объяснение явления в терминах отдаленных причин. Однако для аналитического анализа это совсем не является невозможным.

Конечно, логическая сила, достигаемая анализом графов, не уменьшает неопределенности объяснения, когда эта неопределенность является частью анализируемой системы. (Как в случае неконтролируемых возмущений, действующих на уровень преступности.) Кроме того, мы должны либо оценить, либо угадать значения структурных коэффициентов (т. е. силу влияния), прежде чем ожидать, что анализ будет достаточно точным и ценным.

Прогнозирование

Выражения приведенной формы, полученные из анализа графа, определяют для системы уравнения «входа — выхода». Если на данный момент времени будут заданы индивидуальные значения входных переменных, то окончательные последствия для всех зависимых переменных могут быть легко рассчитаны. Таким образом, теория, дополненная количественными оценками структурных коэффициентов, может быть использована для предвидения будущих событий, которые последуют как отзвук текущих изменений. Использование графов и выражений приведенной формы имеет два преимущества по сравнению с менее формальными подходами к прогнозированию. Во-первых, могут быть легко прослежены все последствия изменений входной переменной, даже если система включает множество запутанных причинных связей. Во-вторых, можно оценить изменения зависимой переменной от всех входных переменных системы. Этим можно избежать чрезмерного упрощения и ошибочных суждений.

Некоторые запреты

Потоковые графы и соответствующие им выражения приведенной формы могут ориентировать принятие практических решений, 1) указывая те исходные переменные, изменение которых может привести к желаемому результату; 2) подсказывая величину действий, необходимых для достижения желательных изменений; 3) указывая те дополнительные эффекты, которые возникнут помимо желаемых.

Определяя входные переменные, манипулирование которыми может привести к нужному результату, анализ потоковых графов особо подчеркивает, что не надо слишком прямолинейно, в лоб,

устранять нежелательные последствия. Пусть выходная переменная Z может принять нежелательные значения в силу ее зависимости от входной переменной X . Тогда один из способов устранить нежелательное значение Z заключается в выполнении этого посредством изменения X . Но если на зависимую переменную влияет еще другая переменная Y , то нежелательное значение Z может быть исключено посредством воздействия также на Y . Выявление всех источников обеспечивает повышение гибкости стратегии с точки зрения практичности, стоимости и этических норм.

При наличии количественной оценки структурных коэффициентов в выражениях приведенной формы становится возможным определение величины изменения в источнике, необходимого для достижения желаемого изменения выходной переменной. Этот тип квантификации сам по себе вполне годится для объективного анализа наилучшего способа действий. Например, может действительно оказаться, что некоторый источник имеет громадное влияние на целевую зависимую переменную, но издержки или этические соображения препятствуют действиям данного рода. Тогда менее эффективное манипулирование другим источником или комбинация действий с несколькими источниками могут быть предпочтительнее, и можно провести до конца сравнительно точный анализ затрат — результатов, используя выражения приведенной формы.

Как только будут выбраны отдельный источник или множество источников, можно будет провести последовательность соответствующих процедур не только относительно данной выходной переменной, но и по отношению к другим выходным переменным системы, используя выражение приведенной формы. Поэтому, проводя аналитическое моделирование процесса до его реализации, мы можем сделать вывод, что совокупная комбинация всех влияний, производимых данными управляющими действиями, оказывается совсем не столь желаемой, как это казалось сначала. В таком случае, конечно, можно продолжать анализ, разыскивая источник или набор источников, манипуляция которыми дает оптимальный выигрыш и в то же самое время влечет минимальный ущерб.

Источники и дополнительная литература

Артуром Л. Стингкомбом предложен другой вариант введения в элементарный анализ потоковых графов, предназначенный для исследователей в области общественных наук, *Constructing Social Theories*, New York, 1968. Кроме того, анализу графов многие учебники по системному анализу посвящают раздел или главу. Например, A. D. Hall. *A Methodology for Systems Engineering*, Princeton, N. J., 1962. Отлично написанное обозрение более сложной техники анализа графов дается в работе W. H. Huggins and Doris R. Entwistle. *Introductory System and Design*, Waltham, Mass., 1968. Этот источник предлагает многочисленные задачи и примеры, но без решений. Чарльз С. Лоренс (C. S. Lorenz).

Flowgraphs for the Modeling and Analysis of Linear Systems, N.Y., McGraw-Hill, 1964) предлагает более математизированную технику и показывает, как применение анализа графов может быть распространено на многочисленные технические проблемы, имеющие общий интерес.

УПРАЖНЕНИЯ

1. Несмотря на то что на графах единицы измерения структурных коэффициентов показываются не часто, о них нельзя забывать. Пусть, например, структурный коэффициент влияния переменной X на переменную Y равен 0,4. Это значит, что причинный оператор порождает 0,4 единицы Y на одну единицу X , и поэтому полный коэффициент выражается как $0,4 \left[\frac{\text{единицы } Y}{\text{единицы } X} \right]$. Предположим, социально-профессиональный статус измерен в шкале рассчитываемого социального положения (такой, какая разрабатывается в Исследовательском центре общественного мнения — National Opinion Research Center): для удобства мы можем обозначать каждую единицу по «шкале NORC». Кроме того, пусть образование будет измеряться числом лет обучения. Наконец, предположим, что известно следующее: если индивидуум A имеет статус на 70 единиц NORC выше, чем индивидуум B , то, как правило, уровень образования у сыновей A будет на 3 года обучения больше, чем у сыновей B . При таких данных что можно сказать о полном коэффициенте и единицах измерения влияния занятий отца на образование сына?

2. Предположим, социально-профессиональный статус некоего человека оценивается в 60 единиц NORC. Используя результат предыдущего примера, определите ожидаемый уровень образования его сыновей. В частности, покажите, как сокращаются единицы размерности показателя статуса отца с теми же единицами в структурном коэффициенте, в результате чего получается размерность образования.

3. Положим, в результате ряда неудачных обстоятельств некоторый индивидуум к достижению своего совершеннолетия потерял в профессиональном статусе 10 единиц NORC. Используя результат первого примера, определите ожидаемое влияние этой ситуации на образование его сыновей. В частности, покажите, как сокращаются единицы размерности показателя потери профессионального статуса с соответствующими единицами размерности в структурном коэффициенте.

4. Пусть структурный коэффициент, соотносящий образование сына с его конечным профессиональным статусом, равен:

$$5,0 \left[\frac{\text{NORC сына}}{\text{год обучения сына}} \right].$$

Используйте этот факт и результат упражнения 1, чтобы предсказать профессиональное положение сыновей человека с социально-профессиональным статусом в 80 единиц NORC. Покажите, как при вычислениях сокращаются единицы размерности.

5. Пусть мы изучаем некоторое общество, отношения в котором описываются условиями задач 1 и 4, в котором обычно отцы используют непосредственно контакты и влияние, связанные с их статусом, для повышения статуса сыновей. Положим, структурный коэффициент прямого влияния статуса отца на статус сына равен:

$$0,25 \left[\frac{\text{NORC сына}}{\text{NORC отца}} \right].$$

(а) Постройте граф связей для этой гипотетической системы мобильности поколений, показывающий структурные коэффициенты в соответствующих единицах измерения (пренебрегая возможными возмущениями).

(б) Используйте правила редукции графа для того, чтобы написать уравнение, связывающее профессиональные уровни отца и сына. Проследите за единицами измерения в процессе редукции.

(в) Какой будет ожидаемый профессиональный уровень сына для данной системы отношений, если статус отца равен 80 единицам NORC?

6. По-видимому, верно, что:

возрастание доли немущих в городе приводит к росту уровня преступности;
высокий уровень преступности требует больше полицейских;
большая плотность полиции в городе сдерживает преступность.

Пусть переменные будут определены следующим образом:

$$I \text{ (доля немущих)} = \frac{\text{число немущих в городе}}{\text{число жителей в городе}};$$

$$C \text{ (уровень преступности)} = \frac{\text{число преступлений в городе}}{\text{число жителей в городе}};$$

$$P \text{ (насыщенность города полицией)} = \frac{\text{число полицейских в городе}}{\text{число жителей в городе}}.$$

Пусть также количественная величина структурных коэффициентов для каждого из этих отношений представляется как d , e и f соответственно.

Какой будет знак у каждого из коэффициентов? Каковы единицы измерения каждого коэффициента? Нарисуйте граф для данной системы и снабдите его единицами измерения, соответствующими каждому коэффициенту.

7. В каких единицах выражается эффект обратной связи в задаче 6? В каких единицах должен был бы выражаться эффект обратной связи, если бы уровень преступности имел другую базу, например число преступлений относилось не к одному жителю, а к 1000 жителей?

8. Пронтерпретируйте смысл приводимых ниже ограничений в терминах системы, определенной упражнением 6, и укажите влияние на уровень преступности и насыщенность города полицией следующих мероприятий:

а) принятая программа повышения минимального дохода, чтобы почти всех людей вывести из состояния бедности;

б) без изменения реальных доходов осуществляется программа роста благосостояния, при которой у немущих удовлетворяются основные материальные потребности и они избавлены от стресса и унижений, толкающих их на преступные действия;

в) полиции приданы повышенные технические средства для того, чтобы увеличить ее мощь и эффективность, а средства массовой информации поддерживают ее своими сообщениями;

г) федеральные фонды распределяются по всем городам так, чтобы довести размер полицейских сил до одного полицейского на 10 000 жителей.

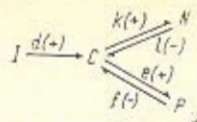
9. Причинная зависимость насыщенности города полицией от уровня преступности в упражнении 6 может возрасти благодаря контролю политических сил за соблюдением закона; т. е. потребность в большей или меньшей (по численности) полиции определяется политическими силами и средствами массовой информации в зависимости от изменений уровня преступности. Можно ли ожидать, что преступность будет непрерывно уменьшаться, как только этот политический оператор будет введен в систему обеспечения действия закона? Почему?

10. Продолжим снова рассмотрение предположений упражнения 6. На этот раз проблема заключается в оценке влияния дополнительной управляющей петли, возникающей в системе при введении разных федеральных программ. По первому плану переменная федеральной поддержки местной полиции присоединяется к системе так, что высокий уровень поддержки ведет к увеличению уровня насыщенности полицией, а высокий уровень насыщенности полицией ведет к уменьшению федеральной помощи:

$$I \xrightarrow{d(+)} C \xleftarrow{e(+)} P \xleftarrow{f(-)} F.$$

Второй план связывает национальную обеспеченность полицией (например, агентами ФБР или Государственного казначейства) непосредственно с уровнем преступности в городах. Чем больше уровень преступности, тем больше агентов

назначается в эти города, и чем больше выделено агентов, тем ниже уровень преступности.



Предположим, что программы разрабатываются так, что степень влияния петли управления одна и та же в обоих случаях, т. е. $j \cdot g = k \cdot l = L$. Какая из двух программ ведет к наибольшему сокращению преступности? Объясните этот результат словами на содержательном уровне.

11. Пусть мощь регулярных вооруженных сил каких-либо трех стран A, B и C обозначена соответственно буквами A, B, C . По предположению все три переменные измерены в единой шкале. Положим, что вооруженные силы любого государства определяются следующими условиями:

1) T — размером территории, которую должно защищать государство: большая площадь порождает большие оборонительные силы с управлениями различных родов войск, действующими как оператор (а);

2) S — уровнем развития фундаментальной науки. Всевозможные научные достижения превращаются в достижения военной техники, при этом военно-промышленный комплекс служит оператором (d). С другой стороны, военные исследования и программы развития, по-видимому, незначительно воздействуют на чистую науку (в отличие от техники), поэтому причинная направленность может считаться односторонней.

В соответствии с известной моделью Ричардсона (см., например, Томас Л. Саати. Математические модели конфликтных ситуаций, М., Сов. радио, 1977) можно положить, что возрастание вооружений одной державы вызывает соответствующий рост и у других. В каждом случае оператор, отвечающий за реакцию на движение другой стороны (f), является комплексом, состоящим из национальной интеллигенции, политических деятелей и военно-промышленного сектора.

(а) Изобразите потоковый граф, представляющий эту систему, используя индексы, когда это необходимо, чтобы различать переменные и операторы. Выпишите также структурные уравнения.

(б) Сколько петель будет у этой системы? Какие петли будут усилительными и какие — управляющими?

(в) Найдите полный эффект влияния (T) уровня науки страны A (S_A) на ее вооруженные силы (A). Опишите на бытовом языке, какой вклад вносит в этот общий эффект каждая петля.

12. Продолжим анализ системы упражнения 11. Предположим, теперь, что возрастание вооруженных сил в одной стране приводит к постоянному их росту в двух других странах и что отношения те же самые, независимо от того, какая страна является инициатором. Другими словами, все f положительны и равны друг другу. Кроме того, предположим, что время реагирования одно и то же, независимо от страны. При данных условиях система будет устойчивой, только в том случае, если величина f будет меньше, чем 0,5. (Это было установлено анализом динамики системы.) Проинтерпретируйте смысл f -коэффициентов значения, меньшего, чем 0,5, и на базе своей интерпретации оцените, устойчивой или неустойчивой будет эта международная система.

13. а) Выведите объем вооружений страны A для системы, предложенной в упражнении 11. Как и в упражнении 12, снова предположите для простоты, что все f равны.

б) Возможно, определенный тип людей действительно воспринимает систему в ее приведенной форме (после «выведения» переменных петель). Если это так, то должны существовать операторы (или их представления) для всех введенных стрелок. Какой тип операторов мог бы объяснить эти пути? Как бы такая личность объяснила изменения в уровне вооружений, возникшие после новых достижений науки страны A ? После новых достижений в стране B ? Что будет сильным и слабым местом восприятия системы в приведенной форме — в предположении, что система с первоначальной обратной связью является более точной моделью?

Глава 3 • СТАТИСТИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ

Потоковый граф описывает причинные процессы, затрагивающие частный случай социальной системы (например, семья, фирма или экологическое окружение), и одно и то же структурное описание может не подходить в каких-либо других случаях. Однако частные ситуации не могут служить обычной целью социального исследования. Более часто отыскивают модель, обобщающую некоторый класс ситуаций (например, все семьи без отца, все банки, все пограничные города), так что одно и то же описание системы может использоваться для объяснения результатов любого конкретного случая в терминах общей причинной структуры, но при индивидуальных входах. Тогда в каждом конкретном случае что-то будет известно о способе действия оператора, какие-то выводы о его истории могут быть сделаны по настоящему положению, как и предсказания относительно реакции при новых входах.

Рассмотрение целого класса событий сразу вводит нас в новую сферу применения причинного анализа, нежели та, которую мы рассматривали раньше. Мы больше не имеем дела с единственной системой, изучаемой с точки зрения ее реакции на те или иные входные сигналы. Теперь наше внимание направлено на совокупности эквивалентных систем, и мы рассматриваем распределение выходов, генерированное в каждом отдельном случае реакцией на свои собственные значения входов.

Такое обобщение придает новую силу теоретическому анализу и в эмпирических исследованиях нацелено на причинное влияние и количественное описание систем, как это обсуждается в гл. 4 и 5. Чтобы реализовать возможности этого подхода, необходимо иметь язык для описания совокупностей, в частности для описания распределений их переменных и для установления соотношений между распределениями разных переменных. На этом концентрирует свое внимание статистика, и в этой главе приводятся некоторые элементарные сведения из статистики.

Распределения

Структурное уравнение задает алгебраическое соотношение, существующее между переменными причины и следствия, когда причина сохраняет свое значение на постоянном уровне, а все причинные действия завершаются порождением значений выхода.

Например, в простой системе $X \xrightarrow{\frac{1}{2}} Y$ значение Y полностью

определяется X , и если X установлен на некотором значении и сохраняет его, то (раньше или позже) Y должен принять значение, равное $\frac{1}{2}X$. Если X установлен на 1, то Y должен быть равным 0,5; если X установлен на 5, то Y должен быть равным 2,5.

Положим, наблюдается случай, когда X устанавливается каким-то образом на значение +1,0; затем обнаруживается, что $X = -3,0$, в другом случае $X = -1,0$ и, наконец, следующий случай, когда $X = 3,0$. Из уравнения $Y = \frac{1}{2}X$ необходимо ожидать, что Y примет соответственно значения +0,5; -1,5; -0,5 и +1,5. Табл. 3.1 построена для того, чтобы изобразить всю эту информацию в сжатой форме:

3.1. Распределение двух переменных: X и Y

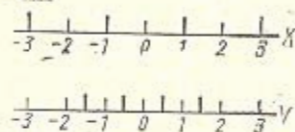
Случай	X значение	Y значение
1	+1,0	+0,5
2	-3,0	-1,5
3	-1,0	-0,5
4	+3,0	+1,5

Столбцы в 3.1 представляют переменные X и Y несколько в новом свете. Переменная принимает различные значения, когда мы идем вниз по столбцам, значение меняется в зависимости от наблюдаемого случая. Несомненно, X (или Y) по-прежнему представляет переменную, причинно связанную с другой переменной системы, но множество значений в каждом столбце подчеркивает, что X (или Y) имеет также *распределение*, в данном случае — по множеству случаев. Так как распределения играют основную роль во всем последующем анализе, полезно выразить идею распределений графически и заранее приготовить идеи для характеристики их как целого.

Графики распределений

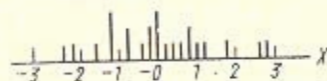
Распределение значений X в примере 3.1 может быть представлено графически, если изобразить шкалу измерения X как горизонтальную ось и представить точками на вертикальной оси число случаев каждого измерения. Например, распределение X (и Y) из примера 3.1 может быть изображено графически следующим образом:

3.2. Графики распределений 3.1.



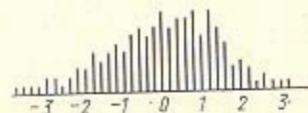
Обычно распределения предполагают больше случаев, чем здесь, — сотни и тысячи — и графики распределения принимают определенную форму по мере увеличения числа случаев. Например, при 25 случаях распределение X может принять форму, показанную ниже на 3.3.

3.3. Возможное распределение переменной X при 25 случаях.



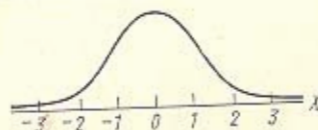
Когда число случаев становится все больше, верхняя часть распределения выглядит все более похожей на гладкую кривую, что иллюстрируется в 3.4.

3.4. Возможное распределение с большим числом случаев.



Как показано на 3.4, распределение переменной не обязано быть симметричным. Обоснованность последующего анализа не зависит от симметричности. В частности, ни одна из дальнейших процедур не зависит от того, что распределение имеет «нормальный» тип.

3.5. Нормальная кривая.



Нормальные распределения играют важную роль в теории вероятностей; так, они полезны, когда при наличии небольшого количества информации хотят сделать статистические выводы относительно всех случаев в совокупности. Однако *предполагать нормальность распределений нет необходимости, когда работают на чисто теоретическом уровне анализа систем*, потому что аналитические процедуры, развиваются применительно к данной совокупности всех существующих случаев. Поэтому делаются выводы невероятной природы. Только когда мы оцениваем значения структурных коэффициентов по выборочным данным, делается предположение нормальности; при этом оно используется в выработке суждения о достоверности результатов оценивания. (Однако в случае когда распределения одних переменных асимметричны, а других — приближенно нормальны, можно считать, что рассматриваемая система не является строго линейной (см. упражнение 2).)

Статистическое изучение статических систем включает другое предположение о распределении переменных. Так, предполагается, что дисперсия каждой переменной конечна, что вероятность появления на шкале переменной значения, выше некоторого большого числа или ниже некоторого очень малого, крайне мала. Прежнее ограничение анализа устойчивыми линейными системами гарантирует это, поэтому с данной точки зрения нет необходимости делать какие-либо новые предположения.

Среднее

Среднее есть центр тяжести распределения — точка на шкале измерения, в которой фигура, ограниченная кривой распределения, точно должна уравновеситься, если ее поместить на оси вращения. Среднее вычисляется суммированием значений всех случаев и делением на число этих случаев.

Среднее является настолько важной математической характеристикой, что используется всегда, во всех случаях и практически для всех типов распределений.

3.6. Алгебра математического ожидания

«Ожидаемое значение» переменной есть ее среднее, вычисленное на совокупности всех возможных исходов. Оно обозначается буквой $\mathcal{E}$ перед соответствующей переменной, заключенной в скобки. Например, $\mathcal{E}(X)$ — среднее значение величины X по всем возможным исходам в совокупности. Если переменная выражена с помощью алгебраических выражений, ожидаемое значение переменной может быть определено с использованием следующих правил.

1. Ожидаемое значение переменной, умноженное на константу, равно этой константе, умноженной на ожидаемое значение переменной

$$\mathcal{E}(aX) = a\mathcal{E}(X).$$

2. Ожидаемое значение суммы двух переменных равно сумме ожидаемых значений

$$\mathcal{E}(X + Y) = \mathcal{E}(X) + \mathcal{E}(Y).$$

3. Ожидаемое значение произведения переменной на себя или другую переменную не может быть представлено в виде разложения на произведение подобно разложению ожидаемого значения суммы:

$$\left. \begin{array}{l} \mathcal{E}(X^2) \\ \mathcal{E}(X \cdot Y) \end{array} \right\} \text{ не могут быть упрощены далее.}$$

Пример. Пусть Z определяется соотношением

$$Z = (a \cdot X) + (b \cdot Y) + (c \cdot X \cdot Y).$$

Тогда ожидаемое значение величины Z равно:

$$\begin{aligned} \mathcal{E}(Z) &= \mathcal{E}(a \cdot X + b \cdot Y + c \cdot X \cdot Y) = \mathcal{E}(a \cdot X) + \mathcal{E}(b \cdot Y) + \mathcal{E}(c \cdot X \cdot Y) = \\ &= a \cdot \mathcal{E}(X) + b \cdot \mathcal{E}(Y) + c \cdot \mathcal{E}(X \cdot Y). \end{aligned}$$

Несмотря на фундаментальное значение средних, здесь они не будут встречаться в выражениях с использованием распределений переменных, потому что с этого момента и далее предполагается, что переменные измерены в таких шкалах, что нулевая точка шкалы является средней данного распределения. Другими словами, предположено, что переменная измеряется в терминах отклонений от среднего или просто в величинах отклонений. Как было

замечено в гл. 2, потоковые графы и полученные из них выражения могут быть упрощены, если шкалы измерений выбраны так, что постоянный член исключен. Именно это здесь и делается. Использование величины отклонений исключает из диаграмм константы, как и из производных выражений, и это условие упрощает статистический анализ. Как указывалось в гл. 2, прибавление или вычитание констант с целью приведения измерительных шкал к более удобной форме не создает дополнительных ограничений для статического анализа. Величины отклонений — это просто измерители, получаемые вычитанием константы — ожидаемого значения переменной — из каждого первоначального значения.

Для того чтобы отразить факт предполагаемого теперь изменения в величинах отклонений, не будет делаться никакого изменения в обозначениях. На это условие лучше смотреть как на уточнение предыдущего обсуждения.

Дисперсия

Распределения, порождаемые отношением $X \xrightarrow{\frac{1}{2}} Y$, показанные на 3.2, имеют то же самое среднее (0,0), но имеется очевидное различие в том, как они развернуты: распределение X имеет большую область, чем распределение Y . Другими словами, значения X в среднем лежат дальше от центра своего распределения, чем значения Y — от центра их распределения. Значения X называют на больший разброс.

Наиболее употребительная мера разброса — дисперсия — средний квадрат расстояний индивидуальных значений от среднего данного распределения. При условии измерения всех переменных в величинах отклонений дисперсии могут быть получены просто возведением в квадрат наблюдаемых значений и нахождением их среднего. Итак, дисперсия, обозначаемая символом σ^2 , есть ожидаемое значение возведенного в квадрат отклонения от среднего.

3.7. Определение дисперсии

$$\sigma_X^2 = \mathcal{E}(X^2),$$

где переменная X измеряется в величинах отклонений.

Пример.

Дисперсия переменной распределения X в 3.1 равна:

$$\begin{aligned} \sigma_X^2 &= [(1,0)^2 + (-3,0)^2 + (-1,0)^2 + (3,0)^2]/4 = \\ &= [1 + 9 + 1 + 9]/4 = 20/4 = 5,0. \end{aligned}$$

Дисперсия переменной Y равна:

$$\begin{aligned} \sigma_Y^2 &= [(0,5)^2 + (-1,5)^2 + (-0,5)^2 + (1,5)^2]/4 = \\ &= [0,25 + 2,25 + 0,25 + 2,25]/4 = 5/4 = 1,25. \end{aligned}$$

Совместные распределения

Причинное отношение между двумя переменными согласовывает их значения. Например, в 3.1 положительные значения X соответствуют положительным значениям Y , а отрицательные

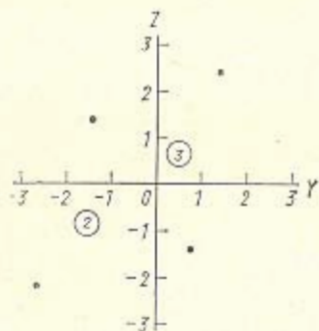
значения X — отрицательным значениям Y . Чтобы исследовать согласованность двух переменных, необходимо рассмотреть их совместное распределение.

Графики совместных распределений

График (или диаграмма рассеивания) совместного распределения строится представлением шкалы измерения одной из переменных в виде горизонтальной оси, а шкалы другой переменной — в виде вертикальной оси. Случаи конкретного сочетания значений двух переменных изображаются точками на плоскости графика. (Если в некоторую точку попало более одного случая, можно это отметить, написав число исходов в этом месте.)

При одном подходе диаграмма рассеивания есть просто нечто вроде усовершенствованного варианта таблицы распределения по реализациям, что легко увидеть, сравнивая табл. 3.9 с графиком 3.8.

3.8. Диаграмма рассеивания, показывающая совместное распределение 10 исходов переменных Y и Z .

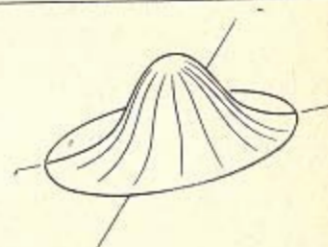


3.9. Совместное распределение Y и Z , представленное в виде таблицы

	Y					
	от -3,0 до -2,1	от -2,0 до -1,1	от -1,0 до -0,1	от 0 до 0,9	от 1,0 до 1,9	от 2,0 до 2,9
Z	от 2,0 до 2,9				1	
	от 1,0 до 1,9		1			
	от 0 до 0,9			3		
	от -1,0 до -0,1		2			
	от -2 до -1,1			1		1
	от -3 до -2,1	1				

При другом подходе диаграмма рассеивания изображается в виде трехмерного графика, рассматриваемого сверху. Каждая точка указывает единичный отрезок прямой, идущий вертикально от плоскости и обозначающий, что в этом месте происходит одно событие; число на графике обозначает, что более длинный отрезок от данного места поднимается вертикально (например, число 2 на графике 3.8 указывает, что отрезок длиной в 2 единицы восходит в этом месте). Когда рассматривается небольшое число событий, поверхность, образованная вершинами этих вертикальных отрезков, оказывается неровной. Однако при значительном увеличении числа событий поверхность выравнивается, обычно принимая вид колокола.

3.10. Возможна гладкая поверхность, образуемая совместной функцией распределения двух переменных, когда рассматривается большое число событий. Обычно такая поверхность имеет вид колокола, хотя этот колокол может быть деформирован самым различным образом.



Когда имеют дело только с несколькими сотнями событий, вертикальное измерение в диаграммах рассеивания можно в значительной степени игнорировать с целью уделить внимание расположению рисунка, образованного точками двухмерного представления. Если значения двух переменных некоторым образом согласованы, их совместное распределение образует на диаграмме рассеивания вполне определенную картину.

3.11a. Диаграмма рассеивания для двух переменных с несогласованными значениями.



3.11b. Диаграмма рассеивания для двух переменных, значения которых согласованы в значительной степени. Прямая линия, проходящая через точки, может быть использована для предсказания ожидаемого значения Z при заданном конкретном значении X .



3.11в. Диаграмма рассеивания для двух переменных, значения которых скоординированы в значительной степени. Кривая линия, проходящая через точки, может быть использована для предсказания ожидаемого значения Y , если задано конкретное частное значение W .

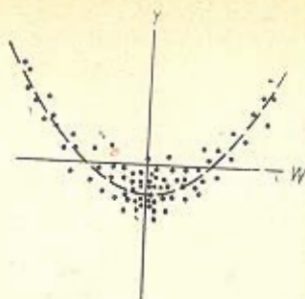
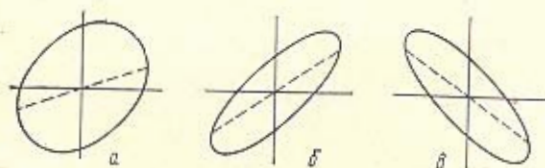


График 3.11а не указывает на какое-либо согласование; две переменные не коррелированы друг с другом. График 3.11в указывает две переменные, значения которых связаны. Тот факт, что точки выглядят сгруппированными вдоль прямой линии, означает возможность линейного описания этой взаимосвязи. Наличие таких линейных связей предполагается во многих статистических исследованиях, как и во всем анализе этой книги.

График 3.11в иллюстрирует согласованные переменные, но взаимозависимость эта нелинейна. В причинном анализе переменные этого графика должны были бы быть подвергнуты специальной процедуре. Например, можно было бы провести два приближенных линейных анализа: один — для значений W , больших или равных нулю, другой — для значений W , меньших или равных нулю.

Дополнительные примеры линейных зависимостей даются в 3.12, где эллипсы дают общее представление распределения.

3.12.



Положительная связь;
низкая степень
согласования

Положительная связь;
средняя степень
согласования

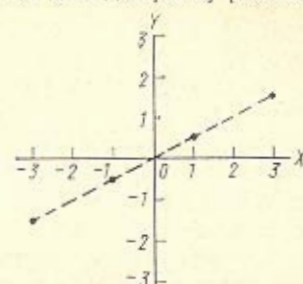
Отрицательная связь;
средняя степень
согласования

Диаграммы рассеивания обычно строят по таблицам наблюдаемых значений. Так, возвращаясь к примеру, введенному в 3.1, мы получаем следующую картину.

3.13.

Событие	X	Y
1	+1,0	+0,5
2	-3,0	-1,5
3	-1,0	-0,5
4	+3,0	+1,5

Эта таблица дает диаграмму рассеивания:



Этот пример интересен тем, что события укладываются строго вдоль прямой линии и предполагают строго линейную координацию между X и Y . И действительно, так и должно быть, потому

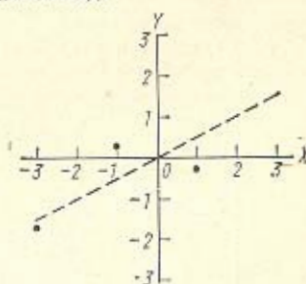
что значения Y были заданы значениями X и отношением $X \rightarrow Y$. Так как ничего, кроме X , не действует на Y , координация их значений должна быть совершенной. Полезно, однако, рассмотреть, что произойдет, когда в систему войдут несколько более сложные причинные отношения.

3.14. $X \xrightarrow{\frac{1}{2}} Y \xleftarrow{\frac{1}{4}} W$ означает $Y = \frac{1}{2}X + \frac{1}{4}W$

Если происходят события в соответствии со следующими значениями,

Событие	X	W	Y
1	-3,0	-1,0	-1,75
2	-1,0	+3,0	+0,25
3	+1,0	-3,0	-0,25
4	+3,0	+1,0	+1,75

то диаграмма рассеивания X, Y имеет вид:



Ясно, что добавление другой детерминанты для причинного влияния на Y ведет к уменьшению степени чисто линейного соответствия между X и Y . Точки на диаграмме рассеивания X, Y больше не ложатся строго на прямую линию, и больше нельзя сделать точный прогноз значения Y для нового случая, задаваемого значениями X . Это показывает, каким способом возмущения искажают прогноз. Они вносят добавки к вариациям, что не может быть учтено в терминах наблюдаемых переменных.

Еще одно обстоятельство, которое можно заметить по 3.14, состоит в том, что предсказуемость снижается, даже если причинные отношения между X и Y сохраняются теми же, что и в 3.13 (влияние X на Y одно и то же в этих двух примерах). Это происходит потому, что во втором примере у Y имеется дополнительный разброс, производимый W , и поэтому X детерминирует меньшую долю общего разброса Y . Более точно, Y имеет большую дисперсию во втором примере (1,56), чем в первом (в котором дисперсия равна 1,25). Так как переменная X вкладывает в вариацию Y в обоих случаях одно и то же количество, оно необходимо составит пропорционально меньшую часть дисперсии Y во втором примере, в связи с чем получается меньшая предсказуемость, чем это было в более простом первом случае.

Ковариация

Диаграммы рассеивания полезны для установления факта существования соответствия между двумя переменными и его формы; однако для точного исследования необходимо оценить степень этого соответствия. Наиболее подходящая мера этого — *ковариация*. Этот показатель соответствия аналогичен дисперсии, но он характеризует совместное распределение пары переменных, а не распределение единственной переменной. Дисперсия получается умножением на себя отклонений от среднего и усреднением результата. Аналогично ковариация получается перемножением отклонений от среднего одной переменной на соответствующие отклонения второй переменной и усреднением произведений. Ковариация обозначается буквой σ , снабженной двумя индексами переменных, для которых она вычисляется. Порядок индексов не играет роли, например $\sigma_{XY} = \sigma_{YX}$.

3.15. Определение ковариации

$\sigma_{XY} = \mathcal{E}(X \cdot Y)$, где X и Y суть отклонения от среднего. Вычисление ковариации проведем для примера 3.1.

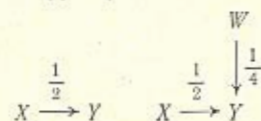
Случай	X	Y	$X \cdot Y$
1	+1,0	+0,5	+0,5
2	-3,0	-1,5	+4,5
3	-1,0	-0,5	+0,5
4	+3,0	+1,5	+4,5
Сумма			10,0

$$\sigma_{XY} = \frac{10}{4} = 2,5$$

Если ковариация равна 0, это означает, что либо две переменные совершенно независимы (как в 3.11а), либо их взаимосвязь нелинейна (как в 3.11в). Ненулевая величина ковариации отражает величину линейной связи подобно той, что показана на первой и второй фигуре в 3.12. Отрицательная величина говорит о взаимозависимости типа изображенной в третьей фигуре в 3.12.

Ковариация зависимой переменной с входной переменной не изменится, если система изменится введением дополнительного входа, не связанного с первичной входной переменной. Например, ковариация между X и Y одна и та же для обеих диаграмм в 3.16.

3.16. Ковариация σ_{XY} одна и та же для обоих случаев при условии, что дисперсия σ_X^2 постоянна, а X и W некоррелированы.



Постоянство ковариации в аналогичных ситуациях иногда весьма удобно, но, с другой стороны, это означает, что ковариация явно не отражает уменьшения предсказательной способности, что происходит при добавлении возмущения. Необходимо ввести еще один показатель для указания степени предсказуемости.

Коэффициент корреляции

Коэффициент корреляции измеряет степень, в которой вариации одной переменной могут быть предсказаны на основании информации о другой. Один из наиболее важных коэффициентов корреляции — коэффициент корреляции Пирсона ρ , представляющий собой отношение действительной ковариации между двумя переменными к максимально возможной ковариации. Если наблюдаемая ковариация равна нулю, $\rho = 0$ и эта величина тем самым указывает, что между двумя переменными вообще нет линейной связи. Если реальная ковариация равна максимальной ковариации, то $\rho = 1$ и это означает, что значения одной переменной могут быть предсказаны точно по значениям другой. Наблюдаемая ковариация может быть отрицательной, указывая, что высоким значениям одной переменной соответствуют низкие значения другой переменной, в этом случае коэффициент ρ также отрицателен, отражая этим факт обратного соответствия.

Максимально возможная ковариация определяется как произведение двух средних отклонений от центра, своего для каждой переменной. Необходимые величины этого вычисления суть *стандартные отклонения*, или корни квадратные из дисперсий двух этих переменных. Дисперсии, конечно, суть средние квадраты отклонений, и поэтому квадратный корень из дисперсии представляет собой среднеквадратическое отклонение.

Итак, формула, определяющая смешанный коэффициент корреляции, имеет следующий вид.

3.17. Определение коэффициента корреляции Пирсона

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2} \cdot \sqrt{\sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

Построение отклонений от среднего путем вычитания среднего общего распределения из первичных измерений является полезным способом для упрощенных формул, и стандартные отклонения можно использовать вполне обычным путем, чтобы получить определенное дополнительное упрощение в определении ρ .

Стандартные величины суть величины отклонений, деленные на стандартные отклонения распределения. Разница между стандартной величиной и величиной отклонения отражена в символах для переменных — для стандартных величин используются строчные буквы.

3.18. Определение стандартной величины. Здесь X уже есть отклонение от среднего

$$x = \frac{X}{\sigma_X}$$

Переменные, измеренные в масштабе стандартных величин всегда имеют нулевое среднее (они по-прежнему остаются отклонениями от среднего) и единичную дисперсию. Это единичное значение дисперсии показано в 3.19.

3.19. Дисперсия стандартизованных переменных равна 1:

$$\sigma_x^2 = \mathcal{E}(x^2) = \mathcal{E}\left(\left(\frac{X}{\sigma_X}\right)^2\right) = \mathcal{E}\left(\frac{X^2}{\sigma_X^2}\right) = \mathcal{E}\left(\left(\frac{1}{\sigma_X^2}\right) X^2\right).$$

Здесь значение величины $1/\sigma_X^2$ является константой, потому что она остается одной и той же независимо от конкретного наблюдения, сделанного в данный момент. Поскольку она — константа, ее можно вынести за знак математического ожидания и получить:

$$\sigma_x^2 = \frac{1}{\sigma_X^2} \cdot \mathcal{E}(X^2) = \frac{1}{\sigma_X^2} \sigma_X^2 = 1.$$

Кроме того, корреляция между двумя переменными, выраженными в стандартных значениях, точно равна их ковариации.

3.20. Корреляция двух стандартизованных переменных — это то же самое, что их ковариация: $\rho_{xy} = \sigma_{xy}$, потому что

$$\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2} \cdot \sqrt{\sigma_y^2}} = \frac{\sigma_{xy}}{\sqrt{1} \cdot \sqrt{1}} = \sigma_{xy}.$$

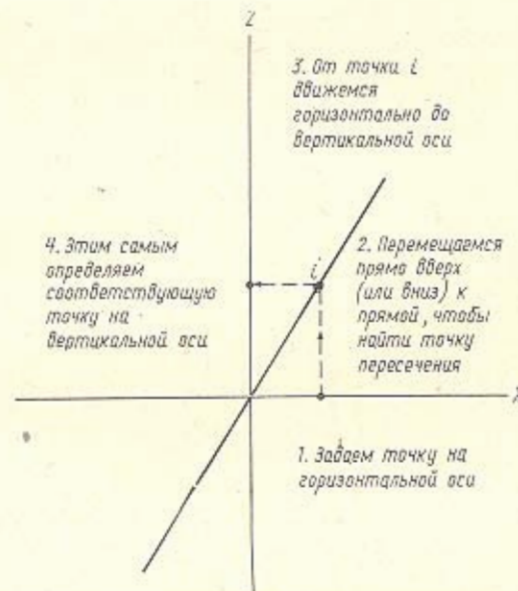
Следуя процедуре, подобной 3.19, можно также показать, что корреляция между двумя переменными в стандартизованной форме равна корреляции переменных в нестандартизованной форме, т. е. $\rho_{xy} = r_{xy}$. Однако ковариация двух переменных, измеренных в терминах обычных величин отклонений, вообще говоря, не есть та же самая, что их ковариация, когда они измерены в стандартных величинах. В общем, не существует числовых границ пределов

для ковариации, но ковариация стандартных величин должна находиться в интервале от -1 до $+1$, так как она эквивалентна коэффициенту корреляции.

ЛИНЕЙНАЯ РЕГРЕССИЯ

Координация между двумя переменными делает возможным преобразование информации, относящейся к одной переменной, в информацию о другой переменной. Например, обращаясь опять к 3.11б, можно увидеть, что значения X подразумевают соответствующие значения Z : положительные значения X обычно связаны с положительными значениями Z , отрицательные значения X — обычно предполагают отрицательные Z . Действительно, линия, проходящая через точки, подсказывает мысль, что частные конкретные значения X могут быть преобразованы во вполне определенные значения Z , ожидаемые при данном X , как это показано в 3.21.

3.21.



Конечно, ожидаемое значение Z , получаемое из значения X , редко будет точно тем же, что и наблюдаемое значение Z . Такое соответствие выполняется на самом деле только для тех событий, которые укладываются на прямую линию. Несмотря на это, предсказываемые значения дают лучшее соответствие значений Z , чем если бы каждый раз полагали в качестве предсказываемого значения Z значение, равное среднему значению Z по всему распределению.

Линии, переводящие значения одной переменной в значения другой, называются *регрессионными линиями*. Если X оказывается причинной детерминантой Z , слово «регрессия» может интерпретироваться следующим образом: эта линия показывает такое значение Z , к которому все события с заданным значением X возвращались бы, если бы возмущения в Z были устранены. Регрессионные линии могут быть прямолинейными (3.11б) или криволинейными (3.11в), однако в соответствии с нашим решением ограничить внимание только на линейных системах, здесь рассматриваются только линейные регрессионные линии. Предположение линейности есть ограничительное предположение; оно означает, что связи определенного типа (подобные 3.11в) либо вообще не могут рассматриваться, либо должны быть тем или иным способом аппроксимированы линейными отношениями.

Регрессионная линия идеально определяет точное математическое ожидание значения предсказываемой переменной при заданной информации относительно предсказываемой переменной, и такая точность не может быть достигнута лишь проведением на глаз линии через множество точек. Необходимы некоторые объективные процедуры, для того чтобы определить единственную наилучшую прямую линию.

Наилучшая линия — это та, которая подходит ближе всего к предсказываемому вертикальному положению каждой точки диаграммы рассеивания; желательно измерять «близость» на языке квадрата расстояния. Кроме того, предполагается, что более удаленные точки диаграммы рассеивания больше влияют на расположение линии, чем те, которые ближе к центру. Поэтому важность каждого случая взвешивается его квадратичным отклонением предсказываемой переменной. Наилучшая линия устанавливается таким образом, что она дает наименьшую взвешенную сумму квадратов расстояний между прямой и всеми точками диаграммы рассеивания. Нахождение такой линии представляет собой задачу, которая может быть решена математически с помощью вычислений. Однако здесь рассматриваются только результаты формального анализа.

Искомая линия всегда проходит через начало координат диаграммы рассеивания, если переменные измерены в отклонениях от среднего. Дополнительно математический анализ определяет *коэффициенты регрессии*, или наклон прямой, которые равны ожидаемому значению предсказываемой переменной, когда предсказываемая переменная равна $+1,0$. Так, мы знаем две точки, через которые проходит наилучшая линия регрессии, и поэтому регрессионная линия может быть вычерчена точно.

Если определять степень подгонки в терминах квадратов отклонений, взвешенных другими квадратичными отклонениями, то не следует удивляться тому, что математическое решение для регрессионных коэффициентов определяется на языке дисперсии и ковариации.

3.22. Регрессионный коэффициент определяется как наклон линии регрессии для предсказываемого Z к значениям X . Порядок индексов в коэффициенте регрессии важен: первый относится к предсказываемой переменной, второй — к предсказываемой

$$b_{zx} = \frac{\sigma_{zx}}{\sigma_x^2}.$$

Буква « b » принята для обозначения коэффициента регрессии. Вот почему мы исключили прописную букву b из списка букв для обозначения структурных коэффициентов: буква b была оставлена в качестве символа для коэффициентов регрессии.

Коэффициент регрессии может быть использован в алгебраической формуле для предсказания одной переменной по другой, поэтому становится ненужным вычерчивание графика.

3.23. Использование коэффициента регрессии в предсказывающей формуле. $\hat{Z}$ представляет математическое ожидание величины Z при заданном значении предсказываемой величины X , где X и Z измерены в отклонениях от среднего

$$\hat{Z} = b_{zx} \cdot X.$$

Для того чтобы решить, какая переменная должна быть предсказываемой, а какая — предсказываемой, имеются нестатистические основания. И действительно, всегда можно получить две регрессии, являющиеся двумя предсказывающими формулами. Вообще говоря, регрессионные коэффициенты в этих двух формулах не равны, т. е. $b_{zx} \neq b_{xz}$. В причинном анализе выбор из двух коэффициентов делается на базе теоретических соображений. Обычно предпочтение отдается такому коэффициенту, который обеспечивает предсказание выходной переменной по значениям входной переменной.

ОСТАТОЧНАЯ ДИСПЕРСИЯ

Положим, мы рассчитали среднеквадратическое различие между наблюдаемыми и предсказываемыми значениями, полученными по формуле регрессии. Итоговое различие есть дисперсия — средний квадрат расстояний между наблюдаемыми и ожидаемыми значениями — и может быть интерпретирована тремя различными способами. Во-первых, она представляет собой меру рассеивания точек около регрессионной линии, т. е. эта новая дисперсия равна нулю, если все точки ложатся точно на регрессионную линию, и она становится больше нуля, как только среднее расстояние между точками и регрессионной линией возрастает. Во-вторых, она является мерой остаточных вариаций, остающихся в предсказываемой переменной после того, как прогноз сделан. Этим она измеряет степень, с которой могут быть предсказаны вариации зависимой переменной. Для третьей интерпретации предположим, что ошибки предсказания будут причинно обусловлены прибавлением

значений переменной e к предсказанным значениям. Тогда новая дисперсия будет равна дисперсии этой переменной e , или σ_e^2 . Переменная может рассматриваться как составная переменная, представляющая все факторы, вызывающие ошибки прогноза.

3.24. Определение остаточной дисперсии или дисперсии ошибок.

$$\sigma_e^2 = \mathcal{E}((Z - \hat{Z})^2).$$

Так как линия регрессии была математически выведена для того, чтобы принять во внимание каждую порцию линейной координации между независимой и зависимой переменной, остатки или ошибки прогноза сами не коррелированы со значением (независимой) переменной. Так, например, в линейном случае все ошибки прогноза не могут быть положительными, когда переменная X положительна, так как в этой ситуации регрессионную линию можно было бы повернуть вверх и получить улучшение предсказания. Эта независимость между предсказывающими переменными и ошибками позволяет установить основной принцип: *общая дисперсия предсказываемой переменной равна дисперсии предсказываемых значений плюс дисперсия остатков.*

3.25. Дисперсия переменной может быть разделена на компоненту, предсказываемую по линейной регрессии, и дисперсию остатков, которые появятся, когда предсказанные значения будут вычтены из наблюдаемых значений

$$\sigma_Z^2 = \sigma_{\hat{Z}}^2 + \sigma_e^2.$$

Более удобная формула для расчетов остаточной дисперсии может быть получена с помощью введения *коэффициента детерминации*.

3.26. Коэффициент детерминации линейной регрессии равен дисперсии предсказываемых значений, деленной на дисперсию наблюдаемых значений. Эта дробь обозначается символом R^2 :

$$R_{Z \cdot X}^2 = \frac{\sigma_{\hat{Z}}^2}{\sigma_Z^2}.$$

Когда Z прогнозируется ровно по одной другой переменной, коэффициент детерминации тождественно равен квадрату коэффициента корреляции:

$$R_{Z \cdot X}^2 = \rho_{ZX}^2.$$

По заданным формулам из 3.25 и 3.26 величина остаточной дисперсии тоже может быть определена.

3.27. Приведем еще одну формулу определения остаточной дисперсии:

$$\sigma_e^2 = \sigma_Z^2 (1 - R_{Z \cdot X}^2).$$

Если для прогноза Z используется только одна переменная (X), эта формула эквивалентна следующей:

$$\sigma_e^2 = \sigma_Z^2 (1 - \rho_{ZX}^2).$$

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Прогнозы обычно могут быть улучшены путем использования нескольких независимых (предсказывающих) переменных вместо одной, т. е. переходом от бивариантного к многовариантному анализу. Пусть, например, прогнозируемая переменная Z коррелирует с переменной Y так же, как и с первой переменной X ; тогда, вероятно, Y может быть использована для прогноза некоторой части остаточных вариаций, остающихся после регрессии Z по X . Если это так, то было бы возможно усовершенствовать однофакторную прогнозную формулу, чтобы выразить математическое ожидание величины Z при заданных значениях X и Y . Дисперсия ошибок, остающихся после применения этой новой формулы, должна быть меньше, чем остаточная дисперсия, после прогноза по одному X .

Основной смысл *многофакторной регрессии* состоит в попытке предсказать остающееся непредсказанным одной регрессией добавлением другой и таким образом построить более точный прогноз на базе многих предсказывающих переменных. Эта весьма простая идея, однако, усложняется требованием, что коэффициенты регрессии должны быть совместно пригодными в единой прогнозной формуле. Все это не составило бы проблемы, если бы предсказывающие переменные всегда были не коррелированы друг с другом, потому что тогда прогноз по одной переменной не был бы коррелирован с прогнозами по другим переменным, и результаты различных прогнозов можно было бы сложить вместе. Но обычно прогнозирующие переменные *коррелированы*. Поэтому их прогнозы в определенной степени коррелированы, и избыточность в прогнозах порождает смещение, если они все будут сложены друг с другом.

Итак, в действительности проблема заключается в следующем: получить единую формулу для предсказаний значений одной переменной (Z) по значениям других переменных (X , Y и т. д.). Примем во внимание некоторую корреляцию между прогнозирующими переменными таким образом, что каждый коэффициент регрессии отражает только отдельные прогнозы, которые были бы сделаны по вариациям одной переменной, когда все другие прогнозирующие переменные имели постоянное значение.

В действительности математическое решение этой проблемы предполагает построение нескольких различных регрессий. Например, коэффициент прогнозирования Z по X при фиксированном Y получается сначала построением регрессии Z по Y , чтобы получить остатки этого прогноза. Затем, если имеется некоторая корреляция между прогнозирующими переменными, находится также регрессия X по Y , чтобы получить множество остаточных значений X , которые не коррелированы с Y . Наконец, определяется регрессия остатков Z на остатки X , чтобы получить искомый коэффициент регрессии с нужными свойствами. Точно такие же шаги повторяются до получения коэффициента регрессии, прогнозирующего Z по Y при фиксированном X .

Таким образом, эти коэффициенты в задаче множественной регрессии представляют собой наклоны регрессионных линий, предсказывающих значения одного множества остатков по значениям другого. Эти наклоны, или *частные коэффициенты регрессии*, также обозначаются буквой b , но теперь с отмеченными точкой индексами; например, $b_{ZX.Y}$ есть коэффициент для прогноза остатков Z по остаткам X , когда оба уже определены регрессией по Y ; $b_{ZY.X}$ — коэффициент для прогноза остатков Z по остаткам Y , когда оба регрессированы на X .

Частные коэффициенты регрессии не искажаются согласованием прогнозов, потому что каждый коэффициент зависит только от согласования между прогнозирующей и прогнозируемой переменными, после того как будут рассчитаны другие прогнозы и статистически устранена корреляция между прогнозируемыми переменными. Взяты вместе в единой формуле они не будут избыточными и дают единый объективный прогноз.

3.28. Приведем пример прогнозной формулы, основанной на множественной регрессии:

$$\hat{Z} = b_{ZX.Y}X + b_{ZY.X}Y, \text{ где } X, Y \text{ и } Z$$

суть величины отклонений.

К счастью для получения частных коэффициентов регрессии нет необходимости выводить множество различных регрессий. Когда имеются ровно две прогнозирующие переменные, формула 3.29, выведенная аналитически, дает тот же самый результат.

3.29. Вычислительная формула для частных коэффициентов регрессии:

$$b_{ZY.X} = \frac{\sigma_X^2 \sigma_{ZY} - \sigma_{ZX} \sigma_{YX}}{\sigma_Y^2 \sigma_X^2 - (\sigma_{YX})^2}.$$

Когда имеется более двух прогнозирующих переменных, вычислительные формулы становятся громоздкими, и, вообще говоря, будет весьма трудно вычислять коэффициенты ручным способом. В этом случае будут рентабельны электронно-вычислительные машины. Большинство библиотек стандартных программ содержат программы множественной регрессии, которые частные коэффициенты регрессии рассчитывают по дисперсиям и ковариациям или по первичным наблюдениям переменных.

В многофакторном случае, так же как и в двухфакторном, можно получить остатки вычитанием прогнозируемых отклонений Z из первичных отклонений Z . И опять сам характер математического решения гарантирует, что остатки будут некоррелированы с каждой из прогнозирующих переменных.

Можно легко получить дисперсию остатков многофакторного прогноза, используя определение многофакторного коэффициента детерминации. Это определение дается формулой 3.30 для случая двух факторов.

3.30. Коэффициент детерминации для случая двух прогнозирующих переменных:

$$R_{Z.XY}^2 = b_{ZY.X}^2 \left(\frac{\sigma_Y^2}{\sigma_Z^2} \right) + b_{ZX.Y}^2 \left(\frac{\sigma_X^2}{\sigma_Z^2} \right) + 2b_{ZY.X}b_{ZX.Y} \left(\frac{\sigma_{XY}}{\sigma_Z^2} \right).$$

При трех и более прогнозируемых переменных формула становится громоздкой. Вычислительные программы обычным образом рассчитывают *множественный коэффициент регрессии*, или R , а коэффициент детерминации есть просто квадрат коэффициента R : R^2 . Поэтому для определения остаточной дисперсии в задаче множественной регрессии можно взять квадрат множественного коэффициента корреляции из верхней формулы 3.27.

СТАНДАРТИЗОВАННЫЕ КОЭФФИЦИЕНТЫ

Регрессии могут быть получены, когда переменные измерены в стандартизированной форме, т. е. стандартизованы. Как правило, коэффициенты регрессии изменяются при изменении шкалы измерения — приведением ее к стандартизированной форме. Это отражается и в использовании других символов для *стандартизованных коэффициентов регрессии*, обозначаемых буквой β вместо b . Значения стандартизованных коэффициентов могут быть получены без реальной переделки шкал первичных измерений и подгонки регрессии. Это показано в формуле 3.31.

3.31. Стандартизованные коэффициенты регрессии могут быть рассчитаны по статистикам для нестандартизованных данных:

$$\beta_{ZX} = b_{ZX} \frac{\sigma_X}{\sigma_Z}.$$

В парном случае (одна прогнозирующая переменная) стандартизованный коэффициент регрессии равен коэффициенту корреляции (ρ); коэффициент, получаемый, когда первичные зависимая и независимая переменные меняются местами, также равен ρ .

3.32. Для случая одной независимой переменной

$$\beta_{XY} = \beta_{YX} = \rho_{XY}.$$

В многофакторном случае эти равенства, вообще говоря, не выполняются. Действительно, в этом случае коэффициенты могут даже быть вне области $-1, +1$ значений коэффициента корреляции.

Частные β -коэффициенты в многофакторном случае могут быть получены прямо по корреляциям между переменными. В 3.33 приводится формула, когда имеются ровно две независимые переменные.

3.33. Вычислительная формула для расчета стандартизованных частных коэффициентов регрессии:

$$\beta_{zx \cdot y} = \frac{\rho_{zx} - \rho_{zy} \cdot \rho_{xy}}{1 - \rho_{xy}^2}$$

Большинство вычислительных программ регрессионного анализа рассчитывают стандартизованные коэффициенты, так же как и нестандартизованные; задачи с числом прогнозирующих переменных более двух вручную решают весьма редко.

Формулы для коэффициентов детерминации, приводимые в 3.26 и 3.30, с помощью стандартизованных коэффициентов регрессии могут быть переписаны в новом виде.

3.34. Когда имеется ровно одна прогнозирующая переменная, $R_{y \cdot x}^2 = \beta_{yx}^2 = \beta_{xy}^2$. Для двух прогнозирующих переменных

$$R_{z \cdot yx}^2 = \beta_{zy \cdot x}^2 + \beta_{zx \cdot y}^2 + 2\beta_{zy \cdot x} \cdot \beta_{zx \cdot y} \cdot \rho_{xy}$$

Вторая формула легко может быть распространена на общий многофакторный случай; когда прогнозирующих переменных больше, чем две, обычно применяется вычислительная техника.

Формула в 3.27 может еще использоваться для определения дисперсии остатков после вычитания прогнозируемых значений из первичных значений; но так как теперь измерительные шкалы стандартизованы, первичная дисперсия прогнозируемой переменной должна быть равна 1, и поэтому дисперсия остатков равна просто 1 минус коэффициент детерминации.

РЕГРЕССИЯ И ПРИЧИННЫЙ ВЫВОД

Уравнение регрессии есть лишь средство для перевода информации одного вида в другой. Причинные процессы могут обуславливать возможность такого перевода; но это не означает, что уравнение регрессии выявляет эти процессы для любого простого случая. Например, система $Y \xleftarrow{a} X \xrightarrow{c} Z$ ведет к координации между значениями X и Y , позволяющей нам предсказать значения Y по информации об X . Уравнение регрессии $Y = b_{yx}X$ имеет некоторое сходство со структурным уравнением, устанавливающим причинную связь между X и Y , и в этом простом случае коэффициент b_{yx} является оценкой коэффициента a . Однако можно использовать то же самое эмпирическое соответствие для предсказания значений X по значениям Y , но коэффициент b_{xy} не будет непосредственной оценкой соответствующего коэффициента a . Далее система скоординирует Y и Z (в силу их обоюдной зависимости от X), обеспечив предсказание одного по другому, несмотря даже на то, что эти две переменные не имеют прямой причинной связи вообще. В этом случае коэффициенты регрессии не будут соответствовать никакому структурному коэффициенту.

Более того, координация между переменными может порождаться избирательными механизмами выбора, которые не относятся к причинным связям между переменными. Предположим, например, что в некоторой совокупности события осуществляются только в том случае, если сумма двух характеристик X и Y выше некоторого критического значения. Тогда внутри этой совокупности X и Y будут в некоторой степени скоординированы (низкие значения одной переменной должны быть связаны с высокими значениями другой), и, вероятно, уравнение регрессии могло бы использоваться для перевода информации об одной переменной в информацию о другой. Это уравнение регрессии, однако, вообще не имеет значения для суждений о причинной связи между X и Y , потому что связь не была генерирована причинными отношениями между этими переменными, а скорее была навязана «пропускными» механизмами.

Позже мы увидим, что в причинном анализе регрессии полезны, но на уравнения регрессии нельзя смотреть шаблонно, как на структурные уравнения, непосредственно представляющие причинные процессы.

ИСТОЧНИКИ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Имеется обширная литература по статистике, слишком обширная, чтобы быть перечисленной. Приведенные здесь наименования упоминаются потому, что они ориентированы исключительно на применение статистических методов в причинном анализе и построение моделей. Элементарные сведения даются Хубертом М. Блэлоком (Hubert M. Blalock, Jr., *Social Statistics*, New York, McGraw-Hill, 1960); по технике регрессионного анализа см. F. Kerlinger and E. Pedhazur, *Multiple Regression in Behavioral Research* (New York, Holt, Rinehart & Winston, 1973); A. Гольдберг (Arthur S. Goldberger, *Econometric Theory*, New York, Wiley, 1964) дает математический аппарат регрессионного анализа при высоком уровне ясности изложения и подходит во всех отношениях. Д. Ван де Гир (John P. Van de Geer, *Introduction to Multivariate Analysis for the Social Sciences*, San Francisco, Freeman, 1971) описывает множество статистических процедур и использование потоковых графов, чтобы показать, как разная аналитическая техника — регрессионный, канонический и факторный анализ — соотносится с различными причинными моделями.

Весь современный аппарат метода множественной регрессии и других многомерных методов зависит в некоторой степени от использования матричной алгебры. Забавное введение в эту область содержится в работе Филиппа Дж. Дэвиса (Philip J. Davis, *The Mathematics of Matrices; A First Book of Matrix Theory and Linear Algebra*, N. Y., Blaisdell, 1965). Книги Ван де Гира и Гольдбергера содержат дополнительный материал.

1. Президент США может рассматриваться как некоторый оператор, производящий определенные отношения связи. Возможно ли изучать правление президента статистически в рамках некоторой статичной структуры?

2. Распределение семей по доходу имеет явно выраженный асимметричный характер — большая часть семей имеет относительно меньшие доходы, а то время как меньшая обладает миллионами. Положим, такое асимметричное распределение доходов возникает в специально подобранной совокупности, в которой размер семейных жилищ распределен приблизительно по нормальному закону. Положим также, что в этой совокупности богатые имеют большее по размерам жилище.

(а) Постройте диаграмму совместного распределения доходов и размеров жилья и приближенно обрисуйте профиль, который должно было бы иметь это распределение.

(б) Могло ли такое распределение быть порождено линейной причинной системой, содержащей влияние дохода на размеры жилища? Может ли быть это распределение доказательством того, что имеет место эволюционный процесс как противоположность причинного процесса (см. гл. 1)?

(в) Пусть изучалась другая совокупность, для которой распределение размера жилья было также скошено, так что большинство семей имело относительно меньший размер жилищ. Могла ли быть в этом случае линейной связь между доходом и размером жилья?

3. Обычно автобусы имеют одного водителя. Предположим, добавлен второй водитель с равными возможностями управления, но отделенный стенкой от первого. При этом их управляющие действия передаются машине в суммированном виде. Был бы такой автобус с двумя водителями, задающими изменения скорости и направления, более безопасен?

4. Что можно сказать в следующих случаях:

(а) Некоторый исследователь установил, что корреляционный момент между коэффициентом интеллектуальности (КИ) и средним баллом по успеваемости (СБУ) равен +1,08.

(б) Обычный (нестандартизованный) коэффициент регрессии для прогноза КИ по СБУ равен +10,00.

5. Приведите следующее выражение к более простому виду, используя алгебру математического ожидания и определение среднего:

$$\mathcal{E}((X - \bar{X})^2).$$

(Эти X указывают первичные измерения, а не отклонения от среднего. Заметим, однако, что величина в круглых скобках являет собой отклонение от среднего, и в действительности все выражение есть определение дисперсии X .)

6. Упростите следующее выражение, как в задаче 5:

$$\mathcal{E}((X - \bar{X})(Y - \bar{Y})).$$

(Заметьте, что это выражение определяет ковариацию между X и Y .)

7. Упростите следующее выражение:

$$\mathcal{E}((\bar{X} + e)^2).$$

Здесь, однако, предполагаем, что и $\bar{X}$, и e измерены как отклонения от среднего. По вашему результату установите точно, что произойдет, если дисперсия переменной в точности равна дисперсии прогнозируемых значений плюс дисперсия ошибок прогнозов.

8. Предположим, что $\beta_{SF.E} = 0,015$, где S означает профессиональный статус сына, E — образование сына, а F — профессиональный статус отца. Проинтерпретируйте «полученные данные» с чисто статистической точки зрения.

9. Для «Страннограда» в США корреляции между переменными задачи 8 равны:

$$\rho_{FE} = -0,60; \quad \rho_{ES} = 0,30; \quad \rho_{FS} = 0,50.$$

Вычислите стандартизованный частный коэффициент регрессии $\beta_{SF.E}$, используя формулу 3.33. Объясните его смысл.

Причинный анализ в социальных науках проводится в основном в рамках «профильной статистики» (cross-sectional statics), т. е. наблюдения в разные моменты времени относят к одной и той же случайной переменной. В этом случае предполагается, что для многочисленных отдельных случаев существует единая основа причинной структуры и что каждый отдельный случай наблюдается после фиксирования его входов и поддержания на постоянном уровне достаточно долго, чтобы все причинные последствия были реализованы. Так, например, социально-экономические успехи индивидуумов могут изучаться в этом разрезе при допущении, что ключевые входы зафиксированы в некоторый ранний момент жизни индивидуума и что наблюдения индивидуумов отложены до тех пор, пока они не реализуют все последствия этих входов.

Такой подход — статический, поскольку наблюдения делаются после того, как реализованы причинные последствия определяющих переменных, когда и входы, и выходы будут поддерживаться на установившемся уровне. Разумеется, это то же самое условие, которое применялось в гл. 2. При этом внимание фокусируется не на единственном примере причинной системы (как в гл. 2), а на множестве случаев эквивалентных причинных структур, наблюдаемых более или менее одновременно.

В рамках такой схемы знание структуры причинной системы может быть использовано для преобразования статистического описания входов в статистическое описание выходов. Данная техника составляет содержание этой главы. С другой стороны, при заданном статистическом описании входов и выходов можно сделать некоторые выводы о причинной структуре, преобразовывающей одно в другое. Это содержание гл. 5.

Всякий анализ на базе выборочной статистики основывается на предпосылке, что причинные системы, действующие в различных случаях, эквивалентны: они имеют одни и те же организационные и структурные параметры, и причинные операторы для каждого функционируют нормально. В данной и следующей главах предполагается, что эти требования удовлетворены благодаря тщательному определению множества исходов событий, которые должны быть рассмотрены, за исключением тех, у которых совсем другие структуры или операторы не действуют.

МОДИФИКАЦИИ ПОТОКОВЫХ ГРАФОВ

Выражения для дисперсий, ковариаций и корреляций, порожденные системой, могут быть получены прямо из причинной диаграммы, соответствующим образом модифицированной. Основные положения, называемые *путевым анализом*, были развиты в 1910 и 1920 гг. биологом Сьюэллом Райтом (Sewall Wright). В следующих разделах описывается, как надо изменить диаграмму, чтобы она соответствовала путевому анализу; затем устанавливаются правила для выражений ковариаций, дисперсий и корреляций в виде произведений и сумм характеристик диаграммы. Условия и правила путевого анализа, излагаемые в этом разделе, несколько отличаются от традиционного подхода, развитого Райтом и др. Осуществлена некоторая корректировка с целью упрощения процедур и чтобы показать, что эти процедуры применимы к изучению ковариаций, как и корреляций, а также распространить традиционные процедуры путевого анализа на изучение систем с петлями.

Редукция петель

В том виде, как он возник, путевой анализ применялся только к системам без петель обратной связи любого типа. С некоторой точки зрения это ограничение все еще сохраняется, и теория потоковых графов расширяет путевой анализ только указанием того, как могут быть убраны петли из диаграммы, чтобы были применимы правила обычного путевого анализа. Эта точка зрения принимается здесь из-за своей методологической простоты.

IV. 1. Путевой анализ систем с петлями требует перерисовывания диаграммы системы для каждой петли или комплекса петель в частично редуцированную форму (см. правило II. 18).

Фактически можно было установить специальные правила путевого анализа для изучения петель в их первоначальной форме, но такие правила увеличили бы методическую сложность путевого анализа. Для первого знакомства, по-видимому, предпочтительнее сделать сложными диаграммы (редуцированием всех петель), но сохранить простой идейную сторону.

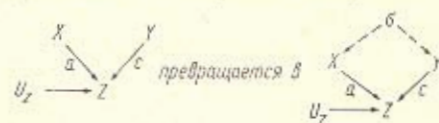
Ковариации входов

Статистические характеристики выходов системы зависят от структурных коэффициентов и статистических характеристик входов. Переходя от анализа графов отдельных случаев к путевому анализу множества случаев, надо произвести необходимые изменения в диаграммах с тем, чтобы учесть определенные статистические характеристики входов. В частности, необходимо предусмотреть явные изображения на диаграмме координат входов, прежде чем изучать систему.

IV. 2. Ковариации входов должны быть представлены на потоковом графе системы, прежде чем диаграмма может интерпрети-

роваться статистически. Ковариация графически изображается как источник для своих двух переменных и связывается с ними пунктирными стрелками, а не сплошными.

4.1. Уточнение представления на диаграмме ковариаций входных переменных.



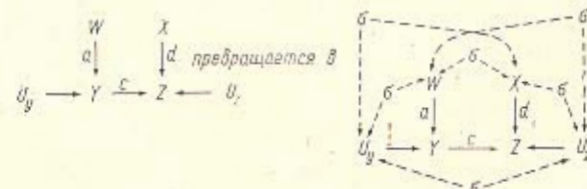
Пунктирные стрелки употребляются потому, что ковариационные члены не совместимы с обычным анализом графов; они выражают характеристики распределений и применяются только в выводах о статистиках распределения.

В традиционных биологических и социологических схемах путевого анализа корреляции между входными переменными изображались искривленными стрелками с двумя концами, отмеченными символом — коэффициентом корреляции. Принятое нами здесь условие обеспечивает более характерное обозначение для этого специального типа путей (что желательно, когда рассматриваются сложные системы с петлями). Обращение с ковариациями как с «псевдоисточниками» помогает упростить формулировки правил путевого анализа. Ради непрерывности, однако, старое условие остается в силе, когда проводится путевой анализ стандартизованных переменных, в соответствии с первоначальным смыслом.

Ковариации возмущений

Возмущающие члены представляют собой неучитываемые источники для каждой зависимой переменной системы; поэтому статистические характеристики возмущений должны быть рассмотрены при любом анализе системных статистических последствий. Ковариации возмущений друг с другом и с явными входами также должны быть добавлены к диаграмме, как это показано в 4.2.

4.2.



Ясно, что ковариации возмущений значительно усложняют диаграмму. Позже будет показано, что если ковариация двух входных переменных имеет нулевое значение, то ковариация не

оказывает никакого влияния на статистические характеристики. Поэтому эти члены могут быть удалены из диаграммы без потери, в результате чего упростится внешний вид. В следующем разделе вообще предполагается, что возмущающие члены не коррелируют ни с установленными входными переменными, ни друг с другом. Поэтому их ковариацию не надо добавлять на диаграмме. Правила путевого анализа не ограничиваются этим предположением. Его принимают только ради удобства, чтобы сохранить диаграммы простыми. С другой стороны, в гл. 5 показано, что некоррелированность возмущения обычно выполняется, когда изучаемая система действует в эмпирических данных.

АНАЛИЗ СТАТИСТИЧЕСКОЙ КООРДИНАЦИИ

Координирующие пути

Причинные отношения могут порождать координацию между переменными системы тремя различными способами.

1. Две переменные могут иметь согласованные значения, потому что зависят от одной и той же входной переменной системы. (Например, $Y \leftarrow X \rightarrow Z$, где Y и Z зависят от X .) Изменения в источнике передаются обоим зависимым переменным, заставляя их значения в некоторой степени согласовываться.

2. Две переменные также могут быть согласованы из-за предшествующей координации между влияющими на них переменными. Эта координация заставляет выходные переменные быть более выравненными, чем это было бы в другом случае. Например, если X и W — входные переменные, и система преобразует значения W , порождая значения Y , и значения X , порождая значения Z , то любая предыдущая координация между W и X имеет тенденцию передаваться к Y и Z .

3. Если одна переменная причинно определяет другую либо непосредственно, либо через промежуточные переменные, значения этих двух переменных координированы причинным преобразованием первичной переменной в значения зависимой переменной.

Координация, порожденная двумя первыми механизмами, приводит нас к ложным корреляциям в том смысле, что они не порождены прямыми причинными связями. Однако в статистическом анализе координация имеет большое значение независимо от источника ее происхождения. Поэтому вводится следующее правило, которое определяет типы путей на причинных диаграммах, порождающих координацию, и не делает специального разграничения между всеми тремя механизмами. Это правило объединяет все три механизма в понятие координирующего пути.

IV.3. «Координирующий путь» между двумя переменными состоит из последовательности стрелок при выполнении следующих трех условий:

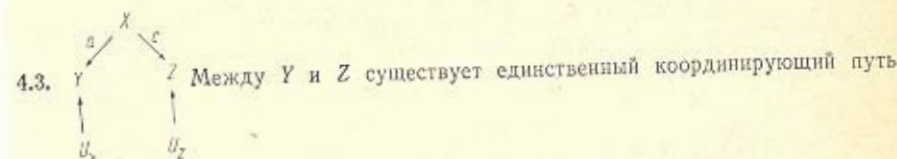
(а) Обе интересующие нас переменные являются конечными точками пути.

(б) Путь состоит из двух субцепей, каждая из которых отходит от переменной или ковариационного члена на графе и последовательно идет в одном направлении в сторону конечных точек. Переменная или ковариационный член, откуда отходят субцепи, называется «началом» пути. Субцепи отходят от начала пути к конечным точкам.

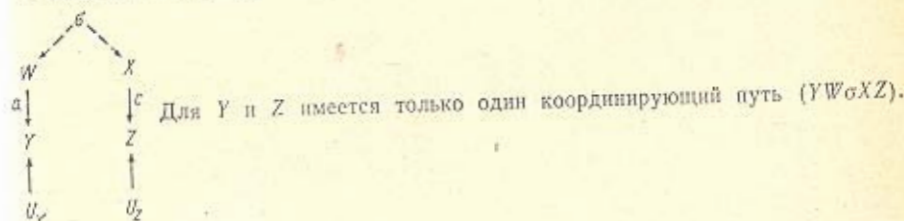
(в) Данный путь пересекает каждую переменную или ковариационный член только один раз.

(г) Начало пути может быть одной из конечных точек пути. В этом случае одна из двух субцепей не различается, и весь координирующий путь состоит из единственной субцепи.

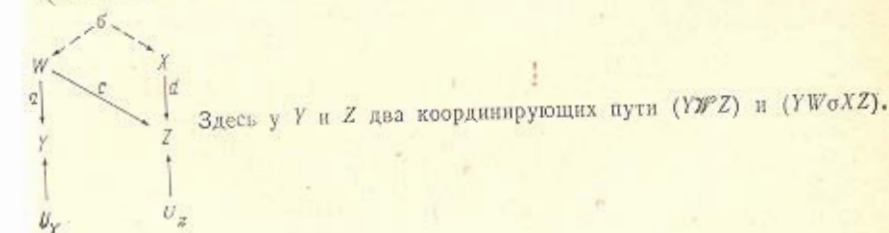
Несмотря на многословность этого правила, координирующие пути устанавливаются весьма легко. Действительно, мы начинаем с одной переменной, движемся вдоль стрелок в обратную сторону к началу пути, затем перемещаемся по стрелкам в прямом направлении, пока не достигнем другой переменной. Координирующий путь обозначается переменными, которые он затрагивает. Начало пути обозначается соответствующей курсивной буквой или ковариационным символом.



($Y\mathcal{X}Z$) или ($Z\mathcal{X}Y$). Оба обозначения эквивалентны, они относятся к одному и тому же координирующему пути.



Знак σ в идентификаторе координирующего пути всегда выражает ковариацию переменных с обеих сторон от него.



$W \xrightarrow{a} X \xleftarrow{b} U_X$ ($XWYZ$) — единственный координирующий путь между X и Z .
 $U_Y \rightarrow Y \xleftarrow{c} U_Y$ ($XYZY$) — единственный координирующий путь между X и Y .

Необходимо строго придерживаться процедуры следования от одной конечной точки пути вдоль стрелок в обратную сторону, затем прямо к другой конечной точке. Нельзя идти сначала прямо, затем в обратную сторону или сначала в обратную сторону, затем прямо, а потом снова обратно. Вообще говоря, как только мы начнем идти по стрелкам прямо, путь не может продолжиться изменением направления.

4.4.
 На этой диаграмме между X и Y не существует координирующего пути. В частности, (XZY) не есть координирующий путь, потому что стрелки не ведут от начала к конечным точкам.

Здесь (XZY) есть единственный координирующий путь между X и Y .

V и Z скоординированы только одним путем ($VWXYZ$).
 ($VWXYZ$) не является координирующим путем, потому что включает изменение направления более чем один раз.

Часть (в) приведенного выше правила, утверждающая, что путь не может затрагивать одну и ту же переменную дважды, еще более ограничивает возможные выделения координирующих путей в некоторых случаях вроде тех, что приводятся в примере 4.5.

4.5.

Y и Z имеют только один координирующий путь

(YZ). В частности, ($YXWYZ$) не может рассматриваться как координирующий путь, потому что проходит через X дважды.

Снова только (YZ) является координирующим

путем для Y и Z . ($YXWYZ$) проходит через X дважды.

Часть (г) правила IV.3 распространяет определение на важный специальный случай, когда начало пути совпадает с одной из конечных точек. Его применение иллюстрируется в 4.6.

4.6. $X \xrightarrow{a} Y \xleftarrow{b} U_Y$. Для X и Y координирующим путем будет (XY).

X и Y скоординированы путем (XY); Y и Z — путями (YZ) и (ZY); X и Z — путями (XYZ) и (ZXY).

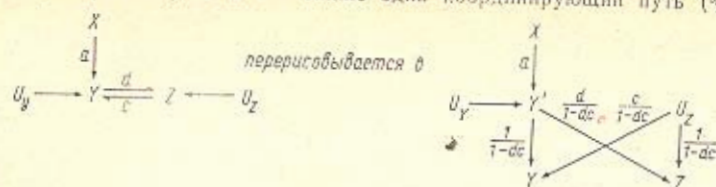
W и Y координируются путем (WY); W и Z — путями (WYZ) и (ZYW); Y и Z — путями (YZ) и (ZY); X и Y — путем (XY). Другие координирующие пути имеются между X и Z .

Идентификация координирующих путей на диаграммах с петлями не представляет собой особой задачи при условии, что петли на диаграммах соответственно переведены в частично редуцированную форму.

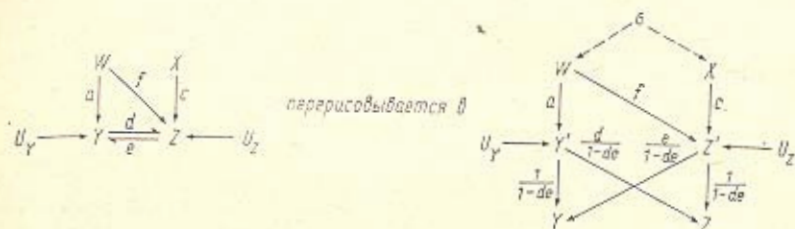
4.7.

перерисовывается в

X и Y имеют один координирующий путь ($\mathcal{E}Y'Y$); X и Z — один координирующий путь ($\mathcal{E}Y'YZ$); Y и Z — также один координирующий путь ($\mathcal{Q}YZ$).



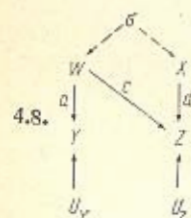
X и Y имеют один координирующий путь ($\mathcal{E}Y'Y$); X и Z — координирующий путь ($\mathcal{E}Y'Z$).



Между W и Z имеются три следующих координирующих пути: ($W\sigma XZ'Z$), ($W'Z'Z$), ($WY'Z$). Между Y и Z имеются шесть координирующих путей ($YY'W'Z'Z$), ($Y\mathcal{Q}YZ$), ($YZ'WY'Z$), ($YZ'Z$) и еще два других.

Ковариационный анализ

IV.4. «Эффект влияния координирующего пути» — это произведение структурных коэффициентов вдоль пути на дисперсию начальной переменной. Если началом пути является ковариационный член, эффект пути равен произведению структурных коэффициентов на этот ковариационный член. Эффект координирующего пути обозначается буквой C с индексами, являющимися идентификатором этого пути. Примеры приводятся в 4.8.

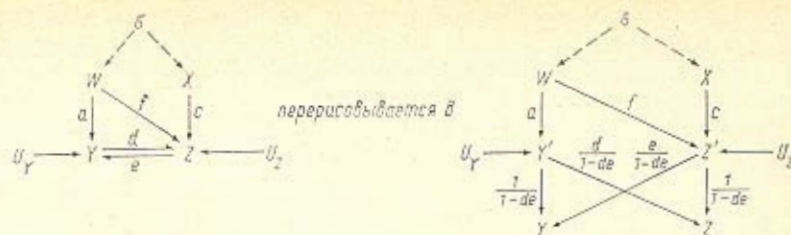


факты равны:

$$C_{YW\sigma XZ} = a \cdot 1 \cdot 1 \cdot d \cdot \sigma_{WX} = a \cdot d \cdot \sigma_{WY}$$

$$C_{Y'W'Z} = a \cdot c \cdot \sigma_{W'Y'}$$

Между Y и Z имеются два координирующих пути. Их эф-



Для координирующих путей между W и Z :

$$C_{W\sigma XZ'Z} = \frac{c}{1-de} \sigma_{WX},$$

$$C_{W'Z'Z} = \frac{f}{1-de} \sigma_{W'Y'},$$

$$C_{WY'Z} = \frac{ad}{1-de} \sigma_{W'Y'}.$$

Для шести координирующих путей между Y и Z :

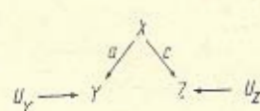
$$C_{YY'W'Z'Z} = \frac{a}{1-de} \cdot \frac{f}{1-de} \sigma_{W'}^2 = \frac{a \cdot f}{(1-de)^2} \sigma_{W'}^2, \quad C_{YZ'Z} = \frac{d}{(1-de)^2} \sigma_{Y'}^2,$$

$$C_{YZ'WY'Z} = \frac{e \cdot fad}{(1-de)^2} \sigma_{W'}^2, \quad C_{YZ'Z} = \frac{e}{(1-de)^2} \sigma_{Z'}^2,$$

$$C_{YY'W\sigma XZ'Z} = \frac{ac}{(1-de)^2} \sigma_{WX}, \quad C_{YZ'X\sigma WY'Z} = \frac{ecad}{(1-de)^2} \sigma_{WX}.$$

Координирующие пути, идущие в обратных направлениях, не различаются, если они затрагивают одни и те же точки. Эффект остается неизменным независимо от того, какая из конечных точек используется в качестве исходной.

4.9.



$$C_{Y\mathcal{E}Z} = a \cdot c \cdot \sigma_X^2,$$

$$C_{Z\mathcal{E}Y} = c \cdot a \cdot \sigma_X^2,$$

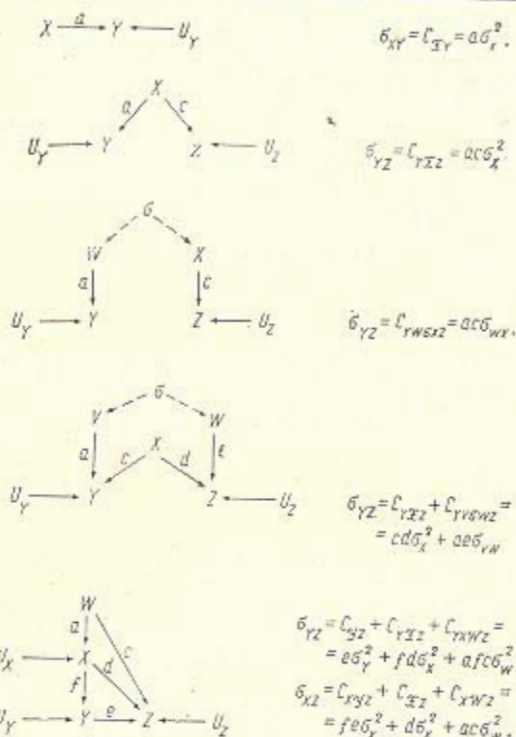
так что $C_{Y\mathcal{E}Z} = C_{Z\mathcal{E}Y}$.

Эффект координирующего пути определяет ковариацию между двумя переменными, вызванную данной цепочкой причинных звеньев. Например, для $X \xrightarrow{a} Y$ мы имеем: $C_{XY} = a \cdot \sigma_X^2$, правильное выражение для ковариации X с Y . (Здесь не доказывается, но 4.17 иллюстрирует, как любой результат путевого анализа может быть подтвержден алгебраически.) В более сложных задачах с большим числом координирующих путей между переменными ковариации определяются правилом IV.5.

IV.5. Ковариация между двумя переменными равна сумме эффектов всех отдельных координирующих путей между переменными. Другими словами, каждый координирующий путь порождает

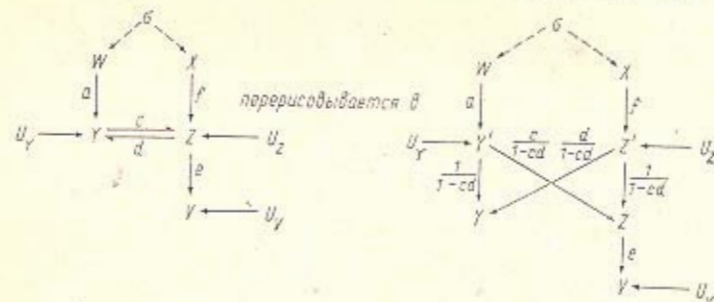
определенное количество ковариации между двумя переменными — количество, указываемое эффектом координирующего пути. Общая ковариация есть сумма ковариаций, производимых всеми различными путями. Примеры в 4.10 иллюстрируют самые различные способы, которыми может аккумулироваться ковариация между двумя переменными.

4.10.



Правило координирующего пути применяется в системах с петлями, как только петли будут редуцированы, что показано в 4.11.

4.11.



Ковариации между переменными могут быть считаны из скорректированной диаграммы, как в следующих примерах:

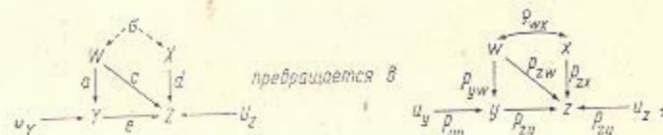
$$\begin{aligned}\sigma_{WZ} &= C_{WY'Z} + C_{WZXZ'Z} = \frac{ac}{1-cd} \sigma_W^2 + \frac{f}{1-cd} \sigma_{WX}, \\ \sigma_{WV} &= C_{WY'ZV} + C_{WZXZ'ZV} = \frac{ace}{1-cd} \sigma_W^2 + \frac{fe}{1-cd} \sigma_{WX}, \\ \sigma_{YZ} &= C_{Y'Y'Z} + C_{Y'Z'Z} + C_{Y'WZXZ'Z} + C_{YZ'X'WY'Z} = \\ &= \frac{c}{(1-cd)^2} \sigma_Y^2 + \frac{d}{(1-cd)^2} \sigma_Z^2 + \frac{af}{(1-cd)^2} \sigma_{WX} + \frac{afcd}{(1-cd)^2} \sigma_{WX}, \\ \sigma_{ZV} &= C_{Z'V} = e\sigma_Z^2.\end{aligned}$$

Корреляционный анализ

Так как корреляция — это просто ковариация стандартизованных переменных, то она также может быть считана прямо с путевой диаграммы с помощью вышеприведенных правил. Однако структурные коэффициенты и входные ковариации должны быть изменены, поскольку значения этих показателей зависят от шкал измерения. Все показатели на диаграмме должны быть переведены в стандартизованную форму.

Стандартизованные значения структурных коэффициентов называются *путевыми коэффициентами* и обозначаются прописной буквой p . Стандартизованные значения входных ковариаций — это просто корреляции между входами. Таким образом, чтобы считать корреляции непосредственно, символы диаграммы должны даваться в этой форме. Кроме того, в стандартизованном случае корреляционные связи представляются сплошной стрелкой с двумя концами в соответствии с традицией путевого анализа.

4.12.



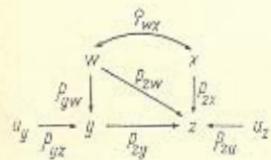
где все переменные стандартизованы.

На измененных диаграммах возмущающие члены также должны быть представлены как стандартизованные переменные. Это означает, что возмущения переменных переведены в отдельные гипотетические переменные с единичной дисперсией точно так, как и все другие переменные системы. Поскольку зависимая переменная частично определяется другими переменными системы, вклад возмущающего члена опять должен быть понижен, прежде чем его можно будет суммировать с другими источниками. Это достигается с помощью новых коэффициентов (p_{YU} и p_{ZU}). Можно было бы оставлять возмущения нестандартизованными, как это

мы делали до сих пор. Однако коэффициенты типа p_{yu} и p_{zu} дают точное указание степени подверженности системы влиянию со стороны неучтенных факторов и они показываются на путевой диаграмме для стандартизованных переменных традиционно.

При стандартизованной форме переменных и параметров системы выражения для корреляций могут быть прочитаны прямо с диаграммы при помощи правил координирующих путей. Существенное упрощение выражений появляется как результат стандартизации. Когда диаграмма системы определена на языке путевых коэффициентов и используется для определения корреляций, нет необходимости показывать в определяющих выражениях дисперсии источников, поскольку все эти дисперсии равны единице.

4.13.



Корреляция между w и z равна их ковариации, так как все переменные стандартизованы. Между w и z имеются три координирующих пути:

$$C_{wz} = p_{zw} \sigma_w^2 = p_{zw} \cdot 1 = p_{zw}, \quad C_{wyz} = p_{yw} \cdot p_{zy}, \quad C_{wpxz} = 1 \cdot 1 \cdot p_{zx} p_{wx} = p_{zx} p_{wx}$$

$$\text{Поэтому } p_{wz} = \sigma_{wz} = p_{zw} + p_{yw} p_{zy} + p_{zx} p_{wx}.$$

Стандартизация и нестандартизация

Путевые коэффициенты упрощают выражения для декомпозиции корреляций. Кроме того, они дают некую основу для сравнения силы различных операторов в системе.

Всякий структурный коэффициент диаграммы указывает число единиц изменения, ожидаемое в зависимой переменной при единичном изменении входной переменной и фиксированных значениях всех остальных переменных. Этим количественное значение структурного коэффициента связывается с единицами измерения обеих переменных. Если шкала измерения какой-либо одной переменной изменилась, то значение коэффициента также должно быть изменено, а если шкалы измерения произвольны (скажем, в разных исследованиях пользуются разными шкалами), то значения коэффициентов также будут произвольными: эти значения нельзя осмысленно сравнить между собой. Стандартизация переменных и коэффициентов на основе дисперсии совокупности может обойти некоторые из этих трудностей путем уменьшения произвола в единицах измерения. Каждая переменная измеряется в шкале, единица измерения которой статистически сравнима с единицами других шкал в том смысле, что дисперсия всех переменных одна и та же (1,0). Единицы измерения различных переменных «сбалансированы», и относительные силы различных коэффициентов одной и той же системы могут сравниваться, потому что каждый из них указывает, как изменения на одной стандар-

тизованной шкале обращаются в изменения на другой. Более того, переход к стандартизованным шкалам позволяет осуществлять сравнение различных исследований одной и той же совокупности, так как процедура стандартизации превращает измерительные приборы, проградуированные по-разному, в приборы со стандартными единицами измерения, которые зависят от распределения совокупности, а не от используемых специальных инструментов измерения.

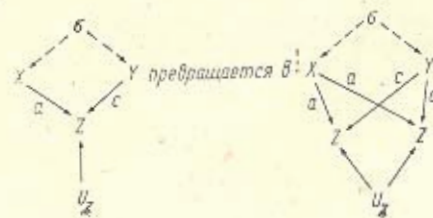
Все же стандартизация подвержена своим собственным недостаткам, проистекающим от тех же самых факторов, которые делают ее полезной, — зависимости от распределения совокупности. Путевые коэффициенты не дают гарантированной основы для структурного сравнения систем, относящихся к совокупностям с различными распределениями. Когда совокупности имеют различные распределения переменных системы, стандартизация внутри каждой совокупности не приводит к эквивалентным единицам сквозного измерения для всех совокупностей; поэтому и структурные коэффициенты, зависящие от этих единиц измерения, оказываются несравнимыми. Эти же ограничения касаются и единой совокупности, изучаемой в течение некоторого времени, если распределения совокупности меняются. В этом случае шкалы, стандартизованные на базе дисперсий разных моментов времени, будут несравнимы, как и структурные коэффициенты, основанные на этих единицах. При сравнительном изучении различных совокупностей или при длительном изучении изменчивых совокупностей необходимо избегать стандартизации.

АНАЛИЗ СТАТИСТИЧЕСКОГО РАЗНООБРАЗИЯ

Дисперсия представляет собой специальный случай ковариации переменной с самой собой. Такой взгляд позволяет прочесть из диаграммы системы выражения для дисперсий почти по тем же правилам, которые уже были введены.

Ковариация переменной с самой собой может быть получена представлением переменной на одной и той же диаграмме дважды и точным дублированием ее связей со всеми другими переменными, как это показано в 4.14.

4.14. Удваивание переменной Z



Ковариация σ_{ZZ} переменных Z и Z из второй диаграммы может быть найдена следующим образом:

$$\begin{aligned} C_{ZZ} &= a \cdot a \cdot \sigma_X^2 = a^2 \sigma_X^2, & C_{ZY\sigma XZ} &= C \cdot 1 \cdot 1 \cdot a \cdot \sigma_{XY} = a \sigma_{XY}, \\ C_{ZY} &= c \cdot c \cdot \sigma_Y^2 = c^2 \sigma_Y^2, & C_{ZX\sigma YZ} &= a \cdot 1 \cdot 1 \cdot c \cdot \sigma_{XY} = a \sigma_{XY}, \\ C_{ZZ} &= 1 \cdot 1 \cdot \sigma_{U(Z)}^2 = \sigma_{U(Z)}^2. \end{aligned}$$

Поэтому $\sigma_Z^2 = \sigma_{ZZ} = a^2 \sigma_X^2 + c^2 \sigma_Y^2 + 2a \sigma_{XY} + \sigma_{U(Z)}^2$.

Перечисление диаграммы с целью определения каждой дисперсии становится весьма обременительным, но при необходимости может быть исключено, если слегка видоизменить правила координирующих путей для определения дисперсии.

IV.6. Дисперсия переменной X равна сумме всех «эффектов рассеивающих путей». Эффект рассеивающего пути — это то же самое, что эффект координирующего пути, но определенный специальными условиями.

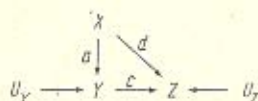
(а) Рассеивающий путь имеет X в качестве обеих конечных точек.

(б) Путь может проходить через непунктирную стрелку дважды, хотя и не должен затрагивать переменных (кроме X) более одного раза. Этот путь может проходить через пунктирную стрелку только один раз.

(в) Рассеивающие пути различны, если их индексы различны или если порядок индексов различен.

4.15. $X \xrightarrow{a} Y \xrightarrow{c} Z$.

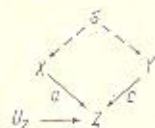
Здесь имеется один рассеивающий путь с переменной Z в качестве обеих конечных точек — $(ZY)Z$, через стрелку с путь проходит дважды. Дисперсия Z равна $\sigma_Z^2 = c \cdot c \cdot \sigma_Y^2 = c^2 \sigma_Y^2$. Путь $(ZY)YZ$ не может быть рассеивающим путем потому, что он проходит через Y дважды.



Здесь пять рассеивающих путей с Z в качестве обеих конечных точек:

$$\begin{aligned} C_{ZZ} &= d^2 \sigma_X^2, & C_{ZY} &= c^2 \sigma_Y^2, & C_{ZZ} &= \sigma_{U(Z)}^2, \\ C_{ZZYZ} &= a \cdot d \cdot c \cdot \sigma_X^2, & C_{ZYZZ} &= a \cdot d \cdot c \cdot \sigma_X^2. \end{aligned}$$

При определении дисперсии два последних эффекта понимаются как различные, потому что порядок их индексов различен. Дисперсия Z равна сумме всех пяти эффектов: $\sigma_Z^2 = a^2 \sigma_X^2 + c^2 \sigma_Y^2 + 2ad \sigma_{WX} + \sigma_{U(Z)}^2$.



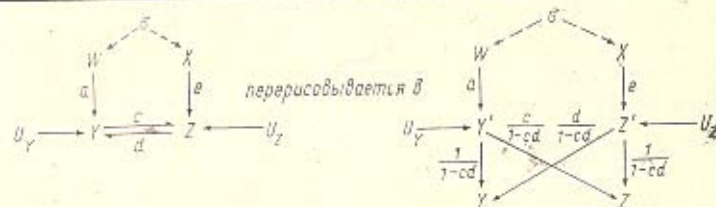
На этой диаграмме $(ZX)YZ$ и $(ZY)XZ$ — различные рассеивающие пути для определения σ_Z^2 , поскольку порядок переменных различен:

$$\sigma_Z^2 = a^2 \sigma_X^2 + c^2 \sigma_Y^2 + 2a \sigma_{XY} + \sigma_{U(Z)}^2.$$

$(X\sigma X)$ и $(Y\sigma Y)$ не являются рассеивающими путями для X и Y , поскольку каждый из них проходит пунктирную стрелку дважды. Таким образом, дисперсии входов X и Y суть базисные данные, которые не могут быть анализированы далее.

Правило рассеивающих путей определяет дисперсии также в системах с петлями, если петли представлены в частично редуцированной форме, как это показано в 4.16.

4.16.



Эффекты рассеивающих путей для дисперсии Y равны:

$$\begin{aligned} C_{YY} &= \frac{1}{(1 - cd)^2} \sigma_Y^2, & C_{YZ} &= \frac{d^2}{(1 - cd)^2} \sigma_Z^2, \\ C_{YY'W\sigma XZ} &= \frac{aed}{(1 - cd)^2} \sigma_{WX}, & C_{YZ'X\sigma WY} &= \frac{aed}{(1 - cd)^2} \sigma_{WX}, \\ \sigma_Y^2 &= \frac{1}{(1 - cd)^2} (\sigma_Y^2 + d^2 \sigma_Z^2 + 2aed \sigma_{WX}). \end{aligned}$$

Обычные процедуры ведут к формулам

$$\sigma_Y^2 = \sigma_{U(Y)}^2 + a^2 \sigma_W^2 \quad \text{и} \quad \sigma_Z^2 = \sigma_{U(Z)}^2 + e^2 \sigma_X^2.$$

Подставляя два последних выражения в формулу для σ_Y^2 , получим:

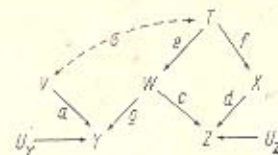
$$\sigma_Y^2 = \frac{1}{(1 - cd)^2} (a^2 \sigma_W^2 + e^2 d^2 \sigma_X^2 + 2aed \sigma_{WX} + \sigma_{U(Y)}^2 + d^2 \sigma_{U(Z)}^2).$$

ПУТЕВОЙ АНАЛИЗ И АЛГЕБРАИЧЕСКИЕ ВЫКЛАДКИ

Правила, предложенные для определения ковариаций, дисперсий и корреляций, определяют те же самые выражения, которые выводятся более традиционной алгеброй, как показывается в 4.17.

4.17. Математика, соответствующая путевому анализу

Небольшое упражнение по алгебре математических ожиданий освещает математические принципы, предполагаемые предыдущими правилами.



Наша цель — получить ковариацию между Y и Z , применяя алгебру, а не анализ графов. Во-первых, мы вспомним, что $\sigma_{YZ} = \mathcal{E}(Y \cdot Z)$. Эта формула будет полезной, если можно определить выражения для Y и Z , которые необходимо подставить в выражение для математического ожидания. Конечно, такие выражения могут быть выписаны из графа $Y = aV + gW + U_Y$; $Z = cW + dX + U_Z$. Теперь перемножение этих выражений для Y и Z эквивалентно перемножению самих переменных, т. е.

$$YZ = acVW + gcW^2 + cWU_Y + adVX + gdWX + dXU_Y + aVU_Z + gWU_Z + U_YU_Z.$$

Математическое ожидание этого выражения определяет ковариацию между Y и Z в соответствии с вышеприведенной формулой:

$$\sigma_{YZ} = \mathcal{E}(YZ) = \mathcal{E}(acVW + gcW^2 + cWU_Y + adVX + gdWX + dXU_Y + aVU_Z + gWU_Z + U_YU_Z),$$

что сводится к формуле

$$\sigma_{YZ} = ac\mathcal{E}(VW) + gc\mathcal{E}(W^2) + c\mathcal{E}(WU_Y) + ad\mathcal{E}(VX) + gd\mathcal{E}(WX) + d\mathcal{E}(XU_Y) + a\mathcal{E}(VU_Z) + g\mathcal{E}(WU_Z) + \mathcal{E}(U_YU_Z).$$

Все математические ожидания в этой формуле сами суть дисперсии или ковариации, поэтому выражение может быть переписано следующим образом:

$$\sigma_{YZ} = ac\sigma_{VW} + gc\sigma_W^2 + c\sigma_{WU(Y)} + ad\sigma_{VX} + gd\sigma_{WX} + d\sigma_{XU(Y)} + a\sigma_{VU(Z)} + g\sigma_{WU(Z)} + \sigma_{U(Y)U(Z)}.$$

Многие члены последнего выражения включают ковариации возмущений с определенным источником или друг с другом. Однако первоначальная диаграмма указывает, что все такие ковариации равны нулю, и члены, их содержащие, могут быть опущены:

$$\sigma_{YZ} = ac\sigma_{VW} + gc\sigma_W^2 + ad\sigma_{VX} + gd\sigma_{WX}.$$

Это не совсем то выражение, которое мы получили бы с помощью путевого анализа, потому что оно содержит члены с σ_{WX} и σ_{VW} , которых не было в выражении путевого анализа. Если применять алгебру математических ожиданий дальше, то можно найти, что

$$\sigma_{WX} = e\sigma_T^2 \quad \text{и} \quad \sigma_{VW} = e\sigma_{VT},$$

и если эти члены подставить в предыдущую формулу, то получится бы

$$\sigma_{YZ} = ace\sigma_{VT} + gc\sigma_W^2 + ade\sigma_{VX} + gde\sigma_T^2,$$

что точно совпадает с формулой путевого анализа. Все дисперсии и ковариации зависимых переменных системы могут быть определены аналогичным образом.

Ничего сколь-нибудь существенного не теряется при использовании графических процедур в сравнении с алгебраическими процедурами, потому что графы сами суть математическое представление систем.

ПРОПУСКНЫЕ МЕХАНИЗМЫ

В предыдущих обсуждениях входные дисперсии и ковариации принимались как данные и интерпретировались как статистические параметры, определяющие дисперсии и ковариации других

переменных системы. Однако можно сделать еще отступление и выяснить, что же определяет значения тех параметров. И тогда по необходимости внимание переместится с интересующей нас системы на силы, действующие во внешней среде системы.

Определение источника разнообразия и координации во множестве входных переменных требует спецификации предшествующей или окружающей системы, для которой эти переменные являются выходными. Фактически первоначальный выбор одного, а не другого множества входов является только делом соглашения. Например, черты характера личности могут быть входом для профессионального поведения, но черты личности будут выходными переменными для системы социализации, и, по-видимому, определение системы может быть расширено прибавлением переменных социализации. Если бы рассмотрение было расширено до источников формирования личности, мы должны были бы включить несколько дюжин переменных социализации, а также генетических факторов. Ради простоты мы оставляем в стороне характеристики личности, но расширение до переменных социализации может быть сделано.

Выход, однако, не всегда оказывается таким простым, потому что некоторые социальные системы не учитывают всех тех явлений, которые охватываются надсистемой. Могут вмешиваться различные пропускные механизмы от личного выбора до специфических процедур пополнения, отбора и отбраковки на границах системы. Пропускные механизмы действуют, чтобы понижать разнообразие явлений, допускаемых системой, за счет того, что такие «ворота» пропускают только те экземпляры, у которых значения выбранных переменных урезаны снизу или сверху. (Пропускные устройства могут также быть предназначены для увеличения входного разнообразия, изменчивости, например, пропуска лишь экземпляры с крайними значениями.) Эти «проходы», ограничивающие область значений переменных, влияют на ковариацию, как и на дисперсию, создавая тенденцию к уменьшению корреляции между переменными системы. Эта важная идея иллюстрируется здесь в примере 4.18.

4.18. Уменьшенная корреляция, вызванная ограничением области



X и Y имеют умеренную степень корреляции, когда рассматриваются их полные области, и корреляция почти нет, когда область значений обеих переменных лежит между 0 и +2,0.

Пропускные механизмы, действуя по большому числу различных входов, одновременно влияют на их статистические характеристики. В результате статистические характеристики выходов также оказываются подверженными их влиянию. Например, конъюнктивный пропускной механизм (не пропускающий, пока не встретится экземпляр с урезанными значениями нескольких переменных) может уменьшать как разнообразие, так и корреляцию входных переменных системы. Поэтому конъюнктивные пропускные механизмы также могут уменьшать дисперсию и корреляцию выходных переменных. Дизъюнктивный пропускной механизм (пропускающий, если хотя бы один критерий удовлетворяется) увеличивает дисперсию входов, в то же время обычно уменьшая ковариацию; и чистое влияние на выходные переменные зависит от самой системы.

Пропускной механизм делит совокупность на два множества: одно с экземплярами, обработанными специальной подсистемой операций, другое с затронутыми элементами. Иначе можно сказать, что основные операции применяются ко всем членам совокупности, но пропускные механизмы «выключают» операторы системы для отделяемой подсовокупности. Это в свою очередь близко понятию ограничителя, если наблюдаемые значения меньше (или больше), чем определенное значение переменной, соответствующий структурный коэффициент будет иметь нулевое значение. Однако пропускной механизм сводит на нет все последующие операции в подсистеме, а не одну, следующую за переменной, по которой действует механизм отсека.

Как «проходы», так и ограничители имеют неприятные последствия для путевого анализа. Мы не можем применять модель подсистемы ко всей совокупности, потому что пропускной механизм (или ограничитель), действуя на входы, приводит к добавлению влияний подсистемы лишь к выходам выделенных экземпляров. Еще хуже то, что влияния, когда они возникают, оказываются смещенными, поскольку входы для рассматриваемых экземпляров имеют систематическое отклонение по сравнению с остальной частью совокупности. Поэтому и причинные влияния этих входов имеют систематические отклонения. Такие смещения могут порождать дополнительные дисперсию и ковариацию между выходами при рассмотрении всей совокупности. Эти увеличенные характеристики совершенно не могут быть объяснены по схеме линейных влияний в подсистеме.

Пропускные механизмы поэтому дают ограничение системы. Система после прохождения через «проход» функционирует в другой совокупности, менее богатой, чем система до действия механизма. Поэтому такая система легко выделяется из своей среды. Более того, специфические дисперсии и ковариации, порождаемые пропускными механизмами, вероятно, могут быть включены в причинный анализ, если затронутые механизмом переменные находятся среди его входов. Статистические характеристики входов

находятся среди данных для анализа, независимо от того, были или нет эти характеристики причинными.

В самом деле, необходимо проявлять определенную осторожность при распространении путевого анализа за границы пропускных механизмов, учитывая, что внимание должным образом ограничено той частью совокупности всех случаев, которая проходит через «ворота». На такие случаи операторы действуют и до, и после прохождения через «ворота», но процессы отбора могут гарантировать, что некоторые возмущения системы вызваны способом, который обычно не может быть отражен в модели. Например, если члены некоторой организации отбираются по их консерватизму, то некоторые из них будут включены, поскольку воспитывались в своих семьях, как консерваторы. Другие, из либеральных семей, будут принадлежать к этой организации только потому, что, уже будучи взрослыми, постоянно подвергались консервативному влиянию. Это означает, что факторы, влияющие на переменную консерватизма во взрослом состоянии, в такой совокупности будут отрицательно коррелировать с семейными детерминантами консерватизма. Анализ путей от социализации в семье к переменной консерватизма будет искажен, пока не будет принята в расчет эта корреляция.

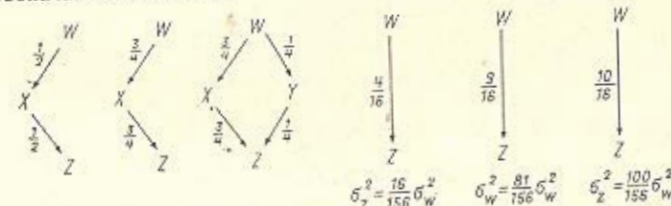
ОБЩИЕ ВЫВОДЫ

Источники разнообразия

Две совокупности с одинаковыми распределениями входных переменных, но различным разнообразием выходов, должны относиться к разным системам. Если у системы нет отрицательных операторов, то больший разброс выходов означает более сильные операции системы и (или) излишние операции.

4.19. Первоначальные системы.

Редуцированные формы



Если мы предположим, что σ_W^2 остается одним и тем же во всех трех случаях, то вторая система порождает большее разнообразие в Z, чем первая, потому что ее операторы больше. Третья — создает еще большее разнообразие, потому что имеет дополнительный путь или лишнюю операцию.

Вообще говоря, разнообразие выходов системы может быть увеличено путем усиления влияний операций системы. Например, если общество решается предложить большее вознаграждение за способности, то появится больший разброс в статусе, пока эта дифференциация не будет сдержана другими механизмами. Излишние операции дают аналогичный результат. Положим, экономические

интересы преобразуются в политические отношения с помощью средств массовой информации. Добавление новых информационных каналов в социальной системе приведет к большей политической дифференциации населения, если одновременно не будет усилен механизм управления.

Структурные коэффициенты с отрицательными значениями усложняют эти принципы в том смысле, что плохо поддаются общему описанию. Понимание может быть достигнуто изучением выражений редуцированной формы для конкретной системы. Все же важные принципы, касающиеся разнообразия, могут быть установлены, когда отрицательные операторы выражены в форме управляющих механизмов или петли с отрицательным обратным эффектом. Назовем управляющий механизм «более сильным», если обратный эффект соответствующей петли имеет большее отрицательное значение. Тогда разнообразие в совокупности будет тем меньше, чем более устойчивыми будут системы, действующие на совокупность, и (или) более сильным устойчивое управление. Добавленное управление с целью сдерживания отклонений — спасительная рабочая идея в политике и законодательных делах, но она является принципом общего значения для всех развивающихся систем, а не только для общественных.

$$4.20. X \xrightarrow{a} Y \xleftrightarrow{c} Z.$$

Предположим, X — талант, Y — образованность, Z — статус; a представляет собой преобразователь таланта в образованность; c — преобразователь образованности в статус; d — административные, социальные и эмоциональные помехи, обращающие высокий статус в снижение образованности.

Если $a = c = 1$, а $d = 0$ (поскольку указанные раздражители отсутствуют), то

$$\sigma_Z^2 = \left(\frac{ac}{1 - cd} \right)^2 \sigma_X^2 = \sigma_X^2.$$

Если мы теперь добавим незначительное влияние этих помех, например $d = -\frac{1}{4}$, то получим

$$\sigma_Z^2 = \left(\frac{1}{1 - (-\frac{1}{4})} \right)^2 \sigma_X^2 = \frac{16}{25} \sigma_X^2.$$

Так что усиление отрицательной обратной связи ведет к уменьшению разброса в статусе.

Разнообразие выходов зависит также от статистического распределения входов. Более вариабельны входы, больше разнообразия у выходов для системы без координации входов, состоящей из положительных операторов. Например, предполагая снова, что экономическая ситуация преобразуется в политические отношения, мы должны ожидать, что как только население обнаружит большее социально-экономическое разнообразие, оно покажет и больший разброс в политическом отношении. Эти же соображения применимы к возмущениям. Большая дисперсия возмущений приводит к большему разнообразию выходов системы.

Влияние координации между входами более многогранно. Если две совокупности подвергнуты действиям одинаковых систем положительных операторов, а дисперсии входов идентичны для обеих совокупностей, но входы одной совокупности положительно более скоординированы, то выходы этой совокупности дадут больший разброс. Положим, например, что две совокупности населения подвержены действиям системы, которая преобразует большую способность к наукам (T) и более высокий статус по происхождению (S) в большее число лет обучения (E): $T \rightarrow E \leftarrow S$. Пусть эти группы идентичны по их разбросу способностей и происхождения. Даже при всем этом сходстве одна популяция будет более вариабельна по уровню образования, чем другая, если у нее в большей степени скоординированы способности и происхождение. (Это различие может возникнуть, если одно общество награждает родителей на основе определенного генетического потенциала, который передается детям и обуславливает их способности, в то время как другое общество поощряет статус родителей на другой основе.) В обществе с координацией будет больше препятствий для соответствия между талантом детей и их происхождением, и эти две силы часто будут действовать вместе, порождая на выходе в результате высокое или низкое образование. В обществе без координации соответствующие браки будут случайны и менее часты, два указанных качества часто будут гасить друг друга, и система даст много примеров образования, близкого к среднему уровню.

С другой стороны, отрицательная координация между входами в системе положительных операторов производит еще более эффективное погашение операций и ведет к еще меньшей вариабельности выходов. Например (в продолжение предыдущего примера), если третья популяция действительно имеет отрицательную координацию между входами — больший статус родителей, меньший талант потомков, — то эти две силы должны погасить друг друга более эффективно, давая этим еще меньшую вариабельность в уровнях образования.

Эти взгляды могут быть резюмированы следующим образом. В системах положительных операторов разнообразие выходов всегда может быть понижено разнообразием входов или изменением координации между входами: уменьшением положительной или увеличением отрицательной. Когда система содержит отрицательные операторы, разнообразие выходов может регулироваться изменением дисперсий и ковариаций, однако необходимые манипуляции должны определяться в каждом конкретном случае.

Координация выходов

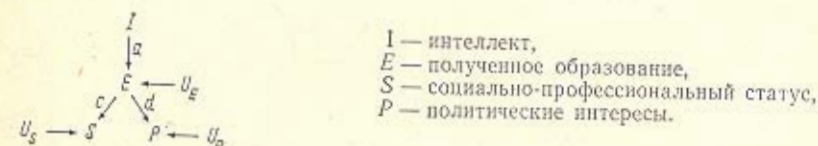
Если две системы различаются только по силе своих операций (скажем, все влияния в одной системе имеют в два раза больший эффект по сравнению с соответствующими влияниями в другой системе), то система с более сильными операторами приведет к

большой координации выходов, независимо от того, измеряется ли координация на языке ковариаций или корреляций. Если система содержит только положительные операторы, координация выходов возрастает, даже если операторы усилены в разной степени. Если система включает отрицательные операторы, то разное усиление может вызвать некоторое погашение эффектов, и влияние на формирование выходов необходимо определять в каждом конкретном случае.

Если две идентичные системы со всеми положительными коэффициентами действуют на совокупности, различающиеся только в дисперсиях входных переменных, выходы должны быть более организованы в совокупности с большим разбросом входов. Однако обычно, чтобы правильно определить влияние дисперсий входов, системы с отрицательными структурными коэффициентами нужно исследовать *ad hoc*.

Усиление возмущения на выходе уменьшает те корреляции между системными переменными, которые не зависят от него, а лишь от входных переменных. В то же время это изменение усиливает корреляцию между переменными, которые подвержены непосредственно ее влиянию, как показано в 4.21.

4.21.



Возрастание возмущений на переменную «образование» уменьшает корреляцию между ним и интеллектом, но это же самое изменение должно *увеличить* корреляцию между профессиональным статусом и уровнем политических интересов. Например, стипендия в колледжах для всех демобилизованных военнослужащих или для всех членов этнических групп в обществе с малым числом окончивших колледж уменьшит корреляцию между переменными интеллекта и образования у населения в целом, но, с другой стороны, возросшая дисперсия образования приведет к более систематическому разбросу в статусе и политических интересах. Следовательно, будет легче по одному из них предсказывать другое.

Степень координации между входными переменными является третьим фактором, действующим на корреляцию между переменными системы. Чем больше величина корреляции между входами, тем больше корреляция между другими переменными системы, по крайней мере если все операции в системе положительны. Если же некоторые входные корреляции и некоторые операции отрицательны, может иметь место погашение влияний, и воздействие входных корреляций необходимо оценивать каждый раз заново.

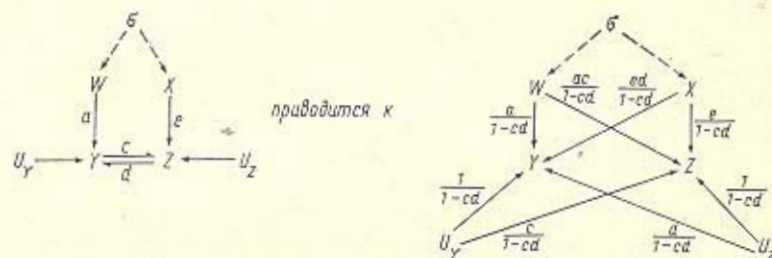
Корреляции в петлях

Переменные петли с положительным эффектом становятся более сильно коррелированными, если операторы петли усилены или если дисперсия переменных, примыкающих к петле, возра-

стает. Более мощные операторы или большая входная дисперсия по-разному порождают более высокую корреляцию между переменными петли с отрицательным эффектом. Действительно, *корреляция в управляющих петлях — явление крайне странное, и неправильное понимание такого явления легко может приводить к конфузу при обсуждении важных научных проблем.*

Чтобы исследовать эти явления, нужна формула, устанавливающая корреляцию между переменными петли. Алгебраические выкладки весьма сложны, поэтому ограничим наше внимание случаем простой петли из двух переменных.

4.22.



Дисперсия Y и Z и их ковариации могут быть получены использованием правил путевого анализа. Подстановка результирующих выражений в формулу для коэффициента корреляции (3.17) дает

$$\rho_{YZ} = \frac{a^2 \sigma_W^2 + e^2 \sigma_X^2 + c^2 \sigma_{U(Y)}^2 + d^2 \sigma_{U(Z)}^2 + ae(1+cd) \sigma_{WX}}{\sqrt{a^2 \sigma_W^2 + e^2 \sigma_X^2 + c^2 \sigma_{U(Y)}^2 + d^2 \sigma_{U(Z)}^2 + 2aed \sigma_{WX}}} \times \sqrt{e^2 \sigma_X^2 + a^2 c^2 \sigma_W^2 + \sigma_{U(Z)}^2 + c^2 \sigma_{U(Y)}^2 + 2aec \sigma_{WX}}$$

Для эвристических целей можно придать многим параметрам удобные значения:

$$a = e = 1; \quad \sigma_{U(Y)}^2 = \sigma_{U(Z)}^2 = 1; \quad \sigma_{WX} = 0.$$

Тогда формула примет вид

$$\rho_{YZ} = \frac{c \sigma_W^2 + d \sigma_X^2 + c + d}{\sqrt{\sigma_W^2 + d^2 \sigma_X^2 + 1 + d^2} \sqrt{\sigma_X^2 + c^2 \sigma_W^2 + 1 + c^2}}.$$

Коэффициенты петли могут задаваться следующими удобными значениями, определяющими отрицательный обратный эффект:

$$c = \frac{1}{2}, \quad d = -\frac{1}{2}.$$

В этом случае формула приобретает вид:

$$\rho_{YZ} = \frac{2(\sigma_W^2 - \sigma_X^2)}{\sqrt{4\sigma_W^2 + \sigma_X^2 + 5} \sqrt{4\sigma_X^2 + \sigma_W^2 + 5}}.$$

Смысл этой формулы может быть виден, если придать дисперсиям W и X различные значения и вычислить значения для ρ :

σ_W^2	σ_X^2	ρ_{YZ}
2,0	0,0	+0,42
2,0	1,0	+0,16
2,0	2,0	0,0
1,0	2,0	0,16
0,0	2,0	-0,42

Ясно, что корреляция между переменными особенно чувствительна к дисперсиям переменных, непосредственно действующих на петлю.

Пример 4.22 показывает, что, изменяя уровень разнообразия входных переменных для управляющей петли, корреляцию между переменными петли можно сделать и положительной, и отрицательной, и нулевой. Аналогично наличие положительной, отрицательной или нулевой корреляции между переменными в управляющей петле само по себе ничего не говорит о природе их причинных взаимоотношений. Рассмотренная многоликость механизмов управления подчеркивает необходимость быть осторожными при выводах причинных структур из наблюдаемых корреляций.

На примере видно, как феномен корреляции в управляющей петле может приводить к затемнению проблемы. Часто утверждают, что установка не влияет на реальное поведение, потому что корреляция между ними весьма мала, во многих исследованиях — практически нулевая. Столь малая корреляция действительно может маскировать существенное влияние отношений на поведение, если учесть, что могут иметь место социальные механизмы контроля, превращающие отклонение направленного на объект поведения (слишком привлекательный или слишком непривлекательный) в изменение отношения на противоположное. Основная идея здесь состоит в том, что проступок не прощается, пока тот, кто его совершил, не проявит нового отношения в предположении, что тот же самый тип отклонения не появится снова. Упрощенное представление такой системы имеет вид

$$U_A \rightarrow A \xrightleftharpoons[c]{a} B \leftarrow U_B.$$

Здесь a выражает преобразование благоприятного отношения в благоприятное поведение, c — преобразование отклонения в обратное изменение установки, а U_A и U_B являются неучитываемыми источниками вариаций в отношениях и поведении. Величина a должна быть положительной, c — отрицательной, чтобы представлять изменение в преобразовании отклоняющихся действий в изменение установки. Такая система соответствует типу, проанализированному в 4.21. Следовательно, корреляция между установкой

и поведением может принимать почти любое значение, и эта корреляция не дает никакого основания для решения о том, определяет ли установка поведение или нет.

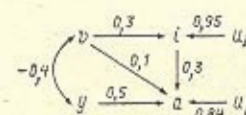
ИСТОЧНИКИ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

С. Райт (S. Wright) суммировал свои классические результаты по путевому анализу в *The Method of Path Coefficients*, *Annals of Mathematical Statistics*, 5 (1934), 161—215.

Основные статьи Райта и других исследователей, которые расширили сферу применения путевого анализа, удачно собраны в книге Н. М. Блалок, Дж., *Causal Models in the Social Sciences* (Chicago, Aldine-Atherton, 1971).

УПРАЖНЕНИЯ

1. Количество осужденных, ведущих себя в заключении агрессивно, зависит от характеристик всей совокупности заключенных и организационных характеристик тюрьмы. В следующей путевой диаграмме отражены некоторые вероятные взаимосвязи.



Переменные имеют следующий смысл:

v — доля заключенных, осужденных за преступления с применением насилия;
 i — доля заключенных, лишенных права посещения их более чем на год;
 u — возрастная характеристика заключенных, доля лиц моложе 21 года;
 a — доля агрессивных заключенных, получивших административное наказание.

- Вычислить ожидаемые корреляции между всеми переменными.
- Какая часть дисперсии в a объясняется другими переменными? Задаст ли приведенная схема границы эффективности программ, нацеленных на уменьшение a путем изменения i , v или u ?
- Какое ожидается влияние на a , если доля молодежи среди заключенных возрастет на одну стандартную единицу, в то время как доля осужденных за насилие останется без изменения? Что произойдет, если доля молодежи возрастет без изменения структуры совокупности правонарушителей?
- Используя введенные здесь три независимые переменные, определить тип коэффициента заключенных тюрьмы, где больше всего проблем по обеспечению внутреннего порядка? Каковы будут характеристики тюрем с минимальным количеством драк среди заключенных?

2. Представим себе некоторую группу населения, в которой имеется дифференциация по образованию, профессиональному статусу и доходу, но нет связи между этими переменными; например, пусть дисперсия всех переменных равна 1 и все три корреляции равны нулю. Положим затем, что введены в действие социальные операторы, в связи с чем занятия индивидуума стали зависеть от его образования, а доходы — от занятий. В такой системе образование будет входной переменной, и начальный разброс по занятиям и доходу вызван неучитываемыми возмущениями. Ради удобства положим, что структурный коэффициент для каждого нового оператора равен 1.

(а) Какими будут ковариации и дисперсия переменных после того, как будет установлена новая социальная система? Чему будут равны корреляции между переменными?

¹ Ellis D., Grasmick H. and Gilman B. Violence in Prisons: A Sociological Analysis, — *American Journal of Sociology*, 80, № 1, July 1974.

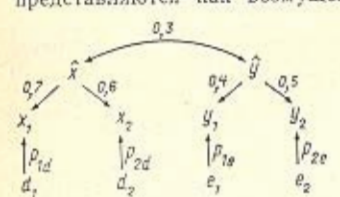
(б) Чему равны значения системных (стандартизованных) путей коэффициентов для новой совокупности? Означают ли эти пути коэффициенты что-нибудь новое по сравнению с первоначальными структурными коэффициентами? (Заметим, что вводимые операторы изменяют характер совокупности посредством увеличения дисперсии профессионального престижа и дохода.)

(в) Совокупность будет обладать социальной стратификацией, если различия по статусу упорядочены так, что множество статусов с точки зрения каждой личности (или семьи) внутренне согласованы. Отправляясь от условий этого упражнения, определите, что требуется для развития стратификации? Какие статусы в социальной системе имеют наибольшее неравенство?

3. Предположим, что социальная система, описанная в упражнении 2, содержит дополнительный оператор, такой, что заданный уровень дохода вызывает мобильность с целью достижения соответствующего престижа (например, было бы именно так, если бы можно было покупать занятия). Предположим, что структурный коэффициент для этого оператора равен 0,5 (по сравнению со значением в 1,0 для двух других операторов). Снова рассчитайте дисперсии, ковариации и корреляции переменных статуса для видоизмененной системы.

Какое влияние оказывает усиливающая петля на дисперсии переменных в петле? На корреляции переменных петли?

4. На приведенной ниже диаграмме $\hat{x}$ обозначает количество стандартизованных единиц, представляющих истинное отношение личности к абиту, а x_1 и x_2 — соответствующие стандартизованные величины, полученные двумя анкетными измерениями этого отношения. Стрелки от истинных величин к индикаторам означают, что истинное отношение личности определяет его реакцию на частное анкетное измерение. Однако реакция на вопросы анкеты зависит от множества других факторов — неправильного понимания, настроения, влияния интервьюера, раздражения. Все такие источники ошибок измерения совокупности представляются как возмущения d_1 и d_2 . Переменная $\hat{y}$ обозначает истинные уровни информированности личности об абите. Опять переменная стандартизована и является входом для двух анкетных измерений y_1 и y_2 . Предполагается, что эти индикаторы имеют ошибки измерения e_1 и e_2 . Коэффициенты при стрелках означают, что каким-то образом нам известны корреляции $\rho_{\hat{x}\hat{y}}$ между истинными величинами и их индикаторами.



(а) Какая корреляция ожидается между обоими индикаторами $\hat{x}$? Между двумя индикаторами $\hat{y}$? Почему корреляция между двумя измерениями одной и той же переменной не равна единице?

(б) Каковы корреляции между индикаторами $\hat{x}$ и индикаторами $\hat{y}$? Если известны значения этих корреляций, но не значение $\rho_{\hat{x}\hat{y}}$, что можно сказать о взаимосвязи отношения и информированности в этом случае?

(в) Какая доля дисперсии каждого индикатора обусловлена ошибками измерения? Если ошибки считать стандартизованными, чему будут равны значения R^2 -коэффициентов при возмущающих стрелках?

(г) Коэффициенты у стрелок от истинных величин к индикаторам могут быть названы коэффициентами обоснованности. Предположим, нам известны только корреляция между двумя индикаторами $\rho_{x_1x_2}$ и коэффициенты обоснованности для x_1 и y_1 . Используйте данный рисунок для оценки корреляции между истинными величинами $\hat{x}$ и $\hat{y}$.

5. В упражнении 4 предполагалось, что ошибки измерения для разных индикаторов были не коррелированы. Если, однако, все вопросы задавались в одном и том же интервью, то этого может и не быть.

Продолжая иметь дело с моделью предыдущего упражнения, но предполагая, что коэффициент корреляции ошибок всех индикаторов друг с другом равен 0,30, ответьте на следующие вопросы.

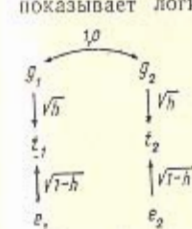
(а) Пересчитайте корреляции между всеми индикаторами, используя параметры из упражнения 4 и дополнительные данные. Какая общая тенденция появится в результатах по сравнению с полученными в упражнении 4? Будет ли это обычным влиянием коррелированности ошибок?

(б) Снова применяя процедуру упражнения 4г, оцените истинную корреляцию $\rho_{\hat{x}\hat{y}}$. Используйте теперь коэффициенты обоснованности для x_1 и y_1 и «наблюдаемую» корреляцию $\rho_{x_1y_1}$, полученную в части (а) этого упражнения. Может ли это привести к ошибочным выводам?

(в) Предположим, что x представляет собой мнение по вопросу, по которому интенсивность согласия или несогласия принимает ровно пять значений — от 1 до 5. Есть основания предполагать, что ошибки d_1 для этой шкалы, возможно, будут отрицательно коррелировать с истинным отношением $\hat{x}$. Объясните, почему?

6. В системе, представленной в 4.20, способности реформировались в образованность (оператор а), а образованность влияла на статус (оператор с). Допускалось также, что управляющий оператор (d) может действовать так, что больший статус породит меньшее образование. Статистические следствия этих предположений были получены, когда структурным коэффициентам а, с и d придавались произвольные различные значения: 1,0; 1,0 и -0,25 соответственно. Теперь положим, что а и d остаются теми же самыми, но системный механизм поощрения с удвоился по силе, так что коэффициент принял значения 2,0. Каким будет действие большего поощрения на варибельность статуса в системе, в которой управляющий механизм совсем не развит (т.е. где $d = 0$)? Каким будет влияние на разброс в статусе для системы, содержащей управляющий механизм ($d = -0,25$)? По этим результатам проинтерпретируйте функции и дисфункции управляющего механизма.

7. Коэффициент наследования h обозначает долю дисперсии наблюдаемых черт, которая может быть объяснена вариациями в генетической конституции. Обычно генетические переменные не могут быть измерены непосредственно, поэтому коэффициент наследования оценивается косвенно. Следующая диаграмма показывает логическую основу многих исследований наследования признаков.



Переменные с индексом 1 относятся к одному члену пары идентичных близнецов, с индексом 2 — к другому. Такой подход трактует пару идентичных близнецов как основную единицу анализа. Переменная g — значение для личности соответствующей генетической переменной. (В действительности почти любая характеристика зависит от многочисленных генетических переменных, и g реально является агрегированной переменной, сравнимой с возмущающим членом.) Переменная t представляет значение интересующей нас характеристики личности (например, рост или интеллект). Переменная e представляет не генетические, а внешние источники вариаций в t . Заметим, что корреляция между g_1 и g_2 равна 1,0. Исследование намеренно ограничивается идентичными близнецами, так что все эти предположения совершенно справедливы. Идентичные близнецы имеют один и тот же генотип. Обратим также внимание, что $\sqrt{h}$ обозначает путьевой коэффициент от g к t . Так должно быть, если исследование есть доля дисперсии характеристик, обусловленных генетически (см. упражнение 16).

(а) При данной выше модели какая наблюдаемая статистика дает прямую оценку коэффициента наследования h ?

(б) Почему исследователи наследственности предпочитают изучать близнецов, которые выросли врозь?

(в) Наследование умственных способностей иногда полагают приблизительно равным 0,80. Значит ли это, что с точки зрения практики воздействием окружающей среды нельзя добиться роста умственных способностей?

(г) Пусть характеристика t измеряется с ошибкой. Что было бы ожидаемым воздействием на оценки наследуемости?

(д) Предположим, что социально-культурная система постепенно становится более разнородной в том смысле, что разнообразие окружающей среды возрастает. Будут ли наследуемости большей части характеристик увеличиваться, уменьшаться или останутся без изменений?

¹ Это упражнение взято из материалов обзора в работе Heise D. Personality: Biosocial Bases. Chicago, Rand McNally, 1973.

8. Предположим, что исследователь, интересующийся эстетическими ценностями, получает данные от всех студентов-гуманитариев в большом университете штата. Он находит, что точная мера эстетических ценностей в этой совокупности только незначительно коррелирует со статусами родителей и многочисленными индикаторами культурного опыта детства. Может ли он сделать вывод, что статус родителей и культурный опыт детства не влияют на эстетические ценности?

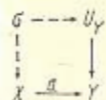
9. Предположим, что в комплекс тестов, широко применяемых для изучения студентов высших учебных заведений в течение нескольких лет, включена действительная проверка увлечения музыкой. Исследователь составляет совокупность из людей, прошедших проверку год тому назад, половина из которых теперь — профессиональные музыканты, а половина выбрана случайно. Он обнаруживает сильную корреляцию между результатами испытания и мерой теперешнего музыкального мастерства. Затем он повторяет исследование на совокупности, составленной только из любителей, и обнаруживает гораздо меньшую корреляцию. Что может быть вероятным объяснением этого типа полученных данных?

Глава 5 • ИДЕНТИФИКАЦИЯ И ОЦЕНКА

Полной способности объяснять, предсказывать и управлять с помощью причинного анализа нельзя достичь до тех пор, пока не оценены численно структурные параметры причинной системы. Иногда, в технике, операторы в соответствии со своим назначением обладают специфическими количественными следствиями, и в этом случае численная оценка коэффициентов не представляет особой проблемы. Однако социальные системы часто являются результатом развития, но не конструирования, и социальные законы, а также основные константы, лежащие в основании функционирования системы, обычно не известны. Таким образом, параметры социальной системы должны оцениваться по эмпирическим наблюдениям над системой, а не дедуктивно из имеющегося знания.

Принципы «путевого анализа», представленные в гл. 4, приводят к важной совокупности методов эмпирической оценки параметров системы. Здесь разрабатывается основная логика *методов наименьших квадратов* по отношению к простой системе, определяемой в 5.1.

5.1. Теоретическая система состоит только из X и Y , причем X воздействует на Y через оператор, линейным действием которого является a . Ковариационный член показывает, что X может согласовываться с другими, точно не указанными факторами, определяющими Y в интересующей совокупности.



Сначала значения X и Y измеряются для всех случаев, которые были подвержены действию системы. По этим результатам измерений вычисляются следующие статистики: дисперсия X (σ_X^2), дисперсия Y (σ_Y^2) и ковариация X и Y (σ_{XY}).

Эти вычисления прямо дают один параметр, σ_X^2 , который нужен для анализа последствий этой простой системы, касающихся распределения. Если бы имелись дополнительные точно указанные причины, тут же были бы известны также их дисперсии и ковариации.

Путевой анализ может использоваться для выражения значений остальных наблюдаемых статистик через основные параметры

системы и совокупности:

$$\begin{aligned}\sigma_{XY} &= a\sigma_X^2 + \sigma_{XU(Y)}, \\ \sigma_Y^2 &= a^2\sigma_X^2 + \sigma_{U(Y)}^2 + 2a\sigma_{XU(Y)}.\end{aligned}$$

Первое уравнение можно разрешить относительно a , что даст формулу, которая выражает значение a через статистики совокупности:

$$a = \frac{\sigma_{XY} - \sigma_{XU(Y)}}{\sigma_X^2}.$$

Две величины справа (σ_{XY} и σ_X^2) известны, потому что их значение было вычислено непосредственно из данных. Однако значение третьей величины $\sigma_{XU(Y)}$ непосредственно не измеримо, и потому, что она неизвестна, формула определяет значение a неоднозначно.

Аналогично второе уравнение, с учетом третьего, можно разрешить, что дает уравнение для дисперсии переменной возмущения $\sigma_{U(Y)}^2$:

$$\sigma_{U(Y)}^2 = \sigma_Y^2 - \frac{\sigma_{XY}^2 - \sigma_{XU(Y)}^2}{\sigma_X^2}.$$

Это уравнение также содержит неизвестную величину $\sigma_{XU(Y)}$ и потому не определяет $\sigma_{U(Y)}^2$ однозначно.

Предположим, что формулируется ограничительное условие. Эти формулы могут применяться только тогда, когда X не согласована с любым возмущением Y . В этом случае неизвестная величина в этих формулах имеет «известное» значение, $\sigma_{XU(Y)} = 0$, а формулы упрощаются, принимая следующий вид:

$$\left. \begin{aligned}a &= \frac{\sigma_{XY}}{\sigma_X^2} \\ \sigma_{U(Y)}^2 &= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}\end{aligned} \right\} \text{ если } \sigma_{XU(Y)} = 0.$$

Таким образом, в ситуациях с ограничением (т. е. когда X и U_Y являются несогласованными) и коэффициент a , и дисперсия возмущения $\sigma_{U(Y)}^2$ однозначно определяются через наблюдаемые статистические величины. Следовательно, формулы могли бы использоваться для оценки этих параметров по эмпирическим статистикам.

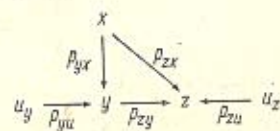
Этот пример освещает типичную трудность в области оценок наименьших квадратов. Путевой анализ дает два уравнения, которые можно решить относительно неизвестных параметров, но эти уравнения содержат три неизвестных. Таким образом, однозначное решение для любого из неизвестных невозможно до тех

пор, пока не определят значение одного неизвестного на эмпирической основе. Общее затруднение при неоднозначном решении известно как *задача идентификации*. Системы, которые дают меньше путевых аналитических уравнений, чем имеется неизвестных, называются *недоопределенными*.

РЕГРЕССИОННЫЙ АНАЛИЗ И ИДЕНТИФИКАЦИЯ

Второй пример иллюстрирует другие особенности логики методов наименьших квадратов. Хотя рассматриваемая система лишь чуть-чуть сложнее, алгебра резко усложняется и два упрощающих предположения делаются заранее. Сначала предполагается, что все переменные измеряются на шкалах со стандартизованными единицами. Это не ограничительное предположение (в сущности, те же самые результаты могут получаться и без него). Однако то, что все дисперсии делаются равными 1,0, вносит упрощение, и дисперсии не должны представляться в формулах явно. Вдобавок предполагается, что возмущения переменной не коррелируют с точно указанными причинами переменной. Это ограничительное предположение в том отношении, что результаты применимы только в ситуациях, в которых верны предположения. Однако именно это частное ограничение требуется для дальнейшего раскрытия логики наименьших квадратов. Рассматриваемая система представлена в 5.2.

5.2. Все переменные измеряются в стандартных единицах, а возмущения не коррелированы с x или друг с другом.



Опять во всех интересующих нас случаях проделываются измерения для каждой переменной и вычисляются определяющие статистики. Так как переменные стандартизованы, дисперсии всех переменных равны 1,0 и ковариации совпадают с корреляциями. Путевой анализ точно описывает корреляции в виде

$$\rho_{xy} = \rho_{yx}, \quad \rho_{yz} = \rho_{zy} + \rho_{yx} \cdot \rho_{zx}, \quad \rho_{xz} = \rho_{zx} + \rho_{yx} \cdot \rho_{zy}$$

и решение относительно неизвестных ρ дает следующие формулы:

$$\rho_{yx} = \rho_{xy}, \quad \rho_{zy} = \frac{\rho_{yz} - \rho_{xy} \cdot \rho_{xz}}{1 - \rho_{xy}^2}, \quad \rho_{zx} = \frac{\rho_{xz} - \rho_{xy} \cdot \rho_{yz}}{1 - \rho_{xy}^2}.$$

Эти формулы, однако, в точности совпадают с формулами, определяющими стандартизованные частные коэффициенты регрессии (см. 3.32 и 3.33). Если мы рассматриваем регрессию y

по x и z по паре x и y , то результаты регрессий оценивают операторы системы:

$$p_{yx} = \beta_{yx}, \quad p_{zy} = \beta_{zy \cdot x}, \quad p_{zx} = \beta_{zx \cdot y}.$$

Путевой анализ дисперсий можно проводить как обычно, если мы помним, что все дисперсии имеют значение 1,0.

$$\sigma_y^2 = 1 = p_{yy}^2 + p_{yx}^2 = p_{yy}^2 + p_{xy}^2,$$

$$\sigma_z^2 = 1 = p_{zz}^2 + p_{zy}^2 + p_{zx}^2 + 2p_{xy}p_{zy}p_{zx}.$$

Путем использования некоторых предыдущих результатов при подстановках можно получить следующие формулы:

$$p_{yy}^2 = 1 - (\beta_{yx}^2), \quad p_{zz}^2 = 1 - (\beta_{zy \cdot x}^2 + \beta_{zx \cdot y}^2 + 2\beta_{zy \cdot x}\beta_{zx \cdot y}p_{xy}).$$

Здесь величины в скобках в точности равны выражениям для коэффициентов детерминации (см. 3.34). Путевые коэффициенты от возмущений можно также получить непосредственно из результатов регрессионного анализа:

$$p_{yy}^2 = 1 - R_{y \cdot x}^2, \quad p_{zz}^2 = 1 - R_{z \cdot xy}^2.$$

Был проиллюстрирован общий принцип. При ограниченных условиях параметры системы могут оцениваться путем многократного проведения регрессионного анализа. Этот параллелизм между системной оценкой и регрессионным анализом имеет силу только тогда, когда переменные-причины и возмущения не коррелированы. Регрессионные процедуры математически определяются для исключения корреляции между независимыми переменными и остатками зависимых переменных. Если переменные-причины и возмущения в действительности коррелируют, коэффициенты регрессии не соответствуют параметрам системы.

Потребность в теории

Данные примеры служат для того, чтобы подчеркнуть, что эмпирические наблюдения могут приводить к выводам о природе системы лишь в контексте теоретических предположений. Формула может использоваться для оценки параметра, если она однозначно определяет этот параметр через измеримые статистики. Вообще это верно, только если некоторые величины исключаются из формулы на теоретической основе; например, то, что переменная-причина не находится под влиянием зависимой переменной, должно быть известно до того, как будет дана причинная интерпретация регрессии одной переменной по другой. Далее, то, что любое ложное согласование между двумя переменными отсутствует или им можно управлять, должно быть известно прежде, чем коэффициент регрессии можно будет интерпретировать как действительную оценку причинного влияния. Различные методы оценки, упоминаемые далее, предлагают некоторое разнообразие

в выборе теоретических предположений, нужных для идентификации системы, но до того, как данные можно будет интерпретировать, понадобится некий тип теоретической информации. Простых наблюдений системы недостаточно для ее идентификации и задачу идентификации никогда нельзя будет решить путем простого собирания наблюдений на большом числе случаев. Недопределенность является теоретической, а не статистической проблемой.

На всем протяжении этой главы предполагается, что дисперсии и ковариации вычисляются на основе всех случаев в совокупности для того, чтобы обойти сложности, создаваемые вероятностным выводом по выборкам. На практике оценки параметра могут вычисляться по выборочным наблюдениям, и эта экономия обычно используется в социальных исследованиях. При работе с выборками увеличение объема выборки обычно действительно улучшает точность оценок. (Проблемы извлечения выборки и статистического вывода кратко обсуждаются в конце главы. Эти вопросы подробно изложены в статистических и эконометрических руководствах.) Тем не менее оценка совсем невозможна, если не существует формулы, которая дает однозначное математическое определение желаемого параметра. Такие формулы получают только путем выдвижения предположений теоретического характера.

РЕКУРСИВНЫЕ СИСТЕМЫ

Следующее определение предлагает понятие, которое пригодно для более подробной разработки методов оценки путем проведения обычного регрессионного анализа.

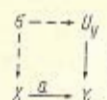
V.1. *Связь между двумя переменными называется рекурсивной, если она линейна, если две переменные не находятся в петле и если переменная-причина не согласована с возмущениями зависимой переменной. Если все причинные связи в системе рекурсивны, рекурсивной называется вся система.*

«Рекурсивная» — просто название для типа причинной связи, при которой параметры определяются на языке обычных регрессионных формул. Благодаря легкости, с которой можно оценить параметры, рекурсивные связи действительно оправдывают некоторое особое внимание.

Вопрос о линейности обсуждался в гл. 1 и больше здесь не рассматривается. Установление того, что две переменные не находятся в петле, равнозначно выводу об отсутствии какого-либо причинного влияния одной переменной на другую. Основания для таких причинных заключений обсуждались в гл. 1 и не нуждаются в повторении. Обсуждение здесь сосредоточено на вопросе согласования между причинами и возмущениями.

С теоретической точки зрения три условия, описанные в 5.3, могут создать корреляцию между переменной-причиной и возмущениями зависимой переменной.

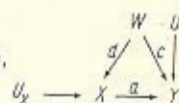
5.3. Линейная связь, направленная от X к Y , рекурсивна, если возмущения зависимой переменной согласуются с точно указанной переменной-причиной; иными словами,



Это согласование может раскрываться тремя способами:

1. Общая причина X и Y явно не рассматривается; она существует неявно в совокупности U_Y . Эту связь можно сделать рекурсивной путем углубления

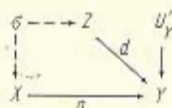
описания системы до явного включения общей причины; например,



Неявная переменная создает рекурсивность, только если она определяет и причину, и зависимую переменную. В частности, вычеркивание в рекурсивной системе промежуточной переменной не создает рекурсивности в уменьшенной системе; например, обе приводимые здесь системы являются рекурсивными:

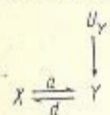


2. Некоторая переменная из совокупности U_Y участвует совместно с X в пропускном механизме, создавая этим искусственное согласование между X и другой переменной, скажем Z , из рассматриваемой совокупности. В этой ситуации можно было бы создать рекурсивность путем удаления Z из совокупности возмущений и рассмотрения ее в явном виде, потому что согласованность между точно указанными причинами не препятствует рекурсивности.



Ограничивающее действие пропускного механизма непосредственно по переменной Y образует отрицательную корреляцию между X и U_Y , как показано в гл. 4. Эту проблему можно снять путем рассмотрения всей совокупности (случаи и подвергшиеся действию пропускного механизма, и не подвергшиеся).

3. Если X и Y вместе входят в петлю, в которой Y непосредственно или опосредованно является причиной X и зависит от X , то возмущения переменной Y должны быть коррелированными со значениями X .



X и U_Y согласуются путем $(U_Y Y X)$. Здесь согласование между X

и U_Y существенно и не может быть исключено дальнейшей детализацией системы.

Таким образом, связь между двумя переменными может считаться рекурсивной, только если (а) причинность односторонняя:

известно, что одна переменная не влияет на другую и нет неясности, какая переменная является причиной, а какая — зависимой; (б) любая третья переменная, которая является причиной обеих, явно включается в анализ; (в) причинная переменная (или одна из них) не включается в пропускной механизм вместе с любой точно не описанной переменной, которая влияет на зависимую переменную, а интересующая совокупность не подвергнута действию пропускного механизма по зависимой переменной.

Обычный метод наименьших квадратов (ОМНК)

Обычный метод наименьших квадратов, описанный в следующем правиле, дает несмещенные оценки параметров, только если связи между зависимой переменной и всеми ее причинами рекурсивны.

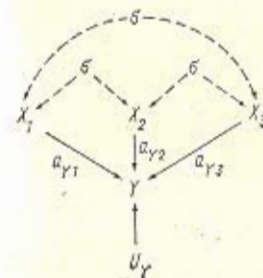
V.2. Если все связи между переменной системы и ее причинами рекурсивны, структурные коэффициенты для операторов, определяющих переменную, можно оценить непосредственно путем регрессионного анализа, а дисперсию возмущения — дисперсией регрессионных остатков. Соответствующая регрессия определяется следующим образом.

(а) Выписывается структурное уравнение, выражающее значение зависимой переменной как функцию ее непосредственных причин, структурных коэффициентов и члена-возмущения, например $Z = aX + dY + U_Z$.

(б) Уравнение определяет модель множественной регрессии — проводят регрессию зависимой переменной (Z) по всем ее точно указанным причинам (X, Y).

(в) Получившиеся в результате коэффициенты регрессии определяют структурные коэффициенты, например $a = b_{ZX \cdot Y}$, $d = b_{ZY \cdot X}$, а остаточная дисперсия, определяемая множественным коэффициентом корреляции для этой регрессии, дает дисперсию возмущения, например $\sigma_{U(Z)}^2 = \sigma_{(Z-\hat{Z})}^2 = (1 - R_{Z \cdot XY}^2) \sigma_Z^2$. Идеи, заключенные в обычном методе наименьших квадратов, детализируются в 5.4.

5.4. Основная парадигма для обычного метода наименьших квадратов



Ограничения: связь Y со всеми ее причинами рекурсивна. Теоретически нет ограничений на количество переменных X или на их взаимосвязи.

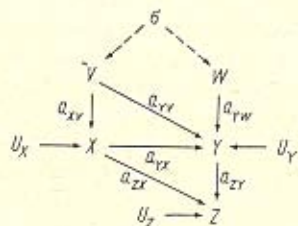
Процедура: провести регрессию Y по X_1, X_2 и X_3 . Частные коэффициенты регрессии определяют структурные коэффициенты: $a_{Y1} = b_{Y1 \cdot 23}$, $a_{Y2} = b_{Y2 \cdot 13}$, $a_{Y3} = b_{Y3 \cdot 12}$.

Дисперсия возмущения определяется множественным коэффициентом корреляции и дисперсией переменной Y :

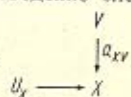
$$\sigma_{U(Y)}^2 = (1 - R_{Y.123}^2) \sigma_Y^2.$$

В рекурсивной системе этот образец может применяться неоднократно для установления и оценки всех неизвестных параметров, что иллюстрируется 5.5.

5.5. Здесь нет петель, а все U не коррелированы с V или W или же друг с другом. Следовательно, для оценки параметров можно использовать обычный метод наименьших квадратов.

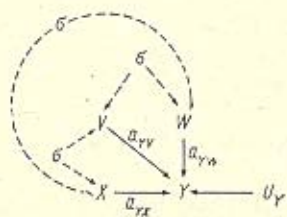


Процедура включает проведение следующих отдельных регрессий для каждой зависимой переменной:



определяет регрессию X по V для

установления a_{XV} и $\sigma_{U(X)}^2$.



определяет регрессию Y по V , W и X для установления a_{YV} , a_{YW} , a_{YX} и $\sigma_{U(Y)}^2$. Обратите внимание, что причинное отношение между V и X просто подытоживается ковариационным членом (и неявно дисперсиями V и X), свидетельствуя о несущественности на этом шаге рода связи между V и X . А именно регрессионная процедура пренебрегает деталями причинных связей между предшествующими переменными и принимает в расчет чистые следствия этих процессов, отраженные в ковариациях и дисперсиях переменных.

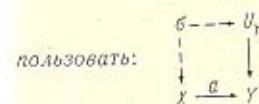
Эта схема определяет регрессию Z по X и Y для установления a_{ZX} , a_{ZY} и $\sigma_{U(Z)}^2$. Опять X и Y трактуются как непосредственные причины Z . Подробности их причинных связей пренебрегают и трактуют эти связи в сжатом итоговом виде по наблюдаемой ковариации σ_{XY} , а также, неявно, по наблюдаемым дисперсиям σ_X^2 и σ_Y^2 .



По существу, обычный метод наименьших квадратов включает в себя разбиение схемы на множество подсхем, причем каждая пригодна для определения дисперсии единственной зависимой переменной через дисперсии и ковариации ее причин. В подсхемы включаются только непосредственные причины зависимой переменной, а связи между этими причинами суммируются ковариационными членами.

В 5.6 приводятся некоторые иллюстративные ситуации, в которых обычный метод наименьших квадратов не применим.

5.6. Случай, в которых обычный метод наименьших квадратов не следует использовать:

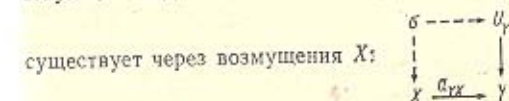


Возмущения переменной Y согласуются с точно указанной причиной X . В этом случае коэффициент регрессии оценивает не a , а скорее

$$b_{YX} = a + \frac{\sigma_{U(Y)}}{\sigma_X^2} \sigma_X^2$$

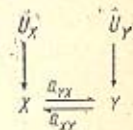
A diagram showing a path from X to Y with coefficient a , and a dashed arrow from U_Y to Y .

Здесь согласуются сами возмущения. Хотя a_{XW} может оцениваться обычным методом наименьших квадратов, a_{YX} однозначно не устанавливается. Соответствующая подсхема должна отражать корреляцию между X и U_Y , которая



существует через возмущения X :

Однако это та же самая схема, которая рассматривалась выше



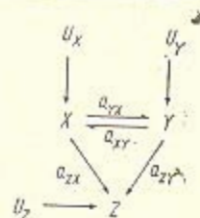
Обе переменные входят в петлю и регрессия Y по X не дает возможности установить a_{YX} . Коэффициент регрессии дает другую величину, не представляющую особого интереса:

$$b_{YX} = \frac{a_{YX} + a_{XY}(\sigma_Y^2/\sigma_X^2)}{1 + a_{YX}a_{XY}}$$

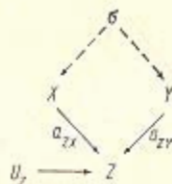
Коэффициенты петли никогда нельзя оценить в точности путем применения обычного метода наименьших квадратов, потому что параметры петли должным образом не устанавливаются по простым регрессионным статистикам. Однако может оказаться возможным установить части системы петли путем использования обычного метода наименьших квадратов. В частности, независимые

переменные в уравнении регрессии могут входить в петли и это не порождает смещения результатов, потому что петля не влияет на открытые пути, ведущие от переменных петли (см. правило «касания» — II.15).

5.7.



Обычный метод наименьших квадратов использовать для оценки a_{YX} или a_{XY} нельзя, но его можно применить для оценки коэффициентов путей, ведущих к Z, потому что изображенная ниже подсистема точно показывает информацию, относящуюся к Z:



Другими словами, X и Y являются рекурсивными причинами по отношению к Z, даже если они являются нерекурсивными причинами по отношению друг к другу.

НЕРЕКУРСИВНЫЕ СИСТЕМЫ

Коэффициенты в нерекурсивных соотношениях также можно установить и оценить при условии, что в описание системы входят некоторые переменные с определенными ограниченными свойствами. Эти «инструментальные переменные» (определяемые в следующем правиле) могут быть частью первоначального описания системы или могут добавляться к описанию системы просто как переменные, участвующие в исследовательских расчетах. Безотносительно к тому, считаются ли инструментальные переменные практически подходящими или нет, их концептуализация во время теоретизирования чрезвычайно важна. Структурные коэффициенты в нерекурсивной системе можно оценивать по данным «поперечного» анализа, только если имеются адекватные инструменты.

Инструментальные переменные

V.3. А. Переменная X является инструментом для Y в нерекурсивном отношении $Y \rightarrow Z$, если

(а) X непосредственно не воздействует на Z;

(б) X влияет на Y, либо непосредственно, либо через промежуточную переменную, которая не воздействует на Z непосредственно;

(в) ни Y, ни Z не действуют на X непосредственно или опосредованно;

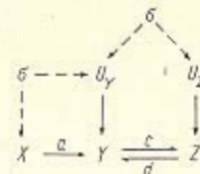
(г) никакой точно указанный фактор не влияет на пару X и Z и, вообще говоря, X не согласована с возмущениями переменной Z.

Б. Переменная X', которая просто коррелирует с Y, также является инструментом для связи $Y \rightarrow Z$ при условии, что она удовлетворяет условиям (а), (в) и (г), приведенным выше.

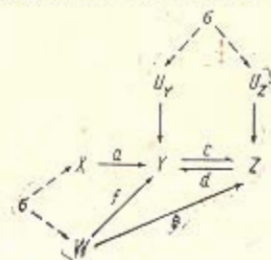
Если X является инструментом для связи $Y \rightarrow Z$, то связь между X и Z подобна рекурсивной в двух отношениях, а именно Z не действует на X, а X не согласуется с не указанными точно переменными, влияющими на Z (возмущениями переменной Z). С другой стороны, инструментальная причина противоположна обычной рекурсивной причине в том, что она определенно не должна оказывать прямого действия на Z; она должна влиять на Z только через точно указанные промежуточные переменные, в частности через переменную Y.

Инструмент X, как определяется в разделе А правила, должен быть причиной для переменной Y, но связь между X и Y не должна быть полностью рекурсивной вследствие того, что X может коррелировать с возмущениями переменной Y. Раздел Б расширяет понятие инструмента до включения любой переменной, которая просто коррелирована с Y, при условии, что корреляция не обусловлена причинным воздействием со стороны Y и что переменная удовлетворяет условиям (а), (в), (г). Примеры инструментальных переменных представлены в 5.8.

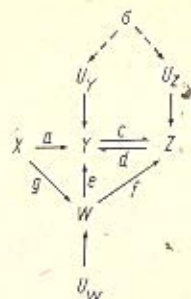
5.8. Примеры инструментальных переменных. В каждом случае X — действительный инструмент относительно связи $Y \rightarrow Z$.



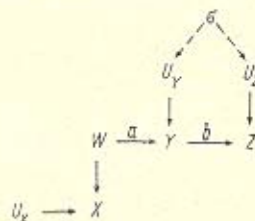
X может быть коррелированной с U_Y , но не с U_Z . U_Y и U_Z могут быть коррелированными, а Y и Z могут быть переменными в петле.



X может быть коррелированной с другими точно указанными переменными-причинами, влияющими либо на Y , либо на Z .



X может влиять на Z через промежуточные переменные, отличные от Y , если они включаются в анализ явно.



X является инструментом, если ее связи с Y и Z опосредуются первичным инструментом W и если ее возмущения некоррелированы с Z .

Дальнейшее истолкование инструментальных переменных можно получить путем рассмотрения некоторых примеров, в которых переменная перестает удовлетворять определяющим условиям.

5.9. В следующих примерах X не является инструментом для связи $Y \rightarrow Z$:

$$U_X \rightarrow X \leftarrow Y \rightarrow Z \leftarrow U_Z$$

так как Y влияет на X .

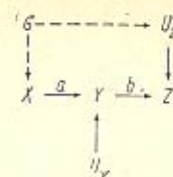
$$U_X \rightarrow X \xrightarrow{a} Y \xrightarrow{c} Z \leftarrow U_Z$$

так как X влияет на Z непосредственно.

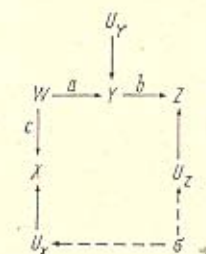
$$U_Y \rightarrow Y \xrightarrow{a} Z \leftarrow U_Z$$

так как Z влияет на X .

$$U_X \rightarrow X \xrightarrow{a} Y \xrightarrow{c} Z$$



так как X коррелирована с U_Z .



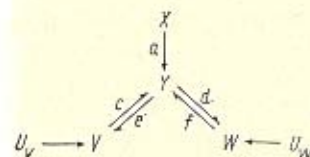
так как X коррелирована с U_Z .

Переменная может служить инструментом более чем для одного отношения, как показано в примерах в 5.10.

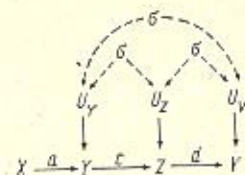
5.10. Примеры многоцелевых инструментов

$$U_X \rightarrow X \xrightarrow{a} Y \xrightarrow{c} Z \leftarrow U_Z$$

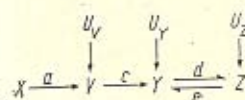
X — инструмент для $V \rightarrow W$ и для $Y \rightarrow Z$.



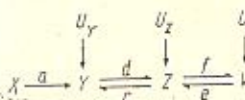
X — инструмент для $Y \rightarrow V$ и для $Y \rightarrow W$.



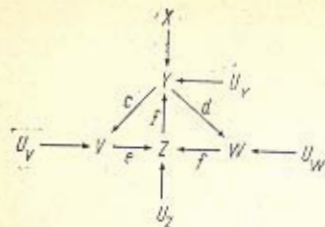
X — инструмент для $Y \rightarrow Z$, а также является инструментом для $Z \rightarrow V$.



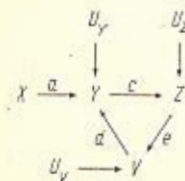
X — инструмент для $Y \rightarrow Z$ и был бы инструментом для $V \rightarrow Y$, но, чтобы иметь дело с этим рекурсивным отношением, никакого инструмента не надо.



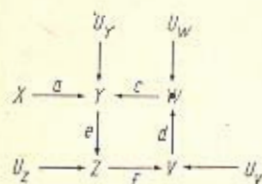
X — инструмент для $Y \rightarrow Z$; это также инструмент для $Z \rightarrow V$.



X — инструмент для $Y \rightarrow V$ и для $Y \rightarrow W$, а также для $V \rightarrow Z$ и $W \rightarrow Z$.



X — инструмент для $Y \rightarrow Z$, а также инструмент для $Z \rightarrow V$.

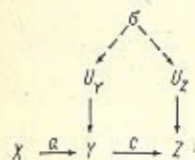


X — инструмент для $Y \rightarrow Z$, $Z \rightarrow V$ и $V \rightarrow W$.

Инструменты и идентификация

При решении задач идентификации инструментальные переменные дают добавочную информацию. Инструмент привносит в анализ некоторые дополнительные дисперсии и ковариации, которые можно подвергнуть путевому анализу для дополнительных уравнений идентификации.

5.11.



Без инструмента X имеются только два уравнения путевого анализа, пригодных для нахождения неизвестных системы:

$$\sigma_{YZ} = c\sigma_Y^2 + \sigma_{U(Y)U(Z)} \quad (1)$$

$$\sigma_Z^2 = c^2\sigma_Y^2 + \sigma_{U(Z)}^2 + 2\sigma_{U(Y)U(Z)} \quad (2)$$

Так как эти два уравнения содержат три ненаблюдаемые величины, система недоопределена. Присовокупление инструмента добавляет одну неизвестную величину a , но оно добавляет также три новые наблюдаемые величины и три новых уравнения путевого анализа.

$$\sigma_{XY} = a\sigma_X^2 \quad (3)$$

$$\sigma_{XZ} = ac\sigma_X^2 \quad (4)$$

$$\sigma_Y^2 = a^2\sigma_X^2 + \sigma_{U(Y)}^2 \quad (5)$$

Можно получить следующие формулы идентификации:

$$a = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{из (3),} \quad (6)$$

$$c = \frac{\sigma_{XZ}}{\sigma_{XY}} \quad \text{из (4) и (6),} \quad (7)$$

$$\sigma_{U(Y)U(Z)} = \sigma_{YZ} - \frac{\sigma_{XZ}\sigma_Y^2}{\sigma_{XY}} \quad \text{из (1) и (7),} \quad (8)$$

$$\sigma_{U(Y)}^2 = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \quad \text{из (5) и (6),} \quad (9)$$

$$\sigma_{U(Z)}^2 = \sigma_Z^2 - \frac{\sigma_{XZ}}{\sigma_{XY}} \left(\frac{\sigma_{XZ}}{\sigma_{XY}} - 2\sigma_Y^2 \right) - 2\sigma_{YZ}. \quad (10)$$

Этот инструмент добавляет достаточно уравнений для определения всех неизвестных. Если бы инструмент X коррелировал с U_Y , он все еще давал бы достаточно информации для определения основных параметров в первоначальном рекурсивном отношении, но параметры, связывающие инструмент с другими переменными, не определялись бы однозначно; например, добавление ненулевой ковариации между X и U_Y дает следующие уравнения для ковариаций и дисперсий:

$$\sigma_{XY} = a\sigma_X^2 + \sigma_{XU(Y)}$$

$$\sigma_{XZ} = ac\sigma_X^2 + c\sigma_{XU(Y)}$$

$$\sigma_Y^2 = a^2\sigma_X^2 + \sigma_{U(Y)}^2 + 2a\sigma_{XU(Y)}$$

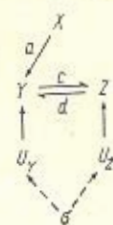
Эти уравнения приводят к тем же самым формулам для нахождения $\sigma_{U(Y)U(Z)}$ и $\sigma_{U(Z)}^2$. Однако коэффициент a и статистические параметры $\sigma_{U(Y)}^2$ и $\sigma_{XU(Y)}$ однозначно определить нельзя.

Добавление инструмента анализа может в достаточной степени увеличивать информацию для определения важных параметров системы, даже если невозможно определить параметры, устанавливающие связь между инструментом и системой.

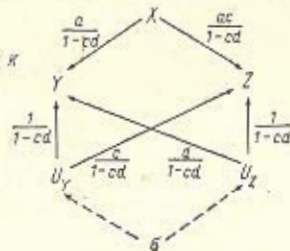
Последняя часть этого примера иллюстрирует важное положение. Коэффициент, связывающий переменную с ее инструментом, нельзя в точности определить, если связь между ними не рекурсивна; иными словами, инструменту, для того чтобы быть полезным, нет надобности быть рекурсивно связанным со своей входной переменной, но его влияние на входную переменную можно правильно оценить, только если отношение между обеими переменными рекурсивно.

Чистый выигрыш информации получается также тогда, когда есть инструмент для одной из связей в петле.

5.12.



можно привести к



Тогда из редуцированной схемы можно получить следующие ковариационные формулы:

$$\sigma_{XY} = \frac{a}{1 - cd} \sigma_X^2, \quad \sigma_{XZ} = \frac{ac}{1 - cd} \sigma_X^2.$$

Деление второго выражения на первое дает формулу для определения коэффициента c :

$$c = \frac{\sigma_{XZ}}{\sigma_{XY}}.$$

Интересно заметить, что тот же самый результат можно получить путем использования регрессии Y по X и регрессии Z по X с последующим выписыванием конечной формулы идентификации через коэффициенты регрессии:

$$c = \frac{\sigma_{XZ}}{\sigma_{XY}} = \frac{\sigma_{XZ}/\sigma_X^2}{\sigma_{XY}/\sigma_X^2} = \frac{b_{ZX}}{b_{YX}}.$$

Обобщенным вариантом этого метода является метод оценки, называемый косвенным методом наименьших квадратов.

Последнее уравнение в 5.12 означает, что если записать уравнения регрессии переменных петли по инструментальным переменным, то коэффициенты петли можно найти по коэффициентам регрессии. Это — главная идея косвенного метода наименьших квадратов. Однако этот метод оценки здесь не развивается, так как у него есть основной практический недостаток. Если для данного отношения имеется более одного инструмента, косвенным методом наименьших квадратов можно получить несколько различных оценок для структурных коэффициентов, но не задана процедура, которая объединит их в одну наилучшую оценку.

К общему методу оценивания неизвестных в нерекursивных соотношениях, даже когда отношение переопределено большим количеством инструментов, чем нужно для единой оценки отдельного коэффициента, приводит другой способ использования инструментальных переменных.

Двухэтапный метод наименьших квадратов

Общий метод оценивания параметров в нерекursивных соотношениях по данным «поперечного» анализа был разработан в 50-х годах специалистом по эконометрии Генри Тейлом (Henry Theil). Этот метод — двухэтапный метод наименьших квадратов — применим, когда возмущения переменных системы коррелированы, он подходит для оценки коэффициентов в петлях и эффективно объединяет информацию от многих инструментов для получения единой оценки каждого структурного коэффициента.

В действительности двухэтапный метод наименьших квадратов включает в себя использование инструментальных переменных в первом цикле проведения множественного регрессионного анализа для определения новых переменных системы, которые свободны от запутывающих влияний со стороны возмущений. Эти

новые переменные используются в конечном цикле проведения регрессионного анализа для оценки структурных коэффициентов, а также дисперсий и ковариаций возмущений. Графическое истолкование двухэтапного метода наименьших квадратов дается позже. Сначала же будут рассмотрены основные шаги, которые можно предпринять для выполнения процедуры.

V.4. Оценка двухэтапным методом наименьших квадратов. Если некоторые из связей между переменной системы и ее причинами нерекursивны, но для них существуют соответствующие инструменты, то можно оценить структурные коэффициенты для операторов, определяющих переменную, следующим образом:

(а) Выписать структурное уравнение, которое выражает значение зависимой переменной как функцию ее непосредственных причин, структурных коэффициентов и члена-возмущения;

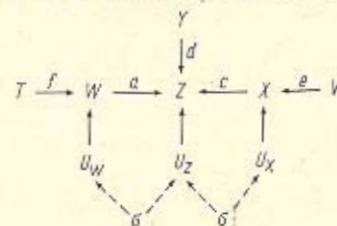
(б) Составить список заранее задаваемых переменных, состоящий из всех рекурсивных причин в уравнении и всех инструментов для нерекursивных причин;

(в) провести регрессию каждой из нерекursивных причин по всем задаваемым заранее переменным для получения множества уравнений регрессии для «предсказания» значений каждой нерекursивной причины по значениям заранее задаваемых переменных;

(г) вернуться к первоначальному множеству наблюдений и в каждом случае вычислить предсказываемое значение для каждой нерекursивной причины, используя формулы, полученные на шаге (в). Эта процедура для каждой из первоначальных нерекursивных переменных-причин порождает по одной новой переменной, называемой «очищенной причиной»;

(д) вернуться к уравнению, определенному на шаге (а), и оценить его коэффициенты обычным методом наименьших квадратов, подставляя очищенные переменные-причины вместо первоначальных нерекursивных причин.

5.13. Задача: оцените a , c и d по наблюдениям T , V , W , X , Y , Z .



Здесь не следует использовать обычный метод наименьших квадратов, потому что X и W — нерекursивные причины для Z : а именно они согласуются с возмущениями переменной Z . Таким образом, мы обращаемся к двухэтапному методу наименьших квадратов.

Шаг 1. Выпишем структурное уравнение для Z :

$$Z = aW + cX + dY + U_Z.$$

Шаг 2. Составим список заранее задаваемых переменных, относящихся к этому уравнению. Этот список включает Y , потому что Y является рекурсивной причиной для Z . Он также включает T и V , потому что они — инструменты для двух нерекурсивных причин Z .

Шаг 3. Проведем регрессию нерекурсивных причин по всем заранее задаваемым переменным.

$$W = b_{WT.VY}T + b_{WV.TY}V + b_{WY.TY}Y + e_W, \quad X = b_{XT.VY}T + b_{XV.TY}V + b_{XY.TY}Y + e_X.$$

Шаг 4. Вернемся к первоначальным данным и образуем очищенные переменные $\hat{W}$ и $\hat{X}$ путем применения следующих формул:

$$\hat{W} = b_{WT.VY}T + b_{WV.TY}V + b_{WY.TY}Y, \quad \hat{X} = b_{XT.VY}T + b_{XV.TY}V + b_{XY.TY}Y.$$

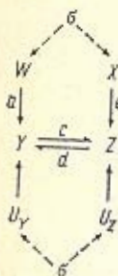
Шаг 5. Изменим уравнение в шаге 1 путем подстановки очищенных переменных вместо первоначальных причин:

$$Z = a\hat{W} + c\hat{X} + dY + U_Z,$$

и оценим структурные коэффициенты обычным методом наименьших квадратов.

Дисперсия возмущения $\sigma_{U(Z)}^2$ оценивается не непосредственно дисперсией остатков в конечной регрессии, а довольно сложным преобразованием остаточной дисперсии. Вычислительные формулы здесь не приводятся.

5.14. Оценка по двухэтапному методу наименьших квадратов коэффициентов в петле



Связи $Y \rightarrow Z$ и $Z \rightarrow Y$ нерекурсивны; однако W и X — инструменты для этих связей.

Шаг 1. Уравнениями для Y и Z являются

$$Y = aW + dZ + U_Y \text{ и } Z = eX + cY + U_Z.$$

Шаг 2. W — рекурсивная причина в первом уравнении и инструмент относительно второго. Аналогично X — рекурсивная причина во втором уравнении и инструмент для первого. Таким образом, W и X составляют множество заранее задаваемых переменных для обоих уравнений.

Шаг 3. Проведем множественные регрессии Y по W и X , а также Z по W и X для получения регрессионных формул:

$$Y = b_{YW.X}W + b_{YX.W}X + e_Y \text{ и } Z = b_{ZW.X}W + b_{ZX.W}X + e_Z.$$

Шаг 4. Для каждого первоначального наблюдения вычисляем предсказываемые значения $\hat{Y}$ и $\hat{Z}$, используя коэффициенты регрессии, полученные на шаге 2. Это определяет две новые переменные:

$$\hat{Y} = b_{YW.X}W + b_{YX.W}X \text{ и } \hat{Z} = b_{ZW.X}W + b_{ZX.W}X.$$

Шаг 5. Используем обычный метод наименьших квадратов для оценки операторов в каждом структурном уравнении. Всегда проводим регрессию для $\hat{Y}$

или $\hat{Z}$ вместо Y или Z . Таким образом, для оценки коэффициентов в структурном уравнении

$$Y = aW + dZ + U_Y$$

проводим следующую регрессию:

$$Y = b_{YW.\hat{Z}}W + b_{Y\hat{Z}.W}\hat{Z} + e'_Y.$$

Получившиеся в результате коэффициенты являются несмещенными оценками соответствующих коэффициентов оператора, а именно: a оценивается величиной $b_{YW.\hat{Z}}$, d оценивается величиной $b_{Y\hat{Z}.W}$.

Аналогично для оценки коэффициентов в $Z = eX + cY + U_Z$ проведем регрессию Z по X и $\hat{Y}$; тогда e оценивается как $b_{ZX.\hat{Y}}$; c оценивается как $b_{Z\hat{Y}.X}$.

Вычислительные процедуры

Процедура для оценки по двухэтапному методу наименьших квадратов, определяемой правилом V.4, требует составления двух отдельных множеств регрессионных уравнений плюс промежуточного шага, на котором мы возвращаемся к первоначальным данным для создания «очищенных» переменных. В действительности этот промежуточный шаг можно довести до конца без обращения к первоначальным наблюдениям путем использования тех процедур, которые даны в упражнении 6 этой главы.

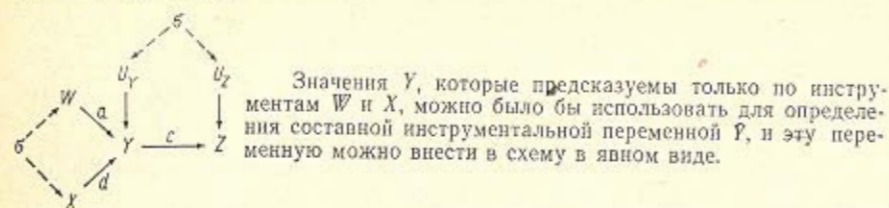
Есть математические формулы, которые определяют двухэтапный метод наименьших квадратов как одношаговую аналитическую процедуру. Матричные уравнения чересчур сложны для того, чтобы их здесь представить, и слишком сложны для использования их при вычислениях вручную. Однако многие вычислительные машины содержат программу двухэтапного метода наименьших квадратов, которая включает математическое решение и обычно может вызываться для работы с различными задачами.

Использование машинных программ двухэтапного метода наименьших квадратов рекомендуется с точки зрения экономии и большей точности. Вдобавок эти программы обычно вычисляют статистики для проверки статистической значимости оценок коэффициентов, когда данные состоят из наблюдений на выборке, а не на всей совокупности. Кроме того, программы ЭВМ обычно предусматривают оценку по двухэтапному методу наименьших квадратов дисперсий и ковариаций возмущений. Формулы для вычисления этих величин здесь не приводятся ввиду их сложности.

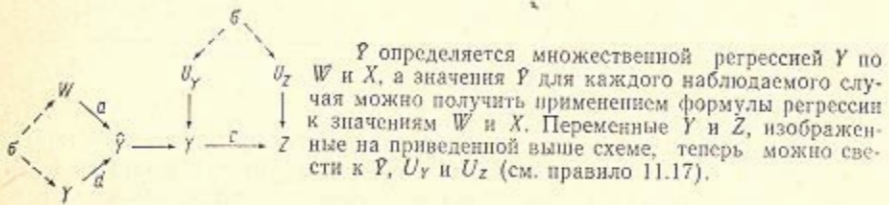
Графическое истолкование двухэтапного метода наименьших квадратов

Граф системы можно преобразовать в граф, соответствующий процедурам в анализе двухэтапным методом наименьших квадратов. Такие графические манипуляции полезны для понимания основной логики этого метода.

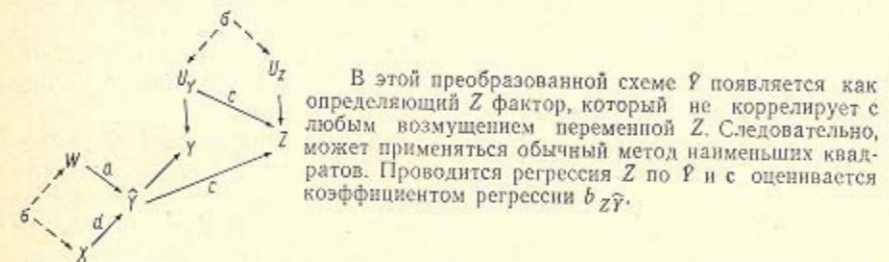
5.15. Задача: Получите единую несмещенную оценку коэффициента при данных инструментах W и X .



Значения Y , которые предсказуемы только по инструментам W и X , можно было бы использовать для определения составной инструментальной переменной P , и эту переменную можно внести в схему в явном виде.



P определяется множественной регрессией Y по W и X , а значения P для каждого наблюдаемого случая можно получить применением формулы регрессии к значениям W и X . Переменные Y и Z , изображенные на приведенной выше схеме, теперь можно свести к P , U_Y и U_Z (см. правило 11.17).



В этой преобразованной схеме P появляется как определяющий Z фактор, который не коррелирует с любым возмущением переменной Z . Следовательно, может применяться обычный метод наименьших квадратов. Проводится регрессия Z по P и с оценивается коэффициентом регрессии b_{ZY} .

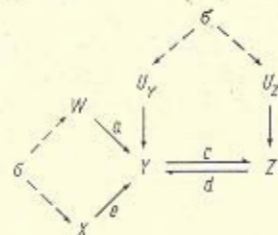
Основной идеей, проявляемой графическим подходом, является то, что значения P представляют собой экстракт действительных колебаний Y , которые ни в каком отношении не испорчены возмущениями переменной Z или коррелируют с ними. Следовательно, связь между Z и этой «очищенной» Y рекурсивна, а коэффициент c можно оценить обычным методом наименьших квадратов.

Этот пример с двумя инструментальными переменными для связи $Y \rightarrow Z$ представляет особый интерес, потому что связь $Y \rightarrow Z$ переопределена. Из первоначального графа можно выписать семь уравнений для выражения значений наблюдаемых ковариаций и дисперсий через неизвестные и наблюдаемые дисперсии переменных W и X . Однако имеется только шесть неизвестных — три структурных коэффициента и три параметра возмущения. Если бы семь уравнений были решены относительно шести неизвестных, получились бы две различные формулы оценивания для коэффициента c . Маловероятно, чтобы две эти формулы всегда давали бы в точности один и тот же результат, если бы мы работали с реальными данными и имела бы задача соединения различных оценок коэффициента c в единое наилучшее значение. Двухэтапный метод наименьших квадратов решает задачу переопределен-

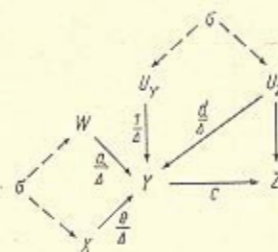
ности созданием единственной составной инструментальной переменной P , которая оптимальным образом объединяет дисперсию, выделенную из Y , используя каждый инструмент.

Оценку коэффициентов петли двухэтапным методом наименьших квадратов можно также представить графически, если интересующая зависимая переменная выводится первой (правило II.19).

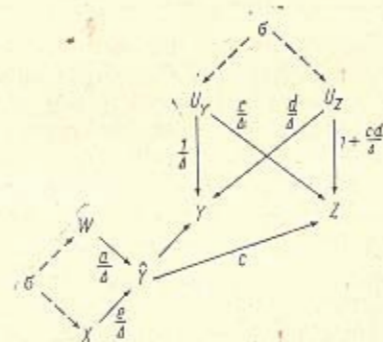
5.16. Задача: получите единую несмещенную оценку коэффициента петли с при данных инструментах W и X .



Сначала из петли выводится переменная Z (по правилу II.19):



где $\Delta = 1 - cd$. Затем проводят преобразование, как в предыдущем примере, что дает



Опять регрессия Y по W и X определяет переменную P , а регрессия Z по P оценивает коэффициент c .

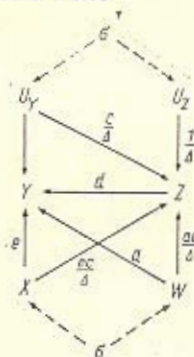
Как показывает пример, выведение (высвобождение) переменных в системе с петлями превращает ее в схему без петель, в которой определяющие связи все еще остаются нерекурсивными.

Таким образом, задача оценивания коэффициента петли по своей природе не отличается от задачи оценивания коэффициента в других типах нерекурсивных связей, хотя при графическом подходе она включает в себя дополнительный шаг вывода.

Задачи идентификации

Петля действительно создает дополнительные трудности путем усложнения задачи идентификации; например, коэффициент c в 5.16 идентифицировался при данных инструментах W и X . В действительности он был переопределен, потому что имелось два инструмента для одной связи $Y \rightarrow Z$. Однако по-прежнему невозможно найти оператор d в системе, определенной в 5.16. Это можно увидеть путем вывода Y вместо Z .

5.17. Выведение Y из системы петли в 5.16



Теперь Z явно коррелирует с U_1 и d нельзя найти обычным методом наименьших квадратов. Кроме того, для связи $Z \rightarrow Y$ нет инструмента; следовательно, невозможно образовать «очищенную» переменную Z , которая может использоваться для оценки d множественной регрессией Y по Z , X и W . Невозможно использовать X и W для создания пригодной Z , потому что обе переменные оказывают непосредственные воздействия на Y . Конкретнее, коэффициент $b_{YZ.WX}$, который, по-видимому, оценивает d при окончательном анализе обычным методом наименьших квадратов, может рассматриваться как коэффициент, полученный проведением регрессии остатков переменной Y (устраняющих все зависимости от X и W) по остаткам переменной Z (устраняющим все зависимости от X и W). Однако, так как Z создается только из X и W , по этим двум переменным она предсказывается полностью, и после ее предсказания по W и X не существует никаких остатков переменной Z . Следовательно, коэффициент $b_{YZ.XW}$ оказывается неопределенным и d не идентифицируется.

Здесь простая идея подсчета уравнений и неизвестных как основа для решения вопроса, идентифицируема ли система, оказалась неудовлетворительной. В приведенном выше примере имеются семь определенных уравнений и всего семь неизвестных, но так как имеется два лишних инструмента для связи $Y \rightarrow Z$ и ни одного для $Z \rightarrow Y$, можно найти только один из операторов петли. Система в целом остается недоопределенной.

Кроме того, теперь ясно, что двухэтапная процедура наименьших квадратов может оказаться неудовлетворительной, когда нет инструментов для нерекурсивной связи. Применимость правила V.4 обязательно должна ограничиваться другим правилом, которое определяет, при каких условиях можно найти коэффициенты в уравнении.

Идентифицируемость

V.5. Если переменная входит в петлю, а ее возмущения коррелируемы со значениями ее причин, то все коэффициенты в ее структурном уравнении определяются, только если имеют силу условия (а) и (б):

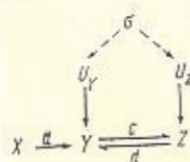
(а) для каждой нерекурсивной связи, представленной в уравнении, имеется по крайней мере один инструмент;

(б) если уравнение включает несколько, скажем M , нерекурсивных причин, то должно быть по крайней мере M различных инструментов для связей между зависимой переменной и ее причинами. (Каждая из инструментальных переменных может служить в качестве инструмента более чем для одной из M связей.)

Если каждая нерекурсивная причина связана с переменной, которая служит инструментом для связей, включающих только эту причину, то структурные коэффициенты всегда теоретически идентифицируемы.

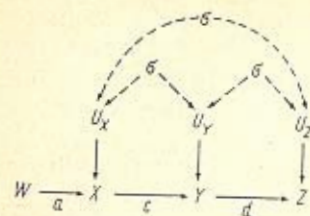
В сущности, условие (а) гласит, что коэффициент в отдельной нерекурсивной причинной связи определяется только в том случае, если для этой отдельной связи имеется по крайней мере один инструмент.

5.18.



Коэффициент c определяется, потому что имеется инструмент (X), который применяется особо к связи $Y \rightarrow Z$. Коэффициент d не определяется, потому что нет инструмента для связи $Z \rightarrow Y$.

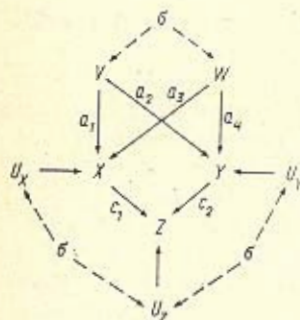
Обратите внимание на то, что один и тот же инструмент может служить несколько раз при анализе различных уравнений.



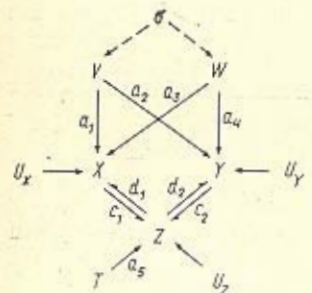
Все коэффициенты математически определены, потому что связь $W \rightarrow X$ рекурсивна, и a определяется по обычному коэффициенту регрессии; $W \rightarrow Y$ — инструмент для связи $X \rightarrow Y$ и c определяется; W также инструмент для $Y \rightarrow Z$ и d определяется.

Другими словами, для каждой нерекурсивной связи в системе необходимо иметь *отдельный* инструмент.

Условие (б) ослабляет требования еще больше, если зависимая переменная включается в несколько нерекурсивных связей. Должен быть инструмент для каждой связи, но не требуется, чтобы каждый инструмент применялся только к одной из них.



V является инструментом для связей $X \rightarrow Z$ и $Y \rightarrow Z$. Точно так же W — инструмент для обеих рекурсивных связей. Тем не менее условие (б) выполняется и обычно возможно оценить все коэффициенты: все a — обычным методом наименьших квадратов, а все c — двухэтапным методом наименьших квадратов.

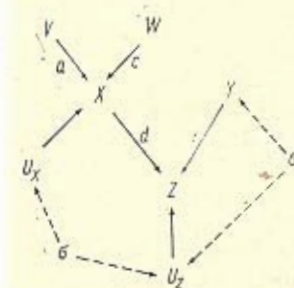


Это изменение приведенной выше ситуации, и все коэффициенты определяются обычным образом. V и W — инструменты, которые определяют все c , а T — инструмент, который определяет все d . Все a также находятся через статистику, полученные путем многократного проведения анализа двухэтапным методом наименьших квадратов.

Далее следуют примеры, в которых одно из необходимых для идентифицируемости условий выполняется, а другое — нет.



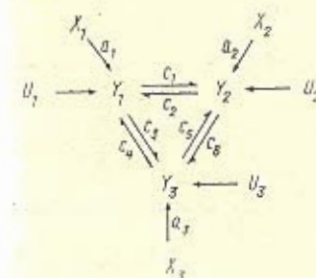
W служит инструментом для обеих нерекурсивных связей $X \rightarrow Z$ и $Y \rightarrow Z$ и условие (а) выполняется, но вследствие того, что для двух нерекурсивных причин переменная Z имеет только один инструмент, условие (б) не выполняется.



Z входит в две нерекурсивные связи и имеют два инструмента V и W ; следовательно, условие (б) выполняется. Однако нет инструмента для связи $Y \rightarrow Z$ и условие (а) не выполняется.

Условия (а) и (б) в правиле V.5 необходимы для идентифицируемости. Все коэффициенты в нерекурсивной системе нельзя определить, если не выполняются эти условия, но они не являются достаточными условиями. Действительно, встречаются случаи, в которых выполняются условия (а) и (б) и тем не менее некоторые структурные коэффициенты совсем нельзя оценить или же оценки содержат такие большие ошибки, что оказываются практически бесполезными; например, это было бы верно, если случайно оказалось, что $a_1 = a_3$ и $a_2 = a_4$ в 5.20.

Последняя часть правила определяет достаточное условие для идентифицируемости. Если каждая нерекурсивная причина является «входным пунктом» для инструмента, который не имеет других входных пунктов, то все структурные коэффициенты определяются математически.



Все коэффициенты определяются математически, потому что с каждой из переменных в совокупности петля связывается единственный инструмент.

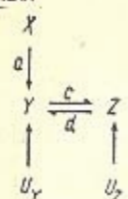
Действительно, это условие *более чем* достаточно. Оно определяет скорее полезную цель при теоретическом описании системы, а не практическую необходимость, потому что бывают случаи (особенно в больших системах), когда все коэффициенты в блоке могут определяться без выполнения этого условия. (Точнее, в эконометрических руководствах обыкновенно представлено «как раз достаточно» условие идентифицируемости, но его формулировка требует знакомства с матричной алгеброй.)

Методы полной информации

Обычный и двухэтапный методы наименьших квадратов являются «методами оценки для одного уравнения»; это подразумевает, что они включают одновременную оценку коэффициентов только в одном структурном уравнении, в противоположность методам полной информации, которые оценивают сразу все коэффициенты в системе. В то время как «методы для одного уравнения» зависят от прослеживания причинных путей только одной или двух петель от зависимой переменной, методы «полной информации» используют информацию о продолженных причинных цепях. Это является одним из главных преимуществ методов полной информации. Они эффективнее используют справедливую теорию, в буквальном смысле включая в оценку каждого параметра системы всю такую информацию, так что оценка оказывается более точной. Главным недостатком методов полной информации является то, что они аналогично включают в расчеты любую ошибочную теорию. Если часть системы описана неправильно, это может затрагивать все оценки коэффициентов, а не только несколько из них, как в методе для одного уравнения. Следовательно, вообще методы полной информации наиболее подходящим образом применяются на последних стадиях исследования, когда имеется высокая степень уверенности в теоретическом описании системы.

Методы полной информации могут также использовать априорное знание о дисперсиях возмущений, что увеличивает способность идентификации. Следующий простой пример иллюстрирует эту логику.

5.23.



Оба коэффициента в петле определяются с единственным рекурсивным инструментом, если известно, что возмущения переменных петли имеют нулевую ковариацию.

Используя инструмент X , мы определяем коэффициент c , как показано в 5.12:

$$c = \frac{\sigma_{XZ}}{\sigma_{XY}}$$

Так как X , U_Y и U_Z некоррелируемы, ковариация σ_{YZ} может определяться исключительно через дисперсии переменных Y и Z . Алгебраические выкладки, показывающие это, здесь не приводятся, но результатом оказывается следующее выражение:

$$\sigma_{YZ} = \frac{c\sigma_Y^2 + d\sigma_Z^2}{1 + cd}$$

Подстановка $(\sigma_{XZ}/\sigma_{XY})$ вместо c и разрешение относительно d дает

$$\sigma_{YZ} = \frac{(\sigma_{XZ}/\sigma_{XY})\sigma_Y^2 + d\sigma_Z^2}{1 + (\sigma_{XZ}/\sigma_{XY})d},$$

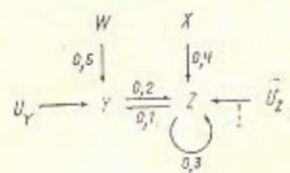
$$d = \frac{\sigma_Y^2\sigma_{XZ} - \sigma_{YZ}\sigma_{XZ}}{\sigma_{YZ}\sigma_{XZ} - \sigma_Z^2\sigma_{XY}}$$

Однако при многих обычных приложениях методы полной информации применяются только для улучшения эффективности оценки, а не для достижения дополнительной способности идентификации. Действительно, одна из процедур — трехэтапный метод наименьших квадратов — включает в себя анализ двухэтапным методом наименьших квадратов в качестве предварительного для его заключительного анализа. Обычно это просто дает более эффективные оценки коэффициентов, которые были оценены двухэтапным методом наименьших квадратов. Аналитические процедуры методов полной информации сложны и могут сжато описываться только с помощью математики более высокого уровня, чем употребляемая в этой книге. Подробные обсуждения этих методов даются руководствами по эконометрике.

Петли с одной вершиной (простые петли)

Не существует никакой процедуры определения коэффициентов для простых петель по данным «поперечного» анализа. Кроме того, когда предполагается, что «простые петли» отсутствуют, если в действительности они присутствуют, то это обязательно влияет на определение некоторых других коэффициентов, как показано в 5.24.

5.24. Влияние простой петли на идентификацию

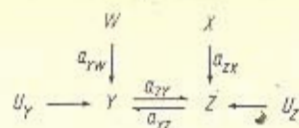


Предположим для удобства, что $\sigma_W^2 = \sigma_X^2 = 1,0$, $\sigma_{WX} = 0$.

В этом случае

$$\sigma_{WY} = 0,515, \quad \sigma_{WZ} = 0,145, \quad \sigma_{XY} = 0,059, \quad \sigma_{XZ} = 0,588.$$

Теперь предположим, что система неправильно описана, как на схеме



Оценками этих коэффициентов двухэтапным методом наименьших квадратов были бы

$$a_{ZY} = 0,281, \quad a_{YZ} = 0,100, \quad a_{ZX} = 0,571, \quad a_{YW} = 0,501.$$

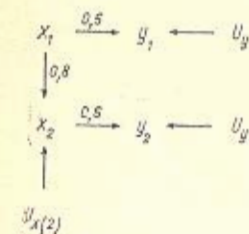
Структурные коэффициенты, определяющие Y, оцениваются точно (в пределах ошибки округления), но структурные коэффициенты, влияющие на Z, слишком велики. Фактически a_{ZY} и a_{ZX} имеют такое систематическое смещение, что для получения правильных значений они должны умножаться на разницу коэффициентов «простой петли» (т. е. $[1 - 0,3] = 0,7$).

Как подсказывает пример, существование неописанной «простой петли» на переменной приводит к смещенным оценкам для коэффициентов в структурном уравнении этой переменной. Однако получающиеся в результате ошибки относительно малы, потому что оцениваемые коэффициенты находятся в надлежащих пропорциях относительно друг друга и действительно точно воспроизводят все статические связи в системе.

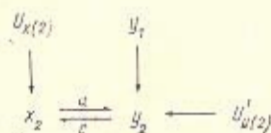
Запаздывающие переменные как инструменты

Более ранние измерения переменной иногда включаются в описание системы в качестве вспомогательного средства для идентификации при предположении, что переменная, измеренная ранее, — инструмент для той же самой переменной, измеренной в более поздний момент. Этот образ действий может приводить к ложным выводам, если стабильность основной переменной поддерживается главным образом стабильностью в других переменных системы, как показано в 5.25.

5.25.



Нижние индексы y и x указывают момент измерения; 0,8 — коэффициент стабильности для переменной x на всем промежутке времени: отсутствие пути из y_t в y_{t+1} показывает отсутствие стабильности в y , не считая стабильности, возникающей из ее зависимости от x . Теперь предположим, что связь между x и y неизвестна и y_t используется как инструмент для определения x на следующей схеме



$\rho_{y(1)y(2)} = 0,2$, $\rho_{y(1)x(2)} = 0,4$. Значение c могло быть оценено как $(0,4/0,2) = 2,0$, что является серьезной ошибкой (правильное значение — нуль). Этой ошибки можно было бы избежать путем включения в анализ x_1 , и в этом случае y_1 оказалась бы неадекватным инструментом для связи $y_2 \rightarrow x_2$, потому что она не имеет корреляции с y_2 при управлении значением переменной x_1 .

Главный урок здесь состоит в том, что запаздывающие значения переменной обычно не могут использоваться для определения ее связей с другими переменными. В частности, мы должны знать заранее, что узловая переменная — это причина для других или что запаздывающая переменная не может быть осмысленным инструментом. Если это неизвестно, то на всех переменных должны проводиться измерения в дополнительное время, для того чтобы извлечь информацию о структуре системы из временного согласования. Такой «продольный анализ» является объединением «поперечного» подхода, который мы уже обсуждали, и анализа временных рядов, обсуждаемого в гл. 6.

ФАКТОРЫ, ВЛИЯЮЩИЕ НА ОЦЕНКИ

Ошибка выбора

Многих сложностей можно избежать, предполагая, что данные собираются для каждого случая в интересующей совокупности. В практических задачах, однако, мы, как правило, рассматриваем выборку случаев из совокупности. Дисперсии и ковариации выборки должны, вероятно, чем-то отличаться от истинных значений в совокупности вследствие особенностей в рассматриваемом определенном подмножестве случаев. Подстановка ошибочных дисперсий и ковариаций в формулы идентификации тоже, естественно, дает до некоторой степени ошибочные оценки структурных коэффициентов.

Выборка из трех случаев с высоким правдоподобием окажется характерной и нерепрезентативной для совокупности. Например, все три случая могут оказаться выше среднего из совокупности по одной переменной и ниже его по другой. У выборки из 10 случаев меньше шансов оказаться особенной, ибо менее вероятно, что так много случаев сместятся в одних и тех же направлениях. При переходе к выборкам, состоящим из сотен и тысяч, возможности характерности и нерепрезентативности уменьшаются еще больше (и это верно, независимо от того, составляет ли размер совокупности 10 тысяч, миллион или миллиард). Таким образом, чем больше выборка, тем вероятнее, что ее статистические характеристики будут соответствовать статистическим характеристикам совокупности в целом. Следовательно, ожидаемая ошибка оценок параметров размеренно убывает с увеличением объема выборки. Минимальный полезный размер выборки зависит от того, насколько большую ошибку в оценке мы хотим допустить, от числа переменных, от силы связей в системе, от того, какими методами оценки пользуются, и от некоторых других факторов, рассматри-

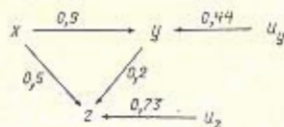
ваемых ниже. Когда имеют дело с социальными системами, прежде чем оценки будут обладать большим правдоподобием, обычно необходима выборка из 100 случаев; число должно быть больше, если система велика, если результаты измерений ничтожны и корреляции слабы и если оценка производится двухэтапным методом наименьших квадратов, а не обычным. Есть ситуации, в которых выборка объемом меньше 100 случаев может давать интересную информацию, но обычно для получения ценных результатов нужна выборка большего объема, даже гораздо большего.

Вдобавок выборка должна извлекаться без смещений, которые могли бы создать неслучайные особенности. Соответствующие выборочные методы и другие методологические аспекты выбора обсуждаются в руководствах по этому вопросу (см. библиографию к настоящей главе).

Сила связей

Структура системы может затруднить точную оценку параметров, если нет больших выборок; например, следующая простая система (все показанные переменные стандартизованы) порождает проблемы при точной оценке некоторых из ее коэффициентов по выборочным данным.

5.26.



Истинными корреляциями являются

$$\rho_{xy} = 0,90, \quad \rho_{xz} = 0,68, \quad \rho_{yz} = 0,65.$$

Множеством гипотетических выборочных корреляций могло бы быть

$$r_{xy} = 0,91, \quad r_{xz} = 0,70, \quad r_{yz} = 0,62.$$

Формулой для определения r_{zy} служит

$$r_{zy} = \frac{\rho_{yz} - \rho_{xy}\rho_{xz}}{1 - \rho_{xy}^2}.$$

При применении к истинным корреляциям она дает

$$\frac{0,65 - (0,90)(0,68)}{1 - (0,90)^2} = 0,2.$$

При применении к выборочным корреляциям формула дает

$$\frac{0,62 - (0,91)(0,70)}{1 - (0,91)^2} = -0,1.$$

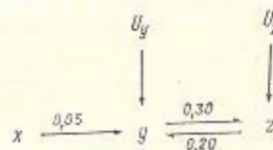
Таким образом, в этом случае выборочные данные дают неточную оценку значения r_{zy} , указывая на отрицательную связь, хотя истинная связь положительна.

Ошибочный результат в этом примере в большой степени обусловлен высокой степенью *коллинеарности* между переменными, определяющими z ($\rho_{xy} = 0,90$). Она преувеличивает влияние ошибок в выборочных корреляциях (которые не являются нетипичными среди ошибок, полученных даже в довольно больших выборках).

Проблема коллинеарности существует, когда корреляции между двумя или более переменными, определяющими зависимую переменную, гораздо выше, чем корреляции между зависимой и определяющими переменными. Точное оценивание структурных коэффициентов тогда требует, чтобы оценки истинных корреляций имели маленькую ошибку, что предполагает необходимость чрезвычайно больших выборок.

Сила связей также влияет на точность оценок, основанных на инструментальных переменных, как показано в 5.27.

5.27.



Истинные корреляции — $\rho_{xy} = 0,053$, $\rho_{xz} = 0,016$, $\rho_{yz} = 0,471$. Множеством «наблюдаемых» корреляций, основанных на гипотетической выборке, является

$$r_{xy} = 0,068, \quad r_{xz} = 0,005, \quad r_{yz} = 0,463.$$

Формулой для определения r_{zy} , использующей x в качестве инструмента, служит (из 5.11)

$$r_{zy} = \frac{\rho_{xz}}{\rho_{xy}}.$$

При применении к истинным корреляциям она дает 0,30 (в пределах ошибки округления). При применении к гипотетической выборочной корреляции она дает

$$\frac{r_{xz}}{r_{xy}} = \frac{0,005}{0,068} = 0,074.$$

Следовательно, слабый инструмент в соединении даже с относительно малыми ошибками выборки в значениях корреляций может давать неточную оценку структурного коэффициента.

Проблема слабого инструмента существует всегда, когда инструментальная переменная имеет близкую к нулю корреляцию со своей зависимой переменной (в приводимом выше примере эта корреляция как раз составляет 0,068); выборочные ошибки в наблюдаемых корреляциях увеличиваются в процессе оценивания параметров и оценки ненадежны. Если в задаче есть лишь один слабый инструмент, точную оценку параметра можно получить только путем использования настолько большого числа случаев, что выборочные ошибки будут малы.

Если для той же самой связи есть дополнительные инструменты, они могут рассматриваться совокупно как множество в процедуре двухэтапного метода наименьших квадратов, а мерой достаточности служит множественная корреляция (R) между инструментами и зависимой переменной. Иногда возможно сделать R достаточно большим путем использования некоторого количества инструментов, каждого из которых в отдельности было бы недостаточно.

Неточность измерения

Даже если рассматривается целая совокупность, можно не получить истинных значений дисперсий и ковариаций, если переменные измеряются неточно. Следовательно, эмпирические оценки параметров могут быть точными только в тех случаях, когда точны измерения.

Анализ проблемы измерений обычно начинают предположением, что ошибки измерений являются непредсказуемыми (случайными) возмущениями наблюдаемых значений переменной. Эту модель легко преобразовать в путевую схему.

5.28.

$$X \rightarrow \tilde{X} \leftarrow e$$

X представляет истинные измерения; $\tilde{X}$ представляет наблюдаемые измерения, получаемые на действующей шкале; e представляет совокупность случайных ошибок измерений. Обычно предполагается, что X и e некоррелированы.

В стандартизованном виде

$$x \rightarrow \tilde{x} \leftarrow \sqrt{1-\rho^2} e.$$

Коэффициент обоснованности измеряет корреляцию между истинными положениями на шкале и наблюдаемыми.

Причинную интерпретацию задачи измерения можно объединить с обычным анализом системы, как показано в простом примере в 5.29.

5.29.



Обе переменные предполагаются измеряемыми с ошибкой, а ошибки предполагаются некоррелированными. Обычно для определения a мы можем использовать формулу из 5.1:

$$a = \frac{\sigma_{XY}}{\sigma_X^2}.$$

Путевой анализ показывает, что наблюдаемая ковариация переменных X и Y равна ковариации истинных положений на шкале.

$$\sigma_{\tilde{X}\tilde{Y}} = \sigma_{XY} = a\sigma_X^2.$$

Однако дисперсия переменной X не оценивается точно эмпирическими результатами измерений:

$$\sigma_{\tilde{X}}^2 = \sigma_X^2 + \sigma_e^2.$$

Следовательно, если бы мы подставили наблюдаемые величины, основанные на ненадежных измерениях, в формулу идентификации, мы бы получили

$$\frac{\sigma_{\tilde{X}\tilde{Y}}}{\sigma_{\tilde{X}}^2} = \frac{\sigma_{XY}}{\sigma_X^2 + \sigma_e^2}.$$

Здесь мы бы недооценили величину коэффициента a тем больше, чем больше ошибки при измерении X . Обратите внимание на то, что ошибки в измерении зависимой переменной не влияют на точность оценивания нестандартизованного структурного коэффициента. Однако такие ошибки смещали бы оценку путевого коэффициента, потому что он стандартизуется на основе и независимой, и зависимой переменных.

Оценки параметров, сделанные по неточным результатам измерений, обычно смещены — проблема, которую никогда нельзя решить только путем увеличения объема выборки. В сущности, неточность измерения — аспект задачи идентификации. Для получения несмещенных оценок параметров мы должны оценивать дисперсии ошибок измерений в дополнение ко всем другим параметрам в системе, и обычно нет достаточной информации.

При решении задачи измерения должны использоваться инструментальные переменные или множественные индикаторы.

5.30.



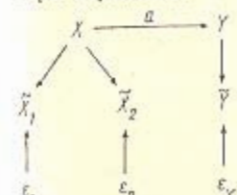
Можно было бы воспользоваться примером в 5.29 путем добавления инструмента — даже такого, который сам измеряется неточно.

С помощью обычной формулы для оценки по инструменту (5.12) мы имеем

$$\frac{\sigma_{\tilde{Y}\tilde{W}}}{\sigma_{\tilde{X}\tilde{W}}} = \frac{ca\sigma_W^2}{c\sigma_W^2} = a.$$

Таким образом, оценка параметра a , полученная с применением инструмента, — несмещенная.

Вместо этого можно иметь дело с задачей, получая дополнительные индикаторы переменной X .



Для удобства при иллюстрации предполагается, что два индикатора эквивалентны в отношении единиц шкалы, так что оба связаны с X коэффициентом 1.0. Теперь несмещенную оценку дисперсии переменной X можно получить из формулы путевого анализа для ковариации индикаторов

$$\sigma_{12} = \sigma_X^2.$$

Это можно применить для получения двух оценок параметра a :

$$\frac{\sigma_{1\tilde{Y}}}{\sigma_{12}} \quad \text{или} \quad \frac{\sigma_{2\tilde{Y}}}{\sigma_{12}}.$$

Эффективные процедуры объединения таких множественных оценок даны в более специальных источниках.

Теоретически для получения несмещенной оценки параметра в ситуации, подобной приведенной выше, надо добавить только одну переменную (т. е. один инструмент или один дополнительный индикатор). Однако самые практические задачи запутываются выбором, как и ошибками измерений, и использование одного слабого инструмента или только двух слабых индикаторов приводило бы к увеличению выборочных ошибок в вычислениях, таким образом вызывая конечные оценки, которые слишком неустойчивы для того, чтобы иметь значимость. Следовательно, проблему измерения надо атаковать путем включения в исследовательские работы множественных инструментов или множественных индикаторов. Анализ должен быть достаточно сложным для включения всей информации; например, применение инструментов ведет прямо к анализу двухэтапным методом наименьших квадратов (или эквивалентной процедуре), а употребление множественных индикаторов вызывает другие многомерные методы, такие, как факторный или канонический анализ (смотрите ссылки).

Как показано выше, можно иногда выходить за пределы задачи измерения с помощью инструментов и двухэтапного метода наименьших квадратов. И обратно, оценки коэффициентов, основанные на анализе двухэтапным методом наименьших квадратов, — обычно несмещенные ошибками измерения при условии, что ошибки измерения не коррелированы через переменные. Таким образом, инструментальные переменные являются сильными инструментами исследования, которые дают несмещенные оценки коэффициентов для сложных и неполно измеряемых систем.

Ошибки описания

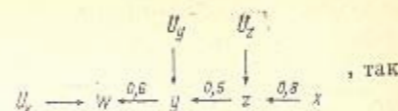
Для обоснованного применения методов оценки, обсуждавшихся в предыдущих разделах, требуются определенные модели причинного устройства. Для обычного метода наименьших квадратов требуются рекурсивные связи; для двухэтапного метода наименьших квадратов нужны обоснованные инструментальные переменные. Ошибка описания происходит, если определенная модель предполагается существующей тогда, когда ее нет. Ошибки описания, которые неверно определяют причинные очередности или ложно исключают корреляции возмущений, могут приводить к серьезным искажениям оценок коэффициентов и к серьезным ошибкам в понимании системы.

5.31. Предположим, что истинной системой является $x \xrightarrow{0,8} y \leftarrow U_y$, так что $p_{yx} = 0,8$ и $p_{xy} = 0$. Если ошибочно предполагается, что y — предшествующая для x , то обычный метод наименьших квадратов приводит к следующим результатам:

$$p_{yx} = 0 \text{ (по предположению)} \text{ и } p_{xy} = 0,8.$$

Эти результаты — серьезная ошибка, потому что они подразумевают причинную модель, прямо противоположную существующей.

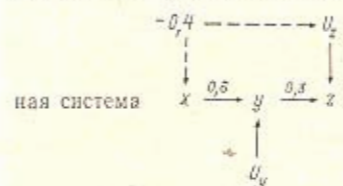
Предположим, что истинная система



что $p_{zy} = 0$ и $p_{yz} = 0,5$. Если ошибочно предполагают, что w — инструмент для $y \rightarrow z$, то двухэтапный метод наименьших квадратов приводит к следующим результатам:

$$p_{zy} = 0,21 \text{ и } p_{yz} = 0,35.$$

Эти результаты являются серьезной ошибкой, потому что подразумевается петля в то время, когда никакой петли не существует. Предположим, что истинная система



$$p_{yx} = 0,6 \text{ и } p_{zy} = 0,3.$$

Если предполагают, что $\sigma_{xU_z} = 0$ вместо $-0,4$, то применение обычного метода наименьших квадратов дает следующие оценки для структурных коэффициентов:

$$p_{yx} = 0,6 \text{ и } p_{zy} = 0,06.$$

Если бы x использовался как инструмент в двухэтапном методе наименьших квадратов для учета возможной корреляции между U_y и U_z (при предположении, что σ_{xU_z} является нулем), оценками были бы

$$p_{yx} = 0,6 \text{ и } p_{zy} = -0,37.$$

Результаты обычного метода наименьших квадратов являются серьезной ошибкой. Результаты двухэтапного метода наименьших квадратов серьезно вводят в заблуждение.

Обычный метод наименьших квадратов не дает осмысленных результатов, если переменные не входят в анализ в соответствии с их истинными причинными очередностями. Результаты анализов двухэтапным методом наименьших квадратов неточны, если переменные, используемые в качестве инструментов, в действительности не удовлетворяют требованиям для инструментальных переменных.

Трудности при использовании в социальном исследовании

Интерес при исследовании часто концентрируется на системе, отражающей суть дела, которая оказывается теоретически и практически наиболее подходящей. Поначалу получение некоторых эмпирических данных и оценка желаемых параметров системы может показаться простым делом. В действительности вопрос обычно не столь прост. Проблема идентификации системы разрастается как снежный ком, ведя от одного к другому, так что для возможности достижения скромных результатов может

оказаться необходимым создать тщательно разработанные вспомогательные средства исследования.

Невключение в описание системы основных переменных системы создает коррелированные возмущения, которые исключают рекурсивность системы и, возможно, нарушают обоснованность инструментов. Во избежание смещенных оценок, которые могут обнаружиться из-за коррелированных возмущений, в анализ должны вводиться либо главные переменные, либо другие переменные, которые могут служить в качестве инструментов для «неочищенных» отношений. В обоих случаях может оказаться необходимым анализировать больше переменных, чем считалось подходящим первоначально. Точно так же коэффициенты в петлях могут определяться, только если имеется достаточно адекватных инструментов. Это может потребовать расширения системы до включения в нее переменных, теоретически и практически малоинтересных, включаемых просто потому, что они отвечают нуждам исследования. Так как социальные переменные редко измеряются точно каким-нибудь одним показателем, в социальном исследовании, как правило, существует задача измерения. Это, может быть, больше, чем другие факторы, требует достаточного расширения числа переменных, которые должны рассматриваться. Если измерения истинны, но все еще обладают довольно большой достоверностью, то для каждой причинной переменной в системе нужны по крайней мере два измерения до того, как появится надежда получить несмещенные оценки параметров системы (а именно каждая переменная должна соединяться с двумя индикаторами или с одним индикатором и одним инструментом). В более вероятной ситуации, когда меры имеют среднюю или низкую степень надежности, чтобы получить надежную оценку параметров системы, для каждой причинной переменной могут потребоваться многочисленные индикаторы и инструменты.

Таким образом, количественные исследования даже довольно простых социальных систем стремятся разрастись в сложные исследовательские программы, которые требуют для получения результатов подробного планирования, собирания больших количеств данных и использования сложных статистических методов.

НУЛЕВЫЕ КОЭФФИЦИЕНТЫ

Анализ может обнаружить равенство нулю некоторых коэффициентов в структурном уравнении; это означает, что какое-то из возможных причинных отношений в системе на самом деле не реализуется в рассмотренной совокупности. Такой причинный вывод вводит в действие правило 1.5, которое утверждает, что одно событие не вызывает другого, если осуществления первого не предполагают осуществлений второго. Нулевой коэффициент означает, что изменения в данной причине не влекут изменений в определенном исходе. Следовательно, можно заключить, что первая переменная не является непосредственной причиной второй.

При данных результатах эмпирического исследования действительно могут быть некоторые сомнения в том, является ли нулем истинное значение коэффициента. Оценки коэффициента сами подвержены ошибке (как обсуждалось выше), и нулевое действие одной переменной на другую может оказаться связанным с оценкой коэффициента, который по значению только близок к нулю, а не является в точности нулем. Из этого следует, что «малые» коэффициенты так же, как и нулевые значения, должны рассматриваться в качестве указания на отсутствие действия. Однако существование ошибок в оценках также открывает возможность того, что оцениваемый коэффициент только случайно мал, причем истинное значение существенно отличается от нуля. Если бы мы трактовали все близкие к нулю коэффициенты как нули, мы могли бы исключить некоторые отношения, которые на самом деле действуют.

Статистические испытания могут применяться для оценки вероятности того, что оцениваемый коэффициент соответствует истинному значению нуль (см. библиографию к главе), и такие испытания обычно должны употребляться для того, чтобы получить информацию для решения, существует ли действие. Однако такие испытания служат только помощью, а не механическим шаблоном для вынесения такого решения. Если мы имеем дело с малой выборкой (или с любой другой ситуацией, которая стремится взвинтить ошибки в оценках коэффициента), статистическое испытание может оказаться слишком нечувствительным для наших целей и приводить нас к отклонению причинных действий, которые на самом деле существуют. Если мы работаем с большой выборкой, статистическое испытание может оказаться столь чувствительным, что будет приводить к сохранению мгновенных коэффициентов, которые могли бы быть ненулевыми только вследствие незначительных смещений в оценках (появляющихся, например, из-за нерешенных задач измерения или из-за выборки, содержащей несколько случаев, в которых операторы перестали действовать). Таким образом, помимо статистической информации в такие решения должен входить прагматический элемент. Отношение не должно исключаться из модели, если коэффициент по-казывает, что величина данного воздействия сравнима с другими рассматриваемыми воздействиями, даже если статистическое испытание свидетельствует о «незначимом» отличии коэффициента от нуля. С другой стороны, даже если коэффициент статистически значим, отношение может исключаться из модели, когда величина воздействия так мала по отношению к другим воздействиям, что она не интересна ни теоретически, ни практически.

Исключение отношений на основе коэффициентов, близких к нулю, дает возможность упростить формулировку системы. Из потокового графа можно удалить стрелки, а из структурных уравнений выбрасываются члены. Остающиеся коэффициенты в упрощенной системе должны заново оцениваться для улучшения их

точности; иными словами, данные должны повторно анализироваться путем использования новых ограничений для получения улучшенных оценок оставшихся параметров системы.

ИСТОЧНИКИ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Несколько статей, знакомящих с применением методов наименьших квадратов для оценивания, переизданы в книге Hubert M. Blalock, Jr. (Ed.). *Causal Models in the Social Sciences* (Chicago, Aldine-Atherton, 1971). Kenneth C. Land. *Formal Theory*, Chapter 7 в Herbert L. Costner (Ed.). *Sociological Methodology: 1971* (San Francisco, Jossey-Bass, 1971) разъясняет дальше связи между теориями, моделями и данными.

Различные методы оценки разрабатывались главным образом эконометриками и руководства по эконометрике содержат самые исчерпывающие обсуждения процедур. Элементарная трактовка (хотя все еще использующая матричную алгебру (см. гл. 3)) дается книгой Ronald J. Wonnacott and Thomas H. Wonnacott. *Econometrics* (New York, Wiley, 1970). Представления в работах Arthur S. Goldberger. *Econometric Theory* (New York, Wiley, 1964) и Henry Theil. *Principles of Econometrics* (New York, Wiley, 1971) математические, но примечательны своей подробностью и фундаментальностью. Гольдбергер и Теил в своих руководствах обращают серьезное внимание на ошибки выбора в оценке (равно как и на другие источники ошибок). Дополнительное обсуждение проблем выбора с точки зрения приложений имеется в Leslie Kish. *Survey Sampling* (New York, Wiley, 1965).

Социологи особенно сосредоточились на способах управления влиянием ошибок измерений при получении оценок структурных параметров. Некоторое количество основных статей входит в сборник Блэлока, упомянутый выше. Дополнительно можно рекомендовать работы: D. R. Heise and George W. Bohrnstedt. *Validity, Invalidity and Reliability*, Chapter 6 в книге Edgar F. Borgatta and G. W. Bohrnstedt (Eds.). *Sociological Methodology: 1970* (San Francisco, Jossey-Bass, 1970); Robert M. Hauser and Arthur S. Goldberger. *The Treatment of Unobservable Variables in Path Analysis*, Chapter 4 в книге Herbert L. Costner (Ed.). *Sociological Methodology: 1971* (San Francisco, Jossey-Bass, 1971) и R. M. Hauser. *Disaggregating a Social-Psychological Model of Educational Attainment*, *Social Science Research*, 1 (1972), 159—188; все они используют в выводах матричную алгебру. Классический психометрический подход к измерению в явном виде вводится в книге Jum C. Nunnally. *Psychometric Theory* (New York, McGraw-Hill, 1967). Подробная авторитетная ссылка на психометрическую традицию дается в Frederic M. Lord and Melvin R. Novick. *Statistical Theories of Mental Test Scores* (Reading, Mass., Addison-Wesley, 1968).

Осмысление множества переменных является классической проблемой общественнонаучного исследования. В книге Morris

Rosenberg. *The Logic of Survey Analysis* (New York, Basic Books, 1968) дается введение в проблемы с численными иллюстрациями. Получаемые математически подходы представлены в упомянутых выше статьях Хаузера — Гольдбергера и Хаузера. Работа K. G. Jöreskog. *A General Method for Estimating a Linear Structural Equation System*, Chapter 5 в книге A. S. Goldberger and O. D. Duncan (Eds.). *Structural Equation Models in the Social Sciences* (New York, Seminar Press, 1973) вводит модель, применимую к большому количеству задач, включая множественные индикаторы и сложные причинные связи между абстрактными переменными. (Книга Гольдбергера — Дункана, кроме того, содержит типичную эмпирическую работу по оценке линейных причинных моделей.)

Многомерный анализ выборки случаев, наблюдаемых в несколько моментов, дает изобилие задач; некоторые из них рассматриваются в D. R. Heise. *Causal Inference from Panel Data*, Chapter 1 в E. F. Borgatta and G. W. Bohrnstedt (Eds.). *Sociological Methodology: 1970* (San Francisco, Jossey-Bass, 1970); Otis Dudley Duncan. *Unmeasured Variables in Linear Models for Panel Analysis*, Chapter 2 в H. L. Costner (Ed.). *Sociological Methodology: 1972* (San Francisco, Jossey-Bass, 1972) и K. G. Jöreskog. *Factoring The Multitest — Multioccasion Correlation Matrix*, RB-69-62 (Princeton, N. J.; Educational Testing Service, 1969). Некоторые модели для объединения данных, относящихся ко времени, представлены математически в Jan Kmenta. *Elements of Econometrics* (New York, Macmillan, 1971), p. 508—517.

УПРАЖНЕНИЯ

1. Можно ли организовать рекурсивную модель из следующих переменных: «образование» отца, «социально-профессиональный статус» отца, «образование» сына, «социально-профессиональный статус» сына? Обсудите различные соображения.

2. Наблюдаемые корреляции между переменными статусов отца и сына в исследовании P. Blau and O. D. Duncan. *The American Occupational Structure* (New York, Wiley, 1967, p. 169) были следующими:

	«Образование» отца	«Профессиональный статус» отца	«Образование» сына	«Профессиональный статус» сына
«Образование» отца	1,000	0,516	0,453	0,322
«Профессиональный статус» отца		1,000	0,438	0,405
«Образование» сына			1,000	0,596
«Профессиональный статус» сына				1,000

Предположим, что в изучаемой совокупности применяется определенная рекурсивная модель. Оцените все подходящие стандартизованные параметры модели. Используйте описание из упражнения 5 гл. 1 (для этой и некоторых последующих задач нужна вычислительная машина).

3. Предположим, что перед некоторым количеством индивидуумов поставлено задание, в ходе которого они могут обмениваться действиями, свидетельствующими об уважении и власти. Пары образуются экспериментально, путем объединения в них лиц, чьи занятия связаны с различными уровнями власти в обществе. Оценки влиятельности этих занятий получаются от участников до начала выполнения задания. После длительного взаимодействия задание завершается, и путем оценки индивидуумом самого себя и своего партнера по той же самой шкале влиятельности получается второе множество оценок. Приведенную модель можно было бы использовать для интерпретации экспериментальных результатов.



O_1 и O_2 — оценки уровня власти для занятий индивидуумов 1 и 2, соответственно, переводятся в начальные оценки межличностной власти (I_1 и I_2) посредством коэффициента k с возможностями относящихся к суждениям возмущений U_1 и U_2 . D является воспринимаемым различием между своей властью и властью другого, подправленным персональной склонностью индивидуума видеть себя более сильным или более слабым, чем другие (P). Акты уважения или доминирования представляются параметром a ; например, если индивидуум воспринимает подправленную разницу как $+1,0$ (сам влиятельнее другого), производится действие преобладания, таким образом добавляющее a единиц к его власти и отнимающее a единиц от власти другого. Если воспринимается разница в $-1,0$, актом уважения вычитается a единиц из его власти и добавляется то же самое количество к власти другого.

(а) Предположим, что исследователь производит регрессию конечных значений переменной I_1 по O_1 и O_2 в смысле обычного метода наименьших квадратов. Какие системные параметры оцениваются коэффициентами регрессии? Можно ли этим путем получить оценки параметра a ?

(б) Предположим, что переменные O и I оцениваются с точки зрения не власти, а привлекательности или добродетели. Тогда a мог бы изменить знак, символизируя действия альтруизма и эксплуатации, а не преобладания и уважения; восприятие положительного различия в добродетели ведет к действию, которое передает что-то из собственных достоинств индивидуума другому; восприятие отрицательного различия ведет к эксплуатации достоинств другого индивидуума для улучшения своего собственного положения. При этом изменении какое ожидаемое действие производили бы коэффициенты регрессии, если провести регрессию I_1 по O_1 и O_2 ? Отличаются ли задачи идентификации в новой системе?

4. Социологи спорят, материалистически или же под воздействием идеологий определяется социальная структура. Предположим, что мы решаем исследовать проблему эмпирически в совокупности малых сообществ, которые были описаны этнографами. (Систематические данные о более чем 1000 культур были опубликованы в журнале «Ethnography».) Ради упрощения давайте предположим, что относящиеся к делу переменные можно представить следующим образом:

S — уровень сложности иерархической социальной структуры, который отражается в существовании классов, каст и высших политических служб;

M — достижения общества в области производства продуктов питания, измеряемые на шкале технологии пропитания, ранжированной от охоты и собирательства до передовых сельскохозяйственных методов, таких, как агрономия хлебных злаков с приращением;

I — разработка авторитетной системы моральных ценностей для упорядочения социальных отношений, отражаемых верой в способность сверхъестественных факторов санкционировать поведение людей.

Задача состоит в оценивании структурных коэффициентов a_{SM} и a_{SI} , в частности в определении, незначителен ли по величине тот или другой из этих коэффициентов. Однако связи могут не быть рекурсивными. Усложнение социальной структуры предположительно могло бы оказывать на технологию пропитания или нравственность обратное действие. Кроме того, прогресса во всех трех областях можно было бы достичь путем распространения культуры из других сообществ, так что возмущения могли бы быть коррелированными. Поэтому мы должны попытаться описать инструменты, которые позволили бы нам определить структурные коэффициенты.

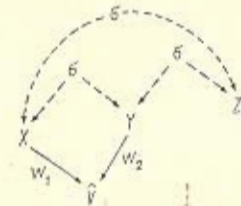
(а) Оцените каждую из следующих переменных, которые характеризуют место действия общества как возможный инструмент для определения структурных коэффициентов a_{SM} и a_{SI} :

- 1) средняя продолжительность периода роста,
- 2) плодородие почвы,
- 3) суровость природы обычного места расселения (т. е. мера условий, угрожающих гибелью от жажды, хищников, голода, воды и падения со скал),
- 4) доступность месторождений металлических руд.

(б) Предположим, мы нашли, что переменная периода роста имеет нулевую корреляцию с переменной социальной структуры S , а также с переменной идеологии I . Поможет ли эта информация при решении вопроса о том, является ли она соответствующим инструментом для связи $M \rightarrow S$?

5. В конце гл. 4 предполагалось, что установка и поведение можно связать в петлю управления. Таким образом, определение воздействия одного на другое по данным обследования потребовало бы применения инструментов. Предположим, что планируются вопросы, которые адекватно измерят отношение человека к алкогольным напиткам и уровень употребления алкоголя. Пункты включаются в опрос студентов большого университета штата, в котором узаконенным возрастом для питья крепких напитков является 21 год, а для пива — 18. Другие пункты определяют, воспитывался ли студент в семье с фундаменталистским религиозным мировоззрением, каковы его еженедельные расходы на развлечения и его возраст. Оцените три последних пункта в качестве инструментов для определения коэффициентов в петле «установка — поведение» для алкоголя.

6. Статистики, характеризующие взвешенную составную величину, можно вычислять по дисперсиям и ковариациям ее компонент-переменных; например, предположим, что мы хотим образовать новую переменную V путем сложения переменных X и Y , взвешенных коэффициентами w_1 и w_2 . Если мы знаем σ_X^2 , σ_Y^2 и σ_{XY} , то легко определить σ_V^2 , $\sigma_X V$ и $\sigma_Y V$. Кроме того, предположим, что в самом начале у нас есть другая переменная, Z , и мы знаем σ_{XZ} и σ_{YZ} . Тогда так же легко определить σ_{VZ} . Логично можно проследить, применяя путь анализ к следующей схеме:



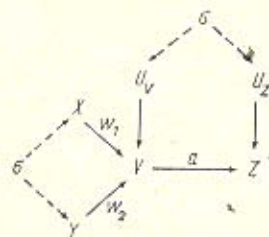
$$\sigma_V^2 = w_1^2 \sigma_X^2 + w_2^2 \sigma_Y^2 + 2w_1 w_2 \sigma_{XY}$$

$$\sigma_X V = w_1 \sigma_X^2 + w_2 \sigma_{XY}$$

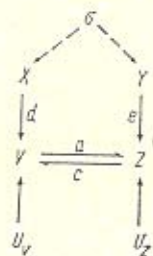
$$\sigma_Y V = w_2 \sigma_Y^2 + w_1 \sigma_{XY}$$

$$\sigma_V Z = w_1 \sigma_{XZ} + w_2 \sigma_{YZ}$$

Это означает, что можно проводить анализ двухэтапным методом наименьших квадратов без возвращения к отдельным наблюдениям после первого цикла проведения регрессионного анализа; например, предположим, что у нас есть следующая задача:



Значения параметров w_1 и w_2 оцениваются проведением регрессии V по X и Y (или использованием формул для коэффициента регрессии, данных в гл. 3). Тогда с помощью приведенных выше процедур мы можем оценить дисперсию предсказываемой величины V и ее ковариацию с Z . Этих диаграмм достаточно для вычисления регрессии второго этапа, оценивающей a (причем опять отсылаем к формулам в гл. 3). Используйте эти процедуры для того, чтобы двухэтапным методом наименьших квадратов оценить структурные коэффициенты в следующей модели:



Дисперсии и ковариации переменных таковы:

	X	Y	V	Z
X	1,000	0,300	0,283	0,235
Y	0,300	1,000	-0,113	0,466
V	0,283	-0,113	1,068	-0,164
Z	0,235	0,466	-0,164	1,051

7. Можно было бы улучшить социально-экономическое положение группы равных меньшинств в Америке, способствуя более высокому уровню образования, после чего последовал бы рост среднего дохода. Можно обосновать также и то, что эти группы должны требовать лучшей работы и более высокой заработной платы, поскольку это может повлечь за собой рост образования и других показателей статуса. Эти доводы означают, что теоретически мы не можем исключить петли между средним уровнем образования и средним доходом в совокупности из национальных меньшинств. Однако, если мы могли бы найти подходящие инструменты, проблему можно было бы исследовать эмпирически. Мы могли бы оценить коэффициент для направлений от образования к доходу и от дохода к образованию для определения того влияния, которое имеет действительную поддержку в американском обществе. Для иллюстрации результатов, которые можно получить, следует использовать данные исследования 63 стандартных столичных ста-

тистических зон [Richard Child Hill, Unionization and Racial Income Inequality in the Metropolis. American Sociological Review, 39, 1974, 507—22.] В исследовании были включены перечисляемые далее переменные: S — наличие стандартной зоны в южной части страны; U — степень объединения в стандартной зоне; P — процент небелого населения зоны; M — процент рабочей силы зоны, занятой в промышленных отраслях; E — среднее образование небелого населения зоны; I — средний семейный доход небелого населения зоны. Сообщалось, что корреляции между этими переменными в 1960 г. были такими:

	S	U	P	M	E	I
S	1,000	1,000				
U	-0,556	-0,258	1,000			
P	0,591	0,510	-0,233	1,000		
M	-0,253	0,226	-0,684	-0,134	1,000	
E	-0,574	0,569	-0,527	0,185	0,709	1,000
I	-0,749					

Автор статьи не сосредоточивает внимание на поднятых здесь вопросах, но в качестве основы для анализа данных можно было бы предложить следующую модель. (Переменные V представляют возмущения.)



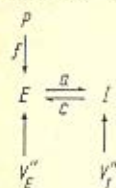
(а) Используйте описанные корреляции и схему для оценки a и c . Используйте процедуры, представленные в упражнении 6, для осуществления подхода с двухэтапным методом наименьших квадратов.

(б) Оцените снова a и c с помощью следующей модели:



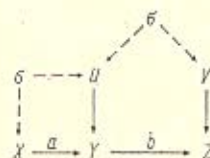
(в) Сравните результаты из (а) и (б). В частности, обсудите адекватность инструментов и смещения, которые они могли произвести. (При обсуждении помните об объеме выборки.)

8. Продолжаем иметь дело с переменными и данными, приведенными в упражнении 7. Предположим теперь, что мы хотим принять следующее описание:



Используйте эту модель для переоценки коэффициентов a и c , а результаты этого анализа вместе с результатами упражнения 7 — для формулировки стратегии улучшения социально-экономического положения городского небелого населения.

9. Ковариации между возмущениями довольно легко можно оценить по наблюдаемым дисперсиям и ковариациям, если система не содержит петель и если имеются инструменты; например, предположим, что у нас есть следующая простая система:



Оценка коэффициента b дается выражением $(\sigma_{XZ}/\sigma_{XY})$. Ковариация возмущений U и V определяется в виде

$$\sigma_{YZ} = b\sigma_Y^2 + \sigma_{UV}$$

или

$$\sigma_{UV} = \sigma_{YZ} - \frac{\sigma_{XZ}\sigma_Y^2}{\sigma_{XZ}}$$

Обычно мы имеем дело с переменными, которые измеряются неточно.



и в этом случае главные ковариации и дисперсии наблюдаемых переменных будут связаны с ковариациями и дисперсиями истинных переменных таким образом:

$$\sigma_{\tilde{X}\tilde{Y}} = \sigma_{XY},$$

$$\sigma_{\tilde{X}\tilde{Z}} = \sigma_{XZ},$$

$$\sigma_{\tilde{Y}\tilde{Z}} = \sigma_{YZ},$$

$$\sigma_{\tilde{Y}}^2 = \sigma_Y^2 + \sigma_F^2.$$

Эмпирически наблюдаемые статистики служат для оценки параметров системы. Следовательно, на практике оценкой ковариации возмущений было бы

$$\tilde{\sigma}_{UV} = \sigma_{\tilde{Y}\tilde{Z}} - \frac{\sigma_{\tilde{X}\tilde{Z}}\sigma_{\tilde{Y}}^2}{\sigma_{\tilde{X}\tilde{Y}}}.$$

Воспользуйтесь данными формулами для определения того, как неточные измерения влияют на оценки ковариаций возмущений. Оправдана ли наша попытка дать теоретические интерпретации оценкам ковариаций возмущений, когда переменные измерены с ошибкой?

Статический анализ требует, чтобы существующие значения переменных на входе системы сохранялись на одних и тех же уровнях в течение периода, достаточно долгого для того, чтобы их полные причинные последствия ощутились по всей системе. Таким образом, наблюдаемые результаты отражают конечные последствия входов. Однако исследователи, занимающиеся социальным анализом, могут интересоваться не только конечными результатами, но и процессами, приводящими к этим результатам. Тренды и отклонения от них переменных социальной системы сами составляют историю и имеют для людей огромное личное значение. Кроме того, может оказаться необходимым рассмотреть динамику системы для нахождения нежелательных процессов, которые происходят между входами и выходами, или же столь экстремальных процессов, что они разрушают систему прежде, чем может достигаться статическое состояние. Некоторое рассмотрение динамики системы нужно и просто для лучшего понимания статического анализа. Требования, чтобы переменные на входе оставались постоянными, а последствия полностью реализовывались, в действительности присутствуют не всегда: входные переменные могут иметь значения, которые сильно меняются во времени, или система со стабильными входными переменными может все же не породить полных последствий этих входных переменных. В каждом из этих случаев обычные статистические оценки параметров системы будут смещаться, и для того чтобы понять проблему, нам надо обратиться к динамическому анализу.

Исследование динамики системы сосредоточивается на периоде, которым пренебрегают при проведении статического анализа, находящемся между началом нового ввода и конечными результатами системы. Как и в статике, имеются и дедуктивные, и индуктивные вопросы, заслуживающие внимания. На дедуктивном уровне динамический анализ дает информацию о явлениях адаптации системы. Результаты могут представлять определенный интерес, потому что изменения значений переменных в линейной системе часто являются сложными нелинейными функциями времени. К тому же исследование динамики раскрывает новую тему: обработка информации в виде переменных во времени входных «сигналов» для определения того, как система реагирует на временное поведение входных переменных и как она образует новые

траектории своих собственных переменных. Со стороны индукции или вывода исследование динамики приводит к дополнительным процедурам выяснения структуры системы посредством использования данных о временных рядах.

Хотя предмет динамики системы, несомненно, важен для ученых, занятых социальными исследованиями, его нельзя рассмотреть здесь систематически. Даже элементарное введение потребовало бы больше страниц и больше математики, чем мы уже отдели в этой книге. (Исследование динамики системы включает целую новую размерность — время, которая требует временной разработки процедур сбора данных и которая усложняет анализ привлечением в математику комплексных переменных.) Эта глава имеет скромную цель — попытаться углубить понимание статических моделей путем краткого обсуждения нескольких избранных проблем динамики систем. Сначала рассматриваются динамические характеристики некоторых простых линейных систем для того, чтобы получить оценку их сложности; затем — некоторые возможные способы затемнения оценок параметров статической модели получением данных из системы, которая находится в динамическом, а не в статическом состоянии.

ДИНАМИКА ПРОСТЫХ СИСТЕМ

Причинное запаздывание

Обычно, пока операторы создают следствия из причин, имеется по крайней мере короткое запаздывание. Поэтому причинно связанные события чаще всего разделяются некоторым периодом запаздывания, и система обычно реагирует на ввод информации некоторым динамическим процессом, хотя простым или кратким; например, начиная с момента, когда педаль газа в автомобиле изменяет положение, имеется малый, но определенный период, в продолжение которого нет никакого изменения вращающего момента, передаваемого на колеса, потому что цепь следствий еще не достигла этой стадии; или вслед за утверждением законов имеется период без влияния их на всех граждан, пока официальные лица приводят в действие соответствующую политику. Для удобства анализа полезно предположить, что причинное запаздывание существует всегда, хотя и допускается, что в некоторых отдельных случаях время запаздывания может становиться практически нулем.

Периоды запаздывания сильно меняются в соответствии с типом рассматриваемой системы. Психологические скрытые состояния обычно могут измеряться в сотые доли секунды, межличностные скрытые состояния — через секунды или минуты, скрытые состояния формальной организации — через часы и дни, а скрытые состояния общины или общества — через месяцы или годы. Таким образом, обычно мы не можем ожидать исследования динамики различных уровней организации в одних и тех же временных рам-

ках. Процессы, которые могли бы интересовать психолога, происходят на временной шкале, которая бесполезна для социолога, и наоборот. Однако с поправкой на различие временных шкал, которые соответствуют различным явлениям, формальные характеристики изменения могут оказаться довольно сходными независимо от уровня анализа. Это и есть один из основных выводов теории систем.

Есть схематическое средство для того, чтобы представить причинные запаздывания на потоковых графах и показать, что периоды запаздывания меняются при различных процессах. Следуя соглашениям, которые изложены в литературе по динамике систем, временную задержку обозначают буквой z , а ее относительную величину — показателем степени со знаком минус. Для представления периода причинной задержки на схеме мы просто «приписываем» соответствующий временной символ обозначению причинного оператора, как показано в 6.1.

$$6.1. X \xrightarrow{az^{-2}} Y.$$

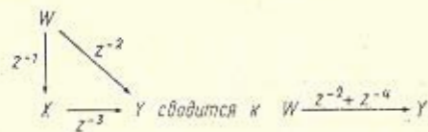
Оператор a для превращения изменения переменной X в изменение переменной Y требует двух единиц времени.

Действительная временная шкала определяется рассматриваемой задачей; например, две единицы времени, показанные в 6.1, могут относиться к двум наносекундам для инженера-электрика и к двум годам для исследователя в области политики.

Это представление временных запаздываний особенно интересно, когда имеют дело с линейными системами, потому что все обычные правила анализа потокового графа могут применяться осмысленно. Временной символ трактуется, как если бы он представлял собой особый линейный оператор с неизвестным значением, которое подчиняется обычным правилам умножения и сложения. Отсюда непосредственно следуют виды интерпретации потокового графа, изображенные в 6.2 (вспомните специальное алгебраическое правило умножения показательных функций, т. е. $z^m \cdot z^n = (z^{m+n})$).

$$6.2. W \xrightarrow{z^{-1}} X \xrightarrow{z^{-3}} Y \text{ сводится к } W \xrightarrow{z^{-4}} Y.$$

Изменение переменной W вызывает изменение переменной X через один период, а изменение переменной X вызывает изменение переменной Y через три периода. Таким образом, изменение W влечет изменение Y через четыре периода



Изменение W влечет изменение Y через два периода и образует дополнительное изменение Y через четыре периода.

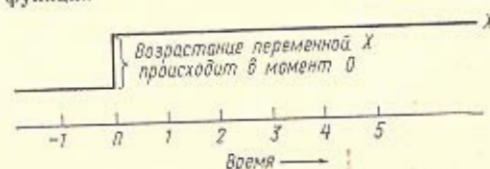
Преобразования графов в 6.2 определяют *передаточные функции* для простых иллюстрируемых систем. Подобным же образом полный набор правил преобразования графов может применяться для определения передаточных функций для более сложных систем с петлями. Осуществление интересных применений требует понимания символа z как комплексного числа и продолжения динамического анализа по всей шкале. Для этих целей достаточно того, что у нас есть соглашения для графического представления связей между переменными системы и временного элемента, относящегося к действиям системы.

Элементарные динамические модели

Тренды и колебания, которые становятся очевидными при откладывании значений переменной системы на графике относительно времени, обыкновенно называют *сигналом*. Это траектория значений, которая составляет сигнал. Одна общая задача в динамическом анализе состоит в описании сложного сигнала относительно простым математическим выражением. Вторая состоит в отыскании относительно простого математического выражения (передаточная функция), которое можно приспособить для описания того, как определенная система преобразует входной сигнал в выходной с другой траекторией. В этом отношении мы имеем дело с реакцией данной системы на различные входные сигналы или с обработкой данного входного сигнала различными системами.

Поскольку здесь нас все еще интересуют системы, которые можно охарактеризовать как находящиеся обычно в статическом состоянии, мы можем обойти большинство аналитических задач, порожденных сложным входным сигналом. Обычное предположение при работе со статическими моделями состоит в том, что входные сигналы внезапно изменились в некоторый более ранний момент, а затем сохранялись постоянными до тех пор, пока операторы системы не генерировали все последствия. Таким образом, входные сигналы, представляющие главный интерес, являются простыми ступенчатыми функциями, иллюстрируемыми в 6.3.

6.3. Ступенчатая функция

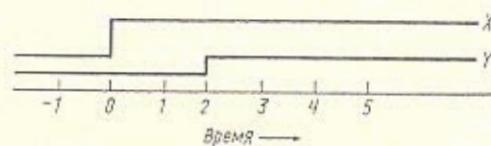


Следовательно, цель состоит в разработке неких грубых представлений о том, как различные типы системы преобразуют ступенчатые изменения входных сигналов во множество выходных сигналов в течение переходного периода до достижения статического состояния.

Простейшая система состоит из *единственного оператора*, который порождает реакцию внезапно через некоторое время после того, как входной сигнал изменил значение. Если в качестве входного сигнала подается ступенчатая функция, в качестве выходного получается запаздывающая ступенчатая функция, как показано в 6.4.

6.4. Основной системой является $X \xrightarrow{\frac{1}{2}z^{-2}} Y$.

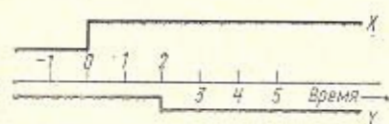
Предположим, что значение переменной X увеличивается в точке на одну единицу в момент, произвольно именованный нулем. Входной и выходной сигналы системы будут иметь следующий вид:



Если структурный коэффициент имеет отрицательное значение, выходной сигнал изменяется в направлении, противоположном направлению изменения входного сигнала, а именно если ступенчатая функция применяется к следующей системе:

$$X \xrightarrow{-\frac{1}{2}z^{-2}} Y,$$

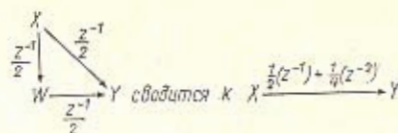
то мы получаем следующую траекторию:



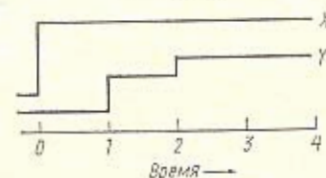
Единственной интересной особенностью в динамике такой простой системы оказывается период ожидания между изменениями на входе и на выходе.

Рекурсивная сеть операторов может порождать более сложный выходной сигнал, потому что следствия, порождаемые по различным путям, будут обычно достигать выходной переменной в разные моменты. Если все операторы в системе оказывают положительные действия, конечные значения выходной переменной будут стремиться монотонно накапливаться во времени. Если же между входом и выходом существуют какие-то отрицательные пути, значение выхода, до того как оно достигнет своего статического состояния, может смещаться и вверх, и вниз. Оба случая иллюстрируются в 6.5.

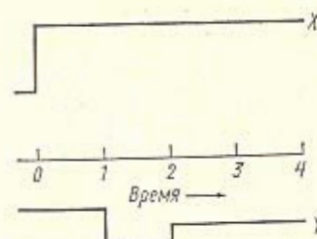
6.5.
Система



Ступенчатое изменение на входе порождает следующие сигналы:

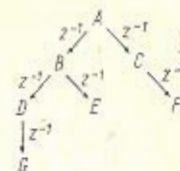


Если бы оператор между X и Y оказывал отрицательное действие, приведенной системой была бы $X \xrightarrow{-\frac{1}{2}(z^{-1}) + \frac{1}{4}(z^{-2})} Y$, а сигналы входа-выхода имели бы следующие траектории:



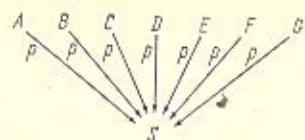
Некоторые социальные процессы можно описывать с точки зрения рекурсивного распространения по социальной сети (например, принятие технологического нововведения), и в этом случае любой выходной сигнал, который зависит от этих изменений (например, производительность), будет иметь тенденцию изменяться нарастающим образом, следуя за ступенчатым возрастанием на входе. Простая иллюстрация дается в 6.6.

6.6. Этот потоковый граф может представлять иерархическую сеть влияний в малой группе фермеров:



Переменные от A до G можно интерпретировать как меру уровня сельскохозяйственной технологии каждого фермера, и все они могут аддитивно связы-

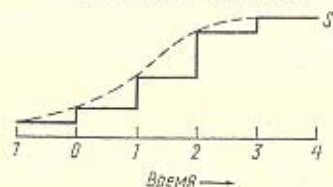
ваться с коллективной переменной, такой, как сельскохозяйственные излишки (S); помимо вышеприведенных стрелок мы имеем следующее:



где p показывает количественную связь между технологией фермера в данном году и общественными излишками в том же году. Когда две схемы объединяются и приводятся, мы получаем между A и S следующую связь:

$$A \xrightarrow{p(1+2z^{-1}+3z^{-2}+z^{-3})} S.$$

Ступенчатый рост уровня технологии A в нулевой момент вызвало бы следующую модель изменения коллективных излишков:



Поведение изменений на выходе приближенно представляется непрерывной «сигмоидальной» кривой, показанной пунктирной линией. Приближение непрерывной кривой в такой малой группе людей неточно, но согласие было бы лучше, если бы излишки составлялись из малых вкладов от сотен фермеров.

Пример в 6.6 иллюстрирует тот факт, что даже те системы, у которых отсутствует обратная связь, могут производить довольно сложные преобразования входных сигналов. Путем включения в систему, рассматриваемую в 6.6, отрицательных связей можно было бы получить даже более сложные выходные сигналы, чем показанные — отрицательные связи могли бы представлять отношения конкуренции, эксплуатации и обмана между фермерами. При таких отрицательных операторах можно было бы определить структуры, в которых возрастание технологии A ведет к колебанию во времени или даже чистому уменьшению излишков сообщества.

Воздействия обратной связи

Системы с обратной связью всегда производят из ступенчатых изменений на входе относительно сложные выходные сигналы. Действительно, системы с обратной связью могут предназначаться для преобразования ступенчатого входного сигнала в выходной сигнал почти любой желаемой степени сложности, хорошо согласованный почти с любой желаемой моделью. Здесь мы ограничиваемся только несколькими основными типами модели, которые производятся относительно простыми устройствами обратной связи.

Усиление — простейший тип явления обратной связи — порождается петлей с положительным обратным эффектом. В гл. 2

рассматривалось чистое следствие или конечный результат такой петли. Теперь интерес сосредоточивается на ряде изменений, которые достигают наибольшего значения в конечном результате. Вообще эти изменения происходят в моделях с монотонным накоплением, иллюстрируемых в 6.7.

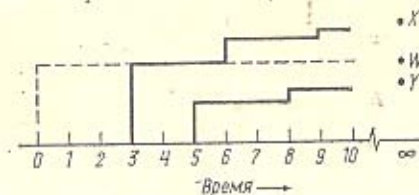
6.7. Предположим, что у нас есть система, в которой изменение переменной W вызывает изменение переменной X через три периода, любое изменение X вызывает изменение в Y через два периода, а любое изменение в Y вызывает изменение переменной X через один период. Допустим далее, что все структурные коэффициенты положительны, и в этом случае у петли есть положительный обратный эффект, а это — усилитель. Задание коэффициентам некоторых произвольных значений дает следующую систему:

$$W \xrightarrow{z^{-3}} X \xrightleftharpoons[0,6z^{-1}]{0,5z^{-2}} Y.$$

Теорема Мэсона (правило II.16) могла бы применяться как обычно для приведения X и Y к W , но тогда для использования результатов при вычерчивании графиков выходных сигналов мы должны были бы понимать z как комплексное число и проводить некий математический анализ повышенного уровня. Вместо этого мы просто прослеживаем действия системы в продолжение нескольких первичных периодов после изменения переменной W для того, чтобы охарактеризовать ее динамику. Предположим, что до изменения W значение всех трех переменных было нулем. В нулевой момент значение W возрастает до 1,0; X и Y все еще остаются неизменными вследствие трехпериодного запаздывания между изменениями переменных W и X . После этого значения X и Y будут изменяться, как показано в следующей таблице. (Δ следует читать как «изменение в», а индекс i читается как «в момент i ».)

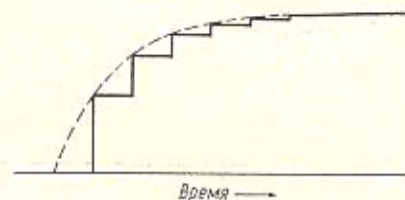
Время	Значение X	Значение Y
1	0	0
2	0	0
3	$0 + 1(\Delta W_0) = 1$	0
4	1	0
5	1	$0 + 0,5(\Delta X_3) = 0,5$
6	$1 + 0,6(\Delta Y_5) = 1,3$	0,5
7	1,3	0,5
8	1,3	$0,5 + 0,5(\Delta X_6) = 0,65$
9	$1,3 + 0,6(\Delta Y_8) = 1,39$	0,65
10	1,39	0,65
$\dots$	$\dots$	$\dots$
∞	1,43	0,71

Значения для ∞ получаются путем применения правила II.16 к статической системе (в этом случае всеми z пренебрегают). Вычерчивание графиков переменных системы относительно времени дает следующее:



Петля в 6.7 является усилителем, потому что она принимает начальное изменение переменной X (или Y) и превращает его в большее конечное изменение. Рост происходит не сразу, а ступенчато. По мере того как изменения циркулируют по петле снова и снова, они на каждом цикле немного увеличивают значения переменных. В примере *большая часть* роста происходит во время нескольких первых циклов. Размер приращений становится с течением времени все меньше и меньше до тех пор, пока в некоторый момент ими можно будет пренебречь и для всех практических целей переменные достигнут своих конечных значений. Это особенно верно для устойчивых усилительных петель (неустойчивые петли рассматриваются позже). Изменения происходят в соответствии с «экспоненциальной кривой», которая приближается к некоторому постоянному значению.

6.8. Устойчивые петли усиления в ответ на вводящееся изменение дают непрерывно убывающий рост. Период роста может приблизительно характеризоваться гладкой экспоненциальной кривой, которая показана пунктирной линией.



Усилительная петля реагирует на убывание входного значения (а не на приращение) процессом того же рода, за исключением того, что экспоненциальная кривая «выпукла вниз», а значение переменной из петли убывает ступенчато.



Устойчивая петля с отрицательным обратным эффектом представляет собой *устройство управления*, потому что она стремится противодействовать изменениям и в конечном счете исходит системы менее затрагиваются изменениями на входе системы, чем в начале. Динамический процесс особенно значителен в петлях управления, потому что восстановление не является гладким монотонным процессом. Управляемые переменные системы восстанавливаются от возмущений посредством колебательного процесса, как показано в 6.9.

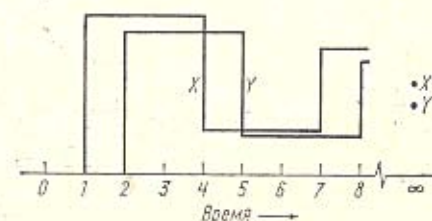
6.9. Следующая система стремится противодействовать влияниям изменений переменной W . Управляющая природа системы отражается в отрицательном обратном эффекте для петли — $(0,9)(-0,8) = (-0,72)$.

$$W \xrightarrow{z^{-1}} X \xrightarrow{0,9z^{-1}} Y \xrightarrow{-0,8z^{-2}} X$$

Предположим, что до того, как W изменяется в нулевой момент на одну единицу, все переменные имеют значение нуль. Тогда значения X и Y в более поздние моменты оказываются следующими:

Время	Значение X	Значение Y
1	$0 + 1 (\Delta W_0) = 1$	0
2	1	$0 + 0,9 (\Delta X_1) = 0,9$
3	1	0,9
4	$1 + (-0,8) (\Delta Y_2) = 0,28$	0,9
5	0,28	$0,9 + 0,9 (\Delta X_4) = 0,25$
6	0,28	0,25
7	$0,28 + (-0,8) (\Delta Y_5) = 0,80$	0,25
8	0,80	$0,25 + 0,9 (\Delta X_7) = 0,72$
...		
∞	0,58	0,52

Таким образом, поведение значений переменных на выходе во времени носит следующий характер:

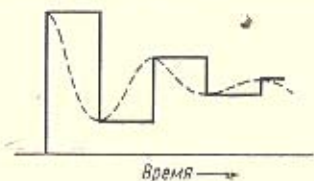


Когда значение переменной в устойчивой управляющей системе изменяется, в этой переменной начинается последовательность адаптаций, достигающая высшего значения в окончательном изменении, которое меньше, чем начальное изменение. Но в процессе достижения конечного состояния значение переменной слишком далеко отклоняется в одну сторону, затем слишком далеко в другую сторону, только постепенно возвращаясь в конечную неизменяющуюся точку. Как и при усилительных петлях, устойчивые петли управления образуют наибольшие сдвиги в начале процесса. После некоторого момента времени колебания оказываются достаточно малыми для того, чтобы ими пренебречь, и для всех практических целей значение переменной достигло своего конечного состояния.

Как показано в 6.10, колебания управляемой переменной соответствуют синусоидальной кривой, амплитуда которой со време-

нем убывает. Этот факт широко используется при математическом анализе динамических систем управления.

10. Синусоидальная кривая, описывающая изменения в петле управления.



Петля, которая изолирована от других петель, *неустойчива*, если ее обратный эффект больше или равен 1,0 (в усилителях) либо меньше или равен $-1,0$ (в петлях управления). Неустойчивые петли показывают те же самые общие кривые изменения их выходных сигналов, как и устойчивые петли, за исключением того, что произведенные изменения становятся со временем больше, а не меньше. (Изменения остаются на одном и том же уровне, если обратный эффект оказывается 1,0 или $-1,0$).

6.11. Возможный выходной сигнал неустойчивой усилительной петли:

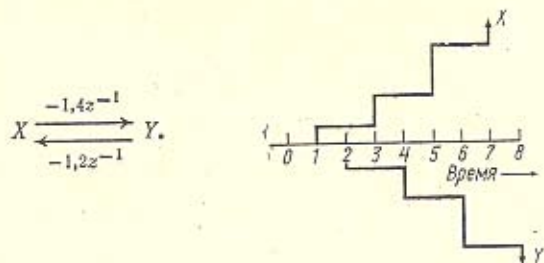
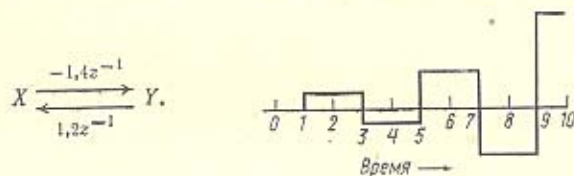


График возможного выходного сигнала в X из неустойчивой петли «управления»:



Рост, производимый неустойчивым усилением, все еще экспоненциальный, но теперь скачкообразно экспоненциальный. Аналогично изменения, производимые неустойчивой петлей с отрицательным обратным эффектом, все еще соответствуют лежащей в основе синусоидальной кривой, но теперь колебания становятся более беспорядочными.

Характеристики устойчивости системы со множеством петель нельзя распознать путем рассмотрения устойчивости каждой петли в отдельности, как иллюстрируется сопровождающим примером. На вопрос об устойчивости можно ответить, если известны все временные и структурные параметры системы. Однако требующийся анализ включает углубленную математическую трактовку передаточной функции системы.

Неустойчивые процессы никогда долго не продолжаются (хотя значение слова «долго» зависит от временной шкалы рассматриваемой системы — секунды для электронного оборудования и, возможно, столетия для обществ). Это материально обусловлено, что можно видеть, рассматривая значения переменной как интенсивности определенных событий, происходящих на более низком уровне — точка зрения, введенная в гл. 1. Когда переменная принимает крайнее значение, операторы более низкого уровня работают вблизи своих ограничений. Соответственно невозможны никакие дальнейшие возрастания интенсивностей событий более низкого уровня, потому что операторы более низкого уровня ничего больше сделать не могут. Далее имеются две возможности. Связи между переменными системы могут стать нелинейными — изменение одной переменной больше не оказывает обычных воздействий на другие переменные, — что, в сущности, останавливает скачкообразный процесс. Как альтернатива, операторы более низкого уровня могут начать выходить из строя, и в этом случае весь ход процесса запнется и система развалится.

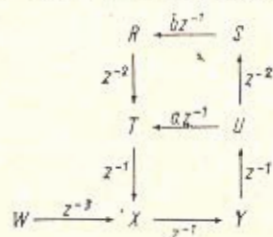
Неустойчивость почти всегда считается серьезной проблемой и чем-то, чего следует избегать. Система, работающая на пределе своих возможностей, больше не является адаптивной; она не может дифференциально реагировать на колебания входных сигналов. Такая система равнозначна интенсивному и бесполезному поглотителю энергии. Неудобства альтернативной возможности разрушения очевидны, особенно если рассматривается собственная социально-экологическая система человека. Таким образом, если система неустойчива, она бесполезна или обречена до тех пор, пока (и если) не исключена неустойчивость. Один способ решения задачи состоит в преобразовании системы для исключения неустойчивой петли или петель. Второй фундаментальный метод состоит в добавлении петель управления более высокого порядка, которые противодействуют неустойчивости, создаваемой петлями более низкого порядка.

Обратная связь более высокого порядка

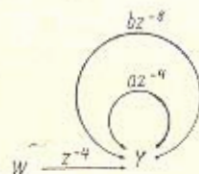
Совокупность петель создает *обратную связь более высокого порядка*, когда времена оборота для различных петель совокупности различны («время оборота» — это время, требующееся для завершения одного цикла). Обратная связь более высокого порядка дополнительно усложняет выходные сигналы. Обсудить здесь все возможности нельзя, но для оценки многообразия

выходных сигналов, которые можно получить при простом добавлении петли второго порядка, полезно рассмотреть несколько примеров. Простая система второго порядка, показанная в 6.12, используется для порождения иллюстративных данных путем применения в W входного сигнала в виде ступенчатой функции и последующего рассмотрения выходного сигнала в Y .

6.12. Представлен пример простой системы с обратной связью второго порядка (вертикальные стрелки могут представлять процессы коммуникации):



Ради упрощения всем коэффициентам, за исключением a и b , задано значение 1.0. Для создания систем различных типов знаками коэффициентов a и b можно сделать $+$ и $-$. Изображенную выше схему можно привести к схеме



которая соответствует «разностному уравнению»

$$Y_k = aY_{k-4} + bY_{k-8} + W_{k-4}$$

в нем нижние индексы показывают момент времени, в который наблюдаются переменные. Например, значение переменной Y в момент восемь равно a -кратному значению переменной Y в момент четыре плюс b -кратное значение переменной Y в момент нуль плюс значение переменной W в момент четыре.

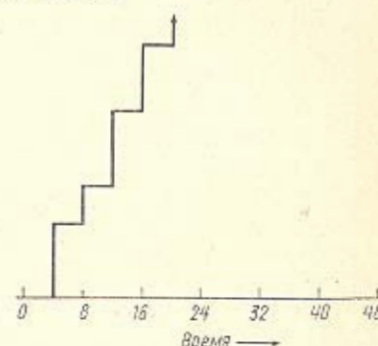
Для образования примера выходных сигналов из этой системы второго порядка абсолютное значение коэффициента a задали как $\frac{1}{2}$, а b — как $\frac{3}{4}$. Предположили, что значения всех переменных были нулями до нулевого момента, когда W возросла на одну единицу и там зафиксировалась. Знаки коэффициентов a и b меняли для того, чтобы образовать следующие типы систем: a положителен и b положителен — усиление первого и второго порядка; a положителен, а b отрицателен — усиление первого порядка с управлением второго порядка; a отрицателен, а b положителен — управление первого порядка и усиление второго порядка; и a , и b отрицательны — управление первого и второго порядка. Выходные сигналы получали подстановкой каждого множества значений a и b в разностное уравнение 6.12, а затем применением несколько раз формулы для образования последовательных значе-

ний переменной Y . Подробности вычислений скучны и в 6.13 приводятся только графические результаты.

6.13. Пример выходных сигналов систем второго порядка

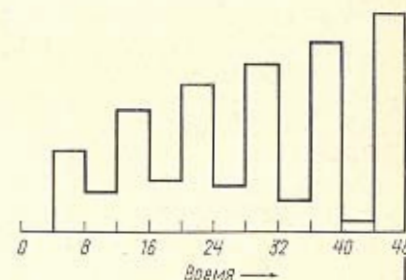
Усиление первого и второго порядков:

$$Y_k = \frac{1}{2} Y_{k-4} + \frac{3}{4} Y_{k-8} + W_{k-4}$$



Управление первого порядка, усиление второго порядка:

$$Y_k = -\frac{1}{2} Y_{k-4} + \frac{3}{4} Y_{k-8} + W_{k-4}$$



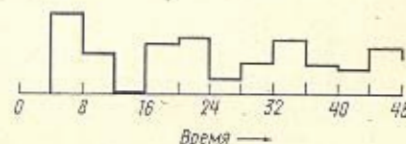
Усиление первого порядка, управление второго порядка:

$$Y_k = \frac{1}{2} Y_{k-4} - \frac{3}{4} Y_{k-8} + W_{k-4}$$



Управление первого и второго порядков:

$$Y_k = -\frac{1}{2} Y_{k-4} - \frac{3}{4} Y_{k-8} + W_{k-4}$$



Два первых графика в 6.13 иллюстрируют тот важный момент, что усиление второго порядка может создавать неожиданные неустойчивости. Два последних графика изображают сложные входные сигналы, которые часто производятся при наличии управления более высокого порядка.

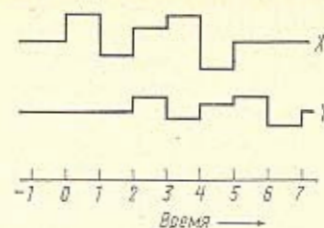
Второй график заслуживает особого внимания, потому что он иллюстрирует, каким образом динамика системы может не быть интуитивно ясной. Система состоит из петли управления первого порядка и усилительной петли второго порядка, объединение которых производит неустойчивое колебание. Этот результат можно было бы считать удивительным по трем причинам. Во-первых, каждая из двух петель в отдельности была бы вполне устойчивой. Неустойчивость бывает только тогда, когда они связаны иерархически, с усилителем на втором уровне. Во-вторых, могло бы показаться вполне разумным создание такой системы в обыденной жизни; например, нормы трудовой группы могут удерживать производительность в ограниченном интервале (петля управления первого порядка), поэтому администрация накладывает свою собственную программу стимулирования (усилитель второго порядка); но если реакция администрации на производительность по скорости равна только половине скорости реакции трудовой группы, результаты, как показывает график, могут сильно отличаться от желаемых. В-третьих, статическая формулировка системы не дает и намека на ее динамическую неустойчивость. Действительно, кажется, что в статической формулировке системы (которую можно получить удалением из схемы в 6.12 всех z) две петли в большой степени погашают друг друга, причем усилитель до некоторой степени преобладает. Неустойчивость можно было бы предсказать заранее только тогда, когда у нас есть информация о временных запаздываниях и мы знаем, в частности, что действие усиления длится вдвое дольше действия управления.

Сложные входные сигналы

В предыдущих примерах рассматривались характеристики сигналов на выходе системы, когда применялись простые ступенчатые изменения значения единственной входной переменной. Результаты демонстрировали многообразие выходных сигналов, сложность которых зависела от рассматриваемой системы. Теперь пора признать, что изменения на входах не всегда соответствуют простым ступенчатым функциям. Действительно, во многих социальных системах такое входное изменение может быть скорее исключением, чем нормальным случаем. Кроме того, большинство интересных систем имеет множество причинных переменных, а не просто единственный вход.

Даже простейшая система может порождать сложный выходной сигнал, если она действует от сложного входного сигнала. Это иллюстрируется путем предназначения однооператорной системе, определенной в 6.4, замысловатого входного сигнала. Результатом является столь же замысловатый выходной сигнал, который показывает ту же кривую изменений, что и вход, хотя они и задержаны до более поздних моментов времени и несколько ослаблены действием системы.

6.14. Возможные входной и выходной сигналы для системы из 6.4.



Если простая система, для того чтобы произвести свой выходной сигнал, суммирует значения двух входных переменных, выходной сигнал обычно будет соответствовать не тому или другому входному сигналу, а взвешенному среднему обоих. Хотя обычно мы ожидаем, что такой выходной сигнал будет более сложным, чем каждый из входных сигналов, возможно также, чтобы выходной сигнал имел более простую кривую изменений, если два входных сигнала стремятся погасить друг друга. (Это основная идея, используемая инженерами при проектировании «фильтров» сигнала: для того чтобы избавиться от нежелательного поведения сигнала, образуйте дополнительную кривую и сложите.)

Система с обратной связью замысловатым способом действует на сложный входной сигнал, производя выходной сигнал, который может иметь мало очевидного сходства с каждым входным сигналом или с реакцией системы в виде ступенчатой функции. Причина этого в том, что система реагирует на каждое изменение входного сигнала целым рядом адаптаций. Следовательно, выходной сигнал в любой момент не является единичным исходом, а представляет собой накопление реакций системы на все предшествующие изменения на входе (причем более поздние реакции в устойчивой системе имеют большие веса). Ясно, что если у системы с обратной связью есть множество входов, выходные сигналы бывают более сложными функциями входных сигналов и характеристик системы.

Несмотря на эти осложнения, линейные системы любой степени сложности сохраняют привлекательную особенность. Если у нас есть полное описание входных сигналов и полное описание системы, оказывается возможным вывести полное описание выходного сигнала. Таким образом, динамический анализ систем остается определенной дисциплиной. Может быть, еще важнее то, что любые два из трех описаний (т. е. описаний входных и выходных сигналов или характеристик реакций системы) можно использовать для вывода третьего. Это принцип дает основание для методов идентификации системы, использующих данные, зависящие от времени; они дополняют статические методы, обсуждаемые в гл. 5. Тот же самый принцип позволяет охарактеризовать прошлые входные сигналы исходя из знания системы и ее выходных сигналов — факт, который является фундаментальным при

общении (позволяя нам реконструировать мысли других по тому, что мы видим и слышим) и мог бы позволить исследователям, проводящим социальный анализ, проделывать исторические реконструкции некоторых переменных, применив свое знание данной социальной системы и ее событий-исходов. (Однако надо заметить, что полезные описания сигналов и систем обычно являются математическими выражениями и что такие выводы обычно включают довольно сложные математические процедуры.)

Идентификация системы по данным, зависящим от времени, — это слишком сложная тема для того, чтобы ее здесь обсуждать (см. библиографию к главе), но один общий момент заслуживает внимания. Такая идентификация возможна только тогда, когда имеется информация сразу и о входных, и о выходных сигналах. Когда имеют дело с личностными или социальными системами, и профаны, и ученые в равной степени иногда грешат диагностикой болезней системы только на основании результатов, которые она произвела. Точно так же часто предполагается, что изучение только продуктов системы — ее поведенческой деятельности или ее исторической записи — позволяет распознать ее устройство, «подразумеваемую» индивидуальность или «необходимую» социальную организацию. При описании такого анализа неявно сделаны предположения о входных сигналах от окружающей среды или же ненадежны выводы. Историческое исследование выходных сигналов системы позволяет делать выводы, касающиеся структуры системы, только когда одновременно рассматривается запись входных сигналов, а затем для распутывания данных может потребоваться проведение довольно сложного анализа.

ЗАПУТЫВАЮЩЕЕ ДЕЙСТВИЕ ДИНАМИКИ ПРИ СТАТИЧЕСКОМ АНАЛИЗЕ

Рассмотрев несколько элементарных особенностей динамики системы, мы можем теперь вернуться к связанному с этим спорному вопросу о ее практическом значении для современной социальной науки. Каким же образом «поперечное» исследование на статических моделях запутывается системами, находящимися в момент наблюдения в переходной фазе, а не в устойчивом состоянии? Удобно разделить задачу на два аспекта. Во-первых, мы рассматриваем системы, которые еще не завершили свою реакцию на множество входных сигналов, поддерживаемых на постоянных уровнях. Во-вторых, мы рассматриваем системы, которые являются в отношении продолжительности динамическими, потому что некоторые из их входных сигналов меняются во времени.

Уравновешивание

Данные «поперечного» исследования собираются на многих примерах одной и той же системы путем измерения значений переменных системы для каждого случая в определенный момент.

Входные сигналы по предположению заданы на различных уровнях для различных случаев, и переменные исходов тоже будут обнаруживать изменения при переходе от случая к случаю. Следовательно, кривые изменений при прохождении по переменным и по случаям должны раскрывать действия системы. Однако имеется период ожидания: с того момента, как заданы входные сигналы, до того, как можно наблюдать связанные с ними последствия. Если наблюдения делаются слишком рано, переменные исходов могут показать слабую связь или отсутствие связи со значениями входных сигналов, потому что операторы системы еще не смогли закончить свою работу.

Задачу можно представить в простой форме путем пересмотра системы из 6.4, в которой имеются две переменные, X и Y , и оператор, который производит свое действие после задержки на два временных периода. Предположим, что у входной переменной X задано новое значение для всех случаев в нулевой момент. Ясно, что нет смысла наблюдать значения двух переменных системы в момент один, потому что в этот момент у X есть ее новые значения, тогда как у Y — все еще ее старые значения (которые являются реакциями на предшествующие значения переменной X). Только при ожидании по крайней мере до момента два, когда динамика закончится и случаи «уравновесятся», можно наблюдать значения переменной X и значения переменной Y , которые осмысленно связаны причинным действием. Итак, *если «поперечный» анализ должен быть осмысленным, все наблюдаемые случаи должны уравновешиваться.*

Отдельный случай готов для наблюдения, когда он завершил все реакции на данные входные сигналы и вследствие этого достиг состояния равновесия. Хотя после того, как заданы входные сигналы, всегда имеется период, во время которого случай не готов для статического анализа, но раз этот случай достиг равновесия, то он остается в равновесии до тех пор, пока снова не зададут его входные сигналы. Следовательно, *все уравновешенные случаи можно рассматривать вместе в одном и том же анализе; нет необходимости, чтобы их входные сигналы были заданы в один и тот же момент.*

Время уравновешивания

Продолжительность времени, необходимого для уравновешивания, зависит от периодов задержки отдельных операторов системы. Однако время уравновешивания может значительно превышать периоды задержки для отдельных операторов системы с цепочками воздействий или петлями обратной связи.

«Поперечные» наблюдения рекурсивных систем должны откладываться до тех пор, пока последствия данных выходов не получат возможности пройти по самому длинному пути в системе. «Срединные» переменные в системе нельзя трактовать просто как

входы для следующих переменных, пренебрегая динамическими процессами в предшествующих переменных. Если система не уравновешена, окружение срединных переменных может меняться до того, как будут наблюдаться их причинные воздействия.

Это можно проиллюстрировать простой системой-цепью: $W \rightarrow X \rightarrow Y$, где каждая причинная связь сопровождается запаздыванием на один временной период. Предположим, что W переопределяется в выборке случаев, а после этого сохраняет постоянное значение. Наблюдения, сделанные в момент один, достаточны для оценки воздействия, направленного от W к X , потому что значения переменной W сохранились до момента один, а значения переменной X являются следствиями, произведенными этими самыми значениями W в нулевой момент. Наблюдения в момент один недостаточны для исследования связи $X \rightarrow Y$. Наблюдаемые значения переменной Y являются реакциями на значения переменной X в нулевой момент, но эти значения переменной X не сохранились до момента один. Действительно, наблюдаемые значения переменной X в момент один не имеют логической связи со значениями переменной Y , относящимися к моменту один. С другой стороны, если наблюдения откладываются до момента два, можно исследовать обе связи в системе. Значения переменных W и X остаются такими же, как они были в момент один, и действие $W \rightarrow X$ можно оценить, как и раньше, а значения переменной Y — реакции на значения переменной X в момент один, которые сохранились до момента два, так что можно также оценить действие $X \rightarrow Y$.

Вопрос о времени уравнивания усложняется, когда система содержит обратные связи, потому что теоретически петли поддерживают адаптивное бесконечно. Это могло бы привести к мысли, что такие случаи никогда не уравниваются, потому что они никогда не заканчивают развитие следствий данного входа. Однако ранее указывалось, что адаптации, производимые устойчивой петлей, завершаются для всех практических целей в относительно короткое время. Таким образом, обычно можно определять период, после которого система с устойчивой обратной связью, по существу, статическая в том смысле, что наблюдения, сделанные после этого периода, дают оценки параметров, содержащие ошибки, меньшие некоторой удовлетворительной величины. Допустимая величина ошибки была бы произвольной величиной, которая для различных целей могла меняться, но теоретически ее можно задавать сколь угодно малой. Именно в этом практическом смысле осуществимы проведения статического анализа систем с обратной связью.

Проведения анализа системы с обратной связью для иллюстрации того, как «поперечные» исследования после изменения на входе приводят к ошибкам в оценках параметров, представлены в 6.15. Система в 6.15 простая, а результаты сложные, но даже в этой ситуации есть возможность проникнуть в главные свойства уравнивания, когда имеют дело с системами с обратной связью.

6.15. Предположим, что мы хотим оценить параметры в следующей системе:

$$X \xrightarrow{ax^{-1}} Y \xrightleftharpoons[cx^{-1}]{bx^{-1}} Z.$$

Это описание дает следующие разностные уравнения (при эвристическом предположении, что у Y и Z нет возмущений):

$$Y_t = aX_{t-1} + bcY_{t-2},$$

$$Z_t = abX_{t-2} + bcZ_{t-2}.$$

Мы предполагаем, что система находится в статическом состоянии до момента один и, в частности, что $X_i = X_0$ для $i < 0$. Тогда в момент один значение X меняется на новое, которое сохраняется в дальнейшем, так что $X_i = X_1$ для $i > 1$. Вследствие причинных задержек на значение переменной Y не будет влиять изменение значения X до момента два, а на значение переменной Z не будет влиять до момента три. Следовательно, до этих моментов Y и Z будут зависеть только от X_0 :

$$Y_i = \frac{aX_0}{1-bc} \quad \text{для } i < 2,$$

$$Z_i = \frac{abX_0}{1-bc} \quad \text{для } i < 3.$$

Эти описания позволяют нам привести следующие формулы для выражения значения переменных выхода в различные моменты:

Момент времени	Y	Z
1	$\frac{aX_0}{1-bc}$	$\frac{abX_0}{1-bc}$
2	$a(X_1 - X_0) + \frac{aX_0}{1-bc} =$ $= \frac{a}{1-bc} [X_1(1-bc) + bcX_0]$	$\frac{abX_0}{1-bc}$
3	$\frac{a}{1-bc} [X_1(1-bc) + bcX_0]$	$\frac{ab}{1-bc} [X_1(1-bc) + bcX_0]$
4	$\frac{a}{1-bc} [X_1(1-b^2c^2) + b^2c^2X_0]$	$\frac{ab}{1-bc} [X_1(1-bc) + bcX_0]$
5	$\frac{a}{1-bc} [X_1(1-b^2c^2) + b^2c^2X_0]$	$\frac{ab}{1-bc} [X_1(1-b^2c^2) + b^2c^2X_0]$
6	$\frac{a}{1-bc} [X_1(1-b^3c^3) + b^3c^3X_0]$	$\frac{ab}{1-bc} [X_1(1-b^2c^2) + b^2c^2X_0]$

Задав данные о группе случаев, мы могли бы оценить значение параметра b с помощью формулы из 5.12:

$$b = \frac{\sigma_{ZX}}{\sigma_{YX}}.$$

Для применения этой формулы в различные моменты после изменения переменной X мы должны получить выражения для ковариаций в моменты два, три,

четыре и т. д. (используя процедуры, подобные описанным в 4.17). Результаты таковы:

Моменты времени	σ_{yx}	σ_{zx}
2	$\frac{a}{1-bc} [\sigma_{X(1)}^2 (1-bc) + bc\sigma_{X(0)X(1)}]$	$\frac{ab}{1-bc} \sigma_{X(0)X(1)}$
3	$\frac{a}{1-bc} [\sigma_{X(1)}^2 (1-bc) + bc\sigma_{X(0)X(1)}]$	$\frac{ab}{1-bc} [\sigma_{X(1)}^2 (1-bc) + bc\sigma_{X(0)X(1)}]$
4	$\frac{a}{1-bc} [\sigma_{X(1)}^2 (1-b^2c^2) + b^2c^2\sigma_{X(0)X(1)}]$	$\frac{ab}{1-bc} [\sigma_{X(1)}^2 (1-bc) + bc\sigma_{X(0)X(1)}]$
5	$\frac{a}{1-bc} [\sigma_{X(1)}^2 (1-b^2c^2) + b^2c^2\sigma_{X(0)X(1)}]$	$\frac{ab}{1-bc} [\sigma_{X(1)}^2 (1-b^2c^2) + b^2c^2\sigma_{X(0)X(1)}]$
6	$\frac{a}{1-bc} [\sigma_{X(1)}^2 (1-b^3c^3) + b^3c^3\sigma_{X(0)X(1)}]$	$\frac{ab}{1-bc} [\sigma_{X(1)}^2 (1-b^2c^2) + b^2c^2\sigma_{X(0)X(1)}]$

Таким образом, оценками коэффициента b по данным, собранным в эти моменты (после упрощения), должны быть:

Моменты времени	Оценка коэффициента b
2	$b \left[\frac{\sigma_{X(0)X(1)}}{(1-bc)\sigma_{X(1)}^2 + bc\sigma_{X(0)X(1)}} \right]$
3	b
4	$b \left\{ \frac{(1-bc)\sigma_{X(1)}^2 + bc\sigma_{X(0)X(1)}}{[1-(bc)^2]\sigma_{X(1)}^2 + (bc)^2\sigma_{X(0)X(1)}} \right\}$
5	b
6	$b \left\{ \frac{[1-(bc)^2]\sigma_{X(1)}^2 + (bc)^2\sigma_{X(0)X(1)}}{[1-(bc)^3]\sigma_{X(1)}^2 + (bc)^3\sigma_{X(0)X(1)}} \right\}$

Сначала надо рассмотреть то, что, по-видимому, является неумышленно найденным результатом. Казалось бы, что если мы надлежащим образом выбираем наш расчет времени, мы можем получить несмещенные оценки параметров петли, даже когда система была совсем неуравновешенной. В 6.15 оценка параметра b в точности та, которая нужна, в моменты три и пять. Беда в том, что мы, как правило, не знаем, делались ли наши наблюдения в

«нужный» момент, потому что не было необходимой информации о причинных задержках в системе. Таким образом, оценки, полученные из таких неуравновешенных случаев, могли бы быть смещенными или несмещенными, но мы бы не знали, какими же именно. Кроме того, в более сложной системе со многими входами и многими петлями таких подходящих моментов для проведения наблюдений даже не существует.

Факт, что несмещенные оценки появляются в дискретные моменты, важен. Он показывает, что наблюдения неуравновешенных случаев могут давать оценки параметров, которые крайне разбросаны. Повторное рассмотрение той же самой группы случаев во время их раннего периода развития могло бы навести на мысль, что сама система нестационарна, ее параметры находятся в состоянии постоянного изменения или колебания. Эта нестационарность не является реальной, а вызывается ошибками при оценке, и мы хотим исключить из анализа неуравновешенные случаи для того, чтобы избежать именно таких проблем. Кроме того, этот же факт означает, что консерватизм оправдан при решении вопроса о том, стала ли система нестационарной. Когда последовательные проведения «поперечного» анализа системы обнаруживают, что система стала меняющейся во времени, возможно, что система осталась стационарной, но входные сигналы переменялись таким образом, что случаи не уравниваются.

Ошибки при оценивании параметров устойчивой системы с обратной связью по «поперечным» данным уменьшаются путем ожидания до тех пор, пока система не пройдет через большое количество циклов адаптации, сопровождающих ступенчатые изменения на входах. Результаты из 6.15 ясно показывают, что улучшение не обязательно монотонно (потому что в этом примере b можно было бы оценить без ошибки в моменты три и пять). С другой стороны, наихудшие возможные ошибки со временем убывают. Это можно увидеть для системы из 6.15 путем приписывания подходящих значений главным неизвестным с последующим рассмотрением значения оценки в различные моменты; если, например, $(b \cdot c) = 0.5$, $\sigma_{X(0)X(1)} = 0$, $\sigma_{X(1)}^2 = 1.0$. Величина в скобках тогда равна нулю в момент два, 0,67 — в момент три и 0,86 — в момент пять. Убывающая разность между этими числами и 1,0 показывает, что максимальная ошибка при оценке b меньше в более поздние моменты.

Количество циклов, необходимое для уравнивания, несколько меняется в зависимости от систем и различных обстоятельств.

Чем ближе обратные эффекты петлей к нулю, тем быстрее уменьшаются ошибки оценивания и системы с сильной обратной связью имеют большие периоды уравнивания. К тому же результаты из 6.15 показывают, что смещения в оценках зависят от статистических величин $\sigma_{X(1)}^2$ и $\sigma_{X(0)X(1)}$. Хотя затронуты подробные толкования этих результатов, один важный результат

можно сформулировать просто. Чем больше изменения, которые сделаны на входах, тем больше времени требуется для того, чтобы случаи уравнились снова.

Входные сигналы, меняющиеся со временем

Логическое оценивание структурных коэффициентов по «поперечным» данным требует, чтобы значения входов сохранялись на постоянных уровнях. Однако в естественных ситуациях значения входов часто не остаются совершенно стабильными во времени. Значения следствия Y , соответствующие значениям причины X , могут наблюдаться в данный момент, но относящиеся к делу значения X имели место в прошлом и могут больше не наблюдаться, когда появляются их следствия Y . Следовательно, «поперечные» данные не содержат необходимой информации для статического анализа. Они не показывают точно, как настоящие значения Y связаны со значением X .

Если значения входов полностью непредсказуемы для одного момента по другому, нет возможности осмысленной оценки параметра по «поперечным» данным, но во многих действительных ситуациях все не так плохо. Например, график меняющейся во времени входной переменной X в 6.14 показывает, что, хотя X действительно меняется, она все еще имеет тенденцию оставаться в ограниченной области. Она сохраняет «по существу постоянное» среднее значение по времени. Фактически то же самое верно для переменной-исхода Y . Она также сохраняет по существу постоянное среднее значение по времени. Кроме того, кажется, что среднее значение переменной Y связано со средним значением переменной X структурным коэффициентом для причинного отношения (в этом примере 0,5).

Давайте говорить, что X в 6.14, по существу, постоянный сигнал с наложенным на него сигналом, меняющимся во времени. Таким образом, наблюдение переменной X в любой момент обычно измеряет среднее значение X не непосредственно, а с некоторой ошибкой. Конечно, шумовой сигнал, наложенный на X , — не ошибка в обычном смысле, он действительно производит колебания в Y . Но переменная во времени компонента не влияет на Y до тех пор, пока мы не завершили наблюдение, так что для практических целей временное изменение X может трактоваться, как будто это просто «ошибка измерения». Принимая эту точку зрения, мы становимся перед более известной задачей: оценивание параметров системы, когда переменные измеряются неточно.

Два использованных здесь основных предположения состоят в том, что сигнал X можно разделить на две компоненты — одна постоянная и одна меняющаяся во времени — и что для определения структуры системы их можно анализировать независимо. Фактически можно наглядно показать математически, что любой сигнал делится на совокупность составляющих сигналов, включающую, в частности, постоянную составляющую. Можно также

показать, что выходной сигнал, порождаемый линейной системой, действующей от составного входного сигнала, равняется сумме выходных сигналов от таких же систем, действующих от компонент этого составного входного сигнала. Доказательства обоих принципов даются в учебниках по теории систем (смотрите библиографию к этой главе). Таким образом, имеется оправдание для попытки выявить устройство системы путем анализа постоянной составляющей входного сигнала. Путем анализа этой постоянной составляющей определяется та же самая структура, которая выявлялась бы, если бы мы имели дело с полным входным сигналом во всей его сложности.

Этой аргументацией задачу входных сигналов, меняющихся во времени, можно перевести в задачу предопределенных переменных, измеряемых с ошибкой, допускаемая при этом, что постоянные составляющие сигналов меняются в зависимости от случаев. Тогда задачу входных сигналов, изменяющихся во времени, можно атаковать несколькими способами. Во-первых, так как постоянные составляющие сигналов равняются средним значениям по времени, мы могли бы фактически расширить план исследования до включения повторных наблюдений входных сигналов, которые затем могли бы усредняться. Конечно, такой план в действительности не «поперечный», и для получения точных мер временного среднего требуется большое число «продольных» наблюдений на входах. Однако обратите внимание на то, что вовсе нет необходимости получать повторные наблюдения выходных переменных. Их временные колебания просто образуют «возмущения», которые не смещают оценок параметров. Кроме того, повторные наблюдения входов могли бы делаться после того, как измерены выходы; пока можно предполагать, что средние значения входов остаются неизменными.

Иногда можно манипулировать окружениями для задания постоянной составляющей переменной на определенном уровне для всех членов совокупности, подвергающейся действию этого окружения. В таком случае значение постоянной составляющей может оцениваться по среднему значению «поперечного сечения» наблюдений; например, можно было бы предъявлять членам однородной группы во время интервью установочные стимулы и измерять их установочные реакции. Постоянная составляющая отношения к стимулу, характеризующая любого члена группы, могла бы тогда оцениваться по среднему значению результатов измерений реакций всех членов группы, что давало бы возможность переходным процессам индивидуумов погашать друг друга. Необходимыми условиями для применения этой тактики являются стандартный генерирующий элемент окружения (например, установочный стимул), уверенность в том, что все наблюдаемые единицы эффективно подвергаются действию генератора (например, при представлении его в явном виде во время интервью), уверенность в том, что все единицы эквивалентны по реакции (например, при исследовании только членов однородной группы) и

отсутствие синхронизации переходных процессов (что можно было бы регулировать в опросе установки путем изменения моментов времени, обстоятельств и формы вопросов интервью).

До некоторой степени аналогичные процедуры служат основой «экологической корреляции», при которой географические популяции используются для получения множества средних значений каждой из нескольких переменных, так что можно анализировать связи переменных. Задачи при географическом подходе к агрегату состоят в том, что популяция в регионе может не быть однородной, все члены ее могут по-разному подвергаться действию стандартных генерирующих элементов и в определенный момент переходные процессы членов могут синхронизироваться. Вдобавок, так как генерирующий фактор неявно предполагается, но часто полностью не описан, нельзя быть уверенным в том, что данный фактор задает значения только одной из исследуемых переменных. Если один и тот же генератор служит фактором у нескольких переменных, наблюдаемые корреляции бывают до некоторой степени ложными. Ссылки на обсуждения агрегата и экологической корреляции приведены в библиографии главы.

Третьим подходом является знакомая стратегия использования инструментальных переменных, когда сталкиваются с неполно измеренной причинной переменной, а именно для каждой из колеблющихся входных переменных мы находим другую переменную, которая предсказывает ее среднее значение по времени (и причинно не связана с выходами системы кроме как через этот вход). Таким образом, путем сдвига множества входных переменных назад на один шаг можно получить несмещенные оценки всех интересующих структурных коэффициентов, несмотря на меняющиеся во времени входные сигналы. Выбор инструментов обсуждался на протяжении гл. 5, те же самые принципы применяются и здесь с учетом того, что значения инструмента должны коррелировать с постоянной составляющей входного сигнала, а не с составляющей, меняющейся со временем.

Динамика запаздывающих переменных

Иногда возможно построить инструмент по запаздывающим наблюдениям самой входной переменной. Хотя эта тактика общая, ее применимость ограничивается особыми обстоятельствами, как иллюстрируется 6.16 (и обсуждением запаздывающих переменных в качестве инструментов в гл. 5).

6.16. При данной простой статической модели

$$X \xrightarrow{a} Y \leftarrow U$$

переменную Y можно выразить как функцию временного ряда значений переменной X :

$$Y_t = \left(\sum_{i=0}^{\infty} a_i X_{t-i} \right) + U,$$

где $\sum_{i=0}^{\infty} a_i = a$, причем для упрощения предполагается, что возмущения для Y постоянны во времени. Предположение, что ряд переменных X имеет «по существу постоянную» составляющую, имеет тот смысл, что значение X в любой момент k можно выразить следующим образом:

$$X_k = \bar{X} + x_k,$$

где $\mathcal{E}(x) = 0$. При этом определении Y_t принимает вид:

$$Y_t = \left(\sum_{i=0}^{\infty} a_i (\bar{X} + x_{t-i}) \right) + U = a\bar{X} + \left(\sum_{i=0}^{\infty} a_i x_{t-i} \right) + U.$$

Предположим, что мы наблюдаем значения переменной X в некий более ранний момент $(t-k)$. Предположим также, что колебания постоянной составляющей переменной X не коррелированы с меняющимися во времени колебаниями переменной X и что возмущения переменной Y не коррелированы с X . Тогда

$$\begin{aligned} \sigma_{X(t) X(t-k)} &= \sigma_{\bar{X}}^2 + \sigma_{x(t) x(t-k)} \\ \sigma_{Y(t) X(t-k)} &= a\sigma_{\bar{X}}^2 + \sum_{i=0}^{\infty} a_i \sigma_{x(t-i) x(t-k)} \end{aligned}$$

Если бы X_{t-k} использовались в качестве инструмента, оценкой коэффициента a была бы

$$\frac{\sigma_{Y(t) X(t-k)}}{\sigma_{X(t) X(t-k)}} = \frac{a\sigma_{\bar{X}}^2 + \sum_{i=0}^{\infty} a_i \sigma_{x(t-i) x(t-k)}}{\sigma_{\bar{X}}^2 + \sigma_{x(t) x(t-k)}}.$$

Это давало бы несмещенную оценку структурного параметра, если бы действие X на Y происходило сразу, так что $a_i = 0$ для всех значений i , кроме одного. Однако оценка была бы смещенной в более общем случае, в котором воздействие переменной X на Y распределены во времени, если меняющаяся во времени составляющая переменной X не имеет нулевой автокорреляции, а выбор моментов времени запаздывающих переменных не таков, что $a_k = 0$.

Таким образом, наблюдения входной переменной в различные моменты не должны использоваться в качестве инструмента, если (а) причинные воздействия не происходят сразу, внезапно или если причинные воздействия распределены во времени, (б) меняющаяся во времени составляющая входного сигнала совершенно непредсказуема для одного момента по другому. В последнем случае мы все еще должны быть уверены в том, что период между измерением зависимой переменной и инструментальной переменной не соотносится какому-нибудь периоду причинной задержки между причиной и исходом.

Статический и динамический разброс

Структурные коэффициенты для системы, которая, по существу, является статической (в том смысле, что ее переменные сохраняются на уровнях постоянных средних значений по времени)

можно оценить по «поперечным» данным с помощью указанных выше методов. Однако точная оценка возмущений зависимых переменных не получается обычным путем, когда входные сигналы имеют составляющие, меняющиеся во времени, потому что необъясненный разброс зависимой переменной представляет не только действия не указанных точно переменных (т. е. обычные возмущения), но и запаздывающие воздействия от меняющихся во времени составляющих точно указанных входных переменных. Теоретически последнее объяснялось бы моделью, если бы она применялась динамически.

Предположим, например, что занятие творческой работой заставляет человека стать более либеральным (как предполагается в социологической литературе). Вследствие различия профессий среднее количество выполняемой творческой работы у людей колеблется, и вследствие этой по существу постоянной составляющей в причинной переменной среди индивидуумов имеется разброс. Его можно было бы использовать для оценки структурного коэффициента, связывающего творческую работу с либерализмом. С другой стороны, количество творческой работы, выполняемой индивидуумом, может сильно меняться от недели к неделе или от месяца к месяцу вследствие меняющихся условий труда, потребностей личной жизни или изменений настроения. Эти колебания тоже должны влиять на либерализм. Действительно, чистое действие такого временного колебания уровня творческой работы определяется тем же самым структурным коэффициентом, который можно оценить по постоянной составляющей данных. Тем не менее, если действие происходит только после задержки, изменяющаяся во времени составляющая либерализма при «поперечном» исследовании, по-видимому, будет необъясненной. К моменту, когда либерализм приспособился к уровню творческой работы, уровень творческой работы мог измениться, так что либерализм индивидуума может оказаться несогласованным с причинной переменной, потому что он отражает действие прошлого уровня творческой работы. Просто при данном наблюдении в этот единственный момент мы не можем отличить это объяснимое отклонение от возмущения, вызванного совершенно посторонними и точно не указанными причинами. Таким образом, изменение зависимой переменной, которое обусловлено прошлыми колебаниями причинной переменной, группируется с изменениями, обусловленными точно не указанными возмущениями, а описанная причинная связь, по-видимому, объясняет меньшую часть изменения зависимой переменной, чем в действительности.

Эта задача труднее решается в более простых системах. Если у нас есть единственная причинная связь с запаздыванием, меняющаяся во времени составляющая входного сигнала будет довольно точно передаваться переменной-следствию. Более сложная рекурсивная система, в которой выходы являются реакциями на входные сигналы, в нескольких моментах времени порождает средние значения входных сигналов по времени, тем самым начиная

исключать их колебания. Следовательно, выходные сигналы содержат меньший разброс, обусловленный входными колебаниями. Выходной сигнал петли обратной связи в любой данный момент является взвешенным средним прошлых значений входных сигналов. Значит, действия меняющихся во времени составляющих входных сигналов имеют тенденцию усредняться, а на разброс следствия колебания входных сигналов влияют лишь умеренно. Однако даже при устойчивых системах с обратной связью в выходных сигналах будет обычно некоторый динамический разброс, связанный с самыми последними колебаниями выходных сигналов.

По той же самой причине система, которая объясняет большую часть колебаний своих переменных-следствий при статических условиях, по-видимому, дает менее адекватную модель, если входные сигналы непрерывно изменяются. Действительно, если временные изменения входных сигналов становятся значительно большими, чем постоянное изменение по случаям, формулировка системы, очевидно, может потерять почти всю свою способность объяснять. Однако возможно, а при таких обстоятельствах даже весьма вероятно, что система продолжает определять большую часть изменений следствий. При использовании только «поперечных» данных проблема состоит в том, что мы не можем отличить временные изменения выходных сигналов от необъяснимых возмущений.

Входные сигналы для многих социальных систем, вероятно, имеют некоторые меняющиеся во времени составляющие. Следовательно, теоретические модели (особенно более простые) могут недооцениваться вследствие их слабой способности объяснения разброса «поперечных» данных. Таким образом, если даже структурные коэффициенты можно оценить из «поперечных» исследований, в конечном счете мы должны обратиться к экспериментальным, «продольным» и модельным исследованиям, в которых время рассматривается в явном виде для того, чтобы определить динамические аспекты устройства системы и с уверенностью оценить теорию.

ИСТОЧНИКИ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Элементарная трактовка динамики систем дается в Alpha C. Chiang. *Fundamental Methods of Mathematical Economics* (New York, McGraw-Hill, 1967), особенно в частях 3 и 5. James A. Cadzow. *Discrete-Time Systems: An Introduction With Interdisciplinary Applications* (Englewood Cliffs, N. J., Prentice-Hall, 1973) — ценная книга, где обсуждаются системы более высокого уровня; подобно всем книгам по динамике систем она включает математику, но автор использует элементарную алгебру и адресуется к начинающим студентам. Методы распространения анализа потоковых графов на область динамики представлены (вместе с примерами) в W. H. Huggins and Doris R. Entwistle. *Introductory Systems and Design* (Waltham, Mass, Blaisdell, 1968).

Статистический анализ временных рядов, для планирования и идентификации системы, рассматривается в Gwilym M. Jenkins and Donald G. Watts. *Spectral Analysis and Its Applications* (San Francisco, Holden-Day, 1968) и George E. P. Box and G. M. Jenkins*. *Time Series Analysis: Forecasting and Control* (San Francisco, Holden-Day, 1970). Краткие сведения по тем же самым вопросам имеются в статьях Douglas A. Hibbs, Jr., *Problems of Statistical Estimation and Causal Inference in Time — Series Regression Models* и Thomas F. Mayer and William Ray Arney. *Spectral Analysis and the Study of Social Change* — гл. 10 и 11 соответственно в Herbert L. Costner (Ed.). *Sociological Methodology: 1973—1974* (San Francisco, Jossey-Bass, 1974). В работе Daniel Graupe. *Identification of Systems* (New York, Van Nostrand Reinhold, 1972) дается более углубленная трактовка методов, используемых в технике.

Недавно психологи разработали модели для описания кривых роста по данным, полученным на группах людей. Некоторые главные результаты в этой области обсуждаются в John R. Nesselroade and Hayne W. Reese (Eds.). *Life-Span Developmental Psychology: Methodological Issues* (New York, Academic, 1973). Множество социологических временных рядов описывается и анализируется в книге: Robert L. Hamblin, R. Brooke Jacobsen and Jerry L. L. Miller. *A Mathematical Theory of Social Change* (New York, Wiley, 1973).

Процедуры агрегирования обсуждаются в работах Mattei Dogan and Stein Rokkan (Eds.). *Quantitative Ecological Analysis in the Social Sciences* (Cambridge, Mass., MIT Press, 1969) и Michael T. Hannan. *Aggregation and Disaggregation in Sociology* (Lexington, Mass., Heath, 1971).

УПРАЖНЕНИЯ

1. Рассмотрим такую специфическую социологическую тему, как теория равновесия установок, социометрический анализ или модернизация. «Объем исследований» (R) по этой теме в данном году можно измерить как общее число исследователей, которые уделили теме некоторое минимальное количество времени. Полное количество публикаций (P) за год приблизительно пропорционально объему исследований по теме около трех лет назад (учитывая время подготовки рукописи и задержку публикации). Публикации имеют тенденцию порождать больший объем исследований примерно через два года (время, в течение которого читатели обнаруживают статьи и планируют свои собственные). Между тем федеральная политика направления денежных средств (F) на расширение исследований по теме учитывает приблизительно двухлетнее освещение этого курса по радио и приготовление и обработку предложений. На соответствие федеральных денежных средств потребностям может влиять оценка того, будут ли дотации определяться с учетом будущих публикаций. При предположении, что это так, федеральные вложения средств могут быть пропорциональными скорости роста (G) литературы по теме, измеряемой, скажем, как разность между числом публикаций за последний год и числом публикаций пятилетней давности. В заключение можно предположить, что главным источником нового исследования

* Имеется русский перевод. Бокс П., Дженкинс Дж., *Анализ временных рядов*. М., Мир, 1974. — Примеч. пер.

по теме служит проникновение (I) новой идеи, метода или точки зрения (нововведение) из довольно несвязанной области исследования. Сам входной сигнал, вероятно, лучше всего рассматривать как относительно кратковременный импульс, который должен подкрепляться деятельностью в новой области применения. Более важными нововведениями (имеющими большие значения I) были бы те, которые порождают больший объем исследований через год или около того после их внедрения.

(а) Составьте потоковый граф для этой системы. (Значение изменения G можно представить на линейном графике.) Снабдите каждый структурный коэффициент α нижним индексом, дайте в скобках его знак и сопроводите оператором задержки с соответствующим индексом; например, подписью к одной стрелке может быть $\alpha_{RI}(+)z^{-1}$. Возмущениями пренебрегите.

(б) На основании рассмотрения схемы опишите явные и скрытые функции федеральных органов по вложению средств в развитие научной литературы по данной теме.

(в) Через какой минимальный промежуток времени исследование, проведенное теперь, стимулирует новое исследование?

2. Пусть для упрощения абсолютное значение каждого коэффициента в системе из упражнения 1 равно 1,0.

(а) Изобразите снова приведенный потоковый граф, в котором исключена переменная G .

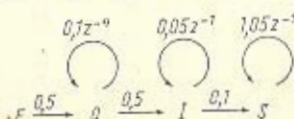
(б) Выпишите разностные уравнения, которые определяют динамическую систему; например,

$$R_t = I_{t-1} + P_{t-2} + F_{t-2}$$

(в) Используйте разностные уравнения для прослеживания 20-летнего влияния единичного импульса в I в нулевой момент (а именно I имеет значение 1,0 при t_0 и значение 0 после этого). Допустим, что R , P и F измеряются как значения отклонений и что тема привлекла «среднее» внимание и имела «средние» количество публикаций и величину вложения средств за последние несколько лет, отражаемые приписыванием значения нуль всем переменным (кроме I) для t_0 и соответствующего числа лет до него. Значения переменных при t_0 и через год показаны ниже:

Время	I	R	P	F
0	1	0	0	0
1	0	1	0	0

3. Предположим, что следующая динамическая модель представляет связи во времени образования индивида (E), его профессионального уровня (Q), годового дохода (I) и сбережений (S) при игнорировании возмущений:



т. е. образование определяет профессиональный уровень, который оказывает прямое воздействие на доход, а годовой доход создает небольшой прирост сбережений. Петля на Q означает, что приблизительно каждые четыре года возрастает профессиональный уровень. Петля на I показывает, что индивид получает основную долю заработной платы в соответствии с профессиональным уровнем плюс процент от дохода прошлого года. Петля на S показывает, что сбережения и реальное имущество сохраняются и увеличиваются процентами на капитал, и повышением стоимости капитала. Допустим, что $t = 0$ является годом, в котором формальное образование и обучение завершаются, и что в этот

момент переменные имеют следующие значения: E_0 — указывается; Q_0 — зависит от E_0 , нуль до $t = 0$; I_0 — зависит от Q_0 , нуль до $t = 0$; S_0 — зависит от I_0 ; для того чтобы представить долги за образование и за домашнее хозяйство, мы делаем ее зависящей от E_0 :

$$S_0 = 0,10I_0 - 0,25E_0.$$

(а) Напишите разностные уравнения, которые выражают значение каждой переменной через настоящие и прошлые значения переменных системы.

(б) Далее выписано уравнение для Q в момент $(t-4)$:

$$Q_{t-4} = 0,50E_{t-4} + 0,10Q_{t-3}.$$

Подставьте это уравнение в уравнение для Q_t , для того чтобы дать другое выражение для значения Q в момент t . Как далеко могли бы заходить такие подстановки? Дает ли это другой способ представления петли на Q ?

(в) Предположим, что образование полагается равным 10,0 в нулевой момент и потом не меняется. Используйте три разностных уравнения для того, чтобы проследить рост Q , I и S в течение 15 лет (вспомните долги). Здесь была бы полезна настоящая вычислительная машина. Теперь допустим, что образование задается как 1,0 в нулевой момент и остается потом без изменений. Вычислите значения переменных на перспективу в 25 лет. Это утомительно, но результаты поучительны и используются ниже. (Ответы для этого второго проведения анализа не даются.)

(г) Предположим, что уровень образования с оценкой 10 есть «доктор философии» и что обычно его достигают в возрасте 26 лет. Предположим, что уровень образования I означает два года средней школы* и достигается в 16 лет. В каком возрасте доктор догоняет по сбережениям выпускника средней школы при данной выше системе? Достигнет ли доктор когда-нибудь значительно большего уровня сбережений, чем выпускник средней школы?

4. Мы продолжаем иметь дело с системой, определенной в упражнении 3. Теперь предположим, что добавочное образование приводит к неудовлетворенности общественной системой, тогда как больший социально-профессиональный статус, больший доход и большие сбережения приводят к позициям поддержки общественной системы. В частности, предположим ради упрощения, что отношение к общественной системе определяется следующим образом:

$$A = -0,2E + 0,2Q + 0,2I + 0,4S.$$

(Предполагается, что эти воздействия на установку почти мгновенны.)

(а) Начертите графики отношений к существующему строю с возраста 26 лет до возраста 40 лет доктора философии и выпускника средней школы (используйте результаты из упражнения 3).

(б) Представим себе тип поселения (характеризуемый возрастом, образованием, профессиональным уровнем и т. д.), в котором наиболее вероятно обнаружить анархиста, «прогрессивного», или либерала, и консерватора. Вероятно ли, что люди более низкого класса — радикалы (правые или левые) в молодости? В старости?

5. Рассмотрим следующую упрощенную систему «классовых конфликтов», в которой коэффициентам приписаны произвольные значения, а возмущениями пренебрегают:

$$P \xrightarrow{0,5} I \xrightleftharpoons[0,42^{-1}]{0,42^{-1}} L \xrightleftharpoons[-0,8]{0,42^{-1}} D.$$

По этому описанию излишки производства (P) увеличивают общественное неравенство (I). Воздействие почти мгновенно по сравнению с временной шкалой других действий в системе (это эвристическое упрощение). Рост общественного неравенства порождает покровительствующее привилегированным слоям законодательство (L), дающее временное запаздывание для того, чтобы привилегированный класс манипулировал составом и мнениями тех, кто исполняет роли вла-

сти. Такое законодательство тоже увеличивает общественное неравенство, опять предоставляя время для планирования и осуществления социальных эксплуататорских программ. С другой стороны, значительное усиление дискриминационных законов вызывает рост возмущения граждан (D), как только понимание законов распространяется в массы. Те, кто находятся у власти, быстро реагируют на такие угрозы общественному порядку смягчением формы или применения законов, чтобы сделать их менее возбуждающими.

Анализ посредством потоковых графов можно использовать для отыскания передаточной функции от P к каждой переменной системы; например, в случае переменной I мы могли бы получить первое выражение из приведенных ниже, которым можно манипулировать алгебраически, как показано:

$$T_{IP} = \frac{0,5(1 + 0,32z^{-1})}{1 + 0,32z^{-1} - 0,16z^{-2}} = \frac{0,50z^2 + 0,16z}{z^2 + 0,32z - 0,16} = \frac{0,50z^2 + 0,16z}{(z - 0,271)(z + 0,591)} = \frac{0,343z}{(z - 0,271)} + \frac{0,157z}{(z + 0,591)}.$$

Это выражение является «разложением на элементарные дроби» передаточной функции (по поводу затрагиваемого метода см. библиографию к данной главе). Оно показывает, что, насколько это касается I , первоначальную систему можно заменить суммой двух более простых систем, а результаты будут в точности теми же самыми. Кроме того, мы сразу же можем увидеть, что обе более простые системы устойчивы: обратный эффект единственной петли в каждой находится между $-1,0$ и $+1,0$. Таким образом, мы знаем, что первоначальная система устойчива. Вышеприведенное выражение дает нам возможность использовать идею «преобразования z » из развитых разделов теории систем (опять см. библиографию). Применение преобразования z дает нам выражение значения I через настоящие и прошлые значения P :

$$I_t = \sum_{i=0}^{\infty} [0,343(0,271)^i + 0,147(-0,591)^i] P_{t-i} = 0,50P_t + 0,00P_{t-1} + 0,08P_{t-2} + 0,03P_{t-3} + 0,02P_{t-4} - 0,01P_{t-5} \dots$$

Тем же методом получаем:

$$L_t = \sum_{i=0}^{\infty} [0,232(0,271)^i - 0,232(-0,591)^i] P_{t-i} = 0,00P_t + 0,20P_{t-1} + 0,06P_{t-2} + 0,05P_{t-3} - 0,03P_{t-4} + 0,02P_{t-5} \dots$$

$$D_t = \sum_{i=0}^{\infty} [-0,499\delta + 0,342(0,271)^i + 0,157(-0,591)^i] P_{t-i} =$$

(где $\delta = 1$, когда $i = 0$, $\delta = 0$ в противном случае)

$$= 0,00P_t + 0,00P_{t-1} + 0,08P_{t-2} - 0,03P_{t-3} + 0,02P_{t-4} - 0,01P_{t-5} \dots$$

При всех этих суммированиях коэффициенты для членов после P_{t-5} при округлении до второго знака после запятой равны нулю.

(а) Используйте приведенные выше формулы суммирования для оценки значений I , L и D в нулевой момент, если P имела значение 1,0 от момента $(t-5)$ до момента t включительно.

(б) Сравните результаты, полученные в (а), с результатами, которые вы получили из статистического анализа той же самой системы, опять предполагая, что P имеет постоянное значение 1,0. Какой вывод вы можете сделать о времени уравнивания для этой системы?

6. Траектории изменения во времени единственной переменной представляют существенный интерес при исследовании физических систем; например, механические колебания с определенной траекторией во времени образуют характерный звук — факт, по важности основополагающий в технике. В общественных науках, вероятно, большее значение придется многомерной конфигурации состояний в

* Что соответствует 10 классам средней школы в СССР. — *Примеч. пер.*

единственный момент или в течение короткого периода; например, это в значительной степени относится к рассмотрению событий «поперечного сечения» общества, относительно которого мы решаем, индустриализуется ли оно, находится ли в состоянии войны, демократическое ли оно и т. д. Однако теоретически многомерное сечение выходных переменных всегда можно переопределить через строго динамические траектории входов системы. Предположим, что мы определяем популистское преобразование, или революцию, как период сильного гражданского возмущения, после которого вскоре следует крушение общественного неравенства. При данной системе с одним входом, описанной в упражнении 5, возможно достичь такого преобразования, только если излишки производства имели определенную историческую траекторию.

(а) Используйте формулы суммирования из упражнения 5 для определения того, какими должны быть свойства этой исторической траектории. (Допустим, что приведенное выше определение требует большого значения D_{-1} и малого значения для I_0 .)

(б) Оказывается, что преобразование, которое определено выше, является в значительной степени эпифеноменом в системе. Крушение неравенства больше обусловлено экономическими обстоятельствами, чем предшествующими гражданскими возмущениями. Предположим вместо этого, что мы определяем преобразование как период гражданского возмущения, сопровождаемый по меньшей мере еще двумя периодами, в продолжение которых законы, защищающие интересы привилегированных слоев, будут ослаблены (D_{-1} велико, L_0 и L_1 малы). Какова теперь необходимая историческая траектория излишков?

7. Можно считать, что установки имеют относительно устойчивую составляющую, основанную на накопленной социализации, а также неустойчивую составляющую, которая меняется в зависимости от последующего опыта. Таким образом, в определенный момент мы могли бы представить чистое отношение индивидуума к объекту как суммирование устойчивой и неустойчивой составляющих:

$$A_S \rightarrow A_N \leftarrow A_T$$

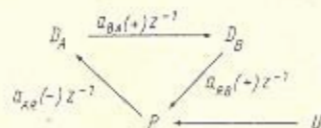
Если бы мы должны были рассматривать того же самого индивидуума через короткий промежуток времени, мы ожидали бы, что устойчивая составляющая будет почти той же самой. Однако неустойчивая составляющая могла бы измениться и, таким образом, чистое отношение тоже могло бы стать другим.

(а) Напишите выражение для дисперсии величины A_N при «поперечном» исследовании совокупности. Дайте словесную интерпретацию компонентов дисперсии и укажите некоторые социальные условия, при которых одна компонента могла бы быть больше другой.

(б) Видоизмените приведенную выше схему для того, чтобы учесть избирательность — люди участвуют именно в тех событиях, которые подкрепляют их склонности. Как это явление влияло бы на наблюдаемую дисперсию при «поперечном» исследовании?

(в) Дайте формулу для корреляции между измерениями установки, сделанными в два момента, разделенные коротким интервалом. (Делая вывод, пренебрегите возможностью избирательности.) Какие социальные условия дали бы относительно большое значение для корреляции? Допустим, что установки в данный момент определяют поведение по отношению к объекту установки в тот же момент. Какие социальные условия повысили бы точность прогнозов по установкам, измеряемым в один момент, для поведения, наблюдаемых позже?

8. В конце гл. 4 говорилось, что установки и поведения могут быть связаны в петле управления. Установки определяют поведение, но поведение, отклоняющееся от норм, вызывает к действию механизмы общественного управления, которые изменяют установки. Пренебрегая направленностью отклонения поведения, мы можем переосмыслить эту систему в терминах отклонений и социальных кар:



а именно отклонение от нормы для установки создает отклонение поведения, которое в свою очередь порождает наказание, но мы допускаем, чтобы наказания не были постоянно связаны с отклонениями путем внесения возмущения U . В конце концов наказание уменьшает отклонения установки. (Произвольные задержки приписаны ради упрощения.) Пренебрегая установками, мы можем сократить эту систему до системы, включающей только отклонения поведения и социальные кары:

$$U_P \longrightarrow P \xrightarrow[\frac{G_{PB}(s)Z^{-1}}{G(-)Z^{-1}}]{} B_B$$

Произведите дальнейшее упрощение системы до системы, включающей только отклонения поведения и возмущения для наказаний. Выразите отклонения поведения в момент t как функцию нарушений в наказаниях в более ранние моменты. К какой теоретической точке зрения приводит результат?

Глава 1

1. Употребление героина — последующее условие — подразумевает употребление марихуаны — предыдущее условие. Это утверждение отражает скорее процессы развертывания, чем причинную обусловленность. Причинная связь может рассматриваться только в том случае, если почти все курящие марихуаны приходят к героину, и тогда мы можем определить некоторый оператор для оценки этой связи.

2. Если привилегированные действия индивидуума определяют статус, то индивидуум будет соответствующим оператором. Его статус может быть изменен как при возрастании материальных и социальных возможностей для привилегированных действий, которое равносильно изменению на входе, так и в случае, когда возрастает склонность к таким действиям на основе доступных ресурсов. Это изменение установки делает индивидуума более сильным оператором, преобразующим ресурсы в статус. Если статус образован подчиненными действиями, то взаимодействующие партнеры являются операторами, порождающими некие собственные статусы. Возможно, иные уступят, так как они приучены к тому, что это вознаграждается или предотвращает наказание. Таким образом, мы понимаем возрастание статуса индивидуума А как изъятие материальной основы для удовлетворения индивидуума В, делая В более уступчивым по отношению к А. С другой стороны, мы можем приучить В проявлять большее уважение к заданному ему уровню потребностей, таким образом, делая его более мощным оператором, порождающим статус.

3. Последующие события не могут обуславливать предшествующие (правило 1.4). Так, социально-профессиональный и образовательный статусы сына не оказывают влияния на какие-либо статусы отца. Если мы готовы предположить, что формальное образование почти всегда заканчивается перед началом карьеры, то это же правило исключает причинную связь от статуса социально-профессионального к статусу образования как для отцов, так и для сыновей. Мы не можем исключить возможность связи от статуса образования к социально-профессиональному статусу (для отцов или сыновей). Учитывая временную упорядоченность, мы можем думать при объяснении обсуждаемых связей, что операторы действительно могут быть в наличии. Мы оставляем связи переменных, описывающих состояние отца, с переменными сына, так как все эти пути выдерживают тест временной упорядоченности. Понятно, что отцы с высоким социально-профессиональным статусом имеют возможности прямо помочь своим сыновьям занять более высокое положение в обществе. Следовательно, мы не можем, вообще говоря, исключить наличие причинной связи. (Заметим, однако, что мы можем допустить применение правила 1.3 при изучении поведения сыновей, которые покидают сферы влияния отцов.) Вполне вероятно также, что отцы с более высоким социально-профессиональным статусом имеют деньги и влияние, чтобы обеспечить своим сыновьям лучшее образование; эта связь не может быть элиминирована. Более образованные отцы могут дать своим сыновьям преимущество в получении образования за счет лучших домашних условий. Они могут также обеспечить лучшее руководство при контактах с образовательными институтами. Возможность прямого влияния образования отцов на образование сыновей нельзя игнорировать. Труднее представить, как образование отцов может прямо влиять на социально-профессиональный статус их сыновей. Наиболее правдоподобным объяснением такого влияния (например, отцовских связей) могут, видимо, служить косвенные воздействия, которые проходят через социально-профессиональ-

ный статус отца или образовательный сына. Мы должны быть готовы применить правило 1.2, чтобы отвергнуть возможность прямого влияния в этом случае.

4. На основе этой дополнительной информации мы можем, применяя правило 1.5, сделать вывод о том, что в современной Америке социально-профессиональный статус отцов не оказывает прямого причинного влияния на этот же статус сыновей.

5. Множество самолетов является потоком в том смысле, что каждый самолет функционирует в течение некоторого определенного периода времени, и постоянство общего их количества может поддерживаться только за счет пополнения. (Разумеется, поток должен измеряться скорее в годах, чем в минутах.) Если мы оцениваем воздушную силу числом самолетов, то она линейно зависит от составляющих: 50 самолетов из страны А и 20 самолетов из нейтральной страны образуют силу в 70 самолетов. Даже когда мы используем более отвлеченную основу для измерения, например огневую мощь, мы также, вероятно, получим линейную зависимость, если предположим, что вклады в огневую мощь пропорциональны количествам самолетов из разных источников. Для целей пропаганды, как установлено наблюдениями, и в сражении с реальным противником самолеты должны рассматриваться в совокупности, безотносительно к их происхождению. В тыловом обеспечении и в единоборстве с технологически развитым противником самолеты из указанных стран следует различать как отдельные силы.

6. Нормативный уровень переменной часто служит практическим нулем в социальной системе. Когда некий феномен находится в норме, типичная реакция людей такова, как если бы ничего не случилось. Нормативный нуль — не истинный нуль в смысле индикации отсутствия потока. Тем не менее это может привести к отсутствию отклика в последующих операторах. Подобным образом «отрицательное» значение (ниже нормативного нуля) не препятствует наличию положительного потока первоначальной переменной, но может быть причиной запрета некоторых последующих потоков.

7. В первом случае более высокое образование приводит к позиции отрицания существующих порядков, учреждений (истеблишмента), а также к отсутствию сдерживающих реакций, порождаемых обладанием богатством; предположительно такие личности настроены радикально против истеблишмента. Во втором случае отсутствует влияние образования, препятствующее удовлетворению порядком, который порождается деньгами; следовательно, эта комбинация характеризует личность, настроенную категорически за существующий порядок. Для личностей, соответствующих своему статусу, положительное влияние денег и отрицательное — образования взаимно погашаются, порождая более или менее нейтральную позицию. Нейтральность эта, однако, психологически различна для личности нищей и неграмотной и для личности образованной и материально обеспеченной.

Предположительно немущая неграмотная личность вряд ли занимает какую-либо позицию, поскольку состояния, формирующие отношение к истеблишменту, почти отсутствуют. Напротив, богатая и образованная личность нейтральна вследствие взаимного сдерживания конфликтующих чувств.

8а. Для удобства предположим, что каждой из M городских семей может быть присвоен номер $i = 1, \dots, M$, а переменные обозначены следующим образом: C — уровень городской торговли в год 1; S_i — излишки i -й семьи в год 2; B_i — число детей, родившихся в i -й семье в год 3; K — численность городской когорты пятилетних в год 8; T — число учителей городских начальных школ в год 9.

Допуская, что число семей может быть очень большим, не представляется реальной попытка описать всевозможные цепи в этом массиве. Сущность системы может быть представлена следующим образом:

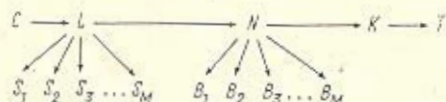


Разумеется, диаграмма отражает только отдельные процессы. Семейные излишки имеют множество источников кроме локальной экономической ситуации, размер K зависит от детской смертности и миграции так же, как и от местной рождаемости пятью годами ранее, число учителей начальных школ определяется размерами всех школьных возрастных когорт, помимо других обстоятельств. Позже подобные «возмущения» в причинных формулировках будут рассмотрены более подробно.

86. Используя причинную аппроксимацию, мы можем определить две новые переменные: L — жизненный стандарт в городе в год 2 (по существу, это средний уровень излишков всех семей города); N — общее число новорожденных в городе в год 3.

Этот индекс рождаемости есть просто сумма числа рождений во всех городских семьях в указанном году.

С этими двумя новыми переменными и с привлечением понятий причинной аппроксимации система может быть представлена следующим образом:



Такое упрощение может оказаться адекватным представлением для многих политических целей. Оно может также быть предпочтительным представлением для анализа, если измерение L и N определенно доступно. С другой стороны, упрощенная система дает новое объяснение для событий в отдельных семьях. Например, почему Смиты получили необычные излишки в год 2? — Потому, что уровень жизни был особенно высоким в этот год (т. е. объяснение не является прямым в терминах недавнего городского экономического бума). Почему Смиты имели два рождения в год 3? Ответ: коэффициент рождаемости был высоким в этом году (предпочтительнее, чем вследствие того, что в предшествующем году Смиты получили беспрецедентно высокие излишки и решили ускорить рост своей семьи). Такие объяснения могут быть адекватными для некоторых целей, хотя они имеют тенденцию маскировать действительные операторы в системе.

9. Мы можем представить соотношения между различными уровнями жизни и различными коэффициентами воспроизводства следующим образом:

Уровень жизни	Деятельность, связанная с воспроизводством	Коэффициент воспроизводства
Очень высокий	→ Врачебная помощь для повышения фертильности	→ Очень высокий
Высокий	→ Повышенная сексуальная деятельность	→ Высокий
Ожидаемый	→ Нормальная сексуальная деятельность	→ Средний
Низкий	→ Сокращенная сексуальная деятельность	→ Низкий
Очень низкий	→ Применение контрацептивов	→ Очень низкий

В крайних случаях связь между уровнем жизни и коэффициентом воспроизводства поддерживается введением новых программ медицинского или фармакологического вмешательства. Такие качественные сдвиги в деятельности оператора не основываются на распространенных механизмах, в которых рост воздействия на входе влечет просто возрастание в деятельности оператора, а он в свою очередь приводит к возрастанию на выходе. (Согласно диаграмме сексуальное поведение действительно функционирует, таким образом, в средней области.)

Несмотря на качественные сдвиги в функционировании, мы можем говорить об одном лишь операторе воспроизводства, который предположительно поддерживает монотонную связь между уровнем жизни и коэффициентом воспроизводства. Такая функциональная непрерывность имеет важное смысловое значение, в свете которого могут быть изучены и другие операторы, подвергаемые системному анализу.

10а. Описанные связи образуют петлю между P и R .

$$P \xrightleftharpoons[d]{c} R.$$

10б. Так как возрастание P влечет убывание R , величина c должна быть отрицательной (т. е. $c < 0$). Подобным образом, поскольку возрастание R влечет за собой убывание P , величина d также отрицательна. Так как оба коэффициента отрицательны, их произведение положительно, и обратная связь является скорее усиливающей, чем регулирующей, т. е. любое смещение в составе популяции будет скорее нарастать со временем, чем сглаживаться под действием обратной связи.

10в. Если абсолютные значения обоих коэффициентов больше 1, а их знаки, как мы уже знаем, отрицательны, то их произведение также больше единицы. Таким образом, первоначальное возрастание R будет усиливаться до тех пор, пока R не станет очень большим. Это, однако, не все. Так как у нас «дважды отрицательный» усилитель, возрастание R сопровождается всегда соответствующим убыванием P . (Проследивая причины и эффекты обратной связи на нескольких циклах, можно увидеть, как это происходит.) В конце концов P станет нулем и популяция будет состоять целиком из R . С этой точки зрения обратная связь саморазрушается в том смысле, что одна из переменных перестает существовать.

Глава 2

1. Сравнение между человеком A и человеком B показывает, что сыновья получают три года обучения в школе на 20 единиц NORC статуса отца. Структурный коэффициент может быть определен как

$$\frac{3 \text{ (год обучения сына)}}{20 \text{ единиц NORC отца}} = \left[0,15 \frac{\text{г. о. с.}}{\text{ед. NORC отца}} \right].$$

Заметим, что «сын» и «отец» упомянуты здесь, чтобы подчеркнуть, какие переменные измеряются.

2. Если F — статус отца, E — образование сына и a — структурный коэффициент между ними, то ожидаемое образование сына кратно F , т. е.

$$\left[0,15 \frac{\text{г. о. с.}}{\text{NORC отца}} \right] \cdot 60 \text{ NORC отца} = 9 \text{ г. о. с.}$$

Таким образом, сын данного индивидуума получит, как можно ожидать, примерно девятиклассное образование.

3. Как упоминалось в гл. 1 (1.6), линейное структурное уравнение, если отбросить его константы, может быть применено как для абсолютного уровня входных воздействий, так и для получения уравнения относительно изменений на входе. Структурное уравнение в этом случае имеет вид:

$$E = \left[0,15 \frac{\text{г. о. с.}}{\text{NORC отца}} \right] F,$$

а уравнение относительно приращений:

$$\Delta E = \left[0,15 \frac{\text{г. о. с.}}{\text{NORC отца}} \right] \Delta F.$$

Применяя его в этом частном случае, получим:

$$\Delta E = \left[0,15 \frac{\text{г. о. с.}}{\text{NORC отца}} \right] \cdot (-10 \text{ NORC отца}) = -1,5 \text{ г. о. с.}$$

Итак, неудачи отца приводят к ожидаемой потере в образовании сына в 1,5 года обучения.

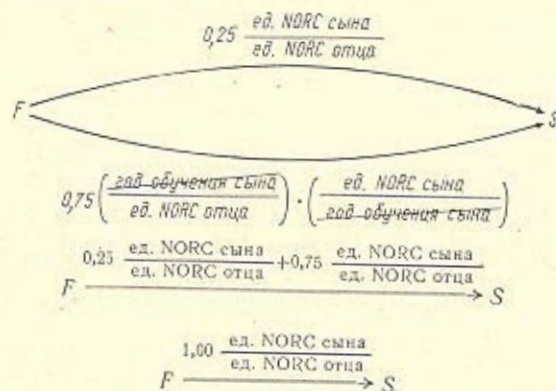
4. По правилу цепи (II.8) социально-профессиональный уровень сына будет:

$$S = \left[0,15 \frac{\text{г. о. с.}}{\text{NORC отца}} \right] \left[5,0 \frac{\text{NORC сына}}{\text{г. о. с.}} \right] 80 \text{ NORC отца} = 60 \text{ NORC сына.}$$

5а.



5б. Предыдущее сводится к



или

откуда получается уравнение:

$$S = \left[1,0 \frac{\text{ед. NORC сына}}{\text{ед. NORC отца}} \right] F.$$

5в. $S = \left[1,0 \frac{\text{ед. NORC сына}}{\text{ед. NORC отца}} \right] 80 \text{ ед. NORC отца} = 80 \text{ ед. NORC сына.}$

6. Согласно предположениям:

$d > 0$ (т. е. положительно),

$e > 0$ (положительно),

$f < 0$ (отрицательно).

Единицы измерения, приписываемые каждому коэффициенту, следующие (отбрасывая выражение «число» и «в городе», которые всегда подразумеваются):

Единица для d есть отношение единицы C к единице I :

$$\frac{\text{преступление/житель}}{\text{неймущий/житель}} = \frac{\text{преступление}}{\text{неймущий}}.$$

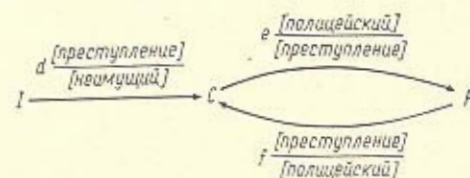
Для единицы e имеем отношение единицы P к единице C :

$$\frac{\text{полицейский/житель}}{\text{преступление/житель}} = \frac{\text{полицейский}}{\text{преступление}}.$$

Для единицы f используется также отношение единицы C к единице P :

$$\frac{\text{преступление/житель}}{\text{полицейский/житель}} = \frac{\text{преступление}}{\text{полицейский}}.$$

Потоковый граф имеет вид:



7. Эффект обратной связи в задаче 6 есть

$$L_{CP} = \left[e \frac{\text{полицейский}}{\text{преступление}} \right] \cdot \left[f \frac{\text{преступление}}{\text{полицейский}} \right] = ef.$$

Поэтому эффект обратной связи является вообще говоря, безразмерной величиной. Так как эффект обратной связи есть безразмерная величина, она не изменится при любых линейных преобразованиях шкал, измеряющих переменные обратной связи (таких, как добавление или вычитание константы либо деление или умножение на константу)*. В частности, процедуры, используемые в проблеме 6, могут быть применены для иллюстрации этого: если уровень преступности измеряется отношением числа преступлений на 1000 жителей, то рассматриваемые три коэффициента будут иметь значения:

$$\begin{aligned} 0,001 \left[d \frac{\text{преступление}}{\text{неймущий}} \right], \\ 1000 \left[e \frac{\text{полицейский}}{\text{преступление}} \right], \\ 0,001 \left[f \frac{\text{преступление}}{\text{полицейский}} \right], \end{aligned}$$

и эффект обратной связи остался неизменным. Итак, эффект обратной связи характеризует ее безотносительно к единицам измерения переменных.

8а. Программа минимального дохода представляет собой попытку свести к нулю значение переменной I (неймущие). Если система составлена корректно, то это должно вести к снижению уровня преступности (хотя, вероятно, не до нуля, в силу того, что имеются, бесспорно, другие источники преступности, помимо указанных здесь). Используя анализ потоковых графов, можно показать, что такой подход к снижению преступности ведет также и к уменьшению доли полицейских.

8б. Мы можем предположить, что, делая жизнь неймущих более комфортной, можно ослабить причинные механизмы между бедностью и преступностью. Если это так, указанная программа уменьшит величину коэффициента d . Уменьшение зависимости преступности от бедности ведет к тем же эффектам, что обсуждалась в 8а.

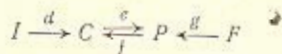
8в. Рост эффективности и заметности действий полиции усиливает ее сдерживающие функции и в силу этого увеличивает отрицательное значение коэффициента f . Действительно, такое уменьшение связи между преступностью и бедностью следует из приведенной формы уравнения:

$$C = \frac{d}{1 - ef} I.$$

* Все величины предполагаются центрированными. — Примеч. пер.

Так, коэффициент преступности будет ниже при заданном уровне бедности. Подобным образом указанные изменения приводят в конце концов к уменьшению доли полицейских при заданном уровне бедности.

8г. На потоковом графе мы можем представить новую программу, добавляя переменные — федеральную помощь полиции — и стрелку от этой переменной к доле полицейских.



(коэффициент g положителен). Пока программа будет вводиться, переменная федеральной помощи изменится от нуля до некоторой положительной величины, которая приведет в свою очередь к возрастанию доли полицейских и к снижению уровня преступности. Эти эффекты будут особенно заметны вначале. Позже, когда система достигнет равновесия, доля полицейских и уровень преступности несколько сдвинутся в сторону их первоначальных значений.

9. Без учета политических аспектов система судебного принуждения может быть представлена следующим образом (в предположениях упражнения б):



Здесь S означает поддержку населения на локальном уровне и h — положительный коэффициент, который представляет трансформацию этой поддержки судебного принуждения в постоянные полицейские силы. Теоретически преступность в этой системе можно искоренить, если поддержка населением полиции будет на уровне, достаточно высоком, чтобы сбалансировать давление преступности в среде неимущих, т. е. если

$$S = \frac{d}{fh} I.$$

Особенность рассмотренной выше системы состоит в том, что она не реагирует на рост преступности в целом. Полиция, с ее постоянным уровнем поддержки, не может мобилизоваться навстречу новым опасностям. Рассмотрение этой проблемы на государственном уровне даст решение. Неизвестно, однако, гарантирует ли создание с помощью государства петли «управления» между преступностью и полицией сохранение проблемы преступности на низком хроническом уровне, если система останется устойчивой и линейной, или в виде отдельных скачков в случае, если полиция сможет исключить преступность целиком. Отсюда и помощь государства как причинный оператор также приводит к демобилизации, когда доля преступности уменьшается.

10. Используя анализ потоковых графов, мы видим, что соотношение между преступностью и ее основным источником по первому плану может быть представлено как:

$$C_1 = \frac{d(1-L)}{1-ef-L} I.$$

По второму плану соотношение будет таким:

$$C_2 = \frac{d}{1-ef-L} I.$$

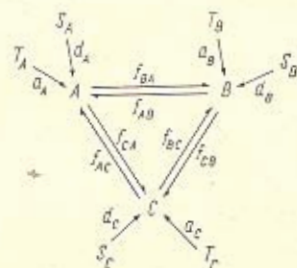
Итак, мы можем видеть, что

$$C_1 = (1-L) C_2.$$

Так как L — отрицательное число (это обратный эффект управляющей петли), величина в скобках больше единицы. Итак, реализация первого плана закончится на более высокой доле преступности по сравнению со вторым планом, даже несмотря на наличие управляющей петли равной силы. На языке потоковых графов это является результатом того обстоятельства, что управляющая петля первого плана не касается открытого пути от бедности к преступности. Следова-

тельно, обратная связь не отражает всего влияния на рассматриваемые отношения. На обычном языке первый план определяет следующий процесс: как только возрастание преступности вызывает в локальной общине мобилизацию большего числа полицейских, часть федеральной поддержки изымается, тем самым прекращая локальную мобилизацию. Следовательно, реакция на волну преступности всегда умеренная. (С другой стороны, локально инспирированное сокращение полиции может иногда вызвать ответную реакцию в виде увеличения федеральной поддержки. Поэтому обратная связь оказывает некоторое управляющее действие, увеличивая численность полиции, когда локальная поддержка слишком низка.) С другой стороны, федеральное вмешательство по второму плану непосредственно управляет преступностью, а не действиями локальных полицейских сил. Его влияние на преступность больше.

11а.



Структурные уравнения имеют вид:

$$A = a_A T_A + d_A S_A + f_{AB} B + f_{AC} C,$$

$$C = a_C T_C + d_C S_C + f_{CB} B + f_{CA} A,$$

$$B = a_B T_B + d_B S_B + f_{BA} A + f_{BC} C.$$

11б. Имеется пять различных петель: (AB), (BC), (AC), (ABC), (ACB). Заметим, что каждая петля касается всех других. При описании системы будем предполагать, что каждый оператор преобразует возрастание одной переменной в возрастание другой. Следовательно, все структурные коэффициенты должны иметь положительные знаки. Так как отрицательных коэффициентов нет, нет и управляющих петель, все петли — усиливающие.

11в.

$$T_{AS(A)} = \frac{d_A (1 - f_{BC} f_{CB})}{1 - f_{BA} f_{AB} - f_{BC} f_{CB} - f_{AC} f_{CA} - f_{BA} f_{CB} f_{AC} - f_{CA} f_{BC} f_{AB}}.$$

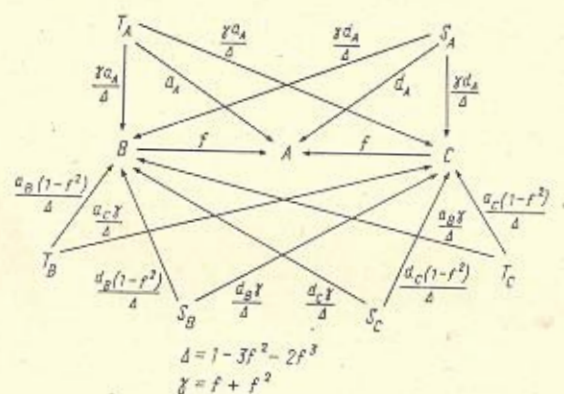
Так как каждый положительный обратный эффект петли вычитается в знаменателе, каждый из них уменьшает его величину и поэтому возрастает значение $T_{AS(A)}$. Рассматривая каждую петлю отдельно, мы видим, что это вызывает следующее: научные достижения в стране A приводят к начальному росту ее военной мощи путем военно-промышленного развития (d_A). Такой начальный рост крайне преувеличивается, вследствие чего возникает цепь ответных реакций двух других стран. Наблюдая рост военной мощи страны A, страна B наращивает свою, что принимается страной A как угроза; страна A должна поэтому подкрепить свои силы новым ростом (петля AB). Далее, военная экспансия страны B тревожит страну C, которая чувствует себя обязанной увеличить собственную военную мощь. Эта очевидная угроза безопасности страны A требует дальнейшей экспансии ее военной мощи (петля ABC). С другой стороны, мы имеем подобные же процессы. Начальный толчок страны A провоцирует страну C к наращиванию своих сил, которое требует ответной реакции A (петля AC). Далее, это же усиление C заставляет B увеличивать свою мощь, что рассматривается уже как новые действия, требующие сильной реакции страны A (петля ACB). Разумеется, каждое из этих наращиваний военной мощи прокручивается

в системе, вызывая еще большее наращивание. Конечный эффект представляется уравненными, приведенными выше. Петля BC образует особый случай. Начальное наращивание военной мощи страны B обуславливает наращивание мощи страны C , что беспокоит не только страну A , но и в свою очередь страну B (которая в конце концов реагирует на эту опасность). Таким образом, они должны наращивать свою мощь в ответ стране C , замыкая тем самым петлю (BC). (Подобно этому образуется петля, трансформирующая начальное военное наращивание C в военную эскалацию B , которая затем должна вызвать ответ в стране C .) Взаимная эскалация между странами B и C имеет некоторое влияние на страну A , так как она повышает возможность провокации со стороны обоих государств. Итак, обратный эффект петли (BC) есть знаменатель приведенного выше уравнения. С другой стороны, взаимное соревнование между странами B и C несколько отдалено от страны A и петля оказывает меньшее действие, чем в том случае, если бы A была непосредственно вовлечена. Это представлено вычитанием обратного (BC) эффекта в числителе; тем самым уточнение числителя сделано в целях отражения того факта, что петля (BC) не действует с полной силой на военную мощь страны A .

12. Если значение коэффициента f не превышает 0,5, то это означает, что, когда одна страна наращивает свою военную мощь на некоторую величину, первоначальная реакция других стран заключается в наращивании их мощи более чем на половину этой величины. Более сложная шкала для измерения военных сил должна принимать в расчет помимо огневой мощи возможности и эффективность систем обеспечения; представление о таком включении может быть получено на примере, содержащем простую шкалу: если страна B добавит 100 танков в свой арсенал, страны A и C в ответ не должны добавлять более чем по 50 танков. Эта система порождает неконтролируемую гонку вооружений, если страны будут отвечать на движения друг друга более энергично. Некоторая нестабильность вероятна, если каждая страна стремится к равенству сил или первенству по отношению к остальным. С другой стороны, большое число факторов направлено на ослабление реакций, что способствует стабильности, например плохая организация разведывательных служб, приводящая к недооценке достижений противников, или альянсы, с помощью которых две страны делают между собой реакцию на третью страну.

Это и другие соображения должны включаться в общее обоснование устойчивости системы.

13а.



13б. Стрелки ($S_A B$) и ($S_A C$) означают, что успехи науки в стране A действуют непосредственно на увеличение военной мощи стран B и C . Возможно, кто-либо предположит, что разведка является средством, объясняющим, как это происходит. Стрелки ($S_B C$) и ($S_C B$) могли бы быть объяснены так же или в терминах альянса и кооперации этих стран. Отметим, что здесь нет стрелок ($S_B A$) и ($S_C A$), если страна A не ведет разведывательных операций и не извлекает пользу из научных достижений других стран. Стрелки ($T_A B$) и ($T_A C$) показы-

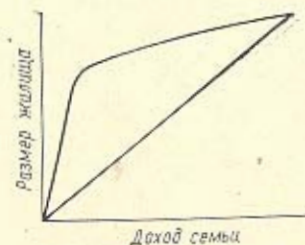
вают, что страна A ответственна за поддержание уровня вооружения стран B и C . Возможно, это следует объяснить алчностью и завистью других, стремлением их оккупировать страну A . Того же рода объяснения могут быть сделаны для стрелок ($T_B C$) и ($T_C B$). Заметим, что страна A не готовит оккупацию других государств — отсутствуют стрелки ($T_A B$) и ($T_A C$). Следовательно, научные достижения страны A , естественно, ведут прямо к ее военному усилению. Однако это не является объяснением причин реакции двух других стран. Последняя связана с похищением научных секретов и использованием их для укрепления своего могущества. Когда это происходит, страна A , как правило, отвечает на вызов дальнейшим наращиванием своей мощи. Научные достижения страны B ведут к усилению ее мощи, а также мощи страны C (за счет обмена информацией или разведки). Страна A не использует секретов страны B , но она должна прямо реагировать на военную эскалацию стран B и C .

Преобразованная модель позволяет делать точное предсказание того, что произойдет за длительный промежуток времени. В этом ее особая сила применительно к рассмотренной выше ситуации. С другой стороны, это может ввести в заблуждение кого-либо, ожидающего простых картин международной динамики. Также будет большим заблуждением предполагать, что успехи контрразведки окажут большое влияние на характер системы.

Глава 3

1. Поскольку в каждый момент имеется только один президент, невозможно провести обычный анализ по множеству реализаций (cross-sectional study), в котором мы изучаем повторяющиеся примеры одного оператора типа. Для этих целей можно удовлетвориться предположением, что президент есть постоянный оператор в течение своего правления, и каждый из законченных эпизодов его деятельности заданного типа может изучаться как статическая конфигурация, характеризующая этого президента. Множество таких эпизодов можно изучать статистически. Этого рода анализ называется «сравнительной статикой».

2а.

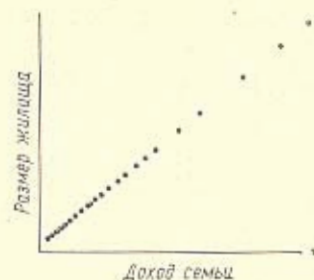


2б. Строго линейная система не может дать такого совместного распределения. Колебания в размерах жилищ для семей с низким доходом много больше, чем для семей с высоким доходом, и нет способа получить такой эффект при строго линейных уравнениях. Такого рода феномен может появиться в причинной системе, если другая переменная, скажем размер семьи, взаимодействует с доходом при определении размера жилища, как показано в следующем уравнении:

$$\text{размер жилища} = a (\text{доход семьи}) \cdot (\text{размер семьи}).$$

Согласно этому положению некоторые большие семьи с низким доходом вынуждены стремиться к увеличению жилища. Обсуждаемое может рассматриваться как доказательство того, что «высокий доход влечет за собой большое жилище». Для интерпретации этого в терминах развертывания мы должны, однако, предполагать, что большое жилище есть необходимая предпосылка для достижения здоровья, но это является сомнительным постулатом в современном обществе.

2в. Линейная связь возможна, если распределение обеих переменных скошено, как представлено на следующем графике:



3. Одним из способов интерпретации этой проблемы является некое множество автобусов, перемещающихся идеально точно в данный момент на безопасной скорости и в безопасном направлении. Водители, как это свойственно людям, вводят некоторые возмущения и отклонения в точные скорости и направления, более или менее отдаляя каждый автобус от идеала. Второй водитель добавит дополнительные возмущения. Таким образом, предполагая действия водителей не координируемым каким-либо способом, можно заключить, что разброс возмущений для двух водителей больше разброса каждого из них порознь. Следовательно, автобусы с двумя водителями будут более далеки от идеала и поэтому более опасными.

4а. Без сомнения, исследователь должен снова провести вычисления. Корреляционный момент не может иметь значение меньше $-1,0$ и больше $+1,0$.

4б. Предполагая вычисления корректными, этот регрессионный коэффициент показывает, что КИ может быть в некоторой степени оценен, если мы знаем СБУ студентов (поскольку коэффициент отличен от нуля), и, в частности, высокий СБУ связан с высоким КИ (так как коэффициент положителен). Тот факт, что коэффициент больше $1,0$, не несет информации в этом случае, ибо значение не стандартизованного коэффициента регрессии зависит от единиц, используемых для измерения переменных. Для не стандартизованных коэффициентов нет, вообще говоря, математически обоснованных границ. Заметим, что здесь мы можем использовать коэффициент регрессии для преобразования одного вида информации в другой, не выражая прямо тем не менее действия причинного оператора, т. е. мы не предполагаем, что существует оператор для преобразования уровня СБУ в уровень КИ.

$$\begin{aligned} 5. \quad E[(X - \bar{X})^2] &= E(X^2 - 2X\bar{X} + \bar{X}^2) = \\ &= E(X^2) - E(2X\bar{X}) + E(\bar{X}^2) = \\ &= E(X^2) - 2\bar{X}E(X) + \bar{X}^2 \end{aligned}$$

Последний переход возможен, так как среднее постоянно:

$$\begin{aligned} &= E(X^2) - 2\bar{X}^2 + \bar{X}^2, \\ &= E(X^2) - \bar{X}^2, \\ &= E(X^2) - [E(X)]^2. \end{aligned}$$

или иначе:

Так как первоначальное выражение определяет дисперсию X , последнее выражение также определяет дисперсию, и мы можем вычислить ее прямо по исходным измерениям без предварительного вычитания среднего.

$$\begin{aligned} 6. \quad E[(X - \bar{X})(Y - \bar{Y})] &= E(XY - Y\bar{X} - X\bar{Y} + \bar{X}\bar{Y}) = \\ &= E(XY) - E(Y\bar{X}) - E(X\bar{Y}) + E(\bar{X}\bar{Y}) = \\ &= E(XY) - \bar{X}E(Y) - \bar{Y}E(X) + \bar{X}\bar{Y} = \\ &= E(XY) - \bar{X}\bar{Y} - \bar{Y}\bar{X} + \bar{X}\bar{Y} = \\ &= E(XY) - \bar{X}\bar{Y}, \\ &= E(XY) - [E(X)][E(Y)]. \end{aligned}$$

или

Итак, ковариация может быть также получена по исходным измерениям без предварительного вычитания среднего.

7. Заметим, во-первых, что $\hat{X}$ — предсказанное значение переменной, не является, вообще говоря, постоянной; ее значение меняется в зависимости от случая.

$$\begin{aligned} E[(\hat{X} + e)^2] &= E(\hat{X}^2 + 2\hat{X}e + e^2) = \\ &= E(\hat{X}^2) + E(2\hat{X}e) + E(e^2) = \\ &= E(\hat{X}^2) + 2E(\hat{X}e) + E(e^2). \end{aligned}$$

Поскольку $\hat{X}$ и e представляют собой, по предположению, оценки отклонения, последнее равенство может быть записано так:

$$= \sigma_{\hat{X}}^2 + 2\sigma_{\hat{X}e} + \sigma_e^2.$$

Это выражение определяет дисперсию переменной $X = \hat{X} + e$. Очевидно, дисперсия X равна сумме дисперсий $\hat{X}$ и e только тогда, когда средний член имеет нулевое значение. Это верно, только в том случае, если ковариация $\hat{X}$ и e нулевая, другими словами, если предсказанное значение некоррелировано с ошибкой предсказания.

8. Так как β — стандартизованный коэффициент, мы можем считать и все другие коэффициенты стандартизованными до начала анализа. Вследствие этого частный коэффициент регрессии может интерпретироваться как результат трех различных моделей регрессионного анализа. Во-первых, мы строим регрессию социально-профессионального статуса отца на образование сына, для того чтобы определить множество остаточных вариаций, наличие которых вызвано действием других факторов. Во-вторых, мы строим регрессию социально-профессионального статуса сына на его образование, для того чтобы определить множество остаточных вариаций статуса сына, которые нельзя предсказать по образованию. В-третьих, мы строим регрессию остатка социально-профессионального статуса сына на остаток статуса отца. Регрессионный коэффициент этой третьей модели есть не что иное, как $\beta_{SF.E}$. Таким образом, мы видим, что $\beta_{SF.E}$ измеряет предсказуемость вариации статуса общественного положения сына, которая не может быть объяснена его образованием; при этом предсказуемая переменная определяется вариацией статуса отца, которая не связана с образованием сына. Поскольку мы начали с согласования шкал всех переменных посредством стандартизации и полученное значение $\beta_{SF.E}$ очень мало, мы можем сделать заключение, что в данном случае между статусами отца и сына относительно мала та ковариация, которая не определяется уровнем образования сына.

9. При этих заданных корреляциях стандартизованный частный коэффициент регрессии от статуса отца к статусу сына имеет значение

$$\beta_{SF.E} = 1,06.$$

Не существует математических условий, ограничивающих значения частных коэффициентов регрессии областью значений коэффициента корреляции. С другой стороны, значение коэффициента β вне интервала $(-1, +1)$ весьма необычно, и

такое может случиться только в том случае, если среди корреляционных коэффициентов между переменными есть отрицательные. Данное значение $\beta_{sf,e}$ показывает, что вариации статуса отца связаны с такими же большими вариациями статуса сына, если образованием сына управляют. В этом своеобразном обществе отцы с высоким статусом имеют сыновей с образованием ниже среднего уровня (что обнаруживается отрицательным значением величины r_{fe}), а плохое образование этих сыновей в свою очередь связано с социально-профессиональным статусом ниже ожидаемого (что выявляется положительным значением r_{es}). Итак, связи, проходящие через образование сына, имеют тенденцию к нейтрализации прямых связей от статуса отца к статусу сына, и общая корреляция часто умеренна. Переменные, подобные образованию сына в этом примере, часто называются «нейтрализующими переменными».

Глава 4

1а. Ожидаемые корреляции между переменными могут быть получены с помощью путевого анализа:

	v	i	y	a
v	1,00	0,30	-0,40	-0,01
i	0,30	1,00	-0,12	0,27
y	-0,40	-0,12	1,00	0,42
a	-0,01	0,27	0,42	1,00

Диагональные члены матрицы корреляций всегда равны 1,00, так как каждый из них представляет корреляцию переменной с самой собой, или, иначе говоря, дисперсию стандартизованной переменной. Заметим также, что корреляции ниже диагонали те же самые, что и выше. Вследствие этого корреляционная матрица часто представляется в виде верхней или нижней треугольной матрицы.

1б. Поскольку возмущения a описаны как некоррелированные с чем-либо, дисперсия a может быть выражена так:

$$\sigma_a^2 = (0,84)^2 \sigma_{u(a)}^2 + S,$$

где S представляет собой дисперсию, определяемую всеми другими источниками. Если вы построите полное выражение для дисперсии a , вы увидите, из чего состоит S , и докажете также, что уравнение верно.

Так как вариация возмущения стандартизована, уравнение может быть переписано в виде:

$$\sigma_a^2 = 0,71 + S \text{ или } S = \sigma_a^2 - 0,71.$$

Так определяется абсолютная величина объясненной вариации. Для того чтобы преобразовать ее в отношение, разделим объясненную дисперсию a на полную дисперсию a :

$$\frac{S}{\sigma_a^2} = \frac{\sigma_a^2 - 0,71}{\sigma_a^2}.$$

Разумеется, для стандартизованной переменной a с дисперсией 1,00 представленное выше отношение равно 0,29. (В общем случае, когда все переменные и возмущения рассматриваются в стандартизованной форме и возмущения некоррелированы ни с чем, доля объясненной дисперсии может быть найдена прямо из диаграммы. Она равна 1,00 минус квадрат путевого коэффициента от возмущения к интересующей нас переменной.) Доля объясненной дисперсии показывает

просто степень различий в значениях переменной, вызванное различием определяющих причин. На чертеже не обязательно представлены какие-либо ограничения на число изменяющихся переменных; например, предположим, что для всех тюрем характерна относительно высокая степень агрессивности заключенных. Вариации степени тюремной изоляции, характера приговора и возраста определяют только часть вариаций относительно общего среднего. Тем не менее уже теоретически возможно изменить, скажем, политику посещения в любой тюрьме, чтобы резко уменьшить степень агрессивности в ней в два, в три и более раза. (Иначе говоря, возможно возрастание i , v или y такое, что степень агрессивности удвоится, утроится, учетверится и т. д.). Поскольку i , v , y объясняют только 29% дисперсии, мы не можем исключить разницу в степени агрессивности заключенных между тюрьмами, исключив только разницу в изоляции заключенных, характере приговоров и возрасте. Тем не менее эти переменные могут быть использованы для исключения дисперсии a путем задания таких вариаций i , v , y , что их действие нейтрализует действие неучтенных источников дисперсии. Это равносильно построению отрицательных корреляций между u_a и i , v или y . Практически это может включать, скажем, перевод молодых заключенных из тюрем с высокой агрессивностью в тюрьмы с низкой агрессивностью.

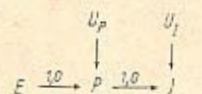
1в. Если возрастная структура заключенных в данной тюрьме изменилась, а состав правонарушителей остался постоянным, влияние агрессивности может быть исключительно на пути ($y \rightarrow a$). Возрастание y на единицу ведет к ожидаемому возрастанию на 0,5 агрессивности заключенных. Если изменение возрастной структуры происходит безотносительно к типу правонарушителя, мы можем ожидать сохранения обычной корреляции между y и v на уровне -0,4. Изменение y на единицу будет вероятно связано с убыванием v на -0,4. Чистое влияние от обоих изменений будет

$$0,5 + (-0,4) [0,1 + (0,3) (0,3)] = 0,42.$$

Итак, некоторому ожидаемому возрастанию агрессивности, вызванному омоложением контингента, противодействует то обстоятельство, что в тюрьме теперь содержатся менее буйные правонарушители.

1г. Ночные кошмары охранников в дальних тюрьмах (далеких от посетителей заключенных) связаны главным образом с молодыми буйными преступниками. Приятным облегчением для тюремщиков была бы тюрьма с правилами посещения типа больничных и с местными правонарушителями среднего возраста и умеренного поведения.

2а. Система с ее обновленным производящим оператором может быть представлена следующим образом (здесь E — образование, P — престиж профессионального положения, I — доход):



где

$$\sigma_E^2 = 1,0, \sigma_{U(P)}^2 = 1,0, \sigma_{U(I)}^2 = 1,0$$

С помощью путевого анализа находим:

$$\sigma_P^2 = (1,0)^2 \sigma_E^2 + (1,0)^2 \sigma_{U(P)}^2 + 2 (1,0) (1,0) (0,0) = 2,0,$$

$$\sigma_I^2 = 2,0 + 1,0 = 3,0,$$

$$\sigma_{EP} = (1,0) \sigma_E^2 = 1,0,$$

$$\sigma_{EI} = (1,0) (1,0) \sigma_E^2 = 1,0,$$

$$\sigma_{PI} = (1,0) \sigma_P^2 = 2,0.$$

Применим формулу для корреляционного коэффициента из 3.17:

$$\rho_{EP} = \frac{\sigma_{EP}}{\sigma_E \sigma_P} = \frac{1,0}{\sqrt{1,0} \sqrt{2,0}} = 0,707,$$

$$\rho_{EI} = \frac{1,0}{\sqrt{1,0} \sqrt{3,0}} = 0,577,$$

$$\sigma_{PI} = \frac{2,0}{\sqrt{2,0} \sqrt{3,0}} = 0,816.$$

26. Путевые коэффициенты суть структурные коэффициенты, которые оценивают корреляцию между переменными. Значения этих корреляций вычислены выше. Следовательно, должны выполняться равенства между корреляционными и путевыми коэффициентами (обозначенными здесь через q):

$$\rho_{EP} = q_{pe}, \quad \rho_{PI} = q_{ip}.$$

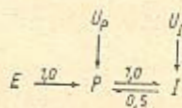
И таким образом, путевые коэффициенты могут быть оценены прямо по корреляционным:

$$q_{pe} = \rho_{EP} = 0,707, \quad q_{ip} = \rho_{PI} = 0,816.$$

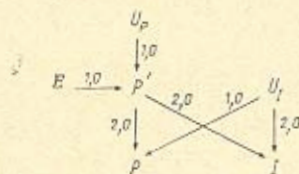
(Отметим тот факт, что $(q_{pe}) \cdot (q_{ip}) = \rho_{EI}$.) Используя первоначальную метрику, мы полагаем, что оба новых оператора одинаково важны, поскольку их структурные коэффициенты имеют значение 1,0. С другой стороны, стандартизуя переменные относительно финального распределения, можно показать, что второй оператор сильнее первого. Разумеется, одни и те же операторы включаются в обоих случаях, которые показывают, что не имеет смысла сравнивать коэффициенты, основанные на разных метриках.

2в. В этом примере предполагается, что корреляции между статусами вызваны специальными механизмами преобразования одного вида статуса в другие, при этом допускается первоначальный разброс причинных статусов. Введение таких механизмов в общество оставляет неизменным разнообразие причинных статусов, но увеличивает разнообразие статусов на выходе. Поэтому мы можем говорить, что наибольшее неравенство проявляется в наиболее удаленных от центра статусах.

3. Система теперь будет такой:



Эффект обратной связи есть $L = 0,5$ и его «разностный эффект» $(1 - L) = 0,5$. Таким образом, частично редуцированная система имеет следующую форму:



Путевой анализ дает следующие результаты (вспомним, что $\sigma_E^2 = \sigma_{U(P)}^2 - \sigma_{U(I)}^2 = 1,0$):

$$\sigma_P^2 = (2,0)^2 \sigma_{P'} + (1,0)^2 \sigma_{U(I)}^2 = (2,0)^2 (2,0) + (1,0)^2 (1,0) = 8,0 + 1,0 = 9,0,$$

$$\sigma_I^2 = (2,0)^2 \sigma_{P'}^2 + (2,0)^2 \sigma_{U(I)}^2 = (2,0)^3 + (2,0)^2 = 12,0,$$

$$\sigma_{EP} = (1,0) (2,0) \sigma_E^2 = 2,0,$$

$$\sigma_{EI} = (1,0) (2,0) \sigma_E^2 = 2,0,$$

$$\sigma_{PI} = (2,0) \sigma_{P'}^2 (2,0) + (1,0) \sigma_{U(I)}^2 (2,0) = (2,0)^3 + 2,0 = 10,0.$$

Заметим, что петля главным образом увеличивает неравенство в профессиональных статусах и доходах.

Корреляции получаются из приведенных выше данных:

$$\rho_{EP} = \frac{2,0}{\sqrt{1,0} \sqrt{9,0}} = 0,667,$$

$$\sigma_{EI} = \frac{2,0}{\sqrt{1,0} \sqrt{12,0}} = 0,577,$$

$$\sigma_{PI} = \frac{10,0}{\sqrt{9,0} \sqrt{12,0}} = 0,962.$$

(Сравните эти значения с корреляциями, полученными в системе упражнения 2). Корреляция между E и P здесь меньше, поскольку P имеет теперь дополнительный источник вариации — U_I . С другой стороны, присутствие усиливающей петли увеличивает корреляцию двух переменных внутри петли.

4а. Путевой анализ определяет корреляцию между индикаторами:

$$\rho_{x(1)x(2)} = \rho_{x(1)x} \sigma_x^2 \rho_{x(2)x} = 0,7 (1,0) 0,60 = 0,42,$$

$$\rho_{y(1)y(2)} = 0,20.$$

Два индикатора переменной (x или y) будут полностью коррелированы (с корреляцией 1,0), если они оба определяются только этой переменной. Однако измерение каждого индикатора подвержено ошибкам. Диаграмма показывает, что ошибки измерения одного индикатора не зависят от ошибок измерения других. Следовательно, ошибки увеличивая дисперсию, не увеличивают ковариацию; тем самым корреляция уменьшается.

4б. Путевой анализ определяет следующие корреляции:

	y_1	y_2
x_1	0,084	0,105
x_2	0,072	0,090

Очевидно, что корреляции между индикаторами x и y много меньше, чем корреляция между истинными значениями. Обычно мы вычисляем такие корреляции по выборке, и есть некоторое подозрение, что их значения будут немного другими для другой выборки. Таким образом, мы склонны рассматривать приведенные корреляции как незначительно отличные от нуля, делая отсюда вывод об отсутствии связи между информированностью и отношением.

4в. Путевой анализ может быть применен для определения дисперсии каждого индикатора, например:

$$\sigma_{x(1)}^2 = (0,7)^2 \sigma_x^2 + (P_{id})^2 \sigma_{d(1)}^2.$$

Поскольку все переменные (включая возмущения) стандартизованы, это уравнение сводится к

$$1 = 0,49 + P_{1d}^2$$

или

$$P_{1d}^2 = 0,51; P_{1d} = 0,71.$$

Итак, 51% дисперсии индикатора определяется ошибкой измерения, и значение коэффициента на стрелке от члена ошибки к индикатору есть 0,71. Вычисления для остальных трех индикаторов производятся аналогично:

$$\sigma_{x2}^2 = 1 = 0,36 + P_{2d}^2 \text{ и } P_{2d} = \sqrt{0,64} = 0,80,$$

$$\sigma_{y1}^2 = 1 = 0,16 + P_{1e}^2 \text{ и } P_{1e} = \sqrt{0,84} = 0,92,$$

$$\sigma_{y2}^2 = 1 = 0,25 + P_{2e}^2 \text{ и } P_{2e} = \sqrt{0,75} = 0,87.$$

4г. Если $\rho_{x\theta}$ неизвестен, то корреляция между x_1 и y_1 может быть определена с помощью путевого анализа:

$$\rho_{x(1)y(1)} = (0,7) \rho_{x\hat{y}}(0,4).$$

Разрешая это равенство относительно неизвестной корреляции между истинными величинами, найдем:

$$\rho_{x\hat{y}} = \frac{\rho_{x(1)y(1)}}{0,7 \cdot 0,4}.$$

Поскольку все величины в правой части предполагаются известными, получается формула для оценки корреляции между истинными величинами. Эта процедура — деление наблюдаемой корреляции на коэффициенты обоснованности для оценки истинной корреляции — хорошо известна в психометрии как «ослабляющая коррекция».

5а. Дополнительная спецификация требует соединения кривыми двухконечными стрелками всех членов ошибок. Каждая имеет корреляцию 0,3, приписанную ей. Применение путевого анализа к пересмотренной диаграмме дает следующее уравнение:

$$\rho_{x(1)x(2)} = (0,7)(1,0)(0,6) + P_{1d}(0,30)P_{2d}.$$

Численные результаты при P_{1d} , P_{2d} и т. д., оцененных по данным упражнения 4в, таковы:

	x_1	x_2	y_1
x_2	0,59	—	—
y_1	0,28	0,29	—
y_2	0,29	0,30	0,44

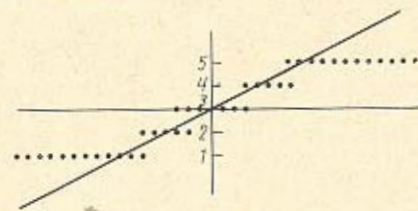
Эти значения больше, чем в случае некоррелированных ошибок. Обычно ошибки отдельных вопросов в различных вопросниках или интервью будут положительно коррелированы, поскольку некоторые из возмущающих факторов действуют в течение всего периода сбора данных. Положительно коррелированные ошибки измерения всегда приводят к увеличению наблюдаемой корреляции между вопросами.

5б. Оцененная корреляция между истинными величинами теперь такова:

$$\rho_{x\hat{y}} = \frac{0,28}{0,7 \cdot 0,4} = 1,00.$$

Этот результат может ввести нас в заблуждение, если мы сделаем заключение о полном соответствии между установкой и информированностью, когда она фактически только умеренная.

5в. Коэффициент обоснованности для этого индикатора равен 0,7, что допускает достаточно тесную связь между истинной величиной и индикатором на большей части шкалы. Мы можем подозревать, однако, что эта связь прерывается на концах индикаторной шкалы. Всем далеким положительным значениям установки — неважно, насколько далеким, — должно быть приписано значение 5,0; аналогично всем далеким отрицательным значениям установки следует приписать значение не ниже 1,0. Таким образом, совместное распределение истинных величин и индикаторов должно выглядеть примерно так:



В средней области имеется небольшая связь между истинными величинами и ошибками измерения, но на краях истинные величины имеют явно отрицательную связь с ошибками (например, чем больше положительные значения величины установки, тем больше она недооценивается). В конце концов это будет создавать отрицательную корреляцию между истинными величинами и ошибками измерения. Это обычно верно, когда шкала измерения ограничена так, что очень низким и высоким отметкам приписываются умеренные значения. С другой стороны, результирующая отрицательная корреляция будет мала или ею можно пренебречь, если лишь в немногих случаях появляются измерения за пределами шкалы.

6. В системе с сильным механизмом поощрения, но без обратной связи дисперсия Z выражается так:

$$\sigma_Z^2 = [(2,0)^2 \sigma_Y^2] = [(2,0)^2 (1,0)^2 \sigma_X^2] = 4,0 \sigma_X^2.$$

Сравнивая это с параллельным результатом в 4.20, мы видим, что сильный механизм поощрения дает четырехкратное увеличение дисперсии статуса. В системе с сильным поощрением и с обратной связью дисперсия Z имеет значение:

$$\sigma_Z^2 = \left[\frac{2}{1 - 2(-0,25)} \right]^2 \sigma_X^2 = \frac{16}{9} \sigma_X^2.$$

Здесь дисперсия увеличилась меньше чем вдвое. В регулируемой системе достижения щедро вознаграждаются, но крайнее расхождение в статусе, которое обычно проявляется, умеряется связыванием успеха с недостатком дальнейших достижений. Эта система служит функцией, порождающей чувство удовлетворения индивида на уровне поощрения и в то же время исключаяющей излишнее социальное неравенство. Однако регулируемая система имеет и отрицательную сторону — растрату таланта:

$$Y = X \text{ без обратной связи } (a = 1, c = 2, d = 0),$$

$$Y = \frac{2}{3}X \text{ с обратной связью } (a = 1, c = 2, d = -\frac{1}{4}).$$

7а. Диаграмма показывает, что корреляция между t_1 и t_2 дает возможность оценить h :

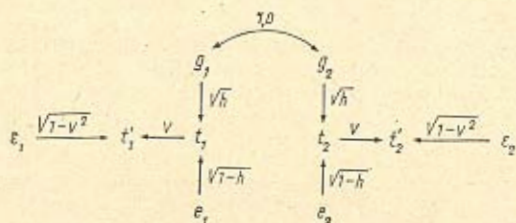
$$\rho_{t(1)t(2)} = \sqrt{h}(1,0) \sqrt{h} = h.$$

Чтобы вычислить корреляцию, мы берем большое число близнецов и проводим измерение всех их по одной схеме. Затем внутри каждой пары мы произвольным образом обозначаем измерение одного из близнецов через t_1 , а другого — через t_2 . Наконец, вычисляется корреляция между парами близнецов.

76. Если близнецы росли вместе, вероятно, что они воспитывались в одинаковом материальном и социальном окружении. Следовательно, мы должны модифицировать диаграмму, чтобы отразить e_1 и e_2 как коррелированные до некоторой неизвестной степени. Тогда, однако, корреляция между t_1 и t_2 есть функция подобия как генетического, так и окружения. Для исключения этого обстоятельства исследователи пытаются найти близнецов, выросших в разном окружении.

7в. Если умственные способности в целом детерминируются окружением, то теоретически возможно изменить их в желательном направлении путем преобразования окружающей среды. Преодоление генетической обусловленности вмешательством окружающей среды может быть кардинальным и дорогостоящим. Будет ли это оправданным — вопрос социальной значимости. (Некоторые близнецы-риальности генетики утверждают, что линейная модель не совсем подходит в этом случае. Улучшение окружающей среды, говорят они, дает рост умственных способностей до определенного предела и дальнейшее улучшение неэффективно.)

7г. Влияние ошибки измерения можно увидеть, перестроив диаграмму и различая истинные величины t и наблюдаемые с ошибкой (как в упражнении 4).



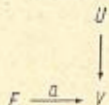
Теперь корреляция между наблюдаемыми отметками есть

$$\rho_{t_1' t_2'} = (v) \sqrt{h} (1,0) \sqrt{h} (v) = v^2 h.$$

Поскольку коэффициент валидности всегда меньше или равен 1,0 (это корреляция между истинной и наблюдаемой отметками), наблюдаемая корреляция будет преуменьшать значение h .

7д. Наследственность проявляется в основном в дисперсии стандартизованного путевого коэффициента. Если характеристики распределения популяции меняются, то меняется и основа стандартизации, а также и наследственность. В частности, возрастание изменчивости окружающей среды обычно проявляется возрастанием изменчивости характерных черт, в то время как генетическая изменчивость остается такой же. Итак, в новых популяциях генетическая изменчивость будет отвечать за пропорционально меньшую часть изменчивости характерных черт, и большинство коэффициентов наследственности будет уменьшаться.

8. Для целей эвристики пусть E обозначает составную переменную, представляющую все детерминанты эстетических ценностей дошкольной жизни, и пусть a — структурный коэффициент, который преобразует E в эстетические ценности школьных лет V . Пусть U обозначает другие детерминанты V , включающие важные события, происходившие в течение десяти школьных лет. Для всей совокупности связи могут быть представлены как

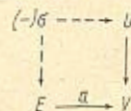


В этом случае корреляция между E и V действительно будет отражать в какой-то степени значение коэффициента a . В частности, квадрат корреляции или

коэффициент детерминации есть

$$\frac{\sigma_{EV}^2}{\sigma_E^2 \sigma_V^2} = \frac{a^2 \sigma_E^4}{\sigma_E^2 (a^2 \sigma_E^2 + \sigma_U^2)} = \frac{a^2 \sigma_E^2}{a^2 \sigma_E^2 + \sigma_U^2}.$$

Эта величина тем больше, чем больше a (в предположении, что изменчивость накопленного опыта дошкольной и школьной жизни примерно одинакова). Представленная выше модель, однако, не описывает изучаемую совокупность. Студенты допускаются к гуманитарному образованию частично на основе повышенных эстетических ценностей, а студенты-гуманитарии с низким значением E имеют, вероятно, высокое значение U . Те же, у кого низкое значение U , вероятно, обладают высоким значением E . Итак, E и U будут отрицательно связаны в совокупности, что отражено в следующей диаграмме:



Отрицательная ковариация между E и U уменьшает наблюдаемую ковариацию между E и V :

$$\sigma_{EV} = a\sigma_E^2 + \sigma_{EU}.$$

Соответственно уменьшается и корреляция между E и V . На самом деле может случиться, что наблюдаемая корреляция в этой совокупности будет пренебрежимо малой, даже когда значение a достаточно велико как теоретически, так и практически.

9. Похоже, что музыкальные интересы подростков выделяют тех из них, кто получил профессиональное музыкальное образование и опыт. Так, высокий уровень интересов предсказывает случай деятельности, приводящие к высокому уровню мастерства. Переменная «ранний интерес» коррелирует с последующим мастерством не обязательно как «причина» мастерства, но вследствие того, что она указывает тех, кто подвергается действию оператора профессионализации, преобразующего талант в достижения. Немного личностей в непрофессиональной совокупности подвергается действию оператора профессионализации. Следовательно, переменная «ранний интерес» дает небольшую информацию относительно позднейшего развития мастерства.

Глава 5

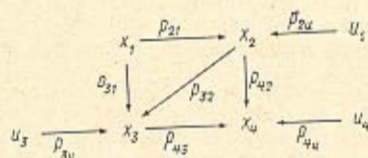
1. Переменные могут быть включены в некоторую рекурсивную модель только в том случае, если мы можем построить некоторое иерархическое упорядочение причин и следствий. К тому же возмущения каждой переменной должны быть некоррелированы со значениями источников тех же переменных. Фактически соотношения могут быть собраны в систему причин и следствий без петель; соответствующее обсуждение представлено в ответе к упражнению 5 гл. 1. Вопрос независимости возмущений порождает несколько различных подходов. Исключая возможность петель, мы уже тем самым исключаем один возможный источник «независимости». Дальнейшее рассмотрение иллюстрируется ниже.

Пропускание Можем ли мы основание предполагать, что отдельные случаи в наблюдаемой совокупности могут быть отсеяны на основе любой из четырех описанных переменных? Скорее всего, да. Сыновья отцов с низким социально-профессиональным положением могут иметь более высокий коэффициент смертности. Сыновья, чьи собственные социально-профессиональное положение и образование низки, могут не попасть в выборку при интервьюировании. Такие проблемы отражают тенденцию порождения отрицательных корреляций между переменными источников и возмущениями переменных пропускания. В идеале они могут быть скорректированы распространением анализа на все соответствующие случаи, но такое решение невозможно, если учтена дифференциальная смертность.

Если проблема оценки смещения параметра представляется достаточно серьезной, то мы должны отказаться от рекурсивности или ограничить совокупность событиями, в которых образование отцов (первичный вход) достаточно высоко, чтобы можно было пренебречь пропуском по другим переменным. Мы также должны рассмотреть возможность того, что переменная некоторого неопи-санного источника совместно с каким-либо описанным источником определяет пропускание. Эта идея может быть иллюстрирована. Предположим, что агрессив-ность есть неопи-санный источник социально-профессионального уровня сына. Предположим также, что агрессивные люди с недостаточным образованием наи-более подходят для отправки на военные операции. Результатом этого является неравномерность выбывания из изучаемой совокупности, порождающая положи-тельную корреляцию между образованием сыновей и нарушением их социально-профессионального состояния. Уцелевшие среди людей с ограниченным образо-ванием окажутся в большинстве довольно пассивными и поэтому уровень их ста-туса будет несколько ниже среднего.

Неопи-санные источники. Если переменная, которая влияет на две или более описанные переменные, окажется неучтенной, то это приведет к проблеме неза-висимых возмущений. Например, умственные способности отцов могут быть детер-минантами их образовательного уровня и они могут быть переданы генетически сыновьям, что также окажет непрямо-е влияние на их образовательный уровень. Такие возмущения в образовании сыновей будут коррелировать с детермина-нтами образования отцов, создавая ложную высокую корреляцию между образо-ванием отцов и сыновей. Если переменная умственных способностей не может быть явно учтена в анализе, необходимо отбросить идею рекурсивности в модели.

2. Модель имеет такой вид:



Переменные: x_1 — образование отца, x_2 — социально-профессиональный ста-тус отца, x_3 — образование сына, x_4 — социально-профессиональный статус сына. Первое структурное уравнение таково:

$$x_2 = p_{21}x_1 + p_{2u}u_2.$$

Это означает регрессию x_2 на x_1 и оценки коэффициентов (в соответствии с 3.32 и 3.34) будут следующими:

$$p_{21} = \beta_{21} = \rho_{21} \approx 0,516, \\ p_{2u} = \sqrt{1 - R_{2,1}^2} = \sqrt{1 - \rho_{21}^2} \approx 0,857.$$

Второе структурное уравнение

$$x_3 = p_{31}x_1 + p_{32}x_2 + p_{3u}u_3$$

определяет регрессию x_3 на x_1 и x_2 и оценки параметров (см. формулы в 3.33 и 3.34) будут иметь значения:

$$p_{31} = \beta_{31,2} = \frac{\rho_{31} - \rho_{32}\rho_{12}}{1 - \rho_{12}^2} \approx \frac{0,453 - (0,438)(0,516)}{1 - (0,516)^2} = 0,309,$$

$$p_{32} = \beta_{32,1} \approx \frac{0,438 - (0,453)(0,516)}{1 - (0,516)^2} = 0,278,$$

$$p_{3u} = \sqrt{1 - R_{3,12}^2} \approx (0,309)^2 + (0,278)^2 + 2(0,309)(0,278)(0,516) = 0,859.$$

Последнее структурное уравнение:

$$x_4 = p_{42}x_2 + p_{43}x_3 + p_{4u}u_4,$$

и коэффициенты оцениваются так:

$$p_{42} = \beta_{42,3} \approx \frac{0,405 - (0,596)(0,438)}{1 - (0,438)^2} = 0,178,$$

$$p_{43} = \beta_{43,2} \approx \frac{0,596 - (0,405)(0,438)}{1 - (0,438)^2} = 0,518,$$

$$p_{4u} = \sqrt{1 - R_{4,23}^2} \approx \sqrt{1 - 0,381} = 0,787.$$

3а. Регрессионные коэффициенты для предсказания I_1 по O_1 и O_2 будут оце-нивать полные эффекты $T_{I(1)O(1)}$ и $T_{I(1)O(2)}$ (см. правило II.17). В статистической литературе они называются «коэффициентами редуцированной формы». Заметим, что вполне можно пользоваться обычной техникой метода наименьших квадратов для оценки полных эффектов, поскольку I_1 и I_2 рекурсивно связаны с O_1 и O_2 в редуцированной модели. Для того чтобы определить, можем ли мы оценить a по коэффициентам редуцированной формы, нам следует вывести выражение для полного эффекта. Это требует применения принципа Мэсона (правило II.16), по-скольку имеется шесть петель в исходной системе:

$$(I_1 D_1) : L_1 = a, \quad (D_2 I_1) : L_4 = a, \\ (D_1 I_2) : L_2 = a, \quad (I_1 D_1 I_2 D_2) : L_5 = a^2, \\ (I_2 D_2) : L_3 = a, \quad (I_1 D_2 I_2 D_1) : L_6 = a^2.$$

Полный эффект I_1 имеет вид (аналогично выглядит и выражение для I_2):

$$T_{I(1)O(1)} = \frac{K(1 - L_2 - L_3)}{1 - L_1 - L_2 - L_3 - L_4 - L_5 - L_6 + L_1 L_3 + L_2 L_4} = \\ = \frac{K(1 - 2a)}{1 - 4a - 2a^2 + 2a^2} = \\ = \frac{K(1 - 2a)}{1 - 4a}, \\ T_{I(1)O(2)} = \frac{K(-2a)}{1 - 4a}.$$

Поскольку коэффициенты регрессии редуцированной формы позволяют оце-нить значения эффектов, мы можем вывести оценку для a , подставляя эти зна-чения в приведенные выражения, деля одно на другое и разрешая относительно a (непрямой подход метода наименьших квадратов):

$$\frac{b_{I(1)O(1) \cdot O(2)}}{b_{I(1)O(2) \cdot O(1)}} = \frac{1 - 2a}{-2a},$$

$$a = \frac{b_{I(1)O(2) \cdot O(1)}}{2(b_{I(1)O(2) \cdot O(1)} - b_{I(1)O(1) \cdot O(2)})}.$$

В действительности мы имеем две оценки a , поскольку может быть исполь-зована также регрессия, включающая I_2 .

3б. Если a обозначает действия альтруизма и эксплуатации и оно отрица-тельно, то L_1, L_2, L_3 — отрицательны и система содержит управляющую петлю. (Силовая система состоит целиком из усилителей.) Общий эффект в системе с управлением будет меньше. Следовательно, коэффициенты регрессии должны быть меньше по абсолютной величине и знаки $b_{I(1)O(2) \cdot O(1)}$ и $b_{I(2)O(1) \cdot O(2)}$ должны быть положительными, а не отрицательными. Тем не менее коэффициент a определен приведенной выше формулой.

4а. Любые заключения о том, что дает или не дает какая-либо переменная в качестве инструмента для выяснения связей, являются умозрительными и тем самым уязвимыми при теоретическом обсуждении. С учетом этого обстоятельства включенные процедуры иллюстрируются ниже. Все предлагаемые инструменты в этом управлении суть различные проявления окружающей среды. Поскольку мы намерены иметь дело с малыми (неиндустриальными) обществами, можно предположить, что никакая переменная окружающей среды не определяется прямо или косвенно переменными состояниями общества S , M или I . Соответственно переменные среды удовлетворяют лишь одному условию — быть инструментом.

Следующий вопрос: будет ли каждая осмысленная цепь связей от предлагаемого инструмента и S всегда проходить через M или I ? Технология пропитания, вероятно, зависит в какой-то степени от продолжительности периода выращивания. Длительный период выращивания, однако, позволяет произвести излишки пищевых продуктов для поддержания системы расслоения общества, даже если оно применяет примитивную технологию. Итак, переменная «период выращивания» может иметь влияние на S , которое не проходит через измеряемые величины M или I , и продолжительность периода выращивания не может служить инструментом в этой проблеме. Подобная же проблема связана с плодородием почвы. Большинство следствий повышения плодородия определяется посредством влияния уровня технологии пропитания, но если бы мы как-то повысили плодородие земли в данном обществе, это могло бы дать некоторую прибавку продовольственных излишков и уровня S даже при неизменной технологии. Итак, переменную плодородия следует также отвергнуть в качестве инструмента. Заметим, что и продолжительность периода выращивания, и плодородие почв могут вернуть свое значение в качестве потенциальных инструментов, если пашни материальным индикатором будет мера продовольственной продукции. (Существующая технология, по-видимому, один из детерминантов количества продовольственной продукции.) К несчастью, уровень производства продовольствия нелегко выявить из этнографических отчетов. Напротив, возможно мы сможем доказать, что влияние «короткого цикла», упомянутое в предыдущем разделе, пренебрежимо мало в обсуждаемом обществе или в некотором специальном образом выделенном подмножестве обществ. В этом случае также упомянутые переменные могут обсуждаться как инструменты.

Суровость места обитания может коррелировать с технологией пропитания, поскольку суровые места обитания связаны иногда с менее развитой технологией (хотя здесь нет строгого детерминизма). Эта слабая связь будет, по-видимому, поглощать какие-либо строго материальные влияния суровых мест обитания на социальную структуру. Однако общество может бороться с постоянными, предсказуемыми опасностями (пустыни, сильные наводнения, скалы), требуя строгой дисциплины, и традиции социальной дисциплины в свою очередь подкрепляют развития иерархических социальных структур. Следовательно, и эта переменная может иметь связь с S , не проходящую через M или I , поэтому она также непригодна как инструмент.

С другой стороны, мы можем обратить внимание на более непостоянные опасности места обитания, защита от которых требует скорее индивидуальной бдительности, чем групповой дисциплины (например, ядовитые и хищные животные, внезапные штормы и опасности типа подвижки или разлома плавучих льдин). Такие условия могут оказывать некоторое сдерживающее действие на развитие технологии пропитания и таким образом влиять на S через M . Эти условия могут также способствовать развитию высокой чувствительности восприятия и способности к анализу [Witkin H. A. A cognitive-style approach to cross-cultural research. *International Journal of Psychology*, 2, 1967, 233—250]. Если это связано с тем, что некоторые социологи называли «чувственной» (sensate) ориентацией, возможно, она препятствует развитию влиятельной идеологии нравственности, поддерживаемой религией, и вследствие этого препятствует усложнению иерархической социальной структуры. Поскольку трудно представить другие цепи связей, посредством которых непостоянные угрозы месту обитания могут быть связаны с социальной структурой, уточненная переменная места обитания представляется подходящим инструментом в проблеме. Ее пригодность зависит тогда от силы связи с M и I .

Доступность рудных месторождений способствует развитию более совершенной технологии пропитания, использующей металлические мотыги, плуги и другой инвентарь, что в свою очередь существенно увеличивает излишки, направляемые для поддержания развитой социальной структуры. Металлургия тоже, вероятно, приводит к развитию совершенных орудий, которые могут стать частью технологии пропитания, но могут также породить новые сильные связи и развитую социальную иерархию. Следовательно, мы имеем прямую связь с S , которая проходит через M и I и устраняет использование запасов руды как средства идентификации коэффициентов в первичных структурных уравнениях. Не отвергая саму идею, можно попытаться расширить уравнение путем представления орудий как детерминанта иерархической социальной структуры. Вместе с переменной, которая явно измеряет развитие орудий в модели, запасы руды пересматриваются как потенциальный инструмент для связи между M и S .

К несчастью, однако, даже этого будет недостаточно. Выплавка металлов скорее всего породит специалистов-металлургов и это доказывает, что любое углубление разделения труда способствует усилению развития иерархической социальной структуры. Разумеется, мы можем явно включить этих специалистов в модель. Однако тогда мы вынуждены будем добавить две новые переменные в структурные уравнения и должны искать дополнительные инструменты для идентификации их эффектов.

Из четырех рассмотренных переменных только одна представляется в качестве возможного инструмента (после уточнения ее определения). Одного инструмента недостаточно, чтобы идентифицировать два коэффициента первоначального структурного уравнения. С этой точки зрения возможны три различных подхода (отвлекаясь от других аспектов проблемы):

- а) продолжить поиск подходящих инструментов для использования в первоначальной форме уравнений;
- б) переопределить переменную M как продовольственную продукцию и посмотреть период выращивания и плодородие почв как инструменты;
- в) разработать уравнения для учета большего числа детерминантов S (например, орудия для обработки металла или специалисты-металлурги). Последнее требует уже большего множества инструментов, но задача может стать легче, поскольку мы уже не ограничиваемся переменными, которые влияют на S только через M или I .

4б. Как мы выяснили в предыдущем обсуждении, уточнение инструмента есть дело скорее содержательных теорий, чем статистики. Так, знание того факта, что продолжительность сезона выращивания имеет нулевую корреляцию с S (или I), несущественно при решении вопроса о пригодности этой переменной в качестве инструмента для связи $M \rightarrow S$. В частности, нулевая корреляция с S обязательно влечет нулевую корреляцию с возмущениями S в указанном уравнении.

5. Бытующие установки или поведение, допускающие алкоголь, не влияют на религиозное воспитание или возраст студентов. Алкоголизм может препятствовать приработкам или ускорять прекращение помощи родителей. Мы можем, однако, предположить, что такая серьезная и очевидная проблема, как пьянство, достаточно редко встречается среди студентов и ее можно игнорировать. (Кроме того, респондента можно спросить, приводило ли пьянство к финансовым санкциям, для того чтобы исключить любого, кто был субъектом такого оператора.) Соответственно три предложенных инструмента не влияют на установки или поведение и поэтому они удовлетворяют одному из обязательных условий. Наконец, необходимо определить, будет ли влияние предложенных инструментов на одну из зависимых переменных всегда осуществляться посредством других.

Фундаментальное воспитание благоприятствует антиалкогольной установке. Таким образом, чтобы стать инструментом, переменная должна влиять на характер употребления алкоголя только через эффект установки. Возможности и образцы, способствующие пьянству, будут отсутствовать в фундаменталистском доме: дети будут изолированы от «обычных» источников формирования пристрастия к алкоголю и будут, вероятно, пить даже меньше, чем это определяется их установкой. Родители, настроенные крайне фундаменталистски, могут пытаться продлить эту изоляцию, аргументируя это ограниченностью удобства проживания в кампусе (студенческом общежитии), но даже определенные успехи родителей в контроле окружения студента могут быть сведены на нет в большом

университете штата после первого года учебы, так что анализ должен быть ограничен студентами старших курсов, чтобы сохранить инструмент. Крайне фундаменталистски настроенные родители могут также отказаться посылать своих детей в большой университет, так что часть студентов отсеется по переменной «религии». Это будет означать, что соотношение между фундаментализмом и установками будет недооцениваться, не предотвращая в то же время использования переменной в качестве инструмента, пока корреляция не станет нулем.

Количество расходуемых денег, по-видимому, не влияет на отношение студентов к алкоголю, хотя допустимо некоторое не прямое влияние на то, сколько он пьет. Главная проблема здесь заключается в том, что количество расходуемых денег нелинейно связано с характером употребления алкогольных напитков. Недостаток средств приводит к уменьшению их потребления, тогда как даже небольшое количество карманных денег может быть использовано при случае для выпивки в компании. Отклонения от этого имеют плохо предсказуемые значения, поскольку лишние деньги могут быть использованы очень многими способами. Итак, переменная, которая может в какой-то степени использоваться как инструмент, должна измерять аскетизм относительно к уровню финансового положения.

Студенты с ограниченным бюджетом, возможно, откажутся от участия в компаниях, склонных к выпивке (или будут отвергнуты ими). Таким образом, их финансовые возможности управляют выбором товарищей, влияние которых может привести к несоответствию между установками и поведением. Здесь даже переменная «аскетизм» не может быть использована как инструмент до тех пор, пока не будут подвергнуты рассмотрению также и нормы друзей.

Студенты могут пить пиво легально, но крепкие напитки недоступны им (в соответствии с законом) до достижения 21 года. Допустимо предположить, что открывающиеся после достижения второго легального возраста возможности увеличивают общее количество потребляемого алкоголя на старших курсах, так что возраст будет иметь отношение к поведению. Здесь, по-видимому, действует чисто психологический механизм, связывающий повзросление с отношением к алкоголю. Мы окажемся в трудном положении при попытке описать, как совершеннолетние будут оказывать влияние на отношение к алкоголю. Однако друзья по колледжу обычно принадлежат к той же возрастной когорте и все однокурсники одновременно получают право употребления крепких спиртных напитков. Следовательно, возраст студента будет коррелировать с отклонением групповых норм, так же как и с индивидуальным поведением. Снова мы имеем переменную, которая может служить инструментом только в том случае, если мы явно рассматриваем нормы друзей.

Механизм пропуска может ставить дополнительные проблемы определения инструментов. В случае со студентами большая доля отсеивания на входе есть следствие выбывания из университета, таким образом, механизм пропуска связан главным образом со средним баллом. По-видимому, весьма сомнительно, что это пропускание имеет какую-либо прямую связь с отношением к алкоголю, но существует некоторая возможность связи с алкогольным поведением. Например, многие студенты-невротики могут быть горькими пьяницами и будут отчислены вследствие плохой успеваемости. Тогда, если невротизм был связан, скажем, с религиозным воспитанием, последняя переменная не будет более полезным инструментом. Если некоторые студенты допиваются до исключения, это явление имеет тенденцию создавать отрицательную корреляцию между возмущением переменной поведения и значением ее источников. Однако ранее уже предполагалось, что это явление не носит всеобщего характера.

В лучшем случае мы имеем теперь только один инструмент, которого недостаточно для идентификации двух коэффициентов в петле «установка — поведение». Другие две переменные могут опять подвергнуться обсуждению, если мы в явном виде рассмотрим групповые нормы. Однако связи между групповыми нормами и установками и поведением индивидуумов обоюдные, так что по меньшей мере один дополнительный инструмент потребуется для идентификации влияния норм.

6. Для того чтобы оценить все коэффициенты, мы должны построить две новые переменные V и Z , посредством регрессии V и Z соответственно на X и Y . Заметим, что переменные не стандартизованы (некоторые дисперсии больше 1,0),

Следовательно, формула в 3.29 применима для получения регрессионных коэффициентов.

$$b_{VX \cdot Y} = \frac{\sigma_Y^2 \sigma_{VX} - \sigma_{VY} \sigma_{XY}}{\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2} = \frac{(1,0)(0,283) - (-0,113)(0,3)}{(1,0)(1,0) - (0,3)^2} = 0,348,$$

$$b_{VY \cdot X} = \frac{(1,0)(-0,113) - (0,283)(0,3)}{(1,0)(1,0) - (0,3)^2} = -0,218,$$

$$b_{ZX \cdot Y} = \frac{(1,0)(0,235) - (0,466)(0,3)}{(1,0)(1,0) - (0,3)^2} = 0,105,$$

$$b_{ZY \cdot X} = \frac{(1,0)(0,466) - (0,235)(0,3)}{(1,0)(1,0) - (0,3)^2} = 0,435.$$

Для оценки a мы построим V , используя первые два из этих коэффициентов:

$$V = 0,348X - 0,218Y.$$

Применяя процедуру, намеченную в формулировке упражнения, мы получим следующие величины:

$$\sigma_V^2 = 0,123, \quad \sigma_{VY} = -0,114, \quad \sigma_{VZ} = -0,020.$$

Эти и другие заданные значения позволяют получить коэффициенты регрессии, оценивающие a и e :

$$b_{ZV \cdot Y} = \frac{\sigma_Y^2 \sigma_{ZV} - \sigma_{ZY} \sigma_{VY}}{\sigma_V^2 \sigma_Y^2 - \sigma_{VY}^2} = \frac{(1,0)(-0,020) - (0,466)(-0,114)}{(0,123)(1,0) - (-0,114)^2} = 0,301,$$

$$b_{ZY \cdot V} = \frac{(0,123)(0,466) - (-0,020)(-0,114)}{(1,0)(0,123) - (-0,114)^2} = 0,500.$$

Статистики для Z определяются подобным образом:

$$\hat{Z} = 0,105X + 0,435Y, \quad \sigma_{XZ} = 0,236,$$

$$\sigma_Z^2 = 0,228, \quad \sigma_{VZ} = -0,019.$$

Регрессионные оценки c и d получаются через регрессию V на Z и X :

$$b_{VZ \cdot X} = \frac{(1,0)(-0,019) - (0,283)(0,236)}{(0,228)(1,0) - (0,236)^2} = -0,498,$$

$$b_{VX \cdot Z} = \frac{(0,228)(0,283) - (-0,019)(0,236)}{(1,0)(0,228) - (0,236)^2} = 0,400.$$

Первоначальные дисперсии и ковариации были получены при следующих значениях параметров: $a = 0,300$, $c = -0,500$, $d = 0,400$, $e = 0,500$.

Разница между этими и вычисленными оценками вызвана ошибками округления.

7а. Для того чтобы оценить a , мы должны построить новую переменную $\hat{E}$ посредством регрессии E на S и U . Так как переменные стандартизованы, можно применить формулу из 3.33:

$$b_{ES \cdot U} = \frac{-0,574 - (0,226)(-0,556)}{1 - (-0,556)^2} = -0,6490,$$

$$b_{EU \cdot S} = \frac{0,226 - (-0,574)(-0,556)}{1 - (-0,556)^2} = -0,1348,$$

$$\hat{E} = -0,6490S - 0,1348U.$$

Соответствующие статистики, содержащие E , получаются, как и в упражнении 6:

$$\sigma_{\hat{E}}^2 = 0,3421, \sigma_{\hat{E}U} = 0,2260, \sigma_{\hat{E}I} = 0,4094.$$

Коэффициенты частной регрессии, оценивающие a , могут быть вычислены из этих и заданных статистик. (Мы обязаны теперь употребить формулу из 3.29, поскольку переменная E не стандартизована.)

$$b_{I\hat{E} \cdot U} = \frac{(1,0)(0,4094) - (0,569)(0,2260)}{(0,3421)(1,0) - (0,2260)^2} = \frac{0,2808}{0,2910} = 0,9649.$$

Подобно этому для получения оценки s мы сначала определим новую переменную $\hat{I}$ и соответствующие статистики:

$$\begin{aligned}\hat{I} &= -0,6262S + 0,2208U, \\ \sigma_{\hat{I}}^2 &= 0,5946, \sigma_{\hat{I}S} = -0,7490, \sigma_{\hat{I}E} = 0,4093.\end{aligned}$$

Взяв эти числа и первоначальные статистики, получаем оценку s :

$$b_{E\hat{I} \cdot S} = \frac{-0,0206}{0,0336} = -0,6127.$$

76. Ключевые величины, необходимые для получения оценки a при использовании второй модели, суть

$$\begin{aligned}\hat{E} &= -0,7563P - 0,3102M, \\ \sigma_{\hat{E}}^2 &= 0,5589, \sigma_{\hat{E}M} = -0,1340, \sigma_{\hat{E}I} = 0,3412, \\ b_{I\hat{E} \cdot M} &= \frac{0,3660}{0,5409} = 0,6766.\end{aligned}$$

Аналогичные величины для оценки s :

$$\begin{aligned}\hat{I} &= -0,5117P + 0,0658M, \\ \sigma_{\hat{I}}^2 &= 0,2819, \sigma_{\hat{I}P} = -0,5270, \sigma_{\hat{I}E} = 0,3412, \\ b_{E\hat{I} \cdot P} &= \frac{-0,0193}{0,0042} = -4,6195.\end{aligned}$$

7в. Коэффициент a имеет довольно большое положительное влияние, а коэффициент s отрицателен в обеих моделях. Оценки не являются, однако, равнозначными. В частности, каждая из конечных оценок s получается делением очень маленького числа на другое очень маленькое число. При таких малых объемах выборки вполне вероятно, что выборочные отклонения статистик будут столь же велики, как и величины, используемые при оценке s . Итак, оценки s могут быть совершенно различными, если мы изучаем различные выборки SMSA. Оценки a менее уязвимы в этом отношении. В частности, поскольку мы в обоих случаях делим на достаточно большое число, они менее подвержены выборочным колебаниям. Следовательно, мы можем быть более уверены в их воспроизводимости.

Используя более развитый анализ, мы можем оценить вероятность того, что истинные коэффициенты отличны от нуля, только по выборочным значениям. Такие процедуры не описываются в этой книге (они обсуждаются в эконометрической литературе). В целом все-таки проверка статистической значимости оценок придает больше объективности при рассмотрении их точности.

Оценки s ненадежны, поскольку ни один из предложенных инструментов не является адекватным с точки зрения статистики. Не совсем очевидно из первичной корреляции, что U и M такие слабые инструменты. Как оказывается, их корреляция с I во многом ложная в контексте этих моделей, т.е. когда мы управляем S или P , остается лишь небольшая связь между U и I или между M и I .

С другой стороны, мы можем увидеть здесь проблему коллинеарности. Корреляция между независимыми переменными (r_{SU} или r_{PM}) слишком высока по сравнению с корреляцией между независимой и зависимой переменными (r_{UI} и r_{MI}). (Мы употребляем букву r для обозначения выборочной оценки истинной корреляции ρ .) Заметим, что проблема коллинеарности в двухэтапном методе наименьших квадратов не обязательно означает, что мы имеем крайне большие корреляции между независимыми переменными.

Одна из оценок s имеет абсолютное значение, намного превышающее 1,0. Это теоретически и статистически возможно. Однако такие значения практически достаточно редки, чтобы служить предостережением. В этих случаях скорее следует видеть сигнал неточности оценки s , как упоминалось выше.

С точки зрения статистики оценки a , видимо, заслуживают некоторого доверия. Обе оценки имеют несколько отличающиеся значения, но это типичное явление и мы можем получить одну улучшенную оценку, используя S и P вместе в одном анализе двухэтапным методом наименьших квадратов. С другой стороны, мы можем изучить различие между оценками, чтобы сопоставить относительные достоинства S и P в качестве инструментов, что иллюстрируется ниже.

При любых статистических данных оценки a и s будут плохи, если пригодность S , P , U и M в качестве инструментов теоретически не обоснована. Теоретическая адекватность каждого инструмента кратко обсуждается ниже.

S : образование и доход группы расовых меньшинств не влияют на их местоположение в городе, поэтому одно условие для использования S как инструмента удовлетворено. Небелые (и белые также) получают небольшое образование на Юге, что подтверждается связью от S к E . Мы можем предположить, что низкие доходы небелых определяются исключительно плохим образованием и это предстает по крайней мере понятным. Хотя, разумеется, расовая дискриминация в области занятости может быть более серьезной причиной для меньшинств на Юге. Это открывает путь от региона к доходу, обходящий образование, отвергая тем самым S как плохой инструмент. Если мы неправильно использовали S в качестве инструмента, нам следует ожидать смещения оценки a . В частности, корреляция между S и I будет слишком сильно отрицательной, чтобы быть объясненной только лишь E , и, если r_{SI} много меньше нуля, оценка a будет иметь завышенное положительное значение. (Это очень легко показать на примере процедуры, описанной в 5.12. Оценка a непрямим методом наименьших квадратов имеет значение:

$$\frac{b_{IS \cdot U}}{b_{ES \cdot U}} = \frac{-0,6262}{-0,6490} = 0,9649.$$

Таким образом, оценка a прямо зависит от величины $b_{IS \cdot U}$, которая отрицательна и чрезмерно велика, если таково же значение r_{IS}). На самом деле, мы установили, что оценка a при использовании S в качестве инструмента в большой мере сопоставима с оценкой, применяющей P .

U : мы теперь знаем, что это неадекватный инструмент, с точки зрения статистики и этому имеется теоретическое объяснение. При спецификации явно предполагалось, что объединение дает значимый эффект изменения доходов меньшинств, подкрепляя путь $G \rightarrow I$. В оригинальной статье Хилл представил два аргумента относительно этой связи. Первый аргумент утверждает, что сплочение увеличивает доходы меньшинств, в противоположность точке зрения классических экономических теорий, делающих упор на конкуренцию между группами. Итак, мы имеем взаимно погашающиеся действия и нет серьезных теоретических оснований ожидать какого-либо пути безусловной связи $U \rightarrow I$.

Пригодность спецификации U включает и другие факторы. Объединение в городе, вероятно, не влияет прямо на уровень образования в расовых меньшинствах; поэтому отсутствие пути $U \rightarrow E$ представляется правдоподобным. Заданная спецификация также определяет, что объединение никоим образом не влияет на уровень образования или доходов в расовых меньшинствах. Но, возможно, образованные члены этих групп проводят работу, чтобы получить защиту, обеспечиваемую союзами ($E \rightarrow U$), или, возможно, растущие доходы меньшинства выглядят как опасность, которая побуждает белых рабочих объединяться для защиты своих трудовых интересов ($I \rightarrow U$). Различные функции союзов

меньшинств подразумеваются в этом обсуждении, но в некоторых случаях не исключена возможность и того, что модель окажется неадекватной. Такая неадекватность является источником ошибок в оценке как a , так и c , т. е. если U не является заранее определенной переменной в связи с E и I , мы не должны включать U на первой стадии определения E . (Если отбросить U , то значения a окажутся большими, чем мы имели ранее.)

P : вторая модель рассматривает процент небелых в городе как причину образовательного уровня небелого меньшинства. Предположение заключается в том, что школы в городах с большой долей небелого населения получают мало помощи, дают образование невысокого качества и не занимаются воспитанием.

Высокий средний уровень образования и дохода среди небелых могут поощрять иммиграцию небелых в город, а низкое образование и доход могут вызывать эмиграцию. Таким образом, мы можем иметь и $E \rightarrow P$, и $I \rightarrow P$. Это может означать тем не менее, что такие эффекты слабы, медлительны и к тому же коррелированы с миграцией белых, так что P может считаться практически не подверженным влиянию E или I . Имеет ли P какое-либо влияние на I , которое не проходит через E ? Один из аргументов состоит в том, что если имеется большая доля небелых, то предложение труда для «небелых занятий» велико. Это снижает доходы небелых. Тем не менее включение E и M должно элиминировать проблему. Низкий уровень E — более прямой индикатор, чем P , превышения предложения дешевого небелого труда, а M — подходящий индикатор спроса на «синие воротнички». Итак, представляется совершенно невероятным, чтобы изменения P влияли на I помимо изменений E и M .

M : сомнительно, что доходы группы меньшинств имеют большое влияние на объем производства в городе. В предприятия вложены существенные капиталы, и они должны располагаться вблизи основных ресурсов и рынков, так что на небольшом отрезке времени объем производства в большом городе не очень влияет на обычные колебания в заработной плате (белых и небелых). С этой позиции уровень образования небелых должен иметь большое влияние на долю лиц, занятых в производстве. Небелые, понятно, могут самообразовываться вне работы на заводе, но работа остается и кто-либо получит ее. Большой объем производства в городе может обеспечить занятость, что отвлечет небелых из школ и снизит таким образом уровень их образования. Тем не менее такие эффекты проявляются через I и подтверждают теоретическую пригодность M в качестве инструмента.

Обсуждение показывает, что инструменты во второй модели теоретически более состоятельны, чем в первой. Однако это подходящий случай еще раз напомнить, что лучшее теоретизирование может привести к неоднозначным выводам.

8. Используя формулы в 5.23, мы получим оценку a :

$$\frac{\sigma_{PI}}{\sigma_{PE}} = \frac{-0,527}{-0,684} = 0,770.$$

Оценка с следующей:

$$\frac{\sigma_E^2 \sigma_{PI} - \sigma_{EI} \sigma_{PE}}{\sigma_{EI} \sigma_{PI} - \sigma_{PE}^2} = \frac{(1,0)(-0,527) - (0,709)(-0,684)}{(0,709)(-0,527) - (1,0)(-0,684)} = \frac{-0,0420}{0,3104} = -0,136.$$

Мы снова приходим к заключению, что a существенно положительна и что c имеет значение, меньшее нуля или равное ему.

Все эти результаты зависят от совсем малой выборки городов, что может быть возражением против наличия вполне удовлетворительных инструментов. Таким образом, придерживаясь абсолютной строгости, мы должны отказаться делать какие-либо заключения вообще. С другой стороны, мы приходим в основном к одним и тем же заключениям, пользуясь тремя различными моделями, которые включают отчасти разные теоретические предположения. Предположим, дополнительно, что это лучшая информация, которая нам доступна, и что важные политические решения уже приняты. Тогда благоразумие требует переступить пределы абсолютной строгости.

Все три анализа допускают, что улучшение образовательного уровня небелых повышает их средний доход. Все анализы также предполагают, что требование повышения заработной платы будет иметь результат, ограниченный только средним доходом в лучшем случае. Если средний доход небелых поднимается — это еще не показатель улучшения образования; фактически имеется намек на противоположное: в американском обществе, как установлено в 1960 г., высокая заработная плата могла соблазнить небелую молодежь прервать свое образование. Таким образом, продолжение работы по повышению образования небелых представляется лучшей стратегией подъема их общего социально-экономического статуса.

9. Оценка ковариации возмущений, основанная на неточных измерениях, может быть определена в терминах статистик истинных переменных:

$$\begin{aligned}\bar{\sigma}_{UV} &= \sigma_{YZ} - \frac{\sigma_{XZ}(\sigma_Y^2 + \sigma_F^2)}{\sigma_{XY}} = \\ &= \sigma_{YZ} - \frac{\sigma_{XZ}\sigma_Y^2}{\sigma_{XY}} - \frac{\sigma_{XZ}\sigma_F^2}{\sigma_{XY}} = \\ &= (\sigma_{UV}) - b\sigma_F^2.\end{aligned}$$

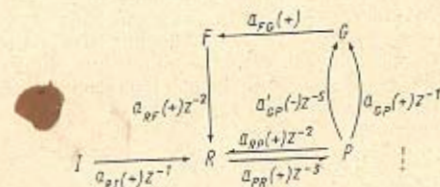
Таким образом, оценка смещена (помимо выборочной ошибки), поскольку она равна истинному значению минус величина $(b\sigma_F^2)$. Дисперсия ошибки измерения Y всегда положительна. Следовательно, оцененная ковариация между возмущениями будет слишком мала, когда коэффициент b положителен (возможна даже ложная отрицательная корреляция). Оценка будет слишком велика, когда коэффициент b отрицателен.

Несмещенные оценки ковариаций возмущений иногда применяются для определения возможности выбрасывания ключевых переменных из описания системы. Однако, если переменные измерены неточно, оценки ковариаций возмущений смещены (даже если оценки структурных параметров не имеют смещения). В частности, их значения зависят от ошибок измерения так же, как и от истинных значений ковариаций возмущений. Таким образом, мы можем сделать ошибку, интерпретируя оценки ковариаций возмущений только с позиций теории.

Этот вывод верен вообще и для более сложных систем, оцениваемых процедурой двухэтапного метода наименьших квадратов. Проблема может быть преодолена за счет использования инструментальных переменных, составных индикаторов и оценочных «методов полной информации» для нахождения одновременно структурных коэффициентов, ковариаций возмущений и значимости индикаторов (см. статью К. G. Jöreskog в библиографии к данной главе).

Глава 6

1а.

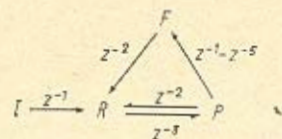


16. Поскольку уровень федеральных вложений прямо зависит от G , мы можем предположить, что очевидная функция инвестиционного органа состоит в том, чтобы поощрять расширение исследований по теме. Но изучение потокового графа обнаруживает, что, хотя такие органы действительно обеспечивают кратковременное усиление исследований (время оборота — 6 лет), они также вводят

долговременную, более высокого порядка, управляющую петлю в систему (время оборота — 10 лет). Таким образом, они выполняют скрытую функцию ограничения объема исследований по теме. Это явление в действительности есть следствие беспокойства за развитие. В какой-то мере это беспокойство скорее в связи с длительными достижениями, чем с коэффициентом роста a_{GP} , приближается по значению к нулю и управляющая петля исчезает.

1в. Кратчайшее время возвращения для R через петлю (RP) — пять лет.

2а.



2б. Другие два разностных уравнения следующие:

$$P_t = R_{t-3}, \quad F_t = P_{t-1} - P_{t-5}.$$

2в.

Период времени	I	R	P	F
2	0	0	0	0
3	0	0	0	0
4	0	0	1	0
5	0	0	0	1
6	0	1	0	0
7	0	1	0	0
8	0	0	0	0
9	0	0	1	-1
10	0	0	1	1
11	0	0	0	1
12	0	2	0	0
13	0	1	0	0
14	0	0	0	-1
15	0	0	2	-1
16	0	-1	1	2
17	0	1	0	1
18	0	3	0	0
19	0	1	-1	0
20	0	0	1	-3

Заметим, что первоначальный импульс в конце концов определяет периоды длительной исследовательской деятельности. Заметим также, что управляющая петля порождает такие колебания, что исследования и публикации иногда падают ниже их «среднего» уровня. Система, по-видимому, неустойчива (ср. года 6, 12 и 18). Тем не менее эвристический выбор значений коэффициентов порождает нереалистичные обратные эффекты в петлях.

3а. Имеются три уравнения:

$$Q_t = 0,50E_t + 0,10Q_{t-4},$$

$$I_t = 0,50Q_t + 0,05I_{t-1},$$

$$S_t = 0,10I_t + 1,05S_{t-1}.$$

3б. Значение Q_t может быть выражено через ранее определенные величины:

$$Q_t = 0,50E_t + 0,10(0,50E_{t-4} + 0,10Q_{t-8}) = 0,50E_t + 0,05E_{t-4} + 0,01Q_{t-8}.$$

Заметим, что этот же процесс может быть осуществлен еще раз:

$$Q_t = 0,50E_t + 0,05E_{t-4} + 0,005E_{t-8} + 0,001Q_{t-12}.$$

Фактически это можно продолжать многократно, что дает простую формулу, которая выражает настоящее значение Q в виде взвешенной суммы прошлых значений E . То же самое может быть сделано и для всех других переменных системы, чтобы получить ее определение в виде временных рядов взвешенных состояний. Определение любой системы может быть приведено к этой форме.

Из этого примера видно, что только первые два члена временных рядов имеют существенные веса. Следовательно, мы можем сказать, что в этой системе достигнутый социально-профессиональный уровень зависит главным образом от квалификации в настоящее время, слабее — от дипломов четырехлетней давности и лишь незначительно — от более ранних условий.

3в. Припоминая, что $Q_t = 0$ для $t < 0$ и $S_0 = -0,25E_0 + 0,10I_0$, мы получим для отдельных моментов времени следующую таблицу (здесь образование принято равным 10 годам):

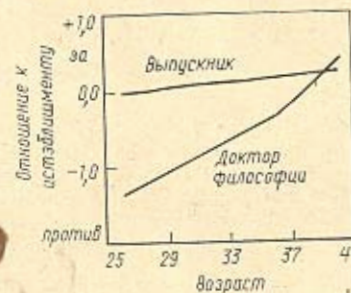
t	E	Q	I	S
0	10	5,00	2,50	-2,25
1	10	5,00	2,62	-2,10
4	10	5,50	2,88	-1,58
10	10	5,55	2,92	-0,15
15	10	5,55	2,93	1,41

3г. В этой системе доктор и выпускники средней школы имеют примерно равные сбережения в возрасте 38 лет. Затем доктор значительно опережает.

4а. Установка в заданный год получается применением формулы к значениям E , Q , I и S , отнесенным к тому же году; например, установка тридцатилетнего доктора философии ($t = 10$) есть

$$A = -0,2(10) + 0,2(5,55) + 0,2(2,92) + 0,4(-0,15).$$

На следующем графике представлены результаты для доктора философии и выпускника средней школы. (Точки были подсчитаны с интервалом два года.)



4б. Предполагая, что эта система действует на совокупность и что анархисты настроены решительно против истеблишмента, мы должны искать их среди образованной молодежи (особенно неработающей). Предполагая, что прогрессивные или либералы ни за, ни против истеблишмента, мы можем найти их почти в любой возрастной группе стабильного рабочего класса или в средневозрастной группе среднего класса. Те индивидуумы, которые хотят сохранения истеблишмента, легче всего могут быть обнаружены среди пожилых представителей

среднего класса. (Богатство без образования в молодом и даже чаще в пожилом возрасте также порождает консервативность.) Согласно этой модели сомнительно, что мы найдем много радикалов любого рода в стабильном рабочем классе. Временные диаграммы показывают, что позиция представителей этого класса почти нейтральна в течение их рабочей карьеры.

5а. Формулы, заданные в задаче, могут быть использованы для оценки значений I_t , L_t и D_t по значениям P_t , P_{t-1} , P_{t-2} , P_{t-3} , P_{t-4} и P_{t-5} . Поскольку P имеет значение 1,0 для всех этих моментов времени, оценки будут просто суммами коэффициентов в каждой формуле:

$$I_t = 0,50 + 0,0 + 0,08 - 0,03 + 0,02 - 0,01 = 0,56,$$

$$L_t = 0,0 + 0,20 - 0,06 + 0,05 - 0,03 + 0,02 = 0,18,$$

$$D_t = 0,0 + 0,0 + 0,08 - 0,03 + 0,02 - 0,01 = 0,06.$$

5б. Для статистического анализа мы пренебрегаем величинами zs в потоковом графе. В этом случае значения переменных таковы:

$$I = \frac{0,5(1 + 0,32)}{1 + 0,32 - 0,16}, \quad P = 0,57,$$

$$L = \frac{(0,5)(0,4)}{1 + 0,32 - 0,16}, \quad P = 0,17,$$

$$D = \frac{(0,5)(0,4)(0,4)}{1 + 0,32 - 0,16}, \quad P = 0,07.$$

Эти результаты подтверждают, что если вход системы поддерживается постоянным в течение шести периодов времени, то мы мало ошибемся, предполагая равновесие.

6а. Значения I_0 и D_{-1} могут быть определены по прошлым значениям, начиная с $t = -5$, с помощью формул упражнения 5:

$$I_0 = 0,50P_0 + 0,08P_{-2} - 0,03P_{-3} + 0,02P_{-4} - 0,01P_{-5},$$

$$D_{-1} = 0,08P_{-3} - 0,03P_{-4} + 0,02P_{-5}.$$

Определение будет удовлетворено только тогда, когда D_{-1} велико, а I_0 мало или когда $(D_{-1} - I_0)$ велико.

$$\begin{aligned} D_{-1} - I_0 &= 0,08P_{-3} - 0,03P_{-4} + 0,02P_{-5} - 0,50P_0 - 0,08P_{-2} + 0,03P_{-3} - \\ &\quad - 0,02P_{-4} + 0,01P_{-5} = \\ &= -0,50P_0 - 0,08P_{-2} + 0,11P_{-3} - 0,05P_{-4} + 0,03P_{-5}. \end{aligned}$$

Эта разность будет велика, когда P_0 , P_{-2} и P_{-4} малы, а P_{-3} , P_{-5} велики. Таким образом, «идеальный» исторический образ реформации в системе таков: период экономического подъема, затем спад, за следующим подъемом следует другой спад, затем период, в котором экономика еще на что-то способна, сменяется глубоким спадом к настоящему времени.

6б. Мы опять хотим сделать величину $(D_{-1} - I_0 - L_1)$ большой. Имеем:

$$\begin{aligned} (D_{-1} - I_0 - L_1) &= 0,08P_{-3} - 0,03P_{-4} + 0,02P_{-5} - 0,20P_{-1} + 0,06P_{-2} - \\ &\quad - 0,05P_{-3} - 0,03P_{-4} - 0,02P_{-5} - 0,20P_0 + 0,06P_{-1} - \\ &\quad - 0,05P_{-2} + 0,03P_{-3} - 0,02P_{-4} = \\ &= -0,20P_0 - 0,14P_{-1} + 0,01P_{-2} + 0,06P_{-3} - \\ &\quad - 0,02P_{-4} + 0,0P_{-5}. \end{aligned}$$

Таким образом, второй тип реформации может ожидаться, когда имеется спад, экономический подъем, затем период, в котором экономика слабо функционирует (пока это еще не глубокий спад), сменяется двумя периодами спада.

Даже здесь очевидно вследствие больших весов при P_0 и P_1 (они не связаны с D_{-1}), что большая часть изменяющегося законодательства обязана экономическим условиям, которые скорее поддаются влиянию элиты, чем возмущениям граждан. В то же время ясно, что общий образ, определяющий «реформацию», здесь, как и в первом случае, вытекает из колебаний излишков продукции. Если излишки могут поддерживаться на постоянном уровне, образ, определяющий реформу, не появится. Разумеется, такие «открытия» связаны только с системой, определенной выше. Включение дополнительных переменных или изменение структурных коэффициентов, или временное запаздывание может привести к разным заключениям.

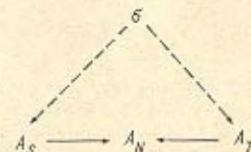
7а. Дисперсия имеет вид:

$$\sigma_{A(N)}^2 = \sigma_{A(S)}^2 + \sigma_{A(T)}^2,$$

т. е. наблюдаемая дисперсия установки есть сумма дисперсий, вызванной стабильными различиями людей, и дисперсии, определяемой различиями людей, обусловленной недавним опытом. Заметим, что последняя не является мерой отклонений от стабильных установок. Скорее она показывает отклонения от стабильных установок после их уточнения с помощью некоторого кратковременного тренда в совокупности, вызванного недавним синхронизированным опытом. Например, каждый мог недавно пережить нечто, вызвавшее положительные кратковременные сдвиги. Тогда все установки будут иметь тенденцию к повышению относительно их стабильных компонент и кратковременная изменчивость около среднего уровня подъема может остаться малой.

Стабильная компонента будет стремиться увеличиться в обществах, где поддерживается разнообразие индивидуальных социализаций и программ создания образов. Нестабильная компонента будет стремиться уменьшиться, если опыт людей относительно объекта, определяющего установку, будет редким, слабым, неличным и синхронизированным. Таким образом, мы можем догадываться, что стабильная дисперсия будет преобладающей в структурированном, плюралистическом обществе, в котором контакты с объектом установки осуществляются главным образом через средства массовой информации. Нестабильная дисперсия будет доминировать в группах с однородной культурой, в которых люди имеют частые, неkoordinируемые, личные встречи с объектом установки.

7б. Избирательное поведение влечет за собой то, что устойчивая установка индивидуума определяет в некоторой степени тип опыта, который он имеет, и, таким образом, получается некоторая координация между устойчивой установкой и кратковременной. Это может быть представлено так:



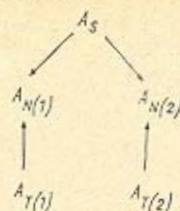
Теперь дисперсия A_N есть

$$\sigma_{A(N)}^2 = \sigma_{A(S)}^2 + \sigma_{A(T)}^2 + 2\sigma_{A(S)A(T)}.$$

Поскольку ковариация положительна, общая дисперсия чистой установки возрастает (в предположении, что изменчивость в недавних переживаниях внутри совокупности не определяется избирательным поведением).

7в. При заданных предположениях в этой проблеме определение чистой установки в момент один и момент два (игнорируя возможность избирательности)

может быть представлено диаграммой:



Корреляция между чистыми установками в оба момента времени есть

$$\rho_{12} = \frac{\sigma_{A(N)(1)A(N)(2)}}{\sigma_{A(N)(1)}\sigma_{A(N)(2)}} = \frac{\sigma_{A(S)}^2}{\sqrt{\sigma_{A(S)}^2 + \sigma_{A(T)(1)}^2} \sqrt{\sigma_{A(S)}^2 + \sigma_{A(T)(2)}^2}}$$

При предположении, что изменчивость в кратковременных изменениях одинакова в оба момента, уравнение сводится к

$$\rho_{12} = \frac{\sigma_{A(S)}^2}{\sigma_{A(S)}^2 + \sigma_{A(T)}^2}$$

Таким образом, корреляция равна доле дисперсии чистой установки, которая определяется стабильной компонентой (при условии, что стабильная компонента действительно не изменяется и кратковременные изменения не коррелированы во времени).

Социальные условия, определяющие большую долю стабильной дисперсии, были представлены в ответе 7а. Эти же условия позволяют дать наиболее точный прогноз поведения, зависящего от $A_{N(2)}$ по измерениям более ранней установки $A_{N(1)}$.

8. График может быть сведен к форме

$$U \xrightarrow{c(-)z^{-2}} D_\theta \circlearrowleft k(-)z^{-1}$$

откуда мы можем выписать и разложить в ряд разностное уравнение для D_θ :

$$\begin{aligned} D_{\theta(t)} &= kD_{\theta(t-3)} + cU_{t-2} = \\ &= cU_{t-2} + kcU_{t-5} + kD_{\theta(t-6)} = \\ &= cU_{t-2} + kcU_{t-5} + k^2cU_{t-8} + \dots \end{aligned}$$

Мы можем интерпретировать последний результат следующим образом: настоящий уровень конформизма поведения есть результат подкрепляющего исторического опыта. Это может навести на мысль, что бихевиористский подход обеспечивает динамические представления в системе, которая изучается на одномоментных множествах событий специалистами по когнитивной психологии или групповой динамике.

● ПРЕДМЕТНЫЙ И ИМЕННОЙ УКАЗАТЕЛЬ

Аверроэс (Ибн Рушд) 13
Агрегированные следствия 204, 208
Алгебра математических ожиданий 88
Алкогольные напитки 175, 237—238
Анализируемость причинности 18—19
Асимметрия 106, 224

Величины отклонений, определение 88—89

Вероятность 88

Вершины см. Потокосые графы, представление переменных

Ветви 51

Взаимодействия 56—57

Возмущение 39—40, 54—55, 58—59, 62, 73, 80, 93, 98, 109—110, 117, 122, 125—126, 128, 131, 133, 205—207, 216

Возраст 131, 175, 227, 237—238

Вооружение 46, 84, 215, 221—222

Вооруженные силы 46, 84, 215, 221—223

Временная шкала 181—182

Временное упорядочение 16—17, 24, 27, 162—163, 214

Временные ряды 204, 208, 245

Время оборота, определение 191

Входы 60, 183—184, 193—194

Входы, меняющиеся во времени 194—196, 202—207

Выборки 87, 123, 139, 153, 163—166, 171—172, 229, 233—234, 237—238

Выборочная ошибка 163—164, 168, 240

Выведение 76—78, 84, 155—156, 222

Выходные (зависимые) переменные 50, 58—60, 73—74, 78—79

Газали (Абу Хамид) 13

Генетика 123, 133, 200

Гипотетические переменные 74—75, 132, 154, 155. См. также Причинное приближение

Гомеостатическое управление см. Управляющие системы

Градиент поля см. Смежность

Графики распределений 86—88

Графы см. Потокосые графы

Группировка наблюдений см. Процедуры агрегирования

«Дважды отрицательный» усилитель 217

Двухэтапный метод наименьших квадратов 76, 150—157, 158, 160—162, 164, 166, 168—169, 176, 243

Диаграмма рассеивания 90—94

Дизъюнктивный пропускной механизм 105—124

Динамика 42—43, 71, 78, 84, 180—213

Динамический разброс 205—207, 247

Дисперсия, возмущения 94, 117, 126, 128, 141—142, 152, 154, 206—207

Дисперсия, определение 89, 106, 176, 225. См. также Статистическое разнообразие

Дисперсия ошибки 99—100, 104, 107, 224—225. См. также Остатки

Доход 46, 55—56, 82—83, 106, 131—132, 176—177, 209—210, 215, 219, 223—224, 227, 241—242, 245—246

Зависимые переменные см. Выходные переменные

Задача идентификации 137—139, 148, 156—157, 167

Задача измерения 166—168, 169, 202—203

Запаздывающие переменные 162—163, 204—205

Идентификация 135—139, 141—142, 160—163, 195, 208

Иерархическое управление 194

Излишки 47, 186, 210, 215, 236

Изменение 15—16, 34—36, 42—43, 52, 66—67, 70, 78—81, 118—119, 183—190, 193—195, 197—202, 227

Изоляция 26, 39, 238

Импликация 16—17, 20, 24, 31

Индикаторы 132, 167—168, 170, 229—231, 243

Инструментальные переменные (инструменты) 27, 144—161, 162—163, 165, 167—170, 204—205, 236—238, 240—243

Интервальное измерение 31—32. См. также Шкалы измерения

Исторический анализ 26—27, 29—30, 78—80, 85, 180, 196, 212, 246—247

Капоический анализ 105, 168
Ковариация, определение 94—97, 106, 175, 225. *См. также* Статистическая согласованность
Колебания 38, 43, 187, 189—190, 194, 201, 244, 247
Коллинеарность 165, 141
Коммуникация 192, 195—196
Комплексные переменные 181, 183, 187
Компоненты 18—22, 43
совместимость *см.* совместимость компонент
смежность *см.* смежность
организация *см.* организующие процессы
Конечные точки пути 110—111, 120
Консерватизм 46, 125, 206, 210, 215, 245—246
Конфигурации событий 15, 19—21, 183
Когьюнктивный пропускной механизм 123—124
Координирующие пути 110—114
Коррелированные возмущения *см.* Корреляции возмущений
Корреляции возмущений 40, 100, 104, 108—109, 122, 124, 135—136, 138—139, 142—145, 147, 150—151, 154, 157, 160, 170, 175, 178—179, 226—227, 233—234, 237, 243
Корреляционная матрица, определение 226
Корреляция *см.* Статистическая согласованность
Косвенный метод наименьших квадратов 150, 235
Коэффициент детерминации 100, 102—103, 104, 138
Коэффициент наследования 194
Коэффициент обоснованности (валидности) 132, 166, 230—231, 232
Коэффициенты приведенной (редуцированной) формы 74, 235
Коэффициент пропуска *см.* Полный эффект
Коэффициент регрессии, определение 98—99
Критические случаи 78—79
Критическое наблюдение 28
Культура 26, 79, 175

Либерализм 46, 206, 210, 215, 245
Линейность 34—38, 50, 57—59, 87—88, 92—94, 95, 97—98, 124—125, 139—140, 181—182, 195—196, 202—203, 223—224
Линии регрессии 97—98
Ложная корреляция 29, 41—42, 104, 110, 138

Математические ожидания 88—89, 91, 93—100, 106, 121—122, 224—225

Материальная основа причинности 16, 18—26, 191
Меггерс 79
Международные отношения 84, 221—223
Методы оценки для одного уравнения, определение 160
Методы полной информации 160—161, 243
Миграция 47, 216
Множественная регрессия 41, 101—104, 140—143, 150—154, 156, 175—176
Множественные причины 17, 38—41, 51, 59, 60—63
Множественные следствия 41—42, 51, 61—62
Множественный коэффициент корреляции 103, 141, 166
Мультиколлинеарность *см.* Коллинеарность
Мэсон 69

Наклон *см.* Коэффициент регрессии
Наука 84, 208—209
Начало пути 111
Недействующие операторы 25—26, 39—40, 107, 124, 170—171, 191
Недоопределенность 137, 148, 157
Независимые переменные *см.* Причинные переменные
Нелинейные связи (отношения) 55—57, 87, 92, 94—95, 98, 191, 232
Ненаблюдаемые переменные 166—167
Непомеченные стрелки 54—55, 74
Нерекурсивные системы 144—161
Нестационарные системы *см.* Стационарные системы
Неустойчивость 43, 70—71, 189—191, 194, 222
Неустойчивые петли *см.* Неустойчивость
Нижние индексы 49—50, 66, 67—68, 72, 94, 100, 114, 120, 162
Нормальное распределение 87
Нормативная нулевая точка 46, 215
Нормы 46, 213, 215, 237—238, 248
Нулевая точка 46, 88, 215

Образование 27, 46, 56, 82, 106, 126—128, 173, 176, 209—210, 214—215, 217—218, 225—226, 228, 233—234, 240—243, 245
Обратная связь 42—43, 52, 186—194, 195, 197—198, 207
Обратная связь более высокого порядка 191—194
Обычный метод наименьших квадратов 141—144, 151—152, 154, 160, 164, 168—169, 235

Объем выборки *см.* Число наблюдений
Объяснение 19, 23, 79—80, 84, 216
Объясненный разброс 93—94, 131, 133, 205—207, 226—227
Ограничение области 123
Ограничитель 55—57, 123—124
Однородные переменные *см.* Однородные потоки
Однородный поток 29—30, 46
Операторы 18—20, 37—38, 39—40, 42—43, 45—47, 49, 84, 107, 125—126, 181—182, 184, 191, 197, 214, 215, 217
Описание (спецификация) системы 140, 144—145, 146, 160, 168—170, 241—242
Опосредованные действия 25, 49, 52, 60—62, 78, 80
Определение причинности 23—24
Организующие процессы 20
Основная составляющая *см.* Сигналы, постоянная составляющая
Остатки 99—103, 104, 141, 152, 225
Открытые пути 66—67, 71—72
Отношения (связи) развертывания 16—17, 27, 106, 214, 223
Отрицательные значения переменных 33—34
Отрицательные коэффициенты 32—33, 37, 50, 63, 71, 126, 128, 184—185, 217
Отрицательные связи (отношения) 33—34, 37, 63, 92, 94—95, 125—126, 128, 184—185, 186, 217
Отсутствие причинности 24—29, 35
Отсутствие стрелок 35, 171
Оценка методом наименьших квадратов 135, 172
Оценка параметров 87, 135—174, 202, 204—206, 243
Очищение, определение 151
Ошибка измерения 132—134, 163—164, 166—167, 168, 172, 178—179, 229—231, 232, 243
Ошибка оценки 159, 161—174, 198, 201, 240
Ошибка описания (спецификации) 168—170

Параметры, определение 34, 136, 142
Передаточная функция 183, 191, 211
Переменные, задаваемые заранее (предопределенные) 151—152, 203
Переменные, определение 34
Переопределенность 154—156
Период запаздывания 181—183, 184—185, 194, 196—201
Петли 52, 60, 64—78, 108, 113, 116, 121, 128—130, 139—140, 143—144,

145, 149—150, 155—157, 160—162, 169—170, 197—201, 219
План исследования 170
Показатель степени 182
Политика 83, 210—211, 220
Полиция 83, 219—221
Полный эффект (полное влияние) 69—70, 72—73, 235
Поля 21—24
«Поперечные» («профильные») данные 86, 107, 144, 161, 196—198, 201—203, 206—207
Постоянная (константа) в уравнениях 32—34, 36, 58—59, 89
Потоки 29—36, 215
Потоковые графы 36, 48—78, 80—82, 85, 108, 182—183, 207, 209, 211
отсутствие стрелок 35
обращение 78
редукция (приведение, сведение) 60—62, 69, 149, 182, 186
представление переменных 48—49
символы 49—50
непомеченные стрелки 55
Правило суммирования 58—59
Правило суммирования (аддитивность) следствий 30, 59, 70, 78
Правило цепи 60—61, 63
Предсказание (прогноз) 80, 93—95, 97—104, 225
Преобразование Z 211
Преступность 26, 43, 79—80, 82—84, 131, 218—221
Приведенная (редуцированная) форма 73—74, 79—81, 125—126
Принцип Мэсона 69, 187, 235
Причинная теория, применения 78—81, 138—139
Причинное приближение (аппроксимация) 42, 47, 216
Причинность 13—30, 34—35, 36—45, 49, 63, 89, 110
Причинные запаздывания 181—183
Причинные (входные) переменные 50, 58—59, 64
Причинные цепи 19, 60—63, 65—68, 79, 110, 115, 197—198
Причинный вывод 15, 24—30, 37, 39—41, 44, 46, 53, 104, 130, 139, 170, 214
Программы роста благосостояния 83
«Продольные» данные 119, 163, 181, 195—196, 203, 207
Производительность 25, 186, 194
Производство следствий 15, 19—20, 23
Промежуточные переменные 19, 60—63, 73—75, 80, 140, 146, 198
Пропорциональность следствий 31—32, 58

Пропускной механизм 105, 122—125, 140—141, 233—234, 238
 Простые петли 52, 64, 70, 161—162
 Профессиональный (социально-профессиональный) статус 46, 82—83, 106, 123, 128, 131—132, 173, 206, 209—210, 214—215, 217—218, 225, 227, 233—235
 «Профильная» статика 107
 Процедуры агрегирования 204, 208
 Процессы распространения 175, 185
 Путевой анализ 108—122, 124, 135—138, 148, 166—167, 175
 Путевые диаграммы (схемы) 108—110, 116—118, 153—155
 Путевые идентификаторы 111, 120
 Путевые коэффициенты 117

 Разнообразие см. Статистическое разнообразие
 Разностные уравнения 192, 199, 209, 244, 248
 Райт 108
 Расовые аспекты 26, 47, 176—178, 240—243
 Распределения 85—96, 119
 Рассеивающий путь 120
 Регрессионный анализ 97—105, 137—139, 141—144, 224. См. также Множественная регрессия
 Рекурсивные системы 139—144, 170, 184—186, 197, 206
 Релевантная обратная связь, определение 69
 Решения о проведении определенной политики 80—81, 83—84, 208—209, 216, 219—221, 242—243
 Рождаемость 47, 216
 Рост 20, 208—209, 244

 Сдерживающие переменные 33, 215, 226
 Сельское хозяйство 79, 174, 185—186, 236
 Сигналы, постоянная составляющая 202—207
 определение 183
 составляющие, зависящие от времени 202—207, 247—248
 Сигналы, по существу постоянные 197—198, 204, 205—207
 Сила связей 37, 117—119, 125—126, 127—128, 133, 163—166, 168, 171, 201, 214, 240—241
 Символ задержки 182
 Синусоидальные кривые 180—190
 Синхронизация 204, 247
 Системы 38—43, 81
 Системы второго порядка 191—194

Словесная (вербальная) теория 36, 45, 48, 52—54, 59, 78, 80, 236, 242—243
 Смежность 21—24, 26
 Смертность 25, 216, 233
 Смещенная оценка 144, 162, 167—169, 180, 200—201, 205, 243
 События 15—31
 Совместимость компонент 20—24
 Совместное распределение 87—94
 Согласованность компонент см. Смежность
 Соприкасающиеся (касающиеся) пути и петли 68—73, 144, 220—221
 Социализация 29, 123, 247
 Социальная мобильность 46, 107
 Социальная стратификация 132, 174—175, 210—211
 Социальная структура 18, 43, 45, 174—175, 196, 236—237
 Социальный статус 45—46, 107, 126—127, 132—133, 214, 228
 Социальное управление 43, 126
 Сравнительная статика 258
 Среднее 88—89. См. также Математическое ожидание
 Стандартизованные переменные 96, 103—104, 117—119, 166—167, 232
 Стандартизованные частные коэффициенты регрессии 103—104, 137
 Стандартное отклонение, определение 95
 Стандартные величины, определение 96
 Статика 58, 70, 88, 107, 162, 180—181, 183, 187, 194, 198—199, 201, 207, 246
 Статистическая согласованность 41, 89—99, 104—105, 108—119, 123—125, 127—130, 139—140, 142—145, 237
 Статистические испытания 87, 153, 171, 240
 Статистический анализ 24, 40, 85—108, 171
 Статистический вывод 87, 139, 170—171
 Статистическое разнообразие 89, 119—121, 123—127
 Стационарные системы 57—58, 201, 207
 Структурные коэффициенты 34, 36—38, 47, 50, 60, 63, 82—83, 87, 99, 104, 114, 117—119, 141, 144, 150—153, 157, 206
 единицы структурных коэффициентов 82—83, 217—219
 Структурные уравнения 32, 34, 36—38, 59—60, 78, 85, 104—105, 151—152, 160, 221, 234
 Ступенчатая функция 183

Таблицы отношений 35
 Таксономия 20
 Тейл 150
 Технология 43, 84, 175, 186, 236
 Типология 20
 Трехэтапный метод наименьших квадратов 161
 Тюрьмы 131, 227

 Управляющие системы 26, 38, 42—43, 71, 126, 129—130, 133, 188—194, 220
 Уравновешивание 196—202, 212, 220, 246
 Уровень жизни 47, 216
 Уровни анализа 19, 23, 30, 41—42, 46, 182, 216
 Усиление 38, 42, 71, 186—188, 189—190, 193—194, 217, 221
 Установка (отношение) 33, 46, 126, 130, 132, 175, 203, 210, 212—213, 215, 237—238, 245, 247—248
 Устойчивое состояние см. Статика и Уравновешивание
 Устойчивость (стабильность) 43, 57—58, 71, 84, 88, 188, 190—191, 194—195, 198, 211. См. также Неустойчивость

 Факторный анализ 105, 168
 Фильтр 195

Характеристики реакции системы 183, 195

 Циклы см. Колебания

 Частичное приведение (редуцирование) петель 74—78, 108, 113, 121
 Частные коэффициенты регрессии 102—103, 141
 Число наблюдений 87, 91, 163—164, 167, 242

 Шкалы измерения 31, 33—35, 54—55, 57—59, 82, 87—88, 89, 103, 117—119, 219, 228
 ограниченные шкалы измерения 231

 Экзогенные переменные 60
 Экологическая корреляция 204
 Эксперименты 15, 24, 26, 39—41, 207
 Экспоненциальная кривая 188, 190
 Эндогенные переменные 60
 Эффект обратной связи (обратный эффект) 66—67, 69—71, 219
 единицы обратного эффекта 219
 Эффект открытого пути 68—72
 Эффект рассеивающего пути 120
 Эффекты координирующих путей 114—116

Предисловие к русскому переводу	5
Предисловие	11
Пролог	13
 Глава 1. Причинность и причинный анализ	15
Причинное упорядочение	15
Операторы	18
Определение причинности	23
Причинный вывод	24
Потоки	29
Линейность	34
Причинные системы	38
Источники и дополнительная литература	43
Упражнения	45
 Глава 2. Причинные диаграммы и анализ потоковых графов	48
Графическое изображение причинных отношений	48
Сложности в диаграммах	50
Анализ потоковых графов	57
Прикладные исследования	78
Источники и дополнительная литература	81
Упражнения	82
 Глава 3. Статистические представления	85
Распределения	85
Совместные распределения	89
Линейная регрессия	97
Множественная регрессия	101
Стандартизованные коэффициенты	103
Регрессия и причинный вывод	104
Источники и дополнительная литература	105
Упражнения	106
 Глава 4. Путевой анализ	107
Модификации потоковых графов	108
Анализ статистической координации	110
Анализ статистического разнообразия	119
Путевой анализ и алгебраические выкладки	121
Пропускные механизмы	122
Общие выводы	125
Источники и дополнительная литература	131
Упражнения	131

Глава 5. Идентификация и оценка	135
Регрессионный анализ и идентификация	137
Рекурсивные системы	139
Нерекурсивные системы	144
Факторы, влияющие на оценки	163
Трудности при использовании в социальном исследовании	169
Нулевые коэффициенты	170
Источники и дополнительная литература	172
Упражнения	173
 Глава 6. О динамике	180
Динамика простых систем	181
Запутывающее действие динамики при статическом анализе	196
Источники и дополнительная литература	207
Упражнения	208
Ответы к упражнениям	214
Предметный и именной указатель	249

Хейс Д.

ПРИЧИННЫЙ АНАЛИЗ В СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Зав. редакцией А. В. Павлюков. Редактор К. М. Чижевская
Мл. редактор И. Н. Горина. Техн. редакторы К. К. Букалова, Г. А. Полякова
Корректоры Г. В. Хлопцева, Т. М. Васильева, А. Т. Сидорова
Худ. редактор Э. А. Смирнов

ИБ № 903

Сдано в набор 18.12.80. Подписано в печать 26.03.81. Формат 60×90¹/₁₆. Бум. тип. № 2. Гарнитура «Литературная». Печать высокая. П. л. 16. Усл. п. л. 16. Уч.-изд. л. 17,56. Тираж 8300 экз. Заказ 937. Цена 1 р. 90 к.

Издательство «Финансы и статистика», Москва, ул. Чернышевского, 7.

Ленинградская типография № 2 головное предприятие ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгения Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.

32

Хейс Д.

X35

Причинный анализ в статистических исследованиях/Пер. с англ. Ю. Н. Гаврильца, Л. М. Кутикова и М. А. Родионова; Предисл. Т. В. Рябушкина и Ю. Н. Гаврильца.— М.: Финансы и статистика, 1981.— 255 с., ил.— (Математико-статистические методы за рубежом).

В пер.: 1 р. 90 к.

Книга отражает современное направление в статистике и социологии. Она базируется на таком описании взаимозависимых переменных, при котором их можно разделить на переменные-причины и переменные-следствия. Для этого понятия причины и следствия выражаются в математической форме, после чего изучение социально-экономических явлений может подкрепляться расчетами. Написана как самоучитель в строго продуманной последовательности с большим количеством примеров и задач. Дана обширная библиография. Для преподавателей статистики, логики, социологии, философии, а также для студентов вузов.

X 10805*-046
010(01)-81 45-81(C) 1702000000

ББК 22.172
517.8

*Второй индекс 10803.