ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

К. Ф. Баум

ЭКОНОМЕТРИКА

ПРИМЕНЕНИЕ ПАКЕТА *STATA*

УЧЕБНИК И ПРАКТИКУМ ДЛЯ ВУЗОВ

Перевод с английского под научной редакцией
С. А. Айвазяна, **Г. И. Пеникаса**

Рекомендовано Учебно–методическим отделом высшего образования в качестве учебника и практикума для студентов высших учебных заведений, обучающихся по экономическим направлениям

Книга доступна в электронной библиотечной системе
biblio-online.ru

Москва ▪ Юрайт ▪ 2016

УДК 519.862.6(075.8)
ББК 65в6я73
Б29

***Автор*:**

Баум Кристофер Ф. — PhD in Economics, профессор Департамента экономики Бостонского колледжа, США.

***Редакторы перевода*:**

Айвазян Сергей Артемьевич — профессор, доктор физико-математических наук, заслуженный деятель науки Российской Федерации, профессор Департамента прикладной экономики факультета экономических наук Национального исследовательского университета «Высшая школа экономики», заместитель директора ЦЭМИ РАН по научной работе.

Пеникас Генрих Иозович — кандидат экономических наук, доцент Департамента прикладной экономики факультета экономических наук Национального исследовательского университета «Высшая школа экономики».

Баум, К. Ф.

Б29 Эконометрика. Применение пакета *Stata* : учебник и практикум для вузов / К. Ф. Баум ; пер. с англ. под науч. ред. С. А. Айвазяна, Г. И. Пеникаса. — М. : Издательство Юрайт, 2016. — 370 с. — Серия : Авторский учебник.

ISBN 978-5-9916-6993-1

Книга содержит как теоретические постулаты эконометрики, так и подробное описание их реализации в современном программном продукте *Stata*. Материал охватывает ключевые темы, начиная от самых простых (линейная регрессия) и заканчивая наиболее сложными (например, оценка моделей панельных данных). Особый акцент делается на непосредственной работе с данными, ее организации, чтобы минимизировать ошибки, которые могут возникнуть при повторных исследованиях или проверке результатов исследования.

Книга будет полезна как студентам, начинающим исследователям, так и имеющим опыт работы с эконометрическими методами, в том числе с инструментом программы Stata, поскольку в ней не только подробно описываются азы работы с программой, но и приводятся тонкости, на которые большинство не обращало внимания.

УДК 519.862.6(075.8)
ББК 65в6я73

ISBN 978-5-9916-6993-1

Оглавление

Предисловие научных редакторов

Успешность вычислительной реализации современных методов анализа данных и, соответственно, эффективность и результативность прикладных социально-экономических исследований в современном мире существенно зависят от того, насколько грамотно ученый задействует доступное программное обеспечение. За последние два десятилетия появилось много программных комплексов, позволяющих решать подобные задачи (*Eviews*, *R*, *Gauss*, *Stata*). Одним из наиболее проработанных является программный пакет *Stata*. Тем не менее до сих пор в России были лишь точечные наработки, описывающие функционал данной программы. Нашей основной целью было представить российским исследователям структурированное описание всех возможностей *Stata*, для чего мы решили подготовить перевод учебного текста профессора Кристофера Баума.

Этот текст дает весьма полное представление как о прикладной эконометрике, так и о возможностях ее непосредственной реализации на данных. В книге освещены ключевые разделы оценки линейных моделей регрессии, работы с панельными данными, построения моделей бинарного и множественного выбора, выявления и поправки на возможные проблемы в случайных остатках модели, учет эндогенности. Приложения в книге позволяют перевести анализ данных с программой *Stata* на новый автоматизированный уровень, благодаря не просто работе с пользовательским интерфейсом и командной строкой, но и с программным кодом, когда вводятся основные понятия о языке программирования *Mata*.

В представленном тексте сохранена авторская нумерация разделов книги.

Польза книги состоит именно в совмещении теории и наглядного представления ее реализации на практике. Любую тему можно отработать на данных, использованных в книге (на что даются соответствующие ссылки), и сравнить результаты. Ценным является описание общей процедуры работы с данными. При этом подчеркивается, что всегда нужно системно смотреть на проводимый анализ, т.е. не как на однократное (авральное) упражнение, а как на процесс, который будет повторяться регулярно вами и вашими коллегами, последователями. Тогда становится особенно важным как минимизировать ошибки, связанные, например, с исправлением параметров, так и обеспечить, чтобы другие максимально быстро поняли ваши данные и разобрались в них, их коде и получили результаты, соответствующие вашим.

Мы благодарим В. А. Банникова, А. П. Грохотову, Г. Грохотову за помощь в подготовке рукописи; отдельно благодарим К. Васильеву и С. Винькова за помощь в доработке текста книги. Мы выражаем благодарность коллективу издательства «Юрайт», тем, кто помог довести книгу до публикации, особенно оказались полезными поддержка С. Г. Дария и правка П. А. Макарова.

Надеемся, что данная книга окажется полезной исследователям, аналитикам и студентам в прикладных исследованиях. Она позволит студентам получить следующие компетенции:

знать

• положения современной теории эконометрики;

• принципы обработки данных в программном пакете *Stata*;

уметь

• проведить исследования реальных данных с развитым программным пакетом *Stata*;

• корректно интерпретировать теоретические положения эконометрики;

владеть

• навыками применения на практике теоретических положений эконометрики.

Работа подготовлена в ходе работы в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ) и с использованием средств субсидии в рамках государственной поддержки ведущих университетов Российской Федерации «5-100».

С. А. Айвазян
Г. И. Пеникас
Москва, 23 марта 2016 г.

Предисловие автора к первому изданию

Эта книга — краткий путеводитель для прикладных исследователей в областях экономики и финансов, позволяющий изучить основы эконометрики и научиться применять эконометрический пакет компьютерного программного обеспечения *Stata*[1] с обучением на примерах, в которых приводятся типичные совокупности данных, рассматриваемые в экономике. Читатели должны владеть прикладной статистикой на уровне методов анализа простой модели линейной регрессии (обычный метод наименьших квадратов — МНК), включая алгебраическое представление этой модели, что соответствует содержанию курса статистики/эконометрики в магистратуре[2, 3]. В книге также используются некоторые элементы дифференциального исчисления (частные производные) и линейная алгебра.

Я предполагаю, что читатель знаком с интерфейсом эконометрического пакета *Stata* для *Windows*-версии, а также с основами ввода данных, их преобразованием и получением описательных статистик. При необходимости повторения данных аспектов рекомендую читателям обратиться к соответствующему руководству *Getting Started with Stata for Windows* («Начальные сведения о пакете *Stata* для операционной системы *Windows*»). Между тем читатели, уже достаточно хорошо знакомые с эконометрическим пакетом *Stata*, могут сразу перейти к гл. 4, где начинается серьезное обсуждение эконометрики.

В любой научно-исследовательской работе большие усилия связаны с подготовкой данных, являющихся неотъемлемой частью эконометрической модели. Несмотря на то что в первую очередь книга сосредоточена на эконометрической практике, мы должны рассмотреть серьезные проблемы, с которыми сталкиваются многие исследователи, когда необходимо преобразовать исходные данные к форме, требуемой для применения эконометрических методов, или хотя бы к виду, на основе которого можно построить таблицы и рисунки для отчета по исследовательскому проекту. Поэтому в гл. 2 описание сосредоточено на возможностях управления данными и на нескольких инструментах, доступных в эконометрическом пакете *Stata*, для проверки того, что все необходимые преобразования данных выполнены точно и эффективно. Если вы знакомы с этими аспектами применения эконометрического пакета *Stata*, то можете бегло просмотреть этот материал, возможно, возвращаясь к нему для освежения вашего понимания применения эконометрического пакета *Stata*. Аналогично гл. 3 посвящена об-

---

[1] Подробную информацию об условиях приобретении любого из трех возможных вариантов *Windows*-версии пакета *Stata*: *Small Stata, Intercooled Stata* и *Stata/SE* (для обучения или профессиональных вариантов) можно найти на сайте: http://www.stata.com (*примеч. перев.*).

[2] Два прекрасных учебника данного уровня: [101, 108].

[3] Из русскоязычных учебников можно порекомендовать читателю [1—5] (*примеч. науч. ред.*).

суждению структуры экономических и финансовых данных и команд эконометрического пакета *Stata*, предназначенных для преобразования данных с целью их представления в одной из стандартных форм (в виде пространственных данных, данных временно́го ряда, панельных или так называемых протяженных данных). Если вы хотите сразу перейти к линейной регрессии, то бегло просмотрите соответствующую главу с возможностью возврата к ней в процессе анализа.

Эконометрическое содержание книги начинается в гл. 4 с рассмотрения наиболее часто используемой в прикладных эконометрических исследованиях классической линейной модели множественной регрессии. В этой главе также обсуждается, как интерпретировать и представлять оценки регрессии и в чем состоит логика тестирования гипотез и наложения линейных и нелинейных ограничений на параметры модели. В последнем параграфе главы рассматриваются анализ остатков регрессии, прогнозные значения и предельные эффекты.

Применение модели регрессии зависит от некоторых предположений, которые часто нарушаются в реальных совокупностях данных. В гл. 5 обсуждается, как критическое предположение о нулевом условном среднем остатков регрессии может не выполняться в присутствии ошибок спецификации. В этой главе также обсуждаются статистические и графические методы обнаружения ошибок спецификации. В гл. 6 обсуждаются другие предположения, которые также могут нарушаться, как, например, предположение о независимости и одинаковой распределенности (НОР) остатков регрессии. В гл. 6 также представляется обобщенная линейная модель регрессии, объясняется, как выявить и учесть два наиболее важных отклонения от предположения о независимости и одинаковой распределенности остатков регрессии: гетероскедастичность и автокоррелированость.

В гл. 7 обсуждается применение переменных-индикаторов (или фиктивных переменных) в линейных моделях регрессии, содержащих как количественные, так и качественные факторы; применение моделей с эффектами взаимного влияния (взаимодействия) регрессоров и моделей со структурными сдвигами.

Во многих моделях регрессии в прикладной экономике нарушается предположение о нулевом условном среднем для остатков регрессии, из-за того что эти остатки одновременно входят в зависимую переменную и в один или более регрессоров или из-за ошибок измерения в регрессорах. Тогда независимо от причины МНК больше не будет давать несмещенные и состоятельные оценки, поэтому вместо МНК следует применять метод инструментальных переменных (МИП). В гл. 8 представлен МИП, его развитие — обобщенный метод моментов (ОММ) и тесты для определения необходимости применения МИП.

В гл. 9 применяются модели к панельным (или протяженным) данным, которые имеют как пространственные, так и временны́е измерения. Такое расширение модели регрессии позволяет получить преимущество за счет богатой информации, содержащейся в панельных данных, относящейся к неоднородности как в пространственном, так и во временно́м измерении.

Во многих эконометрических приложениях моделируются дискретные (категоризованные) и ограниченные зависимые переменные: бинарный ис-

ход, как, например, решение о покупке; или ограниченная переменная, как, например, сумма расходов, которая объединяет решение, покупать или нет, с решением, сколько истратить при принятии положительного решения о покупке. Поскольку методы оценки линейной регрессии исходя из теоретического построения для такого моделирования исходов неуместны, в гл. 10 представлено несколько методов оценивания моделей с дискретными ограниченными зависимыми переменными, доступных в эконометрическом пакете *Stata*.

В приложениях обсуждаются методы импортирования внешних данных в эконометрический пакет *Stata* и объясняются основы программирования в пакете *Stata*. Хотя вы можете использовать эконометрический пакет *Stata* без программирования, знание того, как программировать в нем, позволит вам сэкономить массу времени и сил. Вам также следует научиться создавать основу для проведения регулярных исследований с применением do-файлов, которые вы можете фиксировать, архивировать и запускать повторно. Приводимые рекомендации для пакета *Stata* сделают ваши do-файлы более лаконичными и простыми для использования и изменения.

Благодарности

В процессе создания этой книги у меня образовалось много интеллектуальных долгов. Билл Гулд (Bill Gould), Дэвид Драккер (David Drukker) и Винс Уиггинс (Vince Wiggins) из корпорации *StataCorp* с энтузиазмом поддержали идею о необходимости издания такой книги для экономического и финансового сообщества. Дэвид Драккер (David Drukker), Винс Виггинс (Vince Wiggins), Гейб Уоггонер (Gabe Waggoner) и Брайен Пои (Brian Poi) предоставили неоценимые редакционные комментарии в процессе работы над книгой. Мои соавторы различных подпрограмм пакета *Stata* — Николас Дж. Кокс (Nicholas J. Cox), Марк Шафер (Mark Schaffer), Стивен Стиллман (Steven Stillman), и Винс Уиггинс (Vince Wiggins) — внесли значительный вклад, как и многие другие члены сообщества пользователей пакета *Stata*, с помощью их собственных подпрограмм, советов и вопросов *Statalist*[1]. Доктор Чак Чакрабортай (Dr. Chuck Chakraborty) был полезен в определении тем, интересных для консультационного сообщества. Доктор Петиа Петрова (Dr. Petia Petrova) предоставила содержательный обзор частей рукописи.

В Бостонском колледже (Boston college) я должен поблагодарить Надежду Карамчеву (Nadezhda Karamcheva) за компетентную помощь в тщательных поисках примеров совокупностей данных, а также коллег в Research Services за многие полезные беседы о применении статистического программного обеспечения. Я очень благодарен академическому вице-президенту Джону Нейхаузеру (John Neuhauser) и декану Джозефу Куинну (Joseph Quinn) Колледжа искусств и наук (College of Arts and Sciences) за их принятие этого проекта как достойного использования моего творческого годичного отпуска, выпавшего на 2004 г.

Я адаптировал некоторые материалы в этой книге из записей для студентов университетского курса эконометрики и выпускников. Я благода-

[1] *Statalist* — Организация зарегистрированных пользователей пакета *Stata* (*примеч. перев.*).

рен многим поколениям студентов Бостонского колледжа, которые подталкивали меня на улучшение ясности таких записей и помогли мне понять определенные аспекты теоретической и прикладной эконометрики, наиболее трудны для специалиста.

И последнее, но ни в коем случае не самое малое, я больше всего благодарен своей жене Пауле Арнольд (Paula Arnold) за любезное благосклонное переписывание день за днем материалов вместе со сварливым автором и за поддержку (а иногда и подсказки в грамматике) в процессе творчества.

Кристофер Ф. Баум
Школа дубового сквера Брайтона,
штат Массачусетс, июль 2006 г.

Предисловие автора к русскому изданию

Эта книга была написана для того, чтобы удовлетворить потребности исследователей прикладных задач в формировании устойчивого понимания того, как они могут провести высококвалифицированную обработку данных и эмпирический анализ, используя пакет *Stata*. Она планировалась не как учебник по эконометрике и не как руководство пользователя, но как путеводитель для прикладного исследователя, который ищет способы проведения надежной и воспроизводимой эмпирической работы.

Эта книга была очень благоприятно воспринята исследователями, использующими пакет *Stata* по всему миру. Я надеюсь, что наличие русскоязычного издания расширит потенциальную аудиторию среди студентов, желающих стать прикладными экономистами или профессионалами в области финансов, во многих ведущих российских учреждениях высшего образования. В течение долгих лет преподавания эконометрики в аспирантуре Бостонского колледжа я познакомился со многими талантливыми студентами из России и бывших советских республик. Но практика прикладного эконометрического исследования выходит за рамки уверенного понимания процесса анализа и обязательно включает обучение в ходе непосредственной работы. Я надеюсь, что это русскоязычное издание поможет студентам, ищущим возможности дальнейшего обучения, для выработки дополнительных навыков, чтобы стать эффективными прикладными исследователями в академической среде, исследовательских организациях и частном бизнесе.

Я благодарен профессору Сергею Айвазяну и Генриху Пеникасу из Национального исследовательского университета «Высшая школа экономики», взявшихся за сложную работу по подготовке перевода на русский язык.

Кристофер Ф. Баум
Брайтон, Массачусетс, США
Август 2012

Система обозначений

Я спланировал эту книгу так, чтобы вы учились через практику, поэтому я ожидаю, что вы будете читать эту книгу, сидя за компьютером, и сможете реализовать самостоятельно приводимые к книге команды пакета *Stata*, чтобы повторить мои результаты. Освоив применение команд на примерах, вы сможете адаптировать их под ваши собственные потребности.

Для обращения к командам эконометрического пакета *Stata*, к синтаксису и переменным я использую особый шрифт **command**[1]. Точечная подсказка с последующей командой указывает, что вы можете ввести команду после точки (в контексте), чтобы получить приводимые в книге результаты.

Я придерживаюсь некоторых соглашений о математических обозначениях, имея в виду следующее:

- матрицы обозначаются полужирными заглавными буквами, как, например, $\mathbf{X}$;
- векторы обозначаются полужирными строчными буквами, как, например, $\mathbf{x}$;
- скаляры обозначаются строчными буквами курсивом, как, например, x;
- векторы данных $\mathbf{x}_i$ имеют размер $1 \times k$; думайте о них как о строках из матрицы данных;
- векторы коэффициентов $\boldsymbol{\beta}$ — векторы-столбцы размерности $k \times 1$.

Повсеместное применение буквы N вместо n для обозначения объема выборки вынудило меня сделать исключение из введенных обозначений и обозначить N как объем выборки. Точно так же T обозначает число наблюдений временно́го ряда, M — число кластеров, а L — максимальное число запаздываний (лагов). Я также придерживаюсь универсального соглашения о том, что статистика Льюнга — Бокса обозначается как Q и аналогично тест разности Сагана (*difference-in-Sargan*) обозначается как C.

Чтобы упростить обозначения, я не использую разные шрифты для отличия случайных переменных от их реализаций. Когда моделируется зависимая переменная y, она — случайная переменная. Наблюдения по y являются реализациями этой случайной переменной, и я ссылаюсь на i-е наблюдение по y как на y_i, а на все наблюдения — как на $\mathbf{y}$. Точно так же вектор регрессоров $\mathbf{x}$ — это вектор случайных переменных в общем случае, и я обозначаю i-е наблюдение по этому вектору случайных переменных как вектор $\mathbf{x}_i$, который является i-й строкой матрицы данных $\mathbf{X}$.

Этот текст дополняет, но не заменяет материал в руководствах эконометрического пакета *Stata*, поэтому я часто обращаюсь к этим руководствам,

---

[1] В переводе этой книги для более отчетливого выделения команд будет применяться полужирный шрифт (*примеч. перев.*).

используя обозначения [R], [P] и т.д. Например, [R] xi — это ссылка на пункт **xi** в основном справочном руководстве в пакете *Stata* (*Stata Base Reference Manual*), а [P] syntax — ссылка на пункт **syntax** в Справочном руководстве по программированию в пакете *Stata* (*Stata Programming Reference Manual*)[1].

Перекрестные ссылки на документацию пакета *Stata*[2]

Читая эту книгу и документацию для пакета *Stata*, вы найдете ссылки на соответствующие руководства эконометрического пакета *Stata*. Например:

[U] 26 Overview of *Stata* estimation commands (Обзор команд оценивания в пакете *Stata*);

[R] regress;

[D] reshape.

Первая ссылка — это ссылка на гл. 26 (Обзор команд оценивания в пакете *Stata*) в *Справочнике пользователя пакета Stata* (*Stata User's Guide*), который обозначается как [U], вторая ссылка — это ссылка на пункт regress в *Основном справочном руководстве* (*Stata Base Reference Manual*), которое обозначается как [R], и третья ссылка — это ссылка на пункт reshape в *Справочном руководстве управления данными* (*Stata Data Management Reference Manual*), которое обозначается как [D].

Все руководства в документации эконометрического пакета *Stata* имеют стенографические обозначения, как, например, уже вышеприведенные обозначения: [U] — для справочника пользователя и [R] — для основного справочного руководства.

Полный список стенографических обозначений для руководств следующий:

[GSM] — *Getting Started with Stata for Macintosh* (Начальные сведения о пакете *Stata* для операционной системы *Macintosh*);

[GSU] — *Getting Started with Stata for Unix* (Начальные сведения о пакете *Stata* для операционной системы *Unix*);

[GSW] — *Getting Started with Stata for Windows* (Начальные сведения о пакете *Stata* для операционной системы *Windows*);

[U] — *Stata User's Guide* (Справочник пользователя пакета *Stata*);

[R] — *Stata Base Reference Manual* (Основное справочное руководство по основам пакета *Stata*);

[D] — *Stata Data Management Reference Manual* (Справочное руководство по управлению данными в пакете *Stata*);

[G] — *Stata Graphics Reference Manual* (Справочное руководство по созданию графики в пакете *Stata*);

[P] — *Stata Programming Reference Manual* (Справочное руководство по программированию в пакете *Stata*);

[XT] — *Stata Longitudinal/Panel Data Reference Manual* (Справочное руководство по продольным/панельным данным в пакете *Stata*);

[MV] — *Stata Multivariate Statistics Reference Manual* (Справочное руководство по многомерному статистическому анализу в пакете *Stata*);

---

[1] См. ниже «Перекрестные ссылки на документацию пакета *Stata*» (*примеч. перев.*).

[2] Этот текст дополнен переводчиком.

[SVY] — *Stata Survey Data Reference Manual* (Справочное руководство по работе с данными обследования в пакете *Stata*);

[ST] — *Stata Survival Analysis and Epidemiological Tables Reference Manual* (Справочное руководство по таблицам анализа выживаемости и эпидемиологии в пакете *Stata*);

[TS] — *Stata Time-Series Reference Manual* (Справочное руководство по временны́м рядам в пакете *Stata*);

[I] — *Stata Quick Reference and Index* (Справочник быстрого ознакомления и алфавитного указателя в пакете *Stata*);

[M] — *Mata Reference Manual* (Справочное руководство по программированию на языке *Mata* (матричное программирование)).

Подробную информацию о каждом из этих руководств можно найти онлайн на сайте (доступ платный): http://www.stata-press.com/manuals.

Глава 1
ВВЕДЕНИЕ В ПАКЕТ *STATA*

Данная книга сфокусирована на инструментах, необходимых для выполнения прикладных эконометрических исследований в экономике и финансах. Эти исследования предполагают как знание теоретических основ эконометрики, так и отчетливое понимание того, как использовать эти эконометрические инструменты в процессе исследования. В книге такое понимание достигается за счет интеграции теории и практики, когда эконометрический пакет *Stata* применяется к совокупностям данных, чтобы проиллюстрировать, как такие данные могут быть организованы, преобразованы и использованы для эмпирического оценивания. Мой опыт работы в эконометрике со студентами и аспирантами, использующими эконометрические инструменты в своих исследованиях, привел меня к пониманию того, что научиться применять эконометрику можно только в результате приложения этих инструментов к реальным совокупностям данных. Растущее число учебников начального уровня по эконометрике[1, 2] сосредоточены на теоретических аспектах, которые с высокой вероятностью будут возникать в эмпирической работе. Эта книга предназначена служить дополнением к таким учебникам и отражает практический опыт работы с множеством эконометрических инструментов, которые доступны в пакете *Stata*.

Параграф 1.1 представляет список «одиннадцати лучших», по моему мнению, отличительных особенностей эконометрического пакета *Stata*: пользовательский интерфейс и архитектура пакета *Stata*, а также возможности, которые делают этот программный пакет превосходным инструментом для прикладного эконометрического исследования. В параграфах 1.2 и 1.3 приведена существенная информация для тех, кто хочет самостоятельно численно проанализировать примеры, приведенные в этом тексте. Во многих таких примерах использованы самописные команды *Stata*, которые следует доустановить в копию пакета на вашем компьютере. Удобная программа **itmeus**, описанная в этой главе, сделает такую установку легкой задачей.

1.1. Краткий обзор отличительных особенностей эконометрического пакета *Stata*

Эконометрический пакет *Stata* — это мощный инструмент для исследователей в области прикладной экономики. Пакет может помочь вам проводить анализ исследования легко и эффективно независимо от того, с каким

---

[1] Например, учебники [101, 108].

[2] Из более доступных российскому читателю русскоязычных учебников в этой связи можно упомянуть книги С. А. Айвазяна [1], М. Вербика [3], Я. Р. Магнуса, П. К. Катышева и А. А. Пересецкого [4] (*примеч. науч. ред.*).

типом данных вы работаете — с временны́ми рядами, с панельными или пространственными данными. Эконометрический пакет *Stata* предоставляет инструменты, необходимые вам для организации данных и управления данными, с целью последующего получения и анализа результатов статистических расчетов.

Для одних пользователей эконометрический пакет *Stata* — это статистический пакет с набором меню, которые позволяют пользователям посмотреть на данные, сгенерировать новые переменные, провести статистические расчеты и построить графики. Для других пакет *Stata* — это программа с командной строкой, команды в которую обычно загружаются из do-файла с заранее сохраненными алгоритмами расчетов, которые реализуют все вышеперечисленные шаги без вмешательства со стороны пользователя. Иные же считают, что пакет *Stata* следует рассматривать как язык программирования для разработки ado-файлов, которые задают программы или новые команды пакета *Stata*, дополняющие этот пакет, расширяя возможности управления данными, статистических расчетов или графических построений.

Понимание некоторых из отличительных особенностей эконометрического пакета *Stata* поможет вам использовать этот пакет более наглядно и эффективно. Вы будете иметь возможность избегать ввода повторных команд (или копирования и вставки), постоянно «изобретая колесо». Полезно знать, как написать в вычислительном отношении компактные do-файлы (например, такие, которые позволяют провести расчет за 10 с, а не за 2 мин), но что еще более важно — это уметь писать do-файлы, которые можно легко понять и модифицировать. Эта книга сэкономит вам время, обучив вас готовить понятные и легко адаптируемые do-файлы, которые вы можете запускать повторно с помощью одной команды.

Кратко рассмотрим некоторые из отличительных особенностей эконометрического пакета *Stata*, которые я опишу более подробно позже.

Вы можете легко изучить команды эконометрического пакета *Stata*, даже если не знаете его синтаксис. Пакет *Stata* имеет диалоговое окно для почти каждой официальной команды, и когда вы выполняете команду при открытом диалоговом окне, окно *Review* (*обозрение*) отображает синтаксис команды, так же как если бы вы напечатали эту команду самостоятельно. Хотя вы можете ввести команду без закрытия диалогового окна, часто будет необходимо выполнить несколько последовательных команд (например, создать новую переменную, а затем вывести описательные статистики по данной переменной). Даже если вы пользуетесь диалоговыми окнами пакета *Stata*, вы можете заново переписывать, модифицировать и повторно вводить команды с помощью окна *Review* и *Command* (командной строки). Вы можете сохранить содержимое окна *Review* в файл или скопировать эти команды в окно *Do-file Editor* (*редактор* do-*файлов*) для их модификации и повторного ввода. Чтобы воспользоваться этими опциями, сделайте щелчок правой кнопкой мыши на команде в окне *Review*.

Вы можете применять редактор do-файлов пакета *Stata*, чтобы сэкономить время при проведении вашего исследования. Как только вы познакомились с базовыми командами, вы поймете, что легче их размещать в do-файле и работать с ним, чем вводить команды в интерактивном режиме (исполь-

зуя диалоговые окна или командную строку). Используя вашу мышь, вы можете выбирать любое подмножество команд, появляющихся в редакторе do-файлов, и выполнять только эти команды. Такая возможность облегчает тестирование, приведет ли ввод этих команд к желаемому результату. Если ваш do-файл содержит команды для всего исследования, то он представляет собой наглядный воспроизводимый и документированный отчет о логике вашего исследования (особенно если вы добавляете комментарии для описания того, что получается в итоге выполнения команды, кем и когда это было сделано и т.д.).

Простая команда выполняет все вычисления для всех требуемых наблюдений. Эконометрический пакет *Stata* отличается от некоторых других статистических пакетов своим подходом к переменным. Когда вы считываете совокупность данных в пакет *Stata*, пакет *Stata* вводит в память матрицу со строками, соответствующими наблюдениям, и со столбцами, представляющими переменные. Вы можете увидеть эту матрицу, щелкнув иконку *Data Viewer* (*обозревателя данных*) или иконку *Data Editor* (*редактора данных*) на панели инструментов пакета *Stata*. Большинство команд эконометрического пакета *Stata* не требует, чтобы вы обозначали явно наблюдения. В отличие от других статистических пакетов в программе *Stata* редко возникают случаи, требующие вашего обращения к специфическому наблюдению из целого ряда или массива, и если вы не будете использовать указания на конкретную точку данных, пакет *Stata* будет проводить вычисления намного быстрее. Когда вы должны обратиться к исходному значению наблюдения явно, например когда вы генерируете переменные запаздываний в данных временно́го ряда, *всегда* используйте операторы временны́х рядов пакета *Stata*, как, например, оператор **L.x** — для запаздывания на один шаг переменной **x** или оператор **D.x** — для первой разности данной переменной.

Организация циклов по переменным экономит время и усилия. Одна из самых ценных особенностей эконометрического пакета *Stata* — это возможность повторять в нем шаги (преобразования данных, оценивания или создания графики) по нескольким переменным. Соответствующие команды описаны в руководствах [P] forvalues, [P] foreach и [P] macro[1]; см. также онлайн-помощь (например, **help forvalues**) и приложение B для описания подробностей. Применение этих команд может помочь вам построить do-файл, который будет выполнять цикл команд по переменным, вместо того чтобы писать отдельную команду для каждой переменной; вы можете легко модифицировать ваш файл позже, если вам потребуется повторить команду на ином множестве переменных (подробнее см. гл. 2).

Опция by-groups эконометрического пакета *Stata* уменьшает потребность в программировании. Пакет *Stata* позволяет вам определять группы наблюдений на основе категориальных переменных (с целочисленными значениями), чтобы с помощью коротких и простых команд вы смогли выполнить сложные преобразования данных (подробнее см. гл. 2).

Эконометрический пакет *Stata* имеет много статистических особенностей, которые делают его мощным. Пакет *Stata* может вычислять устойчи-

---

[1] Здесь и далее см. «Перекрестные ссылки на документацию пакета *Stata*» в разделе «Система обозначений» (*примеч. перев.*).

вые (робастные) и устойчивые по группам наблюдений (кластер-робастные) оценки ковариационной матрицы оценок для почти всех команд оценивания[1]. С помощью команды **mfx** рассчитываются предельные эффекты после оценивания. С помощью команд **test**, **testnl**, **lincom** и **nlcom** проводится тестирование по Вальду линейных и нелинейных ограничений и расчет доверительных интервалов для линейных и нелинейных функций от оцененных параметров.

Вы можете избежать проблем с помощью использования актуальной версии эконометрического пакета *Stata*. Если вы имеете выход в Интернет, то опция **update** пакета *Stata* будет периодически бесплатно обновлять расчетные и ado-файлы. Большинство обновлений содержит устранение выявленных ошибок и дополнения к существующим командам (а иногда и совершенно новые команды). Чтобы найти доступные обновления, введите команду **update query** и следуйте ее рекомендациям. Многие проблемы, выявленные пользователями пакета *Stata*, были уже решены в выпущенных обновлениях, поэтому вам следует всегда обновлять расчетные и ado-файлы в пакете *Stata*, прежде чем вы отправите сообщение о какой-либо очевидной ошибке в программе. Убедитесь в том, что вы обновили вашу копию пакета *Stata*, когда повторно установили пакет на новом компьютере или жестком диске, так как инсталляционный компакт-диск содержит только первичный код (т.е., например, версию 9.0 без обновлений заменяете версией 9.2 с обновлениями, которая является самой актуальной на момент написания книги[2]).

Эконометрический пакет *Stata* расширяем бесконечно. Вы можете создавать ваши собственные команды, которые являются неотличимыми от официальных команд пакета *Stata*. Вы можете добавить новую команду в пакет *Stata*, если вы или кто-то другой ее создали, с помощью написания ado- и help-файла. Любые должным образом построенные ado-файлы, размещенные в папке **adopath**, будут определять новые команды с соответствующими именами команд, поэтому возможности пакета *Stata* безграничны (см. [P] sysdir). Так как большинство команд эконометрического пакета *Stata* написано на языке do-файла, то они доступны для просмотра и модификации и отражают лучшую практику в программировании.

Сообщество пользователей эконометрического пакета *Stata* предоставляет многообразие полезных добавлений к пакету *Stata*. Стратегия развития корпорации *StataCorp* заключается в том, чтобы предоставлять пользователям те же самые инструменты для разработок, которые применяются ее собственными профессиональными программистами. Эта практика воодушевила на создание энергичного сообщества пользователей эконометрического пакета *Stata*, которые бесплатно обмениваются своими наработками. Хотя любые пользователи пакета *Stata* могут построить собственное хранилище *net from*, большинство самописных программ доступны из архива *Statistical Software Components* (SSC) (Статистические компоненты программного обеспечения), который я поддерживаю в Бостонском колледже и к которому вы можете получить доступ при использовании

---

[1] Не беспокойтесь, если вы не знаете таких команд, я подробно буду обсуждать эти понятия далее в тексте.

[2] На момент подготовки перевода последней является версия 14 (*примеч. науч. ред.*).

команды **ssc** пакета *Stata* (см. [R] ssc). Чтобы посмотреть архив SSC, вы можете использовать веб-браузер, но следует использовать команду **ssc** для загрузки любого содержимого в архив SSC, чтобы гарантировать, что файлы обработаны должным образом и установлены в соответствующую папку. При вводе команды **ssc whatsnew** перечисляются последние добавления и обновления в архиве SSC. Ввод команды **adoupdate** обновляет пакеты, которые вы установили из архива SSC, из *the Stata Journal* (журнала пакета *Stata*) или с сайтов индивидуальных пользователей, если это необходимо.

Эконометрический пакет *Stata* — межплатформенно совместимый[1] пакет. В отличие от многих статистических пакетов множество особенностей пакета *Stata* не отличается для разных операционных систем (*Windows, Macintosh*, *Linux* и *Unix*), на которых этот пакет используется. Документация эконометрического пакета *Stata* не является специфической только для какой-либо одной платформы (за исключением руководства *Getting Started with Stata — Начальные сведения о пакете Stata*). Do-файл, который читается на одной платформе, будет читаться и на другой (до тех пор, пока система будет иметь достаточно памяти). Эта совместимость позволяет легко перемещать файлы бинарных данных между платформами, т.е. все файлы с расширением **.dta** пакета *Stata* имеют один и тот же формат двоичных данных, и поэтому на любом компьютере одна и та же версия пакета *Stata* может считывать и записывать эти файлы. Пакет *Stata* может также считывать файлы данных, хранимые на веб-сервере, с помощью команды **use http: // ...**, независимой от платформы.

Эконометрический пакет *Stata* может быть забавой. Хотя эмпирическое исследование — это серьезное дело, вам достаточно только изучить несколько тем обсуждений на форумах в сообществе зарегистрированных пользователей *Statalist*[2], чтобы узнать, что многие пользователи, применяя пакет *Stata*, получают незабываемое удовольствие от участия в сообществе пользователей этого пакета. Хотя обучение эффективному применению пакета *Stata*, так же как и изучение иностранного языка, — это трудная работа. Получение навыка решения задач управления данными и статистического анализа вдохновляет. Кто знает? Возможно, однажды ваши коллеги обратятся к вам за помощью по применению пакета *Stata*.

1.2. Инсталляция необходимого программного обеспечения

В этой книге пакет *Stata* применяется для того, чтобы проиллюстрировать многие аспекты прикладного эконометрического исследования. Как упомянуто выше, возможности пакета *Stata* не ограничены официальными командами *Stata*, документированными в руководствах и в интернет-справке, но включают богатство команд, описанных в *the Stata Journal* (журнале пакета *Stata*), в *Stata Technical Bulletin* (техническом бюллетене пакета *Stata*) и в архиве SSC[3]. Такие команды не будут доступны в вашей копии

---

[1] Термин, означающий совместимость с несколькими операционными системами (*примеч. перев.*).

[2] См.: http://www.stata.com/statalist.

[3] Напечатайте **help ssc** для получения информации об архиве SSC в Бостонском колледже.

пакета *Stata*, если вы их не установите самостоятельно. Поскольку в книге используется несколько из таких самописных команд для иллюстрации многообразия инструментов, доступных для пользователя пакета *Stata*, то я запрограммировал команду **itmeus**, которая установит все неофициальные команды, используемые в примерах книги. Чтобы инсталлировать такую команду, следует установить связь с Интернетом и ввести в командную строку *Stata*

```
ssc install itmeus
```

Это восстановит команду из архива SSC. Когда команда **ssc** завершит установку, вы можете напечатать

```
help itmeus
```

Так же следует поступить с любой командой пакета *Stata* или просто ввести

```
itmeus
```

Начнется процедура загрузки. Все необходимые команды будут установлены на вашу копию пакета *Stata*. Тогда можно выполнить любой пример из этой книги (см. следующий параграф, чтобы загрузить do-файлы и совокупности данных, применяемые в примерах).

Сегодня могут быть доступными и более новые версии самописных команд пакета *Stata*, которые вы устанавливаете. Официальная команда пакета *Stata* **adoupdate**, которую можно ввести в любое время, проверит наличие более новых версий самописных команд. Так же как и команда **update query**, определяющая, являются ли расчетные и официальные ado-файлы вашего пакета *Stata* актуальными, команда **adoupdate** выполнит аналогичную проверку на актуальность самописных команд, инсталлированных на вашу копию пакета *Stata*.

1.3. Загрузка баз данных для примеров из книги

За исключением некоторых малых иллюстративных наборов данных, все данные, которые я использую в этой книге, вы можете бесплатно загрузить с веб-сайта *Stata Press* (http://www.stata-press.com). Фактически, когда я ввожу новые совокупности данных, я просто загружаю их в пакет *Stata* тем же самым образом, что и вы (например,

```
use http://www.stata-press.com/data/imeus/traffic.dta, clear
```

Попробуйте сделать это).

Чтобы загрузить совокупности данных и do-файлы для этой книги, напечатайте

```
net from http://www.stata-press.com/data/imeus
net describe imeus
net get imeus-dta
net get imeus-do
```

Материалы будут загружены в вашу текущую рабочую папку. Я советую вам создать новую папку и скопировать данные в нее.

Глава 2
РАБОТА С ЭКОНОМИЧЕСКИМИ И ФИНАНСОВЫМИ ДАННЫМИ В ЭКОНОМЕТРИЧЕСКОМ ПАКЕТЕ *STATA*

Экономическое исследование всегда включает несколько задач управления данными, как, например, задачу ввода данных, задачу проверки достоверности (*validation*) выборки и задачу преобразования данных, что является решающим для того, чтобы получить обоснованные выводы из статистического анализа данных. Эти задачи часто занимают больше времени, чем непосредственно сами статистические исследования, и поэтому обучение эффективному применению пакета *Stata* может помочь вам выполнять эти задачи и получать хорошо документированную (в соответствии с форматами пакета *Stata*) совокупность данных, поддерживающую вашу исследовательскую работу.

В параграфе 2.1 обсуждаются основы работы с данными в пакете *Stata*, в параграфе 2.2 — основные способы преобразования данных[1].

2.1. Основы использования пакета

Для эффективного управления данными в эконометрическом пакете *Stata* вам следует понимать некоторые из основных особенностей данного пакета. Эти особенности будут проиллюстрированы на небольшой совокупности данных из пакета.

2.1.1. Применение команд

Откройте готовый файл данных (.dta) пакета *Stata* (данные переписи в штатах США в северо-восточном (NE) и северо-центральном (NC) регионе) с помощью команды **use**. Вы можете задать только имя совокупности данных, как, например, **use census2c**, или полную траекторию совокупности данных в зависимости от вашей операционной системы, как, например[2],

```
use "/Users/baum/doc/SFAME/stbook.5725/dof/census2c.dta"
```

[1] В оригинале здесь также было ошибочно указано содержание третьего и четвертого параграфов главы, отсутствующих в книге. Этот материал вошел в третью главу книги (*примеч. науч. ред.*).

[2] Скорее всего, на вашем компьютере сразу после установки пакета не будет данного файла. Поэтому лучше его загрузить из Интернета (*примеч. перев.*).

С помощью команды **use** вы также можете открыть файл на сервере сети, например:

```
use http://www.stata-press.com/data/r9/census2
```

В этих форматах требуются кавычки, если имеются пробелы в названии папки или в именах файлов. Если вы используете интерфейс пакета *Stata*, то укажите имя файла. Выбрав *File* > *Open...*, вы можете получить полную траекторию файла из окна *Review* (*обозрения*) и сохранить ее в do-файле.

Далее мы будем использовать файл данных пакета *Stata*, поэтому приведем его содержимое, в том виде, в котором оно выдается при использовании следующих команд[1]:

```
use http://www.stata-press.com/data/imeus/census2c, clear
(1980 Censys data for NE and NC states)
(данные переписи для северо-восточных и северо-центральных
штатов за 1980 г.)
list, sep(0)
```

	state	region	pop	popurb	medage	marr	divr
1	Connecticut	NE	3107.6	2449.8	32.00	26.0	13.5
2	Illinois	N Cntrl	11426.5	9518.0	29.90	109.8	51.0
3	Indiana	N Cntrl	5490.2	3525.3	29.20	57.9	40.0
4	Iowa	N Cntrl	2913.8	1708.2	30.00	27.5	11.9
5	Kansas	N Cntrl	2363.7	1575.9	30.10	24.8	13.4
6	Maine	NE	1124.7	534.1	30.40	12.0	6.2
7	Massachusetts	NE	5737.0	4808.3	31.20	46.3	17.9
8	Michigan	N Cntrl	9262.1	6551.6	28.80	86.9	45.0
9	Minnesota	N Cntrl	4076.0	2725.2	29.20	37.6	15.4
10	Missouri	N Cntrl	4916.7	3349.6	30.90	54.6	27.6
11	Nebraska	N Cntrl	1569.8	987.9	29.70	14.2	6.4
12	New Hampshire	NE	920.6	480.3	30.10	9.3	5.3
13	New Jersey	NE	7364.8	6557.4	32.20	55.8	27.8
14	New York	NE	17558.1	14858.1	31.90	144.5	62.0
15	N. Dakota	N Cntrl	652.7	318.3	28.30	6.1	2.1
16	Ohio	N Cntrl	10797.6	7918.3	29.90	99.8	58.8
17	Pennsylvania	NE	11863.9	8220.9	32.10	93.7	34.9
18	Rhode Island	NE	947.2	824.0	31.80	7.5	3.6
19	S. Dakota	N Cntrl	690.8	320.8	28.90	8.8	2.8
20	Vermont	NE	511.5	172.7	29.40	5.2	2.6
21	Wisconsin	N Cntrl	4705.8	3020.7	29.40	41.1	17.5

Содержание этой совокупности данных (данных переписи), **census2c**, упорядочено в табличном формате, подобно электронной таблице. Строки таблицы — *наблюдения*, или случаи, а столбцы — *переменные*. В таблице

[1] Здесь (см. ниже) и далее командные строки выделены полужирным шрифтом, выводы на печать (распечатки) ниже командных строк имеют обычный шрифт, в круглых скобках (обычный шрифт) содержится соответствующий перевод комментариев или англоязычных выражений. Пакет *Stata* работает с числами в американском формате, поэтому данные загружаются и выводятся только с точкой в виде разделителя десятичной части, выравненные по правому краю. В таком виде они представлены и в данной книге в выводах программы (*примеч. перев.*).

имеется 21 строка, и каждая строка соответствует штату США в северо-восточном (NE) или северо-центральном (N Cntrl) регионе, и семь столбцов, или переменных: **state** (штат), **region** (регион), **pop** (численность населения), **popurb** (численность городского населения), **medage** (медианное значение возраста), **marr** (число браков на тысячу жителей) и **divr** (число разводов на тысячу жителей). Имена переменных должны быть уникальными и подчиняться определенным правилам синтаксиса. Например, имена переменных не могут содержать пробелы или дефисы (-), неалфавитные или нечисловые символы, и они должны начинаться с буквы[1]. Пакет *Stata* чувствителен к нижнему и верхнему регистрам букв, поэтому **STATE**, **State** и **state** — три различные переменные в пакете *Stata*. В пакете *Stata* рекомендуется, чтобы для всех переменных вы использовали строчные имена.

2.1.2. Типы переменных

В отличие от некоторых статистических пакетов пакет *Stata* поддерживает полный диапазон *типов переменных*. Многие типы данных, используемые исследователями, предполагают целочисленные значения. Общее количество таких целочисленных значений может быть малым, как, например, {0, 1} для индикаторных (фиктивных/*dummy*) переменных, или когда переменные имеют значения, ограниченные диапазоном от нуля до девяти. Для экономии памяти и дискового пространства пакет *Stata* позволяет вам по необходимости определять переменные в виде целых чисел, в виде действительных чисел или в виде текста (*string*). Для целого числа в пакете *Stata* существует три типа целочисленных данных: тип **byte** — для присвоения целых чисел в диапазоне одно-двузначного числа; тип **int** — для присвоения целых чисел в диапазоне ±32 740; и тип **long** — для присвоения целых чисел в диапазоне ±2,14 млрд. Для десятичных значений в пакете *Stata* существует два действительных типа данных: **float** и **double**. Переменные, сохраняемые как **float**, имеют семь цифр после запятой; переменные, сохраняемые как **double**, — 15 цифр после запятой. Числовые переменные хранятся как **float**, если вы не специфицируете иное хранение. Для более подробного описания см. **data types** (типы данных).

Текстовые переменные опционально можно объявить как имеющие заданную длину от **str1** до **str244** символов. Если вы сохраняете строку длиннее заданного размера, то пакет *Stata* автоматически увеличивает размер хранения переменной, для того чтобы сохранить эту строку вплоть до максимального размера из 244 символов.

Печать команды **describe** отображает содержание совокупности данных, включая тип данных для каждой переменной. Например, команда

```
describe
```

выдает описание данных, содержащихся в файле census2c.dta:

---

[1] Имя переменной может начаться с символа нижнего подчеркивания (_), но такое начало имени переменной не является хорошей идеей, поскольку многие программы пакета *Stata* создают временные переменные с именами, начинающимися с символа нижнего подчеркивания.

```
    obc:          21      1980, Census data for NE and NC states
(наблюдений)                (данные переписи для северо-восточных
                                  и северо-центральных штатов)
    vars:         7                    9 Jun 2006 14:50
(переменных)
    size:       1,134                99.9% of memory free
  (размер)                     (99.9% памяти свободно)
```

variable name (имя переменной)	storage type (тип хранения)	display format (формат отображения)	value label (название значений)	variable label (название переменной)
state	strl3	Cenreg	%-13s	State (штат)
region	byte		%-8.0g	Census region (регион переписи)
pop	double		%8.1f	1980 Population, '000 (население, в тыс.)
popurb	double		%8.1f	1980 Urban population, '000 (городское население, в тыс.)
medage	float		%9.2f	Median age, years (медианное значение возраста, годы)
marr	double		%8.1f	Marriages, '000 (число браков на тыс.)
divr	double		%8.1f	Divorces, '000 (число разводов на тыс.)

Пакет *Stata* показывает, что совокупность данных содержит 21 наблюдение (**obs**) и 7 переменных (**vars**). Переменная **state** имеет тип **str13**, и, таким образом, в совокупности данных нет имен штатов, длина названия которых превышает 13 символов. Переменные **pop**, **popurb** и **divr** хранятся в формате **double**, а не как целые числа, поскольку они измеряются в тысячах жителей и поэтому содержат дробные части, если какое-либо из их значений не является кратным 1000. Переменная **medage** хранится как **float**, тогда как переменная **region** хранится как **byte**, хотя оказывается, что эта переменная имеет значения **NE** и **N Cntrl**; однако эти значения не являются истинным содержимым переменной **region**, а являются скорее ее значениями меток (***value label***), как описано ниже.

2.1.3. Переменные _n and _N

В списке выше наблюдения в совокупности данных пронумерованы 1, 2, ..., 21, и, таким образом, вы можете обратиться к наблюдению, указав его номер[1]. Для обращения к наибольшему номеру наблюдения вы можете использовать обозначение **_N** — общее количество наблюдений, а для обращения к текущему номеру наблюдения можете использовать обозначение **_n**, хотя эти обозначения могут изменяться по подгруппам в данных (см. подпараграф 2.2.8). Номера наблюдений изменятся, если применяется ко-

[1] Вы можете попросить пакет *Stata* описать, например, наблюдение № 3: **list if [-n] = = 3** (*примеч. перев.*).

манда **sort** (см. подпараграф 2.1.5), которая изменяет порядок наблюдений в совокупности данных в памяти.

2.1.4. Команды generate и replace

Основные команды пакета *Stata* для преобразования данных — это команды **generate** и **replace**, которые работают аналогично, но с некоторыми важными различиями. Команда **generate** создает *новую* переменную с именем, *не существующим* в совокупности данных в текущий момент. Команда **replace** изменяет значения *существующей* переменной, и в отличие от других команд пакета *Stata* для команды **replace** нельзя использовать аббревиатуру[1].

Для иллюстрации команды **generate** давайте создадим новую переменную в нашей совокупности данных, которая измеряет долю населения, живущего в городах в 1980 г., для каждого штата. Нам следует задать только соответствующую формулу, и пакет *Stata* автоматически применит эту формулу к каждому наблюдению, специфицированную командой **generate** согласно правилам алгебры. Например, если бы формула привела бы к результату деления на нуль для данного штата, то результат для такого штата имел бы признак пропущенных данных. Мы создадим переменную **urbanized** и используем команду **summarize**, чтобы отобразить описательные статистики этой новой переменной[2]:

```
generate urbanized = popurb/pop
summarize urbanized
```

Variable	Obs	Mean	Std. Dev.	Min	Max
urbanized	21	.6667691	.1500842	.3377319	.8903645

Средний штат в этой части США имеет долю городского населения на уровне 66,7% (она изменяется в диапазоне от 34 до 89%).

Если переменная **urbanized** уже существовала, но мы хотим выразить ее в процентах, а не в десятичных долях, то мы используем команду **replace** и смотрим описательные статистики переменной:

```
replace urbanized = 100*urbanized
(21 real change made)
(выполняется 21 действительное изменение)
summarize urbanized
```

Variable	Obs	Mean	Std. Dev.	Min	Max
urbanized	21	66.67691	15.00843	33.77319	89.03645

После выполнения команды **replace** сообщается число сделанных изменений — 21 изменение (по числу всех наблюдений).

Вам следует записывать преобразования данных в виде простого, лаконичного множества команд, которые при необходимости вы можете легко

---

[1] Для команды **generate** можно просто ввести **gen** (*примеч. перев.*).

[2] В базовом функционале программы пакет *Stata* выдает результаты в нестандартизованном формате в зависимости от доступной длины ячейки для отображения в рабочем поле, как в таблице ниже. Возможности более удобного представления результатов относятся к работе с сохраненными результатами вычислений, о чем см., например, параграф 4.4 (*примеч. науч. ред.*).