

ГЛАВА 7

ОБРАБОТКА ДАННЫХ

Назначение системы обработки данных МИКС – выдавать первые результаты обследования уже через несколько недель после окончания работы на местах. Данная глава содержит информацию, которая поможет вам осуществить планирование и предварительную подготовку, с тем чтобы превратить эту цель в реальность. Данная глава начинается с общего описания системы обработки данных МИКС. Затем в ней подробно, когда это необходимо – со ссылками на дополнительные источники информации, рассматривается каждая из ее составных частей. В конце главы приведен набор из трех контрольных перечней, которые помогут вам успешно осуществить обработку данных вашего обследования.

ОБЩЕЕ ОПИСАНИЕ

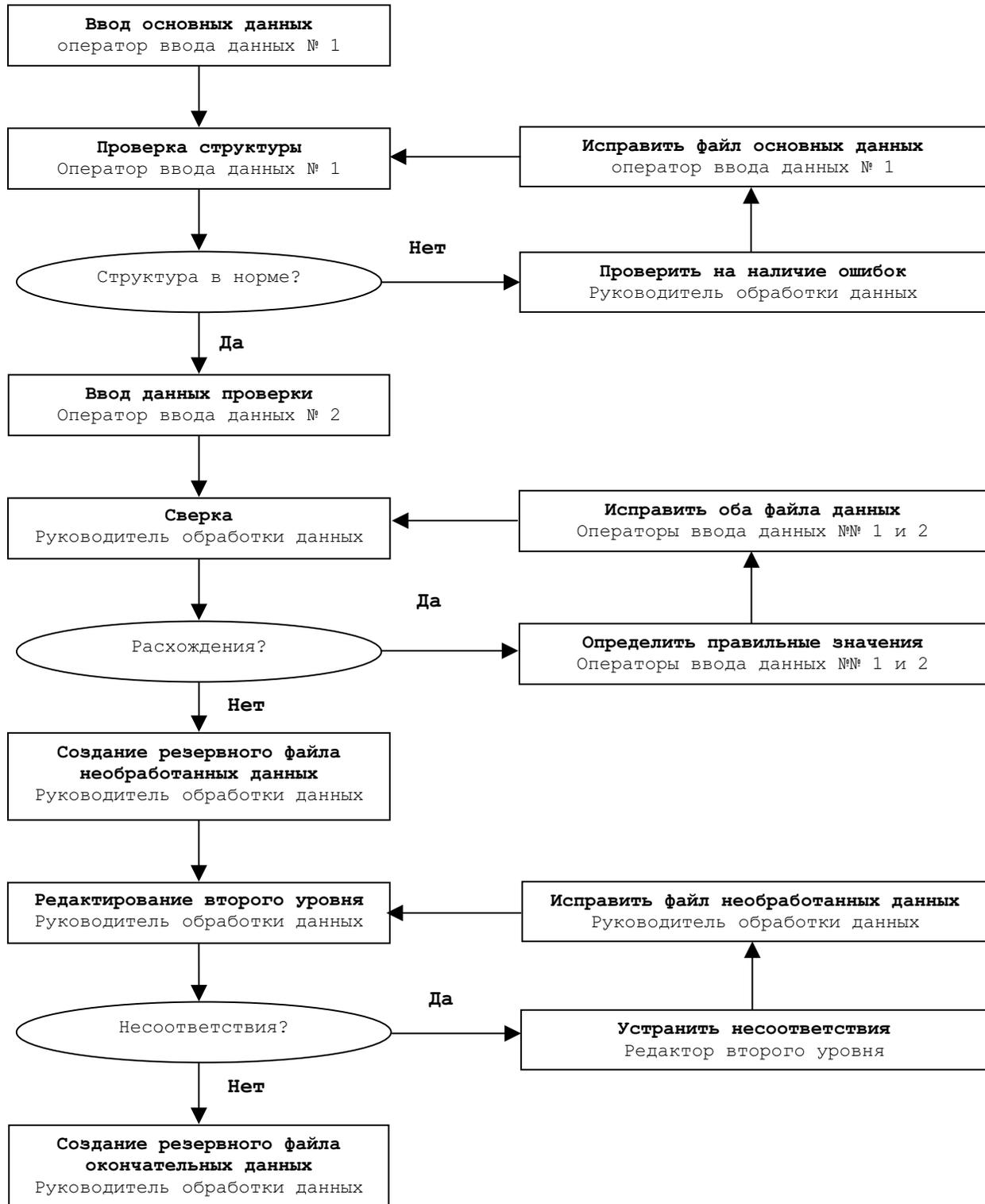
Столь высокая скорость выдачи результатов системой обработки данных МИКС обусловлена тем, что обработка данных ведется параллельно с работой на местах. Для этого данные по каждому кластеру сохраняются в отдельном файле данных. При таком подходе процесс обработки данных разбивается на отдельные сегменты и появляется возможность выполнять обработку данных в то время, когда еще ведется работа на местах. В частности, данные по кластеру обрабатываются, как только вопросники поступают с мест. Таким образом, к тому времени, когда последний вопросник заполнен и возвращен в штаб-квартиру, бóльшая часть данных уже обработана.

Обработка данных по кластерам не представляет трудности, однако требует тщательной организации. Процесс обработки данных может быть разделен на три этапа: подготовка, первичная обработка данных и вторичная обработка данных. Каждый из этих этапов кратко описан ниже в одном из трех подразделов и по каждому в конце настоящей главы прилагается контрольный перечень.

ПОДГОТОВКА К ВВОДУ ДАННЫХ

Цель подготовки к этапу ввода данных состоит в том, чтобы быть готовым начать ввод данных вскоре после начала работы на местах. Этап подготовки включает следующие действия:

- приобретение компьютерного оборудования и подготовка помещения для обработки данных;
- подбор и наем соответствующего персонала;
- адаптация компьютерных программ к вопроснику для конкретной страны;
- создание системы управления файлами вопросников и данных.



ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ

Целью первичной обработки данных является получение проверенных, отредактированных файлов данных. Первичная обработка данных состоит из следующих шагов:

- введение всех вопросников по кластеру в файл данных;
- проверка структуры файла данных;
- повторное введение данных и последующая сверка файла данных;
- создание резервной копии проверенного и сверенного файла данных;
- второй уровень редактирования файла данных;
- создание резервной копии отредактированного или окончательного файла данных.

Последовательность первичной обработки данных представлена в виде блок-схемы на предыдущей странице. Обратите особое внимание на то, что проверка структуры, ввода данных проверки и редактирование второго уровня являются итеративными процедурами, которые выполняются, до тех пор пока не будут разрешены все проблемы или пока не будут найдены приемлемые решения по всем оставшимся спорным вопросам.

ВТОРИЧНАЯ ОБРАБОТКА ДАННЫХ

Цель вторичной обработки данных заключается в создании аналитических файлов данных и составлении стандартных таблиц МИКС. Вторичная обработка данных состоит из следующих шагов:

- соединение всех файлов данных по кластеру в один файл данных;
- экспорт этих данных в систему SPSS;
- расчет весов выборки;
- вычисление индекса благосостояния;
- перекодирование переменных для упрощения анализа;
- создание таблиц, необходимых для анализа данных;
- архивирование и распространение файлов данных.

ПЕРСОНАЛ И ИНФРАСТРУКТУРА

ПЕРСОНАЛ

Команда по обработке данных обследования МИКС состоит из четырех категорий сотрудников: администратор вопросников, операторы ввода данных, редакторы, выполняющие проверку второго уровня, и руководитель обработки данных. Сотрудники каждой категории имеют четкие обязанности, и их совмещение может нанести ущерб качеству ваших данных.

Администратор вопросников (или административный редактор) проверяет и систематизирует вопросники по мере их поступления с мест. Когда в офис обработки данных поступают данные по кластеру, он/она проверяет, все ли вопросники имеются в наличии и готовы к вводу. Если обнаруживается отсутствие каких-либо вопросников, он/она должен разрешать эту проблему с помощью команды на месте (конкретные меры, которые должен предпринять администратор вопросников, подробно изложены ниже).

Операторы ввода данных вводят данные. Они должны иметь опыт ввода данных и знать содержание вопросников. Один из способов выполнения последнего требования – присутствие операторов ввода данных при подготовке интервьюеров. Перед началом ввода данных следует выделить 2–3 дня на подготовку, с тем чтобы ознакомить операторов ввода данных с программой ввода данных и с ритмичностью работы системы обработки данных. К концу подготовки операторы ввода данных должны свободно ориентироваться в программе ввода данных и знать свои повседневные обязанности. Необходимое число операторов ввода данных зависит от количества имеющихся в наличии компьютеров и подробно рассматривается ниже.

Редакторы, выполняющие проверку второго уровня, анализируют и устраняют случаи комплексного несоответствия, обнаруженные программой редактирования второго уровня. Они должны отлично знать содержание вопросников и цели обследования. В помощь редакторам, выполняющим проверку второго уровня, в Дополнении 7 содержатся Руководящие принципы редактирования. Для типичного обследования потребуются один или два редактора для проверки второго уровня.

Руководитель обработки данных является ключевой фигурой в команде обработки данных. Он/она адаптирует типовые программы таким образом, чтобы они соответствовали вопросникам по ее/его стране, и осуществляет надзор за выполнением всех задач обработки данных. Руководитель обработки данных должен иметь опыт управления обработкой данных широкомасштабного обследования или переписи, отлично знать содержание вопросника и иметь навыки программирования в среде прикладных программ CPro и SPSS. Руководитель обработки данных должен быть доступен полный рабочий день в течение всего периода ввода, редактирования и табулирования данных.

Руководитель обработки данных должен быть определен персонально на начальных стадиях планирования обследования, с тем чтобы он/она мог бы принять участие в редактировании вопросника МИКС. С этим сотрудником необходимо консультироваться, для

того чтобы обеспечить последовательность и однозначность используемых в вопроснике схем кодирования и включение всей необходимой идентификационной информации. Руководитель обработки данных также должен быть способен оказать помощь в окончательном редактировании вопросника на основании опыта, приобретенного в ходе введения данных вопросников после предварительных испытаний.

КОМПЬЮТЕРНОЕ ОБОРУДОВАНИЕ И ДРУГИЕ АППАРАТНЫЕ СРЕДСТВА

Ниже перечислено оборудование, необходимое для обработки данных:

- компьютеры для ввода данных;
- компьютер руководителя обработки данных;
- внешнее запоминающее устройство (например, портативное устройство USB или накопитель CD-RW);
- дискеты (или предназначенные для операторов средства передачи файлов руководителю обработки данных, например через сеть);
- принтер;
- бумага;
- картриджи с тонером/ленты для принтера;
- стабилизаторы напряжения;
- источники бесперебойного питания (ИБП);
- ручки с зелеными чернилами.

Компьютеры для ввода данных должны иметь процессоры Pentium, ОС Windows 95 или выше, 32 мегабайт или более оперативной памяти, 1 гигабайт или более свободного пространства на жестком диске и дисководы для гибких дисков 3,5”, или же эти компьютеры должны быть соединены сетью. Количество компьютеров для ввода данных, необходимых для обработки результатов обследования, зависит от размера выборки, количества часов работы оператора ввода данных в неделю, имеющегося помещения и графика проведения обследования. Для расчета количества компьютеров, необходимых для ввода данных, вам следует оценить, сколько времени потребуется оператору ввода данных, чтобы ввести вопросники типичного домашнего хозяйства и умножить это значение на число домашних хозяйств, которое ожидается согласно объему выборки. Если вы не можете оценить время, которое потребуется для ввода вопросников, считайте, что на каждое домашнее хозяйство понадобится ориентировочно 20–30 минут, в зависимости от числа женщин и детей, предположительно приходящегося на каждое домашнее хозяйство. Умножьте это приблизительное значение времени на число домашних хозяйств, чтобы

получить общее количество часов, необходимых для ввода данных. Разделите результаты на количество часов, которое каждый оператор будет работать в неделю, и затем на количество недель, отведенных для полного ввода данных, если вы стремитесь завершить ввод в течение недели после того, как последние вопросники поступят с мест.

Например, если размер выборки для обследования составляет 6000 домашних хозяйств, а на ввод каждого домашнего хозяйства уходит 20 минут, общий объем времени, необходимого для ввода всех этих домашних хозяйств, составляет 2000 часов. Если для полного ввода данных отведено 8 недель, тогда необходимо работать 250 часов в неделю. Если каждый оператор ввода данных работает 40 часов в неделю, вам понадобится 7 компьютеров и 7 операторов ввода данных. Иногда возможно организовать работу по вводу данных в две смены, с тем чтобы на одном компьютере ежедневно работали два оператора. Каждый оператор работал бы, например, 6 часов, и компьютер использовался бы 12 часов в день.

Компьютер руководителя должен иметь более мощный процессор, ОС Windows 98 или выше, 64 или более мегабайт оперативной памяти, 1 гигабайт или более свободного пространства на жестком диске, дисковод для гибких дисков 3,5” (или он должен быть соединен сетью с компьютерами ввода данных) и внешнее запоминающее устройство.

Если в стране, где вы работаете, часто случаются перебои в электроснабжении, необходимы источники бесперебойного электропитания и стабилизаторы напряжения. Когда член команды по обработке данных изменяет данные в вопроснике, он должен пользоваться ручкой с зелеными чернилами. Зеленый цвет делает эти изменения отличными от исходных данных, записанных интервьюером (синие чернила), и любых изменений, произведенных командой на месте (красные чернила).

ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ

Стандартные программы обработки результатов обследований МИКС разрабатывались в среде CSPPro 2.6 и SPSS. Пакет CSPPro, который применялся при обработке результатов как обследований, так и переписей, был разработан совместно Бюро переписи Соединенных Штатов Америки, ORC Macro International и SerPro Ltda. Пакет можно загрузить бесплатно с веб-сайта Бюро переписи Соединенных Штатов Америки¹. SPSS является коммерческим пакетом прикладных программ, который можно приобрести через ЮНИСЕФ и у многих поставщиков программного обеспечения.

РАБОЧЕЕ ПОМЕЩЕНИЕ ОФИСА

Для ввода и редактирования данных необходимы отдельные помещения. Помещение для ввода данных должно быть достаточно просторным, чтобы у каждого оператора ввода данных было место для размещения ее/его компьютера и вопросника, с которым он/она работает. Необходимы планшетные или письменные столы для работы и достаточное количество электрических штепсельных розеток. Помещение должно быть прохладным, хорошо освещенным и, насколько

¹ Адрес веб-сайта: <http://www.census.gov/ipc/www/CSPPro/download/CSPPro26.exe>.

это возможно, в нем не должно быть пыли и высокой влажности. В странах с жарким климатом для этого необходимо, чтобы помещение было оборудовано кондиционером. Каждый компьютер должен быть подсоединен к источнику бесперебойного питания. Если часто или надолго происходят перебои в электроснабжении, необходим аварийный источник электроэнергии, такой как генератор.

Помещение для редактирования данных предназначено для администратора вопросников и редакторов, осуществляющих проверку второго уровня. Оно также должно быть прохладным и хорошо освещенным, и в нем должно быть достаточно места для редакторов, работающих с вопросниками. В идеальном случае в помещении для редактирования должно быть достаточно полок или шкафов для хранения вопросников, систематизированных в определенном порядке. Если нет возможности хранить вопросники в помещении для редактирования, они должны храниться неподалеку, так чтобы доступ к ним не был затруднен, поскольку они требуются на разных стадиях в течение всей обработки. Исключите вероятность недооценки размеров площади, которая понадобится для хранения тысяч вопросников, скапливающихся в офисе к концу работы на местах.

АДАПТАЦИЯ СТАНДАРТНЫХ ПРОГРАММ

Как отмечено в общих чертах в главе 3, типовой вопросник МИКС должен быть адаптирован к ситуации в каждой стране. Это означает, что типовые программы ввода, редактирования и табулирования данных также должны быть модифицированы, для того чтобы соответствовать изменениям, внесенным в вопросник. Чем больше изменений внесено в типовой вопросник, тем больше времени должно быть отведено на адаптацию и тестирование программ. Например, если в вопросник были добавлены новые вопросы, соответствующие добавления должны быть внесены в программы ввода, редактирования и табулирования данных.

Этот процесс существенно облегчается, если сохраняются порядковые номера вопросов в типовом вопроснике. При включении новых вопросов к существующей нумерации следует добавлять букву (например, вопрос, включенный между WS4 и WS5, следует нумеровать как WS4A). Точно так же, если вопросы исключаются, остающиеся вопросы не следует перенумеровывать. Кроме того, когда к уже имеющимся в типовом вопроснике категориям кодов добавляются новые, их следует добавлять в конце существующего перечня, оставляя нетронутыми другие коды. Адаптация программ ввода данных и редактирования должна быть завершена до предварительных испытаний. Вопросники после предварительных испытаний вводятся и редактируются с использованием этих программ. Такая процедура преследует две цели. Она выявит проблемы в кодировании и порядке переходов в вопросниках, и также обнаружит любые ошибки в программах. Как только предварительные испытания завершатся и вопросник примет окончательную форму, в программы могут быть внесены окончательные изменения. В последующих разделах приводятся основные установки по модифицированию типовых словарей данных и типовых прикладных программ пакета CPro. Более подробное резюме содержания прикладных программ CPro приводится в отдельных документах.

Даже если вы не добавляете вопросы к типовым вопросам, типовые словари данных и прикладные программы содержат определенные пункты, которые должны быть обновлены (например, допустимый диапазон для сроков опроса, допустимый диапазон для номера кластера и т. д.). Содержание этих пунктов зависит от специфики конкретной страны и должно определяться вами. Таким образом, вам придется адаптировать стандартные программы, даже если ваша страна использует типовой вопросник.

СЛОВАРИ ДАННЫХ

В системе МИКС группы связанных между собой вопросов сводятся в модули (например, материнская смертность, контрацепция и вакцинация), которые затем составляют вопросники (т. е. для домашних хозяйств, женщин и детей в возрасте до пяти лет). В пакете CSPro для описания этой структуры данных используются словари: группа связанных между собой переменных (вопросы) входит в запись (модуль), а группа записей образует уровень (вопросник). Они хранятся в файле словарей (расширение имени файла: *dcf*). В дополнение к словарю данных для ввода данных используются формуляры, подключенные к словарю. Обычно существует один формуляр для каждой записи. Эти формуляры хранятся в файле формуляров (расширение имени файла: *fmf*). Файлы *dcf* и *fmf* могут быть модифицированы непосредственно. Наилучший способ сделать это – открыть файл формуляров в среде CSPro. Это предоставит вам одновременный доступ к словарю данных и к формулярам и обеспечит сохранение синхронизации между ними. Для справок рекомендуется иметь резервную копию типового словаря данных и файла формуляров.

Существуют три типа вопросников МИКС: для домашних хозяйств, женщин и детей до пяти лет. Каждый вопросник для женщин и детей посвящен единственному элементу анализа: женщине и ребенку, соответственно. Вопросник для домашних хозяйств содержит два элемента анализа: домашнее хозяйство и члены домашнего хозяйства. Все типы вопросников хранятся в файлах *micx3.dcf* и *entry.fmf*.

ИДЕНТИФИКАЦИОННЫЕ ПЕРЕМЕННЫЕ И УРОВНИ

В CSPro каждый вопросник должен иметь набор переменных, однозначно его идентифицирующих. Например, домашнее хозяйство идентифицируется своими номером кластера и номером домашнего хозяйства. Переменные, которые идентифицируют вопросник, известны как идентификационные переменные. В нижеприведенной таблице перечисляются типы вопросников и их идентификационные переменные.

Вопросник	Номер кластера	Номер домашнего хозяйства	Номер строки
Домашнее хозяйство	НН1	НН2	
Женщины	НН1	НН2	LN
Дети	НН1	НН2	LN

Как видно из таблицы, женщины и дети имеют одинаковые идентификационные переменные. Поскольку каждый член домашнего хозяйства занесен в отдельную строку в списке проживающих в домашнем хозяйстве, никогда две женщины или два ребенка не будут иметь одинаковый номер строки, даже если они принадлежат к одному домашнему хозяйству. Таким образом, номер строки в сочетании с номером кластера и номером домашнего хозяйства однозначно идентифицируют женщину или ребенка.

В словаре CSPro уровень определяется набором идентификационных переменных. В словаре МИКС существуют два уровня: домашние хозяйства и индивидуумы (т. е., отвечающие критериям женщины и отвечающие критериям дети). Домашние хозяйства являются первым уровнем, а женщины и дети образуют второй уровень. Эта иерархическая структура является естественной, поскольку в вопроснике МИКС каждая женщина или ребенок принадлежат к тому или иному домашнему хозяйству, а в данном домашнем хозяйстве может быть много женщин или детей.

Вопросники для женщин и детей хранятся на одном уровне, поскольку каждый из них является вопросником, применимым к члену домашнего хозяйства. Прикладная программа ввода данных содержит логическую схему, которая пропускает формуляры, относящиеся к детям, когда вводятся данные из вопросника для женщин, и пропускает формуляры, относящиеся к женщинам, когда вводятся данные из вопросника для детей. Таким образом, хотя и вопросники для женщин, и вопросники для детей хранятся как элементы второго уровня, они не имеют общих переменных, кроме идентификационных переменных.

Модули

Словарь данных составлен для отображения модульной структуры вопросников МИКС. Каждый модуль хранится в собственной записи (исключение: как описание домохозяйства, так и модуль по детям, которые стали сиротами и уязвимыми, имеют две записи каждый в силу их необычной структуры) в файле *mics3.def*, и каждая запись имеет формуляр (или два, в случае с модулем характеристик домохозяйства), ассоциированный с ней, в файле *entry.fmf*. Таким образом, если ваша страна не использует конкретный модуль, вы можете изъять его путем удаления его записи и его формуляра (и удалив любую другую логическую схему, которая вызывает его из прикладной программы ввода данных).

Модулями, применяемыми для вопросников домохозяйства (код/коды модулей перечислены в скобках), являются: информация по домохозяйству (HH), описание домохозяйства (HL и TO), образование (ED), водоснабжение и санитария (WS), характеристики домохозяйства (HC), инсектицидная сетка (TN), дети, которые стали сиротами и уязвимыми (OV и OR), детский труд (CL), детская дисциплина (CD), инвалидность (DA), материнская смертность (MM) и употребление йодированной соли (SI).

Модулями, применяемыми для вопросников для женщин, являются: информация о женщинах (WM), детская смертность (CM), противостолбнячная прививка (TT), здоровье матерей и новорожденных (MN), семейное положение/союз (MA), безопасность владения и долговечность

жилища (ST), использование контрацептивов (CP), обрезание женских половых органов (FG), отношение к домашнему насилию (DV), сексуальное поведение (SB) и ВИЧ/СПИД (HA).

Модулями, применяемыми для вопросников по детям, являются: информация о ребенке до пяти лет (UF), регистрация рождения и раннее обучение (BR), детское развитие (CE), витамин А (VA), грудное вскармливание (BF), лечение болезней и уход (CA), малярия (MLA), иммунизация (IM) и антропометрия (AN).

ПРАВИЛА НАИМЕНОВАНИЯ ПЕРЕМЕННЫХ

Переменные именовываются по модулю вопросников, в котором они расположены, и по номеру вопроса, ответ на который они содержат. Например, ответ на вопрос 9 в описи домохозяйства хранится в переменной с именем HL9. Некоторые вопросы разбиваются на две или более частей, и их отдельные части идентифицируются присвоенной только им буквой. Каждая часть таких вопросов хранится в отдельной переменной. Имена этих отдельных переменных включают буквы, которые определяют части этого вопроса. Например, вопрос 11 модуля по здоровью матерей и новорожденных имеет две части. Ответ на первую часть этого вопроса хранится в переменной *MN11A*, а на вторую часть – в *MN11*.

Некоторые вопросы имеют две или более частей, принадлежащих к разным категориям ответов. Ответы на эти вопросы хранятся в единой переменной, и категории ответа определяются как подэлементы. Когда эти вопросы касаются дат, к основному имени переменной добавляются буквы d (день), m (месяц) и y (год), чтобы создать имя подэлементов. В вопросе 6 модуля информации о женщинах, например, требуется указать день, месяц и год рождения женщины. Ее ответ хранится в WM8, который имеет три подэлемента: *wm8d*, *wm8m* и *wm8y*.

Некоторые вопросы имеют структуру, в которой первая часть ответа является формой ответа, а вторая часть является ответом по существу. Эти вопросы хранятся в единой переменной, а форма и ответ определяются как подэлементы. Именем подэлемента, хранящего форму ответа, является имя переменной с добавленной к нему буквой u (единица), а именем подэлемента, хранящего ответ, является имя переменной с добавленной к нему буквой n (количество). Например, в вопросе 13 в модуле по здоровью матерей и новорожденных спрашивается, через какое время после рождения ребенок впервые получил грудное молоко. Респондент в ответе может назвать часы или дни. Этот ответ хранится в переменной *tn13* с подэлементами *tn13u* и *tn13n*.

ВОПРОСЫ С МНОЖЕСТВЕННЫМИ ОТВЕТАМИ И БУКВЕННО-ЦИФРОВЫЕ ПЕРЕМЕННЫЕ

В вопроснике содержится ряд вопросов, которые допускают множественные ответы. Эти вопросы различаются в вопроснике буквенно-цифровыми кодами ответов (т. е. буквами от A до Z). В словаре данных ответ на вопрос с множественными ответами хранится в буквенно-цифровой переменной, длина которой равна максимальному числу возможных ответов. Они являются единственными буквенно-цифровыми переменными в словаре. Каждая буквенно-цифровая

переменная имеет один подэлемент для каждого кода ответа в вопроснике. Имя такого подэлемента – это имя переменной плюс код ответа, который представляет подэлемент.

Например, во втором ответе в модуле по здоровью матерей и новорожденных регистрируются все данные о физических лицах, которые оказывали женщине дородовой уход перед ее последними по времени родами. Возможные коды ответа – А, В, С, F, G, H, X и Y. Поэтому переменная *mn2* имеет длину 8 знаков и существуют восемь подэлементов: *mn2a*, *mn2b*, *mn2c*, *mn2f*, *mn2g*, *mn2h*, *mn2x* и *mn2y*.

ПРАВИЛА КОДИРОВАНИЯ

Для типовых словарей используется стандартное кодирование конкретных ответов. Сначала мы рассмотрим правила кодирования для числовых переменных. Ответ "Другое" всегда кодируется шестеркой с предшествующими девятками. Несоответствующие ответы всегда кодируются семеркой с предшествующими девятками. Ответ "Не знаю" всегда кодируется восьмеркой с предшествующими девятками. Вопросы, оставшиеся без ответа (т. е. интервьюер не записал ответ на применимый вопрос), всегда кодируются девяткой с предшествующими девятками. Вопросы, которые не применимы к респонденту, всегда кодируются пробелом. В нижеприведенной таблице суммируются стандартные правила кодирования.

Ответ	Длина переменной				
	Символьная	Один знак	Два знака	Три знака	Четыре знака
Другое	X	6	96	996	9996
Нет/Никакой	Y	na	na	na	na
Несоответствующий	na	7	97	997	9997
Не знаю	Z	8	98	998	9998
Отсутствующий ответ	?	9	99	999	9999
Неприменимый вопрос	Пробел	пробел	пробел	Пробел	Пробел

Поскольку коды 6–9 зарезервированы для специального использования, любой вопрос, который требует более 6 категорий ответа, должен иметь двузначное обозначение категории ответа, начинающееся с нуля (например, 01, 02, 03, 04, 05, 06, 96, 97, 98 и 99).

Для буквенно-цифровых переменных ответ "Другое" всегда кодируется как X, ответ "Не знаю" всегда кодируется как Z, отсутствующее значение всегда кодируется с использованием вопросительного знака ?, а неприменимое кодируется как пробел.

ДИАПАЗОНЫ

Большинство вопросов в вопросниках МИКС имеют определенный диапазон ответов. Диапазоны определяются для переменных в словаре *mics3.dcf*. В процессе ввода данных CSPro контролирует, чтобы любое значение, введенное в переменную, находилось в определенном для этой переменной диапазоне. CSPro допускает большое число диапазонов для каждой переменной, поэтому вопросы с диапазонами ответа, которые не следуют один за другим (например, 1–8, 96, 98 и 99), должны определяться с использованием нескольких диапазонов (например, 1–6, 96, 98, 99 вместо 1–99). Хотя диапазоны словаря полезны для проверки простых диапазонов, более

сложные или условные диапазоны (например, согласованность между днем и месяцем в переменной даты) должны проверяться в прикладных программах ввода данных или редактирования.

ПРИКЛАДНАЯ ПРОГРАММА ВВОДА ДАННЫХ

Прикладная программа ввода данных – это объемная и сложная программа. Ограниченные размеры данной главы не позволяют описать эту программу подробно. Вместо этого настоящий подраздел будет посвящен ряду важных общих вопросов, касающихся прикладной программы ввода данных.

ПЕРЕХОДЫ

В вопросниках МИКС широко применяются переходы. Переходы – это указатели в вопроснике, которые предписывают интервьюеру пропустить все вопросы между текущим вопросом и тем или иным вопросом, расположенным далее в вопроснике. Переходам в вопроснике должны соответствовать переходы в соответствующей программе ввода данных. Переходы в программе ввода данных определяют путь ввода данных. CSPro строго выдерживает путь ввода данных, всякий раз когда используются команды *skip* (перейти) или *skip to next* (перейти к следующему).

СООБЩЕНИЯ ОБ ОШИБКЕ

Если оператор ввода данных вводит значение для переменной, которое не соответствует введенной ранее информации, полезно вывести на дисплей сообщение об ошибке. Это сообщение об ошибке должно объяснить характер проблемы и предоставить любую информацию, которая может помочь устранить это несоответствие. В CSPro функция *errmsg* каждый раз при вызове выводит на дисплей сообщение об ошибке с текстом, определяемым пользователем. Сообщения об ошибках для программы ввода данных нумеруются и хранятся в файле *entry.mgf*. Текст, номер и несоответствия, которые привели к выводу каждого из этих сообщений на дисплей, перечислены в Дополнении 7, как и руководящие указания для устранения обнаруженных несоответствий.

Вам следует проанализировать вопросник, с тем чтобы решить, требует ли какой-либо из добавленных вопросов проверки на соответствие. Если проверка требуется, вам следует добавить логическую схему для выполнения проверки соответствия в программе ввода данных, программе редактирования, или в обеих. Когда вы добавляете проверку соответствия, обязательно добавьте соответствующее сообщение об ошибке к файлу сообщений для ввода данных или редактирования. Если вы добавите сообщение об ошибке, удостоверьтесь также, что вы не используете существующий номер сообщения об ошибке.

Некоторые сообщения об ошибке сопровождаются командой *reenter* (повторить ввод), которая возвращает оператора в поле вводимых в настоящий момент данных. Это заставляет оператора ввода исправить ошибку, прежде чем идти дальше. Поскольку оператор ввода данных иногда должен будет вводить исправления, необходим тщательный контроль. Когда вы

добавляете собственные сообщения об ошибке, тщательно проанализируйте, хотите ли вы, чтобы оператор ввода данных разрешал проблему, прежде чем продолжить ввод. В таком случае сопровождайте свое сообщение об ошибке командой *reenter*.

БУКВЕННО-ЦИФРОВЫЕ ПЕРЕМЕННЫЕ

Прикладная программа ввода данных проверяет правильность введения буквенно-цифровых переменных. Программа выполняет четыре проверки каждой буквенно-цифровой переменной. Во-первых, она сверяет, что введенное значение содержит только те коды, которые перечисляются в вопроснике (т. е. проверяет диапазон). Во-вторых, программа контролирует ввод ответов в алфавитном порядке (т. е. *ACG*, а не *GAC*). В-третьих, она проверяет, чтобы в случае включения в ответ кодов "Не знаю" или "Никто" (обычно буква *Y*), не присутствовало другого ответа (т. е. она не допустит ответа *ACY*). В-четвертых, программа проверяет, чтобы в том случае, когда в ответ включен код отсутствия ответа (?), не присутствовало никакого другого ответа (т. е. она не допустит ответа *AC?*).

Прикладная программа ввода данных также переставляет значения, введенные в буквенно-цифровые переменные, таким образом, чтобы каждый ответ хранился в той ячейке памяти, которая определяет его подэлемент. Например, для переменных *mn2* ответ *ACG* будет преобразован в *A C G*, где между *A* и *C*, а также между *C* и *G* имеется по одному пробелу и три пробела после *G*.

ФУНКЦИИ, ОПРЕДЕЛЯЕМЫЕ ПОЛЬЗОВАТЕЛЕМ

Привлекательной чертой *CSPro* является возможность определения собственных функций на уровне программ. Такие функции известны как функции, определяемые пользователем, и они могут оказаться полезными. В частности, такие функции позволяют избежать многократного написания часто используемого кода. Функции, определяемые пользователем, всегда определяются на верхнем уровне прикладной программы *CSPro*. Прикладная программа ввода данных *entry.app* содержит 14 функций, определяемых пользователем. От вас не требуется модифицировать эти функции, но если вы хотите понять принцип работы прикладной программы ввода данных, вы должны понимать, для чего именно эти функции применяются.

Функция *valid* проверяет, является ли значение переменной одним из конкретных значений: "Несоответствующий", "Не знаю", "Отсутствие" или "Не применимо". Если значение переменной "не применимо", функция *natozero* меняет ее на 0, позволяя ей быть добавленной к другой переменной (пример ее применения см. в процедуре *cm9*).

Следующие 7 функций, определяемых пользователем (*zscoef*, *dabs*, *zspct*, *zseval*, *zscr*, *zsanth* и *agemth*), используются в вопроснике по детям, чтобы рассчитать антропометрические показатели, которые находятся в конце антропометрического модуля. Функция *agemth* вызывается для расчета возраста ребенка в месяцах. Затем вызывается функция *zsanth*. Эта функция вызывает *zseval*, *zscr* и *zspct*. Функция *zseval* вызывает *zscoef*, а *zspct* вызывает *dabs*. Вы встретите эти функции только в антропометрических переменных, и если вы их встретите, то будете знать, что они рассчитывают и затем проверяют антропометрические показатели.

Программа функции *agemth* рассчитывает возраст ребенка в месяцах. Поскольку антропометрия чрезвычайно чувствительна к возрасту, возраст ребенка должен базироваться на значении возраста ребенка в днях. Программа сначала рассчитывает количество дней, истекших с начала года до рождения ребенка. Затем она рассчитывает количество дней, истекших с начала года до даты опроса. Наконец, количество дней в годах между годом рождения и годом опроса складывается с количеством дней с начала текущего года до даты опроса. Разница между этими двумя количествами дней является возрастом ребенка в днях. Затем это число переводится в возраст ребенка в месяцах путем деления на 30,4375 (среднее количество дней в месяце в течение четырех лет). Поскольку необходима точность, возраст ребенка в месяцах рассчитывается с точностью до двух знаков после запятой.

Функции *vdvalid*, *vdoi* и *vdob* проверяют, чтобы даты вакцинации, введенные в модуль по иммунизации, были непротиворечивыми, предшествовали дате опроса и не предшествовали дате рождения, соответственно. Функция *endmess* (сокращение от "end message" – "конец сообщения") выводит в конце вопросника на дисплей сообщение, в котором оператору ввода данных задается вопрос, хочет ли он/она просмотреть данный вопросник или перейти к следующему. Наконец, функция *alphachk* осуществляет проверку буквенно-цифровых переменных, подробно описанную в предыдущем подразделе.

ДАТЫ И КОД МЕСЯЦА ПО СТОЛЕТИЮ

Типовые программы (включая прикладную программу ввода данных) используют для большинства дат код месяца по столетию (СМС). СМС даты – это количество месяцев с декабря 1899 года. Например, СМС января 1900 года равен 1, СМС марта 2000 года равен 1203. СМС даты рассчитываются следующим образом: вычитается 1900 из года даты, это число лет умножается на 12, затем к произведению прибавляется порядковый номер месяца даты. Например, СМС марта 2000 года рассчитывается как $(2000 - 1900) \times 12 + 3$.

Прикладная программа ввода данных использует четыре функции, чтобы упростить работу с кодами месяца по столетию. Две из этих функций, *setlb* и *setub*, рассчитывают нижние и верхние границы, соответственно, для СМС даты события. Две другие функции, *adjlba* и *adjuba*, корректируют нижние и верхние границы, соответственно, СМС даты события (т. е. дня рождения ребенка), когда также указывается возраст. В нижеприведенной таблице представлено резюме этих функций.

<pre>1 cmc = setlb (month, yea, minimum) ;</pre>	<p>Аргументы функции – <i>month</i> (месяц), <i>year</i> (год) и <i>minimum</i> СМС. Если <i>year</i> (год) и <i>month</i> (месяц) указаны верно, рассчитывается и возвращается СМС. Если <i>year</i> (год) указан неверно, возвращается <i>minimum</i>. Если <i>month</i> (месяц) указан неверно, возвращается СМС января <i>year</i> (год).</p>
<pre>ucmc = setub (month, year, maximum) ;</pre>	<p>Аргументы функции – <i>month</i> (месяц), <i>year</i> (год) и <i>maximum</i> СМС. Если <i>year</i> (год) и <i>month</i> (месяц) указаны верно, рассчитывается и возвращается СМС. Если <i>year</i> (год) указан неверно, возвращается <i>maximum</i>. Если <i>month</i> (месяц) указан неверно, возвращается СМС декабря <i>year</i> (год).</p>
<pre>t = adjlba (lcmc, ucmc, di, di, age) ; if t < 0, then e = errmsg (2171) ; else lcmc = t; endif ;</pre>	<p>Аргументы функции – минимум СМС даты (<i>lcmc</i>), максимум СМС даты (<i>ucmc</i>), СМС даты опроса (<i>di</i>) и <i>age</i> (возраст). Эта функция повышает <i>lcmc</i>, используя <i>age</i> (возраст) и <i>di</i>. Указанная функция возвращает результирующую дату, если она больше или равна <i>lcmc</i> и меньше или равна <i>ucmc</i>. Если результирующая дата больше <i>lcmc</i>, функция возвращает <i>lcmc</i>. Если результирующая дата превышает <i>ucmc</i>, функция возвращает –1.</p>
<pre>t = adjuba (lcmc, ucmc, di, di, age) ; if t < 0, then e = errmsg (2171) ; else ucmc = t; endif ;</pre>	<p>Аргументы функции – минимум СМС даты (<i>lcmc</i>), максимум СМС даты (<i>ucmc</i>), СМС для даты опроса (<i>di</i>) и <i>age</i> (возраст). Функция понижает <i>ucmc</i>, применяя <i>age</i> (возраст) и <i>di</i>. Функция возвращает результирующую дату, если она больше или равна <i>lcmc</i> и меньше или равна <i>ucmc</i>. Если результирующая дата больше <i>ucmc</i>, функция возвращает <i>ucmc</i>. Если результирующая дата меньше <i>lcmc</i>, функция возвращает –1.</p>

ПОЛУЧЕНИЕ ВОПРОСНИКОВ С МЕСТ

Когда вопросники по кластеру поступают с мест, руководитель опроса должен сверить количество вопросников с контрольным планшетом по кластеру. Если какие-либо вопросники отсутствуют, руководитель опроса должен обратиться к команде, работающей на месте, и установить, можно ли найти эти вопросники. Если их не удастся найти, команда, работающая на месте, должна заново заполнить отсутствующие вопросники, если это вообще возможно, (т. е. если эта команда, работающая на месте, все еще находится рядом с данным кластером). Если это невозможно и отсутствующий вопросник – это вопросник домохозяйства, руководитель опроса должен добавить незаполненный вопросник (кроме идентификационных переменных) с итоговым кодом б к вопросникам данного кластера. Если отсутствующий вопросник является вопросником по женщинам или детям, руководитель опроса должен изменить общее количество заполненных вопросников на титульном листе домашнего хозяйства, к которому относится женщина или ребенок.

Поступление вопросников с мест должно регистрироваться руководителем опроса в формуляре отслеживания кластеров. Формуляр отслеживания кластеров существует как в электронной, так и в бумажной форме. Пример бумажной версии формуляра представлен в конце настоящей главы, а электронная версия формуляра доступна через меню руководителя. До получения каких-либо вопросников с мест кластеры должны быть перечислены в возрастающем порядке в формуляре отслеживания кластеров. По мере поступления кластеров вопросников информация по каждому из них может регистрироваться в соответствующем месте. Кроме того, администратор вопросников должен обеспечить по каждому кластеру соблюдение следующих условий:

- налицо все вопросники, перечисленные в контрольном планшете по кластеру;
- по каждому домашнему хозяйству вопросники по домашнему хозяйству, детям и женщинам находятся вместе;
- вопросники отсортированы по номерам домашних хозяйств в возрастающем порядке;
- все вопросники по женщинам сложены вместе, в возрастающем порядке номеров строк женщин, и за ними следуют все вопросники по детям, также в возрастающем порядке номеров строк;
- все географические коды и коды информации об опросе указаны в информационной панели для домашнего хозяйства в начале каждого вопросника домохозяйства.

Когда проверка завершена и все проблемы урегулированы, руководитель опроса должен зарегистрировать количество вопросников по домашним хозяйствам, женщинам и детям в формуляре отслеживания кластеров. Он/она должен/должна также составить итоговый лист кластера. Итоговый лист кластера должен быть листом плотной бумаги, которым обертывается край пакета с вопросниками таким образом, чтобы был виден номер кластера, записанный жирным шрифтом. В итоговом листе кластера должно также суммироваться количество вопросников в кластере, как показано ниже.

Как только составлен итоговый лист кластера, руководитель должен распределить кластер оператору ввода данных (обратите внимание, что оператор ввода данных, которому распределен кластер, должен ввести все вопросники этого кластера). Затем руководитель опроса указывает фамилию оператора ввода данных в формуляре отслеживания кластеров.

101	
Общее количество вопросников домохозяйства	_____
Общее количество опрошенных домохозяйств	_____
Общее количество вопросников по женщинам	_____
Общее число опрошенных женщин	_____
Общее количество вопросников по детям	_____
Общее число опрошенных детей	_____

Каждый раз, когда вопросники извлекаются с места их хранения для ввода данных или редактирования, они должны быть вновь уложены в том же порядке и возвращены на прежнее место. Четкое, аккуратное обращение с вопросниками в бумажной форме будет способствовать быстрой и эффективной обработке данных.

Когда все кластеры внесены в формуляр отслеживания кластеров, руководитель должен просмотреть вопросники и проверить, готовы ли они для ввода данных. Конкретные проверки, которые следует произвести, перечислены в нижеприведенной таблице. После завершения проверочных действий и разрешения всех обнаруженных проблем руководитель распределяет кластер тому или иному оператору ввода данных.

ВВОД ОСНОВНЫХ ДАННЫХ

Начать ввод данных предпочтительнее всего вскоре после того, как интервьюеры начали работать на местах. Это даст вам возможность обнаружить и исправить ошибки, которые, возможно, допускают некоторые интервьюеры или команды. Можно будет вовремя выявить серьезные проблемы, ускользнувшие от внимания руководителя на месте, с тем чтобы провести дополнительную подготовку персонала на местах и исправить серьезные ошибки. Руководители на местах отвечают за проверку вопросников на полноту и соответствие и за классификацию всех ответов, которые не смог закодировать интервьюер. После возвращения вопросников в офис необходимой должна быть только минимальная проверка, как описано выше. Как только данные по одному кластеру поступают в штаб-квартиру, следует начать ввод данных.

Проверки в офисе

1. Убедитесь, что все вопросники домохозяйств рассортированы в возрастающем порядке номеров домашних хозяйств в пределах кластера.
2. Убедитесь, что все географические коды и коды информации об опросе внесены в модуль информации о домохозяйстве в начале каждого вопросника домохозяйства.
3. Соответствие критериям для опроса каждого члена домашнего хозяйства должно быть проверено в формуляре описи домохозяйства. Чтобы отвечать критериям модулей для женщин, лицо должно быть женского пола в возрасте 15–49 лет. Чтобы отвечать критериям модуля по детскому труду, лицо должно быть в возрасте 5–14 лет (или в возрастном диапазоне, соответствующем данной стране). Чтобы отвечать критериям модулей по здоровью детей, лицо должно быть младше 5 лет. Для каждого отвечающего критериям лица код соответствия критериям должен быть обведен в кружок в соответствующем номере строки или номере строки его матери или основного воспитателя (HL6–HL8, Бланк описи домохозяйства). Для других членов домохозяйства столбцы соответствия критериям в описи домохозяйства должны оставаться незаполненными и будут введены как 00. Для разрешения каких-либо проблем следуйте инструкциям, содержащимся в руководящих принципах редактирования (Дополнение 6), для сообщений 0101–0132.
4. Общее число отвечающих критериям женщин и детей, внесенных в модуль информации о домохозяйстве, должно соответствовать числу отвечающих критериям женщин и детей, внесенных в Бланк описи домохозяйства.
5. Общее число отвечающих критериям женщин и детей, внесенных в модуль информации о домохозяйстве, должно соответствовать количеству вопросников для женщин и детей для домашнего хозяйства.
6. Количество полностью проведенных опросов женщин и детей, внесенных в модуль информации о домохозяйстве, должно соответствовать числу вопросников для женщин и детей с итоговым кодом 1 (завершенный) для домашнего хозяйства.
7. Убедитесь, что все вопросники для женщин сложены вместе, в возрастающем порядке номеров строк женщин (WM4), а за ними следуют все вопросники для детей в возрастающем порядке номеров строк ребенка (UF4).

Ход процесса ввода данных определяет прикладная программа CSPro *entry.ent*. Эта прикладная программа выполняет две основных задачи. Во-первых, для данного домашнего хозяйства она удостоверяет, что введены все вопросники (и только эти вопросники) и что введена вся информация из вопросников (и только эта информация). В частности, прикладная программа ввода данных начинается со ввода всех данных из вопросника домохозяйства. Как только эти данные введены, программа анализирует список проживающих в домашнем хозяйстве и определяет, какие члены домашнего хозяйства отвечают критериям для применения вопросников для женщин. Выявив этих членов, прикладная программа ввода данных запрашивает данные по каждому лицу из соответствующего ему вопросника для женщин в порядке возрастания номеров строк. Затем прикладная программа ввода данных выполняет ту же процедуру для членов

домашнего хозяйства, которые отвечают критериям для применения вопросника по детям. После введения всех вопросников домохозяйства данные о домашнем хозяйстве в целом сохраняются, и прикладная программа переходит к следующему домашнему хозяйству в кластере, если таковое имеется.

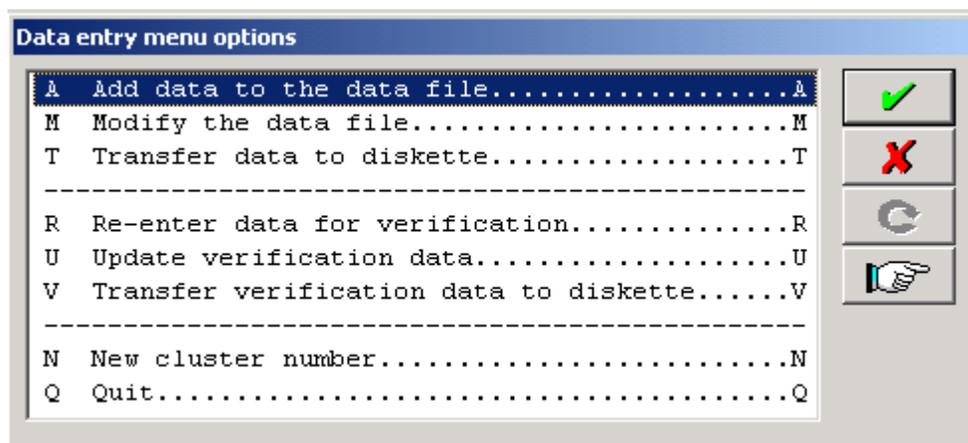
Поскольку данные сохраняются только после ввода вопросников домохозяйства и всех отвечающих критериям женщин и детей, операторы ввода данных не должны покидать свои компьютеры, не завершив ввода данных по домашнему хозяйству. До перерыва или завершения рабочего дня должны быть полностью введены все вопросники для того или иного домашнего хозяйства. Далее в качестве меры предосторожности рекомендуется скопировать данные на компьютер руководителя или на дискету. Кроме того, каждый вечер руководитель должен копировать содержимое каталога *c:\mics* и всех его подкаталогов на внешнее запоминающее устройство. Эта защитная мера позволит руководителю восстановить запись, если ее/его компьютер выйдет из строя.

В дополнение к контролю за тем, какие вводятся вопросники, прикладная программа ввода данных строго контролирует порядок перехода внутри вопросников. Так, она запросит ответы только на вопросы, которые должны были быть заданы, исходя из ответов, полученных на предыдущие вопросы. Например, если в переменную *cml* введено значение 2 (т. е. никогда не рожавшая женщина), прикладная программа ввода данных далее запросит значение переменной *mal*, пропустив все переменные, которые имеют отношение только к рожавшим женщинам.

Вторая задача прикладной программы ввода данных состоит в минимизации количества ошибок при вводе данных. Прикладная программа ввода данных выполняет эту задачу путем проверки в ходе ввода данных. Если введенное значение ответа на вопрос находится за пределами значений, определенных для вопросника, или если обнаружено другое существенное несоответствие, прикладная программа ввода данных выводит сообщение об ошибке на дисплей и, прежде чем продолжить работу, требует от оператора ввода данных ликвидировать это несоответствие. Более сложные несоответствия, разрешение которых существенно замедлит ввод данных, в процессе ввода данных не отслеживаются, но контролируются во время редактирования второго уровня.

МЕНЮ ВВОДА ДАННЫХ

Меню ввода данных упрощает задачу введения данных. Меню ввода данных создается прикладной программой *CSPRO entry_menu.bch* (вам не нужно модифицировать эту программу). Текст, приведенный ниже, описывает меню ввода данных.



Подписи к рисунку:

Опции меню ввода данных

A	Добавить данные к файлу данных.....A
M	Изменить файл данных.....M
T	Перенести данные на дискету.....T

R	Ввести данные повторно для проверки.....R
U	Обновить данные проверки.....U
V	Перенести данные проверки на дискету.....V

N	Новый номер кластера.....N
Q	Выход.....Q

Опции *A*, *M* и *T* используются только в том случае, если оператор ввода данных входит в файл основных данных. Опции *R*, *U* и *V* используются только в том случае, если оператор ввода данных входит в файл данных проверки. Опции *N* и *Q* могут использоваться в обоих случаях.

Опция *A* позволяет оператору ввода данных добавлять блоки в файл основных данных, опция *M* позволяет ей/ему модифицировать существующие блоки в файле основных данных, а опция *T* позволяет ей/ему копировать заверченный файл основных данных на дискету, с тем чтобы его можно было передать на компьютер руководителя. Опция *R* позволяет оператору ввода данных добавлять блоки в файл данных проверки, опция *U* позволяет ей/ему модифицировать существующие блоки в файле проверки данных, а опция *V* позволяет ей/ему копировать заверченный файл данных проверки на дискету, с тем чтобы его можно было передать на компьютер руководителя.

Опция *N* позволяет оператору ввода данных изменить номер кластера так, чтобы он мог вводить данные по другому кластеру. Опция *Q* позволяет выйти из меню ввода данных.

СТРУКТУРА КАТАЛОГОВ В КОМПЬЮТЕРЕ ВВОДА ДАННЫХ

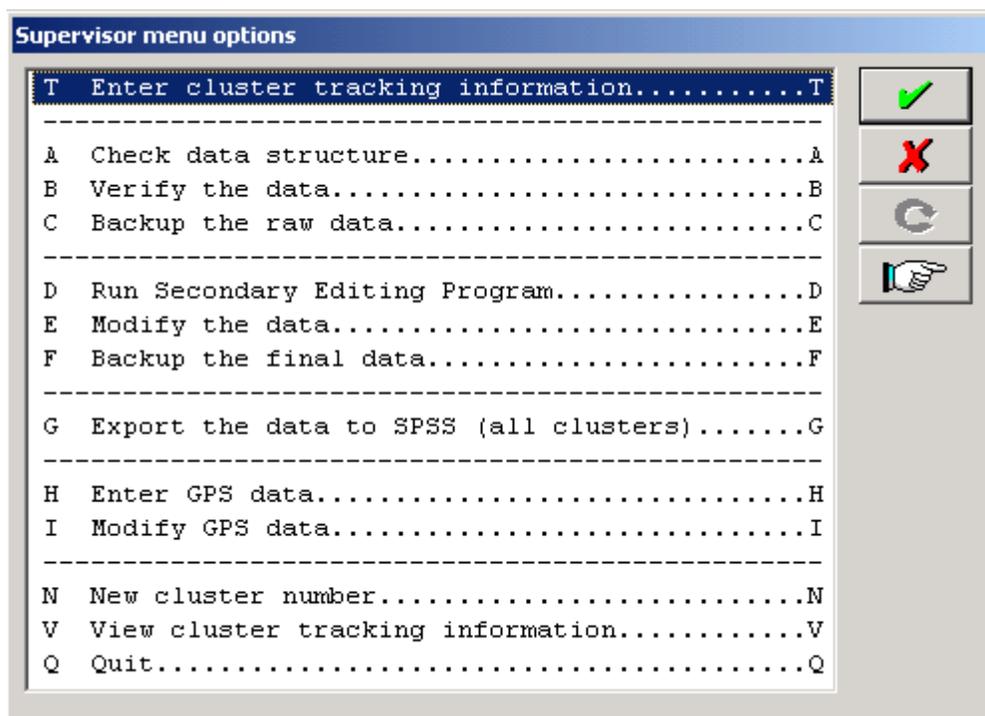
В компьютере ввода данных все файлы и программы перечислены в каталоге *c:\mics\CSPro* или в одном из его подкаталогов. Подкаталоги имеют имена: *data*, *dicts*, *entry* и *veri*. Каталог *data* содержит все файлы основных данных, которые были введены в компьютер. Каталог *dicts* содержит все словари данных. Каталог *entry* содержит прикладную программу ввода данных и прикладную программу, которая создает меню ввода данных. Каталог *veri* содержит все файлы данных проверки, которые были введены в компьютер.

Структура каталогов компьютера оператора ввода данных:

MICS		
	CSPRO	
		DATA
		DICTS
		ENTRY
		VERI
		Каталог основных данных
		Каталог словарей данных
		Каталог прикладной программы ввода данных
		Каталог данных проверки

МЕНЮ РУКОВОДИТЕЛЯ

Меню руководителя помогает руководителю обработки данных управлять системой обработки данных МИКС. Меню запускается выполнением прикладной программы *CSPro super menu.pff*. Меню сначала запрашивает номер кластера, который предстоит обработать. Как только введен номер кластера, появляется нижеприведенный текст меню руководителя, сопровождаемый кратким описанием каждой опции.



Подписи к рисунку:

Опции меню руководителя

```

T   Ввести информацию об отслеживании кластера..... T
-----
A   Проверить структуру данных..... A
B   Сверить данные..... B
C   Создать резервную копию необработанных данных..... C
-----
D   Запустить программу редактирования второго уровня... D
E   Изменить данные..... E
F   Создать резервную копию окончательных данных..... F
-----
G   Экспортировать данные в SPSS (все кластеры)..... G
-----
H   Ввести данные GPS..... H
I   Изменить данные GPS..... I
-----
N   Новый номер кластера..... N
V   Просмотреть информацию об отслеживании кластеров... V
Q   Выход..... Q

```

ОПЦИЯ T. ENTER CLUSTER TRACKING INFORMATION (ВВЕСТИ ИНФОРМАЦИЮ ОБ ОТСЛЕЖИВАНИИ КЛАСТЕРА)

Эта опция выводит на дисплей электронный формуляр отслеживания кластеров, с тем чтобы руководитель обработки данных мог ввести информацию об отслеживании кластера. Информация должна быть введена после получения с мест первых вопросников по кластеру, когда этот кластер распределен оператору ввода данных для ввода основных данных и когда данный кластер распределен оператору ввода данных для ввода данных проверки. Другая информация в электронном формуляре отслеживания кластеров обновляется автоматически, по мере того как руководитель обработки данных продвигается в меню руководителя.

ОПЦИЯ A. CHECK DATA STRUCTURE (ПРОВЕРИТЬ СТРУКТУРУ ДАННЫХ)

Эта опция проверяет структуру файла данных кластера путем выполнения прикладной программы *check.bch*. После проверки структуры файла данных *check.bch* создает файл результата (автоматически появляющийся на экране), который суммирует количество вопросников каждого типа и показывает, сколько вопросников было заполнено. Руководитель обработки данных должен сверить эту информацию с формуляром отслеживания кластеров и убедиться, что данные обоих источников совпадают. Если они не совпадают, руководитель обработки данных должен выявить проблему (например, оператор ввода данных забыл ввести домашнее хозяйство) и решить ее, внимательно проверив вопросники по кластеру. После устранения этой проблемы (либо путем обновления файла данных, либо путем обновления формуляра отслеживания кластеров) руководитель обработки данных должен вновь запустить *check.bch*. Только когда *check.bch* выдаст то же количество вопросников, что и формуляр отслеживания кластера, руководитель обработки данных может распределить вопросники второму оператору ввода данных для ввода данных проверки.

ОПЦИЯ B. VERIFY THE DATA (СВЕРИТЬ ДАННЫЕ)

Эта опция сопоставляет файл основных данных кластера с его файлом данных проверки, используя инструмент сравнения программы CSPro. Если имеются какие-либо расхождения между файлами данных, они будут отражены на экране. Эти результаты должны быть распечатаны и переданы операторам ввода данных, ответственным за ввод кластера. Работая вместе, операторы ввода данных должны обращаться к вопросникам и определять правильное значение в каждом случае, когда их файлы данных не совпадают. После того как операторы определяют правильные значения, каждый оператор должен обновить свой файл данных. После этого вновь следует сопоставить файлы. Если между этими двумя файлами более не обнаруживается расхождений, можно продолжать обработку кластера.

ОПЦИЯ C. BACKUP THE RAW DATA (СОЗДАТЬ РЕЗЕРВНУЮ КОПИЮ НЕОБРАБОТАННЫХ ДАННЫХ)

Эта опция создает резервную копию необработанных данных путем копирования выверенного файла основных данных в подкаталог *backup* на компьютере руководителя. Это следует сделать

после завершения проверки и выверки структуры и до выполнения какого бы то ни было редактирования второго уровня. Необработанные данные копируются, с тем чтобы документировать состояние данных до их редакции.

ОПЦИЯ D. RUN SECONDARY EDITING PROGRAM (ЗАПУСТИТЬ ПРОГРАММУ РЕДАКТИРОВАНИЯ ВТОРОГО УРОВНЯ)

Эта опция осуществляет проверку на комплексные несоответствия путем выполнения прикладной программы *editing.bch*. Результаты работы этой программы выводятся на экран и должны быть распечатаны, если они содержат какие-либо сообщения об ошибке. В таком случае редактору, осуществляющему проверку второго уровня, должны быть переданы перечень несоответствий и вопросники по данному кластеру. Редактор второго уровня, используя свои знания вопросника и пособие по редактированию в Дополнении 7, устраняет каждое из внесенных в перечень несоответствий. Когда редакторы второго уровня завершают свою работу, они передают перечень несоответствий и необходимых действий руководителю обработки данных, который вносит требуемые редакторами изменения (см. опцию E ниже). Затем руководитель обработки данных вновь запускает *editing.bch*. Если сообщений об ошибке более не выводится, можно продолжать обработку кластера; если сообщения об ошибке появляются, их перечень и вопросники должны быть переданы редактору второго уровня для дальнейшего редактирования. В некоторых случаях редактор второго уровня считает сообщения приемлемыми, и дальнейшей необходимости исправлять данные не возникает. Когда остаются только те сообщения, которые редактор второго уровня считает приемлемыми, процесс редактирования второго уровня завершен.

ОПЦИЯ E. MODIFY THE DATA (ИЗМЕНИТЬ ДАННЫЕ)

Эта опция выполняет прикладную программу *entry.ent*, с тем чтобы руководитель обработки данных мог внести изменения, требуемые редакторами второго уровня. После использования этой опции руководитель обработки данных должен вернуться к опции D, с тем чтобы удостовериться в устранении желаемым образом несоответствий после внесения изменений в данные и в отсутствии новых несоответствий.

ОПЦИЯ F. BACKUP THE FINAL DATA (СОЗДАТЬ РЕЗЕРВНУЮ КОПИЮ ОКОНЧАТЕЛЬНЫХ ДАННЫХ)

Эта опция копирует файлы окончательных данных в подкаталог *final* на компьютере руководителя обработки данных. Хранящиеся в этом каталоге данные будут позднее объединены и затем экспортированы в SPSS.

ОПЦИЯ G. EXPORT THE DATA TO SPSS (ЭКСПОРТИРОВАТЬ ДАННЫЕ В SPSS)

Эта опция соединяет все файлы данных в подкаталоге *final* в один файл и затем экспортирует этот файл данных путем выполнения программы *export.bch*. Эта прикладная программа создает четыре текстовых файла ASCII и соответствующий файл описаний SPSS для каждого текстового файла. Один текстовый файл содержит данные по домашним хозяйствам, один – по членам домашних хозяйств, один – по женщинам и еще один – по детям.

Опция H. ENTER GPS DATA (ВВЕСТИ ДАННЫЕ ГПС)

Эта опция позволяет руководителю обработки данных вводить данные ГПС о местоположении путем выполнения программы *gpsentry.ent*. В отличие от программы ввода основных данных эта программа позволяет руководителю обработки данных вводить столько кластеров одновременно, сколько он/она пожелает. Эта прикладная программа требует, чтобы руководитель обработки данных вводил данные ГПС дважды в качестве меры предосторожности от ошибок ввода.

Опция I. MODIFY GPS DATA (ИЗМЕНИТЬ ДАННЫЕ ГПС)

Эта опция позволяет руководителю обработки данных изменять данные ГПС о местоположении путем запуска прикладной программы *gpsentry.ent*.

Опция V. VIEW CLUSTER TRACKING INFORMATION (ПРОСМОТРЕТЬ ИНФОРМАЦИЮ ОБ ОТСЛЕЖИВАНИИ КЛАСТЕРОВ)

Эта опция выводит на дисплей хранящуюся в формуляре отслеживания кластеров информацию для всех кластеров.

Опция N. SELECT NEW CLUSTER (ВЫБРАТЬ НОВЫЙ КЛАСТЕР)

Эта опция позволяет изменить номер кластера, с тем чтобы руководитель обработки данных мог начать обработку нового кластера.

СТРУКТУРА КАТАЛОГОВ В КОМПЬЮТЕРЕ РУКОВОДИТЕЛЯ ОБРАБОТКИ ДАННЫХ

Система обработки данных МИКС использует особую структуру каталогов как в компьютерах ввода данных, так и в компьютере руководителя обработки данных. Структура компьютеров ввода данных рассматривается ниже. В компьютере руководителя все связанные с SPSS файлы и программы хранятся в каталоге *c:\mics\spss* или в одном из его подкаталогов. Все связанные с CSPro файлы и программы в компьютере руководителя хранятся в каталоге *c:\mics\CSPro* или в одном из его подкаталогов. Подкаталоги имеют имена *backup, dicts, entry, export, gps, final, raw, super* и *weights*.

Структура каталогов компьютера руководителя:

MICS		
	CSPRO	
	BACKUP	Резервный каталог
	DICTS	Каталог словарей данных
	ENTRY	Каталог программы ввода данных
	EXPORT	Каталог прикладной программы экспорта
	FINAL	Каталог окончательно отредактированных данных
	GPS	Каталог программы ввода данных ГПС
	RAW	Каталог необработанных данных
	SUPER	Каталог программы редактирования руководителем
	WEIGHTS	Каталог программы весов выборки

Каталог *backup* содержит резервную копию файлов данных, которые были структурно проверены и выверены, но не отредактированы. Каталог *dicts* содержит все словари данных. Каталог *entry* содержит программу ввода данных и программу, которая создает меню ввода данных. Каталог *export* содержит программы, используемые для экспорта данных из CSPro в SPSS. Каталог *final* содержит резервную копию файлов данных, которые были структурно проверены, выверены и отредактированы. Каталог *raw* содержит файлы данных, которые были переданы из машин ввода данных. Каталог *super* содержит программы, которые выполняют структурные проверки, выверку и редактирование второго уровня, и программу, которая создает меню руководителя. Каталог *weights* содержит электронную таблицу, которая рассчитывает веса выборки.

ПРОВЕРКА СТРУКТУРЫ

Важно, чтобы структура данных не была нарушена. Программа ввода данных устанавливает жесткую структурную согласованность, но она не может проверить все, не испытывая серьезных перезагрузок. Поэтому необходимо запускать программу проверки структуры, после того как завершен ввод основных данных. Программа проверки структуры контролирует совпадение количества вопросников в файле данных и количества вопросников, которые поступили с мест, а также выполняет несколько дополнительных проверок структуры индивидуального вопросника.

Программа проверки структуры называется *check.bch*. В настоящем разделе акцент делается на то, что делает программа, а не на то, как она это делает. Это сложная программа, работу которой нельзя объяснить "на пальцах". Наилучший способ понять логику в этой программе – внимательно изучить ее, после того как вы поняли, что делает программа. Пример основных результатов выполнения *check.bch* в случае обнаружения несоответствия показан ниже.

Проверка структуры данных МИКС
Кластер: 3

Домашние хозяйства			Женщины				Дети			
Общее			Отвечают критериям		Опрошенные		Отвечают критериям		Опрошенные	
число	Заверш.	Не заверш.	НН12	ВВЕДЕНО	НН13	ВВЕДЕНО	НН14	ВВЕДЕНО	НН15	ВВЕДЕНО
2	1	1	5	5	4	4	4	4	3	3

Первый блок результата – это итог общего количества домашних хозяйств и их кодов ответов. Второй блок показывает результаты четырех подсчетов числа отвечающих критериям женщин. Подсчеты в столбцах *НН12* и *НН13* – число отвечающих критериям и опрошенных женщин согласно модулю информации о домохозяйствах. Числа в столбцах *found* являются, соответственно, количеством вопросников для женщин и количеством заполненных вопросников для женщин в файле данных.

Третий блок показывает результаты четырех подсчетов числа отвечающих критериям детей в возрасте до 5 лет. Подсчеты в столбцах *НН14* и *НН15* – число отвечающих критериям и опрошенных детей согласно модулю информации о домохозяйстве. Числа в столбцах *found* являются, соответственно, количеством вопросников для детей и количеством заполненных вопросников по детям в файле данных.

Результаты работы *check.bch* должны быть распечатаны руководителем, а содержащуюся в них информацию следует сопоставить с данными формуляра отслеживания кластеров. Если существует расхождение между двумя подсчетами вопросников, руководитель и оператор ввода данных должны использовать перечень ошибок и вопросники кластера, с тем чтобы определить, что вызвало нарушение структуры. После того как эти причины установлены, они должны быть устранены оператором ввода данных. Затем программа проверки структуры должна быть запущена вновь, чтобы удостовериться, что при решении проблемы не было допущено новой ошибки. Только когда произведенные *check.bch* подсчеты совпадут с цифрами в формуляре отслеживания кластеров, можно начать ввод данных проверки.

Программа *check.bch* также генерирует перечень всех домашних хозяйств в кластере. На дисплей выводятся номер каждого домашнего хозяйства и его итоговый код вместе с подсчитанным числом вопросников для женщин и отвечающих критериям детей, если опрос домашнего хозяйства завершен (т. е. итоговый код домашнего хозяйства равняется 1). Образец вывода результатов по одному домашнему хозяйству показан ниже.

```
Сверка структуры данных МИКС
Домашнее хозяйство: 1
Результат: 1
```

Женщины				Дети			
Отвечающие	критериям	Опрошенные		Отвечающие	критериям	Опрошенные	
nn12	ВВЕДЕНО	nn13	ВВЕДЕНО	nn14	ВВЕДЕНО	nn15	ВВЕДЕНО
4	4	3	3	2	2	1	1

Перечень домашних хозяйств может пригодиться для установления причины ошибки на уровне кластера. Допустим, например, что в таблице отслеживания кластеров в данном кластере указано 20 домашних хозяйств, но в файл данных введено только 19. Сопоставляя перечень домашних хозяйств с вопросниками кластера, вы можете установить, какое домашнее хозяйство не было введено.

СВЕРКА

Сверка данных, введенных дважды, производится программой сравнения CSPro. Программа сравнения называется *compare.cmp*. Она содержит перечень всех переменных (элементов). Согласно текущей конфигурации программы, во время сверки будут сопоставляться проверенные элементы, а непроверенные элементы во время сверки сопоставляться не будут. Непроверенной является только одна переменная (код оператора ввода данных, *HH16*) и рекомендуется, чтобы *никакие другие переменные* не оставались непроверенными, поскольку расхождения в других переменных могут снизить качество данных.

Программа сравнения сопоставляет файл ввода основных данных (который был скопирован на компьютер руководителя) с файлом ввода данных проверки (который был скопирован на дискету (или сетевой диск, если используется сеть) и выдает перечень расхождений, если таковые обнаружены. Если расхождений нет, руководитель должен создать резервную копию необработанных данных и затем перейти к редактированию второго уровня.

Если обнаружены расхождения, их перечень должен быть распечатан и передан обоим операторам ввода данных. Затем операторы ввода данных используют перечень расхождений и вопросники кластера, с тем чтобы проверить каждое расхождение и зарегистрировать в перечне, какой из файлов нуждается в корректировке. Когда все расхождения проанализированы, операторы ввода данных исправляют все ошибки в своих файлах. Затем они вновь копируют файлы данных на свои дискеты (или по сети), и файлы снова сравниваются. Этот процесс продолжается до полной идентичности всех файлов.

РЕДАКТИРОВАНИЕ ВТОРОГО УРОВНЯ

Опыт показывает, что простые несоответствия можно легко выявить и скорректировать во время ввода данных. Однако более сложные случаи несогласованности должны разрешаться путем тщательного изучения вопросника. Этот вид проверки на соответствие лучше всего выполнить в качестве отдельного шага, с распечаткой ошибок, которую можно использовать для отметки исправлений. Этот шаг известен как редактирование второго уровня.

Программа редактирования второго уровня называется *editing.bch*. Она генерирует длинный перечень проверок на соответствия (например, соответствуют ли значения возраста и даты рождения) и выводит перечень несоответствий, обнаруженных в файле необработанных данных. Руководитель обработки данных должен распечатать этот перечень и передать его и вопросники кластера одному из редакторов второго уровня. Редактор второго уровня анализирует перечень ошибок и ответы в вопроснике. Руководствуясь принципами редактирования (см. Дополнение 7) и на основании своего знания вопросника редактор второго уровня либо записывает в перечень ошибок, что должно быть исправлено, либо записывает, что не следует предпринимать никаких действий. После анализа всех без исключения сообщений об ошибке редактор второго уровня возвращает аннотированный перечень ошибок руководителю обработки данных. Затем руководитель обработки данных вносит предлагаемые изменения в необработанные данные. После внесения изменений руководитель обработки данных вновь запускает программу редактирования. Только после того как программа редактирования перестает выдавать сообщения об ошибке, можно продолжить обработку кластера.

Критически важными являются три аспекта процесса редактирования. Во-первых, в отношении каждой ошибки редактор второго уровня должен тщательно проанализировать соответствующий вопросник. Во-вторых, прежде чем выработать решение проблемы, редактор второго уровня всегда должен обращаться к руководству по редактированию. В-третьих, процесс редактирования должен повторяться до устранения всех ошибок.

После выполнения третьей задачи руководитель обработки данных может создавать резервную копию отредактированных данных. Эти данные теперь считаются чистыми и могут использоваться для создания аналитических файлов.

ЭКСПОРТ ДАННЫХ В SPSS

После завершения первичной обработки данных у вас есть файл чистых данных для каждого кластера в вашей выборке. Если первичная обработка данных выполняется в среде CPro, вторичная обработка данных осуществляется в основном в SPSS. Поэтому первым шагом в процессе вторичной обработки данных является преобразование данных из формата CPro в формат SPSS. Это делается с помощью опции меню руководителя "Экспортировать данные в SPSS".

Когда вы выбираете эту опцию, все файлы данных в подкаталоге *final* (т. е. все файлы данных, которые были проверены, сверены и отредактированы) соединяются в единый файл данных с именем *all.dat*. Этот файл данных затем экспортируется в SPSS с помощью прикладной программы *export.bch*. Эта прикладная программа создает четыре файла данных ASCII (*mych.dat*, *myhh.dat*, *myhl.dat* и *mywm.dat*) и четыре файла описания SPSS (*mych.sps*, *myhh.sps*, *myhl.sps* и *mywm.sps*) в каталоге *c:\mics\spss*. Хотя файлы описания данных SPSS будут считывать файлы данных ASCII в SPSS, они их не сохраняют. Для того чтобы задать файлам описания данных сохранение данных в формате SPSS, в конце каждого файла должна быть добавлена команда SPSS

```
save outfile = 'filename.sav'.
```

Слово *filename* следует заменить на *hh*, *hl*, *wm* или *ch*, в зависимости от типа файла данных. Когда эта команда будет соответствующим образом записана и добавлена к каждому файлу описания данных, при выполнении файлов описания данных SPSS будут создаваться файлы данных SPSS *hh.sav*, *hl.sav*, *wm.sav* и *ch.sav*.

СОЗДАНИЕ АНАЛИТИЧЕСКОГО ФАЙЛА

Структура файла данных, которая существует в процессе первичной обработки данных, упрощает процесс ввода данных. Однако эта структура не является оптимальной для анализа собранных данных, поэтому после передачи этих данных в SPSS в качестве первой задачи следует перекодировать переменные, с тем чтобы сделать анализ более простым и эффективным. Эта задача известна как создание аналитического файла. В данном разделе будут подробно описаны действия по созданию аналитических файлов МИКС. Аналитические файлы, которые будут созданы в результате этих действий, могут использоваться для типовых планов табулирования, и они удобны для передачи исследователям.

ПЕРЕКОДИРОВАНИЕ ПЕРЕМЕННЫХ

Программы SPSS *makehl.sps*, *makewm.sps* и *makech.sps* выполняют перекодирование существующих переменных для создания новых. В этих программах перекодируются переменные, которые используются в нескольких табуляциях и затем сохраняются на диск; все другое перекодирование производится в программах табулирования и является временным.

Для перекодирования большинства переменных применяются стандартные команды SPSS, и они здесь не рассматриваются. Однако существует один часто применяемый подход, который следует пояснить: перекодирование переменных в 0 или в 100. Это необычное перекодирование производится только для целей представления. Когда программа SPSS отображает проценты в табличной форме, она выводит все категории переменных. В отношении многих таблиц в плане табулирования нас интересует лишь одна категория. Если мы присвоим значение 100 этой категории и значение 0 всем другим категориям, средним значением переменной будет процент респондентов в этой категории. Таким образом, запросив SPSS вывести на дисплей среднее значение новой переменной, мы получим только значение в процентах, которое хотели бы увидеть.

Например, в переменной *hal* записывается, осведомлена ли женщина о СПИДе. Если женщина слышала о СПИДе, переменная принимает значение 1, если не слышала – значение 2. Нас интересует процент женщин, которые осведомлены о СПИДе. В программе *makewm.sps* переменная *hal* перекодируется в переменную *aids*. Переменная *aids* принимает значение 100, если женщина осведомлена о СПИДе, в ином случае – значение 0. Среднее значение переменной *aids* – это процент женщин, которые информированы о СПИДе. Чтобы понять, почему это так, рассмотрим пример, приведенный ниже.

Женщины, осведомленные о СПИДе	10
Общее число женщин	20
Процент женщин, которые информированы о СПИДе	$10 / 20 * 100 = 50$
Среднее значение переменной AIDS	$(10*100 + 10*0)/20 = 10*100/20 = 10/20*100 = 50$

РАСЧЕТ И ДОБАВЛЕНИЕ ВЕСОВ ВЫБОРКИ

Если для различных регионов (или доменов) на первой стадии отбора применялись отдельные выборочные схемы, национальная выборка не была определена с вероятностью, пропорциональной размеру (ВПР). Это также может произойти, если вы стратифицировали выборку согласно какому-либо другому фактору (например, городская местность/сельская местность или трущобы/не трущобы) и взяли различные доли отбора (пропорции) в каждой страте. Эти выборки не являются самовзвешенными, и вы должны взвешивать свою выборку при производстве национальных оценок. Так, вы должны обеспечить, чтобы каждая отдельная подвыборка – например, каждый отдельный регион (или домен) – вносила только то, что она вносила бы, если бы выборка обследования была определена с помощью ВПР на национальном уровне.

Если ваша выборка не является самовзвешенной, вы должны рассчитать веса выборки и добавить их к вашим аналитическим файлам. Эта задача выполняется с использованием электронной таблицы *weights.xls* и программ SPSS *weights_table.sps*, *weights.sps* и *weights_merge.sps*. Электронная таблица используется для расчета весов выборки. Она имеет две рабочие таблицы: *calculations* (расчеты) и *output* (результаты). Рабочая таблица *calculations* выполняет расчеты. Рабочая таблица *output* содержит только веса выборки и перечень номеров

кластеров; этот формат является удобным для считывания данных в SPSS. Программа *weights_table.sps* генерирует данные, необходимые для расчета весов выборки. Программа *weights_merge.sps* добавляет соответствующие веса выборки к аналитическим файлам. Программа *weights.sps*, которую вы никогда не будете выполнять непосредственно, описывает структуру данных в рабочей таблице результатов.

Процесс расчета весов выборки и добавления их к вашим аналитическим файлам можно разбить на семь шагов:

Шаг 1 – Скорректируйте количество строк в рабочих таблицах *calculations* и *output* таким образом, чтобы в вашем обследовании была одна строка для каждого кластера. После того как вы добавили или убрали строки, обязательно проверьте, не изменили ли вы общее количество строк в рабочей таблице *calculations*, выполняя это действие.

Шаг 2 – Введите веса, которые были встроены в схему выборки, в *weights.xls*. Если веса разнятся по кластерам внутри конкретной страты или домена, вы должны заполнить как столбец доли выборки кластера, так и столбец доли выборки страты (или домена) информацией, предоставленной экспертом по созданию выборки вашего обследования. Если веса разнятся по стратам (или доменам), но не по кластерам в пределах страт (т. е. выборка является самовзвешенной внутри страт или доменов), введите значение 1 в столбец доли выборки кластера и заполните столбец доли выборки страты или домена, используя информацию, предоставленную экспертом по созданию выборки вашего обследования.

Шаг 3 – Обновите определение страт (или доменов) в строках 3–10 программы *weights_table.sps*. Стандартные программы предполагают, что страты образуются всеми сочетаниями местности (т. е. городской и сельской) и региона и что существуют четыре региона (чтобы отражать страты или домены, использованные в вашей выборке, программа должна быть модифицирована).

Шаг 4 – Выполните программу *weights_table.sps*.

Шаг 5 – Скопируйте информацию в таблице и вставьте ее в рабочую таблицу *calculations* программы *weights.xls*. Когда вы завершите это действие, *weights.xls* автоматически рассчитает веса выборки.

Шаг 6 – Сохраните рабочую таблицу *output* в качестве файла со значениями, разделенными запятыми (*.csv), под именем *weights.csv* в каталоге *c:mics\weights*.

Шаг 7 – Выполните программу *weights_merge.sps*. Как только вы завершите седьмое действие, обязательно сверьте результирующий перечень сообщений об ошибке, откройте аналитические файлы и убедитесь, что веса были должным образом слиты.

РАСЧЕТ И ДОБАВЛЕНИЕ ИНДЕКСА БЛАГОСОСТОЯНИЯ

План табулирования МИКС включает индекс благосостояния домашнего хозяйства в качестве одной из основных переменных. Такой индекс благосостояния рассчитывается программой *wealth.sps*, которая создает файл данных *wealth.sav*, содержащий идентификационные

переменные, переменную, содержащую количественный показатель благосостояния каждого домашнего хозяйства, и переменную, содержащую индекс благосостояния каждого домашнего хозяйства. Сначала программа *wealth.sps* выдает плотность распределения всех переменных домашнего хозяйства, связанных с благосостоянием или имуществом. Затем она перекодирует переменные, описывающие домашнее хозяйство и личное имущество, в дихотомические переменные. Затем эта программа использует факторный анализ (в частности, анализ основных компонентов) для расчета количественного показателя благосостояния для каждого домашнего хозяйства. Наконец, она использует количественный показатель благосостояния, чтобы создать квинтили благосостояния домашнего хозяйства (т. е. индекс благосостояния) и затем сохраняет их в файле данных SPSS. Выбор переменных, подлежащих включению в факторный анализ, имеет критическое значение, и его нельзя делать, не справляясь самым внимательным образом с построенной плотностью распределения. Решения о том, какие переменные включать, и способ их перекодирования будут окончательно определены на третьем этапе работы. После расчета индекса благосостояния, запустив программу *wealth_merge.sps*, мы добавим его к аналитическим файлам. Непременно проверьте результирующий перечень на предмет сообщений об ошибках, откройте аналитические файлы и убедитесь, что индекс благосостояния был слит должным образом.

ДОБАВЛЕНИЕ ДАННЫХ ГПС

В рамках некоторых обследований МИКС в ходе работ на местах в отношении кластеров используют данные Глобальной системы определения местоположения (ГПС). Показания ГПС, которые точно определяют местоположение кластеров, могут быть использованы для подключения других географических наборов данных (например, данных об осадках) к набору данных МИКС. Если в вашем обследовании используются показания ГПС, вы захотите подсоединить их в свои аналитические файлы. Эта задача выполняется двумя программами SPSS – *gps.sps* и *gps_merge.sps*. Программа *gps.sps*, которую вы никогда не будете выполнять непосредственно, описывает структуру файла данных *gps.dat* (который создан прикладной программой ввода данных CPro *gpsentry.bch*). Если вы изменили словарь CPro *gps.dic*, вы должны обновить *gps.sps*, чтобы отразить внесенные изменения. Программа *gps_merge.sps* включает показания ГПС в аналитические файлы. Вам не нужно модифицировать эту программу. Для включения данных ГПС в аналитические файлы запустите *gps_merge.sps*. Обязательно проверьте результирующий перечень на предмет сообщений об ошибках, откройте аналитические файлы и убедитесь, что данные ГПС включены надлежащим образом.

ТАБУЛИРОВАНИЕ

В главе 8 подробно описывается процесс анализа данных и подготовки отчетов. Сведение данных в таблицу является главным компонентом этой деятельности. Типовой план табулирования (Глава 8 и Дополнение 6) и программы табулирования для SPSS прилагаются к данному Руководству. Для каждой таблицы в плане табулирования МИКС имеется одна программа табулирования SPSS. Именем каждой программы является буква t, за которой следует номер таблицы в плане табулирования. Например, программа *t1.sps* создает таблицу 1 в плане табулирования.

До создания таблиц для отчета важно создать набор (невзвешенных) плотностей распределения для каждой переменной в файле данных. Эти плотности распределения следует проверить на необычные значения, такие, которые находятся за пределами диапазонов большинства ответов, и такие, которые представляются неправдоподобными ответами на соответствующий вопрос. Например, ответ "53" на вопрос о количестве часов, в течение которых ребенок выполнял работу по дому на прошлой неделе, кажется и завышенным, и слишком точным. Для таких случаев должна быть выписана идентификационная информация, а значения в файле данных сверены с подлинниками вопросников.

ПЕРЕСМОТР ТИПОВЫХ ПРОГРАММ

Каждую программу табулирования следует внимательно проанализировать. Важно проверить, существуют ли в вашем файле данных переменные, используемые в программе табулирования. Если их не существует, проверьте, имеет ли эта переменная первостепенное или второстепенное значение. Если в вашем файле данных нет какой-либо переменной первостепенного значения, вы должны либо полностью убрать таблицу, либо попросить аналитика перестроить эту таблицу. Если отсутствует какая-либо переменная второстепенного значения, уберите все отсылки к этой переменной в программе табулирования и выполните все другие корректировки, которые ее отсутствие делает необходимыми.

Вся деятельность по перекодированию также подлежит тщательной проверке. Если в вашем вопроснике имеются переменные, которые относятся к нестандартным категориям, должны быть изучены любые шаги по перекодированию, связанные с этими переменными. Если ваш вопросник содержит нестандартные переменные, они должны быть перекодированы в тех случаях, когда планируется их использование в каком-либо табулировании.

Вы должны также проверять все операции по слиянию, если в вашем вопроснике используются идентификаторы, отсутствующие в стандартном вопроснике. Существует ряд слияний в программе табулирования, которые будут выполняться только при использовании уникальных идентификаторов.

ПРИМЕНЕНИЕ ВЕСОВ ВЫБОРКИ

Взвешивание в программах табулирования является непосредственным, кроме тех случаев, когда задействована команда SPSS *aggregate*. Если целью команды *aggregate* является накопление случаев для расчета числителя и знаменателя, веса должны применяться до команды *aggregate*. Их не следует применять во время работы с результирующим файлом; он уже был взвешен.

Например, таблица НН.1 содержит коэффициент ответов домашних хозяйств. Коэффициент ответов домашнего хозяйства трудно рассчитать, потому что для этого требуется разделить одну переменную на другую внутри таблицы. Одним из решений этой проблемы является создание агрегированного файла, который содержит количество домашних хозяйств в выборке заселенных домашних хозяйств и опрошенных домашних хозяйств. Агрегированный файл будет содержать один случай для каждой категории определенной переменной "разрыва" (например, городской район/сельский район).

Весы должны применяться при создании агрегированного файла для генерирования взвешенного числителя (количество опрошенных домашних хозяйств) и взвешенного знаменателя (количество заселенных домохозяйств). После создания агрегированного файла коэффициент ответов домашнего хозяйства для каждой категории переменной "разрыва" – это числитель, деленный на знаменатель.

Если целью команды *aggregate* является создание общей статистики по индивидуальным блокам, веса должны применяться после команды *aggregate*. Например, таблица 4 содержит информацию о проценте домашних хозяйств, в которых есть по крайней мере один ребенок в возрасте до 15 лет.

Эта информация не содержится в файле данных домашнего хозяйства, но ее можно создать с помощью агрегирования файла списка проживающих в домашнем хозяйстве. Переменные "разрыва" являются номером кластера и номером домашнего хозяйства. Весы применяются после агрегирования, потому что мы интересуемся взвешенным процентом домашних хозяйств, имеющих по крайней мере одного ребенка в возрасте до 15 лет, а не взвешенным числом детей в возрасте до 15 лет в каждом хозяйстве.

КОМАНДА INCLUDE (ВКЛЮЧИТЬ)

Программа SPSS *tables.sps* одновременно запускает все программы табулирования. Она состоит из ряда команд SPSS *include* (включить), которые выполняют программы табулирования индивидуально. Если SPSS обнаружит ошибку во включенной программе (т. е. в запущенной командой *include*), она немедленно прекратит выполнение этой программы и вернется к той программе, которая включала указанную программу (т. е. к программе, которая содержала команду *include*).

Ввиду этого вам следует использовать *tables.sps* только после того, как вы проверили, изменили и протестировали все индивидуальные программы табулирования. Обязательно удалите также команду *include*, выполняющую программу табулирования, которой вы не пользуетесь.

Команда *include* налагает четыре ограничения на программы, которые она выполняет. Первое ограничение состоит в том, что каждая команда должна начинаться в первом столбце этой программы. Это ограничение появляется, чтобы установить предел отступа в программе, но строка может иметь отступ, если она начинается со знака "+". Приведенные ниже команды иллюстрируют применение знака "+".

```
do if (cage >= 6 and cage < =9)
+ compute solids = 0.
+ if (BF3G = 1) solids = 100.
end if.
variable labels solids "Solid foods".
```

Второе ограничение, налагаемое командой *include*, состоит в том, что, если команда занимает несколько строк, столбец 1 продолжающихся строк должен быть пустым.

Нижеприведенный пример иллюстрирует многострочную команду, в которой соблюдено это ограничение.

```
add files
  /file=*
  /file='tmp6.sav'.
```

Заметьте, что подкоманды на второй и третьей строках сдвинуты на два столбца. (Хотя, чтобы соблюсти ограничение, их нужно было сдвинуть вправо только на один столбец, но они сдвинуты вправо на два столбца, чтобы сохранить совместимость со стилем программирования МИКС).

Третье и четвертое налагаемые командой *include* ограничения состоят в том, что командные терминаторы являются факультативными и что звездочка (*) в первом столбце строки обозначает строку комментариев. Ни одно из этих ограничений не влияет на наши программы табулирования.

В дополнение к *tables.sps* существует программа SPSS, которая автоматизирует создание аналитических файлов. Эта программа называется *CSPro.sps*. Ее следует использовать только после того, как были запущены все составные программы и они оказались работоспособными. Эта программа полезна для обновления аналитических файлов после внесения изменения в одну из программ создания файла. Это обеспечивает выполнение всех программ создания аналитических файлов в надлежащем порядке.

АРХИВИРОВАНИЕ И РАСПРОСТРАНЕНИЕ ДАННЫХ

Важным компонентом обработки данных, которому зачастую не уделяется достаточного внимания, является архивирование и документирование файлов данных. Кроме того, крайне важно заблаговременно установить некоторые руководящие принципы распространения данных независимо от того, будут ли файлы данных доступными для широкой аудитории или только в пределах отдельно взятого учреждения. Эти меры – архивирование, документирование и распространение – требуют временных и трудовых затрат. Однако эти затраты оправданы в силу ряда причин:

- **Повышение экономической эффективности сбора данных.** Сбор данных обследования является дорогостоящей и трудоемкой задачей. Чтобы оправдать эти затраты, следует как можно шире использовать собранные данные. Доступность файлов данных для других исследователей повысит экономическую эффективность деятельности по проведению обследований.
- **Расширение прав собственности страны на данные и признание результатов обследований.** Если файл данных доступен для использования другими лицами, повышается доверие к процессу сбора данных. Сборщики данных воспринимают как людей, уверенных в полученных результатах, а доступность файла данных для других исследователей означает, что результаты могут быть воспроизведены и проверены другими.

- **Способность исследовать тенденции.** Зачастую опубликованные результаты различных обследований непосредственно не сопоставимы. Например, в одном отчете об обследовании взрослые респонденты могут определяться как лица в возрасте 15 лет и старше, в то время как в другом взрослые определяются как лица в возрасте 18 лет и старше. Не имея файлов данных, все, что можно сделать – провести неточное сравнение этих двух наборов результатов. Однако когда файлы данных двух обследований доступны, результаты, как правило, могут быть перетабулированы, и, таким образом, они станут непосредственно сопоставимыми, что позволит сделать выводы относительно тенденций.
- **Возможность сравнения между странами или внутри стран.** Часто полезно сопоставлять результаты между странами – в пределах подрегиона или между регионами. Такие сравнения способствуют выявлению районов, где необходимо уделить внимание конкретным программам или где программы оказались особенно успешными. Далее, может быть полезным сравнение результатов различных обследований внутри одной страны. Иногда это делается, с тем чтобы подтвердить достоверность неожиданных результатов (например, младенческая смертность ниже ожидаемой) или чтобы оценить эффективность конкретной методологии сбора данных (например, карты вакцинации в сравнении с сообщениями матерей о вакцинации). Для проведения таких видов анализа исследователям требуется доступ к файлам данных, с тем чтобы можно было рассчитать непосредственно сопоставимые значения.
- **Возможность проведения углубленного анализа в конкретных областях специалистами в какой-либо сфере.** Вследствие необходимости оперативного представления результатов обследования содержащаяся в отчетах информация включает только основные выводы обследования. Хорошо документированный и доступный файл данных позволит провести углубленный анализ по конкретным тематическим областям, и этот анализ может быть проведен специалистами в какой-либо сфере, которых может не быть среди персонала учреждения по сбору данных.

Аналитические файлы МИКС подлежат архивированию, документированию и распространению. Сопровождающая аналитические файлы документация должна включать, по меньшей мере, копию полного отчета, копию вопросника и описание схемы выборки. Документация должна также включать кодировочную книгу, содержащую данные о местоположении и описание каждой переменной в аналитическом файле (книгу легко можно создать в SPSS). Копии всех программ и файлов, применявшихся в ходе обработки данных обследования, также должны быть архивированы и представляться по запросу. Копия аналитических файлов и их документация должны быть направлены в Региональное отделение ЮНИСЕФ и в штаб-квартиру ЮНИСЕФ в Нью-Йорке (Секция статистики и контроля). Наконец, должны быть установлены практика и процедура распространения файла данных среди других получателей.

КОНТРОЛЬНЫЕ ПЕРЕЧНИ

До начала работы на местах:

- получить компьютеры и другое оборудование для обработки данных;
- оборудовать комнату или помещение для обработки данных;
- нанять руководителя обработки данных и прочий персонал;
- установить систему организации деятельности по обработке данных;
- адаптировать программы для соответствия результатам предварительных испытаний вопросника;
- ввести и отредактировать вопросники после предварительных испытаний;
- завершить адаптацию программ в соответствии с результатами предварительных испытаний и окончательным вариантом вопросника.

В ходе работы на местах:

- получать вопросники с мест;
- распределить обязанности по вводу основных данных;
- проверить структуру файла ввода основных данных;
- распределить обязанности по вводу данных проверки;
- сверить файлы основных данных и данных проверки на идентичность;
- создать резервную копию файла необработанных данных;
- выполнить редактирование второго уровня;
- создать резервную копию окончательного файла данных.

По окончании работы на местах:

- экспортировать данные в SPSS;
- перекодировать переменные;
- рассчитать и добавить веса выборки, индекс благосостояния и данные ГПС;
- запустить программы табулирования;
- архивировать данные и разработать практику и систему распространения данных (например, веб-сайт);
- отправить аналитические файлы, документацию на них и все программы в ЮНИСЕФ.

Формуляр отслеживания кластеров

Номер кластера	Дата получения	Отобранные домо-хозяйства	Количество опрошенных			Ввод данных			Дата проведения проверки	Дата проведения сверки	Дата создания резервной копии файла неотреботанных данных	Дата резервного копирования окончательных данных
			Домо-хозяйства	Женщины	Дети	Фамилия оператора	Номер оператора	Дата создания				
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						
						M:						
						V:						

M = ввод основных данных
V = ввод данных проверки