

ОСНОВЫ ЕРІ INFO

РУКОВОДСТВО ДЛЯ ЭПИДЕМИОЛОГОВ



Санкт-Петербург
2001

Учебное пособие «Основы Epi Info. Руководство для эпидемиологов» подготовлено сотрудниками кафедры эпидемиологии СПбГМА (зав.кафедрой профессор Л.П.Зуева) доцентом С.Р. Ереминым и к.м.н. А.В. Любимовой. Пособие предназначено для обучения медицинских работников различных специальностей, занимающихся обработкой и анализом эпидемиологических данных.

ОСНОВЫ EPI INFO: РУКОВОДСТВО ДЛЯ ЭПИДЕМИОЛОГОВ

С.Р.Еремин, А.В.Любимова

Содержание:

Содержание:	1
<i>Список иллюстраций:</i>	2
Предисловие	3
<i>Введение в Epi Info</i>	3
Компьютерные программы для эпидемиологов	3
Почему Epi Info?	4
Epi Info 6 или Epi Info 2000?	4
Где взять Epi Info?.....	5
<i>Структура Epi Info</i>	5
Основные модули	5
Вспомогательные модули	6
Дополнительные модули	6
<i>Первые шаги</i>	6
Инсталляция <i>Epi Info</i>	6
Начало работы	6
<i>Создание вопросников. Программа EPED</i>	7
Текстовый редактор EPED	7
Создание вопросников.....	9
Типы полей данных в Epi Info	10
Создание вопросника в редакторе EPED.....	12
<i>Ввод данных. Программа ENTER</i>	15
Создание файлов данных	15
Ввод данных	16
Команды редактирования в программе ENTER:	17
Последствия изменения QES-файла	18
<i>Анализ данных. Программа ANALYSIS</i>	19
Общие сведения о программе ANALYSIS. Команда READ.	19
Основные команды анализа данных	21
LIST и BROWSE.....	22
FREQ.....	23
TABLES	24
MEANS и DESCRIBE.....	26
REGRESS	28
Графическое представление данных	28
Работа с переменными.....	29
Создание новых переменных. Команда DEFINE.....	29
Вычисляемые переменные.....	29
Операторы IF-THEN.....	30
Команда RECODE	31
Команда SELECT	32
Сохранение данных в новом файле	34
Программы	34
Связанные базы данных. Команда RELATE.....	35
Сохранение результатов анализа	37

Наиболее часто применяемые арифметические и логические операторы.....	38
Управление вводом данных. Программа CHECK.....	38
Основные команды программы CHECK.....	41
Задание диапазона допустимых значений.....	41
Автоматический ввод значения предыдущей записи.....	41
Обязательный ввод данных.....	41
Переходы.....	41
Допустимые значения.....	42
CHK-файлы: сохранение, редактирование.....	42
Статистический калькулятор STATCALC.....	43
Импорт и экспорт данных.....	45
Литература:.....	46
Список иллюстраций:	
Рис. 1. Главное окно программы Epi Info.....	7
Рис. 2. Текстовый редактор EPED.....	8
Рис. 3. Список форматов полей данных в редакторе EPED.....	10
Рис. 4. Текст вопросника в редакторе EPED.....	14
Рис. 5. Стартовый экран программы ENTER.....	15
Рис. 6. Форма ввода данных в программе ENTER.....	16
Рис. 7. Пример заполненной формы ввода данных.....	18
Рис. 8. Стартовый экран программы ANALYSIS.....	20
Рис. 9. Список основных команд программы ANALYSIS.....	20
Рис. 10. Экран программы ANALYSIS после выполнения команды READ.....	21
Рис. 11. Список переменных, содержащихся в файле oswego.rec: результат нажатия F3.....	22
Рис. 12. Экран программы ANALYSIS после выполнения команды BROWSE.....	22
Рис. 13. Результат выполнения команды PIE.....	29
Рис. 14. Окно команды ANALYSIS после выполнения команды SELECT.....	33
Рис. 15. Стартовый экран программы CHECK.....	39
Рис. 16. Экран программы CHECK при работе с открытым файлом данных.....	40
Рис. 17. Экран программы ENTER со списком допустимых значений поля.....	42
Рис. 18. Главный экран программы STATCALC.....	43
Рис. 19. Окно ввода табличных данных программы STATCALC.....	44
Рис. 20. Главный экран программы IMPORT.....	45
Рис. 21. Главный экран программы EXPORT.....	46

Предисловие

Данное руководство предназначено для начинающих пользователей Epi Info и не претендует на исчерпывающую полноту. Подробные сведения по работе с программой Epi Info 6 можно найти в официальном 600-страничном руководстве пользователя. Список литературы для самостоятельного изучения Epi Info предлагается в конце этой книги.

Введение в Epi Info

Компьютерные программы для эпидемиологов

В XXI веке эпидемиологу не обойтись без компьютера и соответствующего программного обеспечения, прежде всего для создания баз данных и статистического анализа.

Среди прикладных компьютерных программ, позволяющих создавать и вести базы данных, следует назвать в первую очередь такие компоненты ведущих офисных пакетов, как СУБД Microsoft Access и электронная таблица Excel, входящие в состав Microsoft Office (пожалуй, наиболее популярный в России среди конечных пользователей программный продукт), Paradox и Quattro Pro в составе Corel WordPerfect Office, а также Lotus Approach и Lotus 1-2-3 (Lotus Smart Suite). Перечисленные пакеты универсальны, обеспечивают графическое представление данных, а электронные таблицы предоставляют широкие возможности для статистической обработки данных¹[]. Кроме этого, на практике широко используются СУБД и АРМы, разработанные профессиональными программистами в различных средах для решения конкретных задач.

Перечень специализированных компьютерных программ по биостатистике слишком длинен, чтобы привести его здесь. Наряду с широко известными «монстрами» статистической обработки данных (SAS, SPSS, Statistica, NCSS, Systat, Statgraphics, Minitab, и т.д.), стоимость которых достигает иногда нескольких тысяч долларов, имеются неплохие пакеты, распространяющиеся по более скромным ценам (прежде всего, Stata), а также целый набор условно-бесплатных (shareware) и бесплатных (freeware) программ². Существует даже целый ряд Web-сайтов, обеспечивающих возможность выполнения биостатистической обработки данных on-line.

Особого внимания заслуживает пакет программ статистической обработки эпидемиологических данных EPI (распространяется бесплатно), включающий десятки полезных (в том числе довольно редких) программ. Этот богатый по возможностям пакет особенно ценен при проведении исследований в области клинической эпидемиологии. Поддерживает работу с файлами в формате .REC (Epi Info). Единственным недостатком является разрозненность отдельных модулей в бесплатной версии. Последняя полноценная версия (4.0) работает под DOS, однако в настоящее время тестируется Windows-версия (программы COMPARE2 и WHATIS³).

¹ Неплохое руководство по применению Excel для статистического анализа можно найти, например, в книге «Прикладная медицинская статистика» под ред. В.М.Зайцева и В.Г.Лифляндского[1]

² Подробные списки статистических пакетов с описанием и указанием сайтов разработчиков можно отыскать, например, на www.doktor.ru/doctor/biometr/list/statleo.htm (здесь же большая коллекция ссылок на статистические ресурсы Интернет) или www.statpages.net.

³ Эти программы можно скачать для тестирования на www.myatt.demon.co.uk

Почему Epi Info?

Всесторонний анализ доступного программного обеспечения и опыт работы с ним показали, что наиболее рациональным выбором для эпидемиологов являются пакеты программ Epi Info, сочетающие в себе возможности СУБД и специализированного текстового процессора, обеспечивающие графическое представление данных и возможность применения широкого спектра биостатистических методов (анализ таблиц 2x2 и 2xN, дескриптивная статистика, дисперсионный анализ, регрессионный анализ и т.д.).

Epi Info разрабатывается и поддерживается Центрами по контролю и предупреждению болезней⁴ и Всемирной Организацией Здравоохранения. В 1999 г. Harbage и Dean подсчитали, что со времени появления в 1985г. первой версии Epi Info было распространено как минимум 145320 копий программы в 117 странах мира[2]; пакет переведен на 13 языков, включая русский⁵, в разных странах организованы многочисленные пользовательские группы поддержки. Обучение основам Epi Info является обязательным элементом большинства образовательных программ по эпидемиологии во всем мире.

Epi Info справедливо считается стандартом компьютерной программы универсального назначения для эпидемиологов. Ее высокая популярность объясняется целым рядом явных достоинств, к которым, наряду с доступностью (бесплатная программа) относятся многообразие функций, высокая степень адаптации к решению эпидемиологических задач, гибкость (возможность настройки без особого труда для выполнения конкретных проектов), наличие опции экспорта/импорта данных, а также возможность полной русификации (Epi Info 2000).

Epi Info 6 или Epi Info 2000?

В настоящее время параллельно существуют два независимых варианта программы: Epi Info 2000 и Epi Info 6.

Epi Info 6[3], работающая под DOS, может устанавливаться практически на любом, самом примитивном IBM-совместимом компьютере. Хотя текущая версия Epi Info переведена на русский язык лишь частично (только главное меню), это обстоятельство, как показывает опыт, не является серьезным препятствием для эффективного использования программы. Надежность программы, интуитивная простота создания баз данных и разнообразие аналитических возможностей позволяют считать ее лучшим выбором для эпидемиолога. Несмотря на появление Windows-версии, программа продолжает развиваться. Последняя на момент издания этой книги версия (6.04d) вышла в 2001г.; появились сообщения о возможности выхода в будущем версии 7 (также для DOS), что в эпоху Windows и Linux может означать лишь признание несомненных заслуг этого программного продукта.

Epi Info 2000⁶ для Windows[4] имеет, наряду с достоинствами графического интерфейса, множество новых функций (добавились множественная логистическая регрессия, анализ кривых выживаемости, картографирование и ряд других). Однако в качестве основного формата файлов данных используется формат Microsoft Access (.mdb); очень похожа на реализованную в Access процедура создания форм для ввода данных, что, по нашему мнению, лишает данную версию программы львиной доли привлекательности ее предшественницы. Программа полностью переведена на русский

⁴ Centers for Disease Control and Prevention (CDC)

⁵ Полностью на русский язык переведены версия 5 для DOS и Epi Info 2000 для Windows версия 1.0.

⁶ Последняя на момент выхода этой книги версия – 1.1

язык, но, несмотря на постоянные обновления, она до сих пор ненадежна в эксплуатации и поэтому не рассматривается в данном руководстве.

Компромиссом между желанием сохранить простоту и изящество DOS-версии и стремлением к использованию достоинств работы в среде Windows является программа Epidata[5]⁷, которая также не обсуждается в настоящем руководстве. Epidata не является полноценной эпидемиологической программой, поскольку не имеет важнейшего компонента – программы анализа, ее основное достоинство – сочетание привычных форматов файлов .QES и .REC с возможностями графического интерфейса Windows.

Где взять Epi Info?

Дистрибутивы Epi Info можно бесплатно скачать с официального сайта программы www.cdc.gov/epiinfo. Установочные копии различных версий Epi Info (в том числе русифицированных) и некоторые дополнительные модули можно также найти на CD-ROM «Рабочее место госпитального эпидемиолога»[6] и одноименном Web-сайте по адресу www.aiha.com/ws.

Структура Epi Info

Epi Info состоит из нескольких компонентов, каждый из которых может запускаться автономно:

Основные модули

- Текстовый процессор (**EPED**). Используется для создания вопросников (анкет), которые применяются для сбора информации, а также служат формами, описывающими структуру базы данных.
- Модуль ввода данных (**ENTER**). Генерирует файл данных (.REC⁸) из вопросника, созданного в текстовом процессоре (QES-файла⁹), и обеспечивает ввод данных в этот файл. Модуль **ENTERX** служит для ввода данных с использованием верхних адресов памяти (более 640K).
- Модуль анализа (**ANALYSIS**). Обеспечивает вывод данных на экран/принтер/в файл, расчет частотных показателей, кросс-табуляцию и многое другое. Кроме файлов в формате Epi Info (.REC), может работать непосредственно с файлами dBase (.DBF). Поддерживается работа с несколькими связанными файлами одновременно. При создании таблиц вычисляются соответствующие эпидемиологические и биостатистические показатели (отношение преобладаний, относительный риск, точные доверительные интервалы, уровни значимости p и т.п.). Программа позволяет проводить линейный регрессионный анализ, однофакторный дисперсионный анализ, стратификационный анализ по *Mantel-Haenszel* и т.д.. Записи в базе данных могут быть отобраны или отсортированы по различным критериям с использованием новых, генерируемых в ходе анализа переменных, и математических и логических операторов. Модуль обеспечивает графическое представление данных, создание отчетов, имеется собственный язык программирования для создания подпрограмм.
- Модуль управления вводом данных (**CHECK**). Позволяет устанавливать допустимые диапазоны и значения данных, проводить автоматическое кодирование,

⁷ www.epidata.dk

⁸ от английского *record* (запись, в данном случае имеется в виду запись в базе данных)

⁹ от английского *questionnaire* (вопросник)

математические и логические операции с полями, обеспечивает доступ к нескольким связанным файлам и ряд других функций.

Вспомогательные модули

- Статистический калькулятор (**STATCALC**). Вычисляет статистические показатели для данных, введенных в табличные формы с клавиатуры. Выполняет расчет соответствующих показателей для одиночных и стратифицированных таблиц 2x2, позволяет рассчитать объем выборки, необходимый для исследования и провести анализ тренда по *Mantel-Haenszel*.
- Эпидемиологический калькулятор (**EPITABLE**). Способен вычислять большое количество эпидемиологических и статистических показателей по данным, введенным в табличные формы с клавиатуры.
- Модуль экспорта данных (**EXPORT**). Экспортирует файлы данных .REC из Epi Info в 12 других форматов, используемых различными базами данных и статистическими программами.
- Модуль импорта данных (**IMPORT**). Позволяет использовать в Epi Info файлы, созданные другими программами.
- Модуль слияния файлов (**MERGE**). Обеспечивает слияние файлов данных, созданных из вопросников, имеющих одинаковый или различный формат. Это позволяет объединять файлы с данными, введенными на разных компьютерах и дополнять старые записи новыми.
- Модуль сравнения файлов (**VALIDATE**). Служит для верификации ввода путем сравнения двух файлов данных, введенных разными операторами из одного и того же источника.

Дополнительные модули

- Модуль анализа кластерных выборок (**CSAMPLE**). Выполняет анализ данных популяционных исследований, имеющих сложный дизайн. Учитывает кластерные выборки, стратификацию, взвешенные переменные
- Калькулятор для вычисления антропометрических показателей (**EPINUT**) - прикладная антропометрическая программа.

Кроме того, в пакет входит гипертекстовое электронное руководство и набор обучающих программ и учебных материалов с примерами.

Первые шаги

Инсталляция *Epi Info*

Инсталляция программы, как правило, не вызывает особых проблем. Дистрибутив Epi Info 6.04d включает 3 самораспаковывающихся архивных файла (epi604_1.exe, epi604_2.exe, epi604_3.exe), которые следует открыть по очереди в отдельной папке. После разархивации нужно запустить программу инсталляции (install.exe) и следовать инструкциям.

Начало работы

После запуска Epi Info появляется главное окно программы, представленное на рис. 1 (так выглядит главное окно русифицированной версии после выбора из меню пункта **Программы**). Верхняя строка экрана содержит основные пункты меню: выбранный

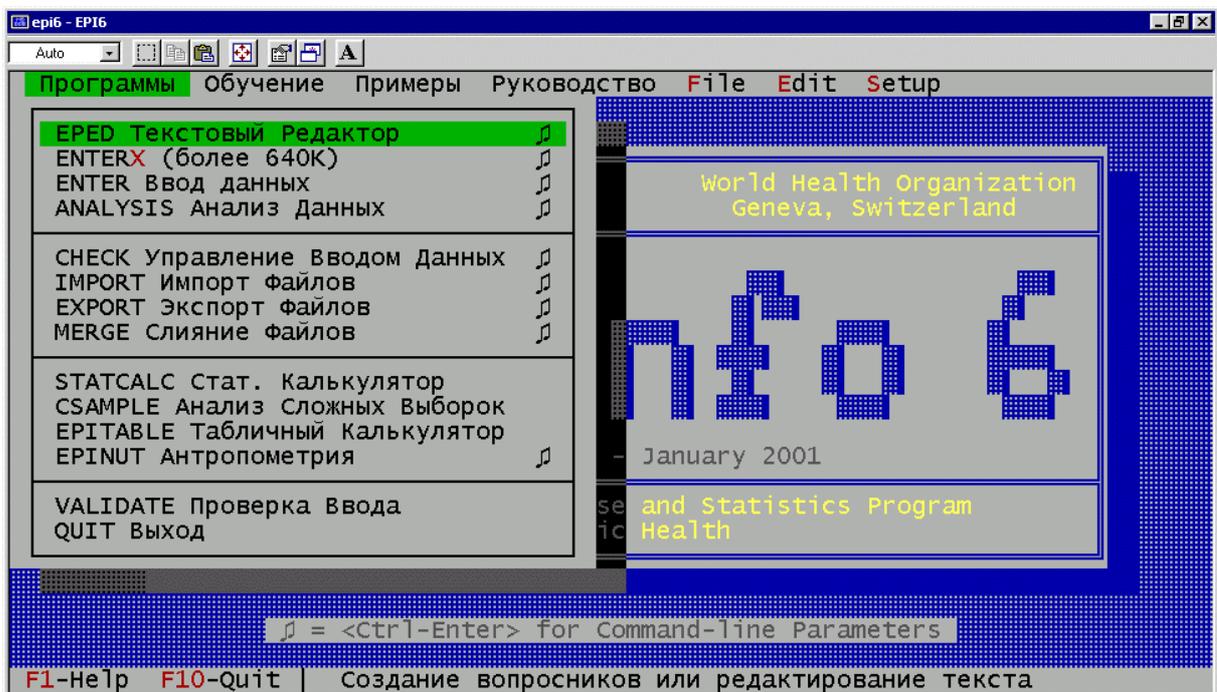
пункт главного меню (**Программы**) и разворачивающегося субменю (**EPED Текстовый Редактор**) на экране компьютера выделены зеленым.

Нужный пункт меню можно выбрать, перемещая курсор с помощью клавиш , , , . О мыши в Epi Info лучше забыть: хотя она и работает в главном окне и некоторых других окнах программы, в большинстве случаев приходится пользоваться клавишами управления курсором и клавишей .

Запуск выбранной программы производится нажатием клавиши . Значок  рядом с пунктом меню означает, что нажав $\langle \text{Ctrl} \rangle + \langle \text{Enter} \rangle$ ¹⁰, можно запустить программу со вводом параметров в командной строке (например, ввести имя файла, с которым Вы собираетесь работать).

Выход из программы (так же как из любого модуля Epi Info) осуществляется путем нажатия функциональной клавиши F10 (**NB!**).

Рис. 1. Главное окно программы Epi Info



В нижней строке любого окна Epi Info всегда можно найти подсказку о назначении функциональных клавиш (**NB!**). В главном окне, кроме того, в правой части нижней строки выводится контекстная подсказка по всем пунктам меню.

Создание вопросников. Программа EPED

Текстовый редактор EPED

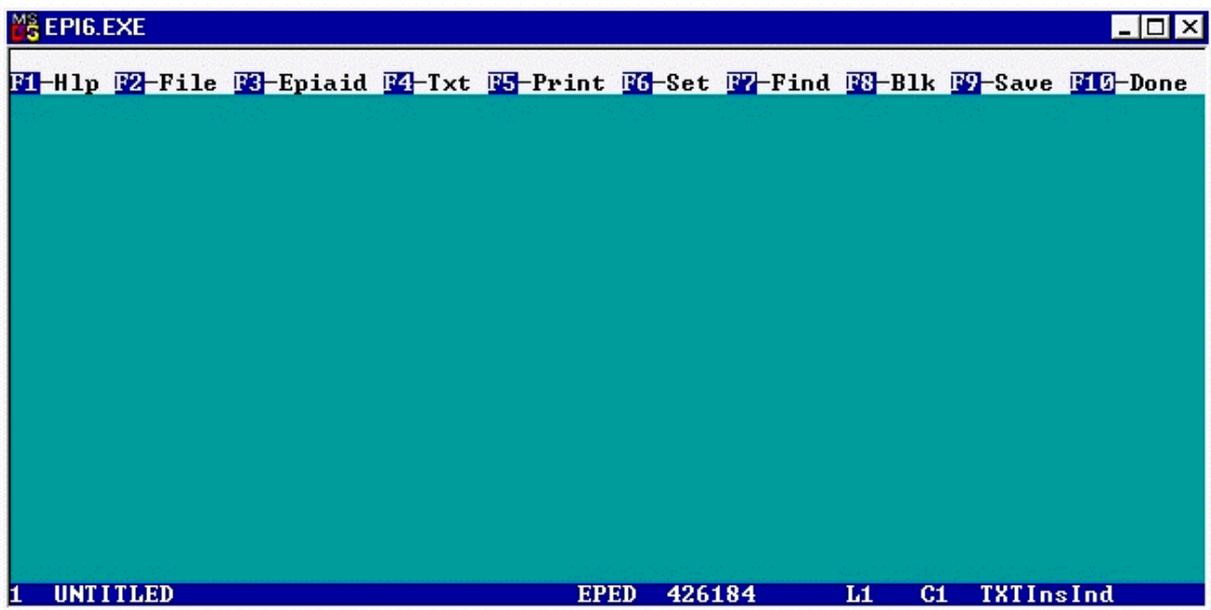
Текстовый редактор EPED позволяет создавать текстовые файлы, прежде всего так называемые вопросники, которые можно распечатать и применить для сбора данных. Заполненные вопросники затем используются для ввода данных в программу ENTER, при этом формат формы ввода данных в точности воспроизводит формат вопросника.

¹⁰ Нажать клавишу , удерживая нажатой клавишу .

Для выполнения различных действий в редакторе EPED имеется система меню, доступ к которому осуществляется с помощью функциональных клавиш F1-F10. Меню и перечень функциональных клавиш расположены в верхней строке экрана (рис.2) Наиболее часто приходится нажимать следующие клавиши (пункты меню):

-  открыть файл (**open file this window**)
-  сохранить файл
-  ВЫХОД

Рис. 2 Текстовый редактор EPED



В нижней строке экрана выводится следующая информация:

«**UNTITLED**» – название документа, который редактируется

«**EPED**» – название программы, которая используется

«**426184**» – Объем доступной памяти. На эту строку можно не обращать особого внимания, вряд ли Вы столкнетесь с ситуацией, когда этот объем станет недопустимо низким.

«**L1**» – Строка текста, на которой расположен курсор

«**C1**» - Колонка текста, на которой расположен курсор

Набор букв справа указывает на текущие установки, изменяемые по нажатию клавиши F6. «**ТХТ**» означает режим ввода текста с неограниченной строкой, который используется, например, для написания компьютерных программ (может быть «**WW**», что означает текстовый режим ввода с переносом строки или «**QES**» – режим создания вопросника). «**Ins**» означает включенный режим вставки (может быть «**Ovr**», что означает режим замены). «**Ind**» означает режим автоматического отступа (может быть «**Jst**», что означает выравнивание по левому/правому краю).

Чтобы избежать утомительной навигации по довольно неудобному и запутанному меню, следует запомнить несколько «горячих» клавиш (комбинации клавиш, дублирующих пункты меню):

^Y – удаление строки¹¹

^Q Y – удаление текста до конца строки

^T – удаление слова

^K B – начало выделения текстового блока

^K K – окончание выделения текстового блока

^K C – копирование выделенного блока в место, на котором установлен курсор

^K M – перемещение блока

^K Y – удаление блока

^Q Q – вызов списка допустимых полей вопросника (об этом чуть ниже).

Создание вопросников

Вопросники Epi Info представляют собой обычные текстовые (ASCII) файлы, обязательно имеющие расширение имени .QES (**NB!**), например, HOSPITAL.QES¹². Помимо того, что напечатанная на бумаге анкета может служить инструментом для сбора данных, вопросники выполняют в программе Epi Info важную функцию: файлы .QES служат шаблонами для создания файлов данных (REC-файлов), которые используются для ввода и хранения данных в электронном виде и их анализа.

В упрощенном виде базу данных можно представить в виде одной или нескольких двумерных таблиц. Каждая строка в такой таблице — конкретная реализация (значение) какого либо понятия (события, предмета изучения) — называется записью базы данных. Столбцы таблицы — свойства (атрибуты) данного понятия. Они называются полями базы данных. Совокупность изменяющихся значений данного поля базы данных (формальное описание атрибута изучаемого объекта, его признака) называется переменной.

Вопросник Epi Info представляет собой список полей, относящихся к одной записи базы данных, например, к одному пациенту. В этом случае для каждого пациента заполняется свой вопросник, который и превращается в запись о данном человеке. Каждая запись при этом может содержать информацию о конкретных значениях множества переменных (имя, возраст, дата рождения и т.п.). Файл .REC в Epi Info – это совокупность записей, введенных с помощью одного и того же шаблона (формы) ввода данных, хранящегося в файле .QES. Для каждого поля задается тип данных, которые могут находиться в нем. Записи в базе данных отличаются значениями своих полей.

Вопросники Epi Info состоят из трех основных элементов:

- Текст – описание содержимого вопросника (формулировки задаваемых вопросов, заголовки, подсказки и т.п.). Этот элемент формально не является обязательным, но, учитывая ограничения на размер имени поля, как правило необходим. В тексте вопросника не должно быть (**NB!**) символов _ (знак подчеркивания), <, > и #: эти символы зарезервированы для обозначения формата полей данных (см.ниже).
- Имена полей данных – названия полей длиной не более 10 символов. Название поля может состоять из любых символов, не должно начинаться с цифры и должно быть уникальным для каждого поля. Называть поля рекомендуется с использованием

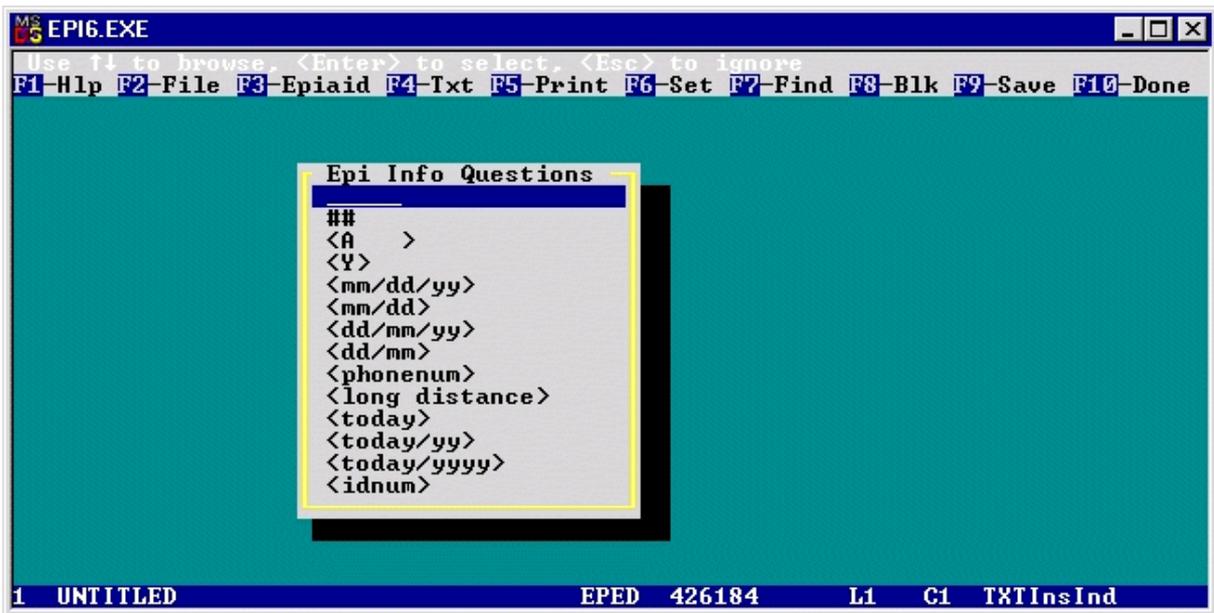
¹¹ Символ ^ означает, что буква (комбинация букв) вводится при нажатой клавише .

¹² Поскольку Epi Info создавалась для работы в операционной системе DOS, действуют соответствующие ограничения: длина имени файла не должна превышать 8 знаков (допустимы комбинации цифр и букв, следует избегать специальных символов), а длина расширения файла не может быть более 3 знаков.

латинского алфавита. Хотя применение кириллицы теоретически возможно и иногда не вызывает проблем, последствия могут быть непредсказуемыми.

- Типы полей данных – описание формата поля данных, по которому программа ENTER создает файл данных REC. В зависимости от характера собираемых данных, поля данных могут иметь различный формат. В программе EPED есть специальная функция, позволяющая вывести на экран список обозначений допустимых форматов полей и ввести маску поля в текст вопросника. Вид окна программы EPED после нажатия ^Q Q¹³ представлен на рис. 3.

Рис. 3. Список форматов полей данных в редакторе EPED



Типы полей данных в Epi Info

- Текстовое поле _____

В это поле можно вводить буквы, цифры и другие печатаемые символы. Длина переменной (поля) определяется количеством знаков подчеркивания: так, например, два знака подчеркивания (__) после названия поля означают, что в это поле можно ввести текст длиной не более 2 символов. Максимальная длина текстового поля в Epi Info - 80 символов (включая пробелы). При использовании списка форматов полей в редакторе EPED, вызываемом по нажатию ^Q Q, Вам придется ответить на вопрос, какова требуемая длина поля.

- Числовое поле ##

В программе ENTER в это поле можно вводить только цифры или пробелы. Если не вводится ничего, результатом будет «пустое» значение поля, которое в программе ANALYSIS интерпретируется как отсутствующее значение и обозначается точкой (.). Количество цифр в поле определяется количеством символов #. Если используется десятичная точка¹⁴, такое поле будет в «фиксированном десятичном» формате, количество символов # после десятичной точки определяет количество допустимых знаков после запятой. Например, для ввода чисел в диапазоне от 000.00 до 999.99, формат поля будет выглядеть как ###.##. Числовое поле может быть длиной до 14 знаков, считая десятичную точку. Не следует использовать символы «больше» и «меньше» (>, <). Выбирая подходящий тип поля для ваших данных,

¹³ Удерживая нажатой клавишу , дважды нажать клавишу .

¹⁴ В Epi Info используется именно точка, не запятая.

следует помнить, что это поле называется «числовым», а не «цифровым». Если вам нужно вводить цифры, с которыми вы не собираетесь производить никаких вычислений (например, номер школы или номер истории болезни), лучше вводить такие данные в текстовое поле. При использовании списка форматов полей в редакторе ERED, вызываемом по нажатию $\wedge Q Q$, Вам придется ответить на вопрос, сколько потребуется знаков до и после запятой.

- Заглавный текст <A >

Это поле общего назначения, однако все введенные буквы преобразуются в заглавные. Длина поля определяется количеством символов (пробелов) между символами «меньше» и «больше» (< >), включая заглавную латинскую «А», используемую для обозначения типа поля.

- Yes/No («Да/Нет») <Y>

При вводе данных в это поле будут восприниматься только Y (что означает “Yes”, т.е. «Да») и N (что означает “No”, т.е. «Нет»), а также пробел или нажатие , которые в программе ANALYSIS интерпретируются как отсутствующие значения. Независимо от того, вводите ли Вы “Y” или “y” (“N” или “n”), введенные буквы преобразуются в заглавные. Длина полей Yes/No всегда составляет один символ. При вводе данных в вопросник на русском языке следует помнить о необходимости переключения раскладки клавиатуры.

- Дата <mm/dd/yy>, <mm/dd>, <dd/mm/yy>, <dd/mm>

Даты могут вводиться в формате США, когда сначала указывается месяц, затем день (<mm/dd/yy>, <mm/dd>), а также в привычном европейском формате (<dd/mm/yy>, <dd/mm>). Epi Info проверяет ввод дат автоматически: Вам никогда не удастся убедить программу в том, что событие произошло 32 июля или 30 февраля. При вводе данных в программе ENTER достаточно вводить только цифры: знаки </> вводить не нужно.

- Местный телефонный номер <phonenum>

Вводится как xxx-xxxx, где «x» обозначают цифры номера.

- Междугородный телефонный номер <long distance>

Вводится как (xxx)xxx-xxxx, где «x» обозначают цифры номера.

- «Сегодняшняя» дата <today>, <today/yy>, <today/yyyy>

В это поле при сохранении записи автоматически вводится текущая дата из системных часов. Если запись редактируется и сохраняется неоднократно, в этом поле хранится запись о дате последнего сохранения. Рекомендуется создавать это поле в формате <today/yyyy>.

- «Счетчик» <idnum>

Специальное поле, позволяющее автоматически нумеровать записи по порядку. Первая запись в файле всегда имеет номер 1, каждая последующая запись имеет значение поля IDNUM на единицу больше. Длина поля определяется количеством знаков между символами «меньше» и «больше» (< >), включая буквы IDNUM, т.е. должна быть длиной 5 и более знаков.

В программе ENTER поле IDNUM заполняется автоматически, при вводе данных курсор в это поле не становится. Если Вам понадобится установить отличное от 1 значение этого поля для первой записи в файле данных, нужно поставить курсор в этом поле, используя клавишу , ввести нужное число и сохранить запись.

Создание вопросника в редакторе EPED

Рассмотрим процедуру создания вопросника в редакторе EPED на простейшем примере. Допустим, Вам необходимо создать электронную базу данных для расследования вспышки, которая на бумажном носителе могла бы выглядеть примерно так:¹⁵

Возраст (полных лет)	Пол	Время ужина	Наличие заболевания	Дата заболевания	Время заболевания	Употребление подозреваемых пищевых продуктов													
						Ветчина	Шпинат	Картофельное пюре	Салат из капусты	Фруктовое желе	Рулет	Черный хлеб	Молоко	Кофе	Вода	Пирожные	Ванильное мороженое	Шоколадное мороженое	Фруктовый салат
13	ж		нет															✓	
63	ж	19 ³⁰	да	18.04	22 ³⁰	✓	✓		✓	✓					✓		✓		
70	м	19 ³⁰	да	18.04	22 ³⁰	✓	✓	✓		✓	✓	✓		✓	✓		✓		
40	ж	20 ³⁰	да	19.04	2 ⁰⁰													✓	✓

Для создания электронной базы данных нужно запустить EPED, выбрав из меню «Программы» пункт «EPED Текстовый Редактор» и нажав клавишу . После появления окна EPED следует ввести текст вопросника.¹⁶

Сначала нужно набрать заголовок (например, «Расследование вспышки в г.Освего»), который из эстетических соображений можно центрировать:

РАССЛЕДОВАНИЕ ВСПЫШКИ В Г.ОСВЕГО

При избытке эстетических соображений можно отделить заголовок от остального текста, например так:

РАССЛЕДОВАНИЕ ВСПЫШКИ В Г.ОСВЕГО

***** ,

или так:

РАССЛЕДОВАНИЕ ВСПЫШКИ В Г.ОСВЕГО

----- ,

но ни в коем случае не так:

РАССЛЕДОВАНИЕ ВСПЫШКИ В Г.ОСВЕГО

_____ ,

поскольку, как уже упоминалось выше, знак подчеркивания зарезервирован для обозначения формата полей.

¹⁵ Примером (довольно упрощенным) служит пищевая вспышка в г.Освего, материалы расследования которой являются стандартным компонентом дистрибутива Epi Info (файлы oswego.*). Вопросник, приведенный в данном руководстве, переведен на русский язык и несколько модифицирован. В последующих главах мы также будем обращаться за примером к материалам расследования этой вспышки.

¹⁶ Переключение раскладки клавиатуры с кириллицы на латиницу и обратно в русифицированной версии, входящей в состав «Рабочего места госпитального эпидемиолога», осуществляется путем нажатия правой клавиши .

С новой строки начинаем создавать шаблоны полей базы данных, соответствующие столбцам таблицы. Сначала вводится пояснительный текст (формулировка вопроса, разъяснения для оператора и т.п.) произвольной (разумной) длины, затем название поля латинскими буквами, заключенное в фигурные скобки¹⁷, и, наконец, шаблон поля, который можно ввести вручную или с помощью команды ^Q Q.

Для поля «Возраст»¹⁸ выбираем числовой формат, 3 цифры без десятичной точки:

```
ВОЗРАСТ (ПОЛНЫХ ЛЕТ) {AGE} ###
```

Следующее поле (выбираем формат «заглавный текст»¹⁹) может быть длиной в 1 символ, если мы собираемся обозначать пол как «м» или «ж», или, например, в 3 символа («муж» или «жен»):

```
Пол {SEX} <A> или Пол {SEX} <A >
```

Специального формата «время» в Epi Info нет, поэтому используем числовой формат, длина поля должна быть 4 символа (например, время 13²⁴ вводится как 1324). Для оператора в тексте вопросника можно поместить подсказку:

```
ВРЕМЯ УЖИНА {TIMESUPPER} #### (13 ч 24 мин вводится как 1324)
```

Точно также задаем поле «Время возникновения заболевания»:

```
ВРЕМЯ ПОЯВЛЕНИЯ СИМТОМОВ ЗАБОЛЕВАНИЯ {ONSETTIME} ####
```

Поле для ввода даты возникновения заболевания создаем в европейском формате:

```
ДАТА ВОЗНИКНОВЕНИЯ ЗАБОЛЕВАНИЯ {ONSETDATE} <DD/MM/YYYY>
```

При расследовании вспышки можно ввести и короткую дату:

```
ДАТА ВОЗНИКНОВЕНИЯ ЗАБОЛЕВАНИЯ {ONSETDATE} <DD/MM>
```

Для оставшихся полей используем формат “Yes/No”:

```
ЗАБОЛЕЛ? {ILL} <Y>
```

Перечень блюд в меню можно предварить подзаголовком или объяснительным текстом, например:

```
КАКИЕ ИЗ ПЕРЕЧИСЛЕННЫХ БЛЮД УПОТРЕБЛЯЛ (А)? (ОТМЕТЬТЕ Y ИЛИ N)
```

```
ВЕТЧИНА {BAKEDHAM} <Y>
```

```
ШПИНАТ {SPINACH} <Y>
```

```
КАРТОФЕЛЬНОЕ ПЮРЕ {MASHEDPOTA} <Y>
```

```
САЛАТ ИЗ КАПУСТЫ {CABBAGESAL} <Y>
```

```
ФРУКТОВОЕ ЖЕЛЕ {JELLO} <Y>
```

```
РУЛЕТ {ROLLS} <Y>
```

```
ЧЕРНЫЙ ХЛЕБ {BROWNBREAD} <Y>
```

```
МОЛОКО {MILK} <Y>
```

```
КОФЕ {COFFEE} <Y>
```

```
ВОДА {WATER} <Y>
```

```
ПИРОЖНЫЕ {CAKES} <Y>
```

```
ВАНИЛЬНОЕ МОРОЖЕНОЕ {VANILLA} <Y>
```

```
ШОКОЛАДНОЕ МОРОЖЕНОЕ {CHOCOLATE} <Y>
```

¹⁷ Epi Info способна автоматически генерировать название поля длиной до 10 знаков из пояснительного текста, стоящего перед шаблоном поля. Однако при создании вопросников на русском языке пользоваться такой возможностью не следует.

¹⁸ Названия полей в рассматриваемом примере даются по-английски, поскольку они соответствуют названиям полей в файле oswego.rec из стандартной поставки Epi Info, к которому мы еще будем обращаться. В другом случае можно было бы написать вместо {AGE}, например, {VOZRASST}.

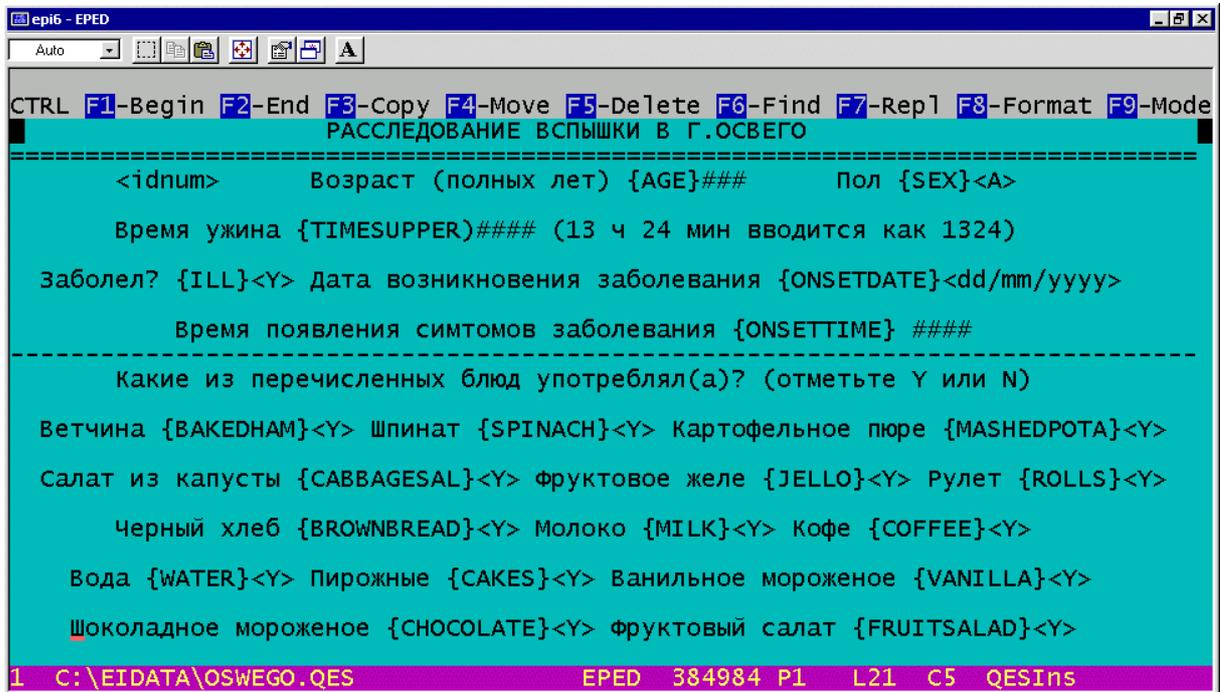
¹⁹ Только для примера, можно с таким же успехом выбрать обычный текстовый формат.

ФРУКТОВЫЙ САЛАТ {FRUITSALAD} <Y>

Для нумерации записей следует вставить поле в формате <IDNUM>.

Получившийся текст вопросника целиком представлен на рис.4:

Рис. 4. Текст вопросника в редакторе EPED



Получившийся вопросник нужно сохранить, нажав клавишу F9. В появившемся окне следует ввести название файла с расширением .QES:

OSWEGO.QES

В этом случае после нажатия  файл сохранится в рабочей папке (директории), по умолчанию C:\Еріб.

Если нужно сохранить файл в другом месте диска, следует ввести имя файла с указанием пути к нему, например

C:\DATA\OUTBREAK\OSWEGO.QES

После завершения работы нужно выйти из EPED по нажатию клавиши F10.

Безусловно, EPED, напоминающий древний текстовый редактор WordStar (Epi Info 2000 снабжена полноценным текстовым процессором), морально устарел и довольно неудобен в работе. Он имеет только два преимущества: возможность автоматического определения форматов полей по нажатию ^Q Q и отсутствие необходимости выхода из Epi Info для создания вопросника. Существует и другой путь создания вопросников: файлы QES можно создавать в других текстовых процессорах (например, MS Word), сохраняя их с тем же расширением имени (.QES) в формате «MS-DOS текст».

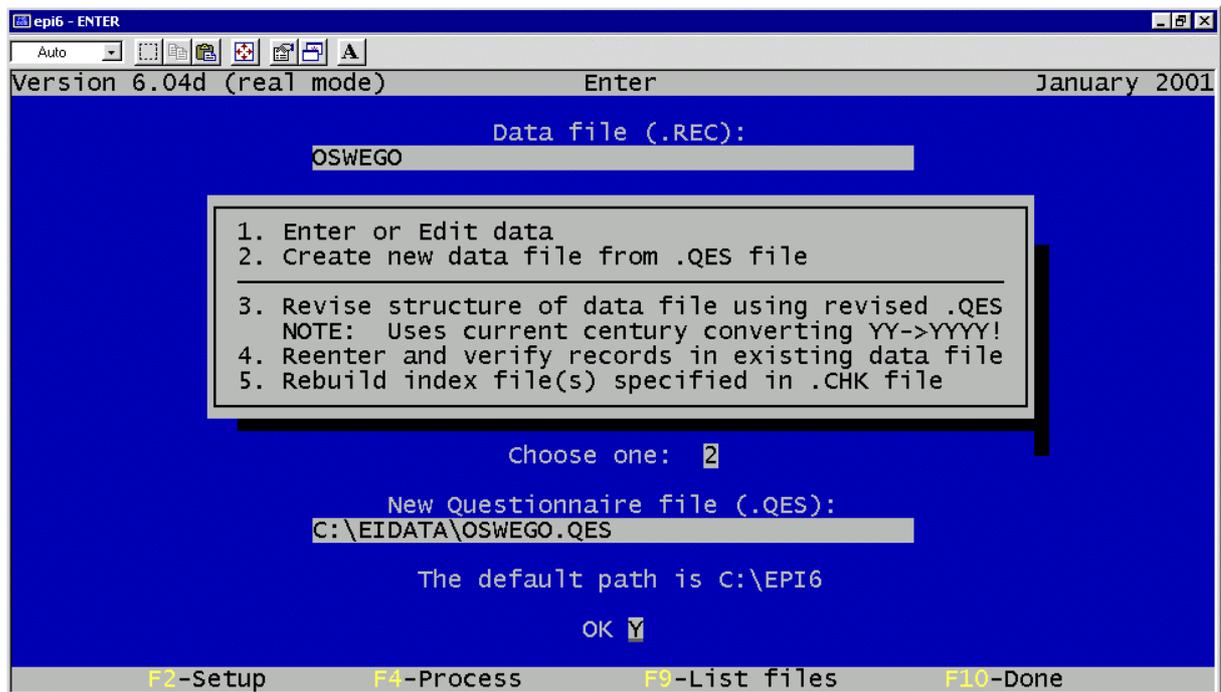
Ввод данных. Программа ENTER

Создание файлов данных

Для того, чтобы по шаблону вопросника создать файл базы данных, (т.н. REC-файл), нужно запустить программу ENTER (или ENTERX, если у Вас мощный компьютер), выбрав из меню «Программы» пункт «ENTER Ввод данных» и нажав клавишу .

В верхней части стартового экрана (см. рис. 5) под заголовком «Data file (.REC)» расположено окно, в которое нужно ввести имя создаваемого файла (можно без расширения, оно будет присвоено автоматически). Файл данных должен иметь то же имя, что и файл вопросника: например, если файл вопросника называется oswego.qes, файл базы данных должен получить название oswego.rec.

Рис. 5. Стартовый экран программы ENTER.



После того, как Вы ввели имя файла и нажали , курсор переместится в следующее окно (небольшое окно с заголовком «Choose one»). В это окно нужно ввести номер опции программы ENTER, которой Вы собираетесь воспользоваться.

Выше этого окна расположена подсказка, в которой перечислены опции выбора (функции программы ENTER):

1. Enter or Edit data	Ввод или редактирование данных
2. Create new data file from .QES file	Создание нового файла данных из файла .QES
3. Revise structure of data file using revised .QES NOTE: uses current century converting YY->YYYY!	Изменение структуры файла данных по измененному .QES
4. Reenter and verify records in existing data file	Повторный ввод и верификация данных в существующем файле
5. Rebuild index file(s) specified in .CHK file	Перестройка индексных файлов, определенных в файле .CHK

Для создания нового файла данных следует выбрать опцию 2 и снова нажать **Enter**. Сразу после этого появится еще одно окно (его не было на стартовом экране программы ENTER) под заголовком «New Questionnaire file (.QES)».

В это окно нужно ввести название вопросника, который служит шаблоном для создания нового файла данных (например, oswego.qes). Если файл находится в папке, установленной по умолчанию, достаточно ввести только имя файла, если вопросник находится в другой папке – нужно указать путь к нему. Если Вы не помните имя файла вопросника и/или путь к нему, следует нажать F9 и выбрать нужный файл, перемещаясь по появившемуся списку файлов с помощью клавиш **Up**, **Down**, **Left**, **Right**. После нажатия **Enter** имя файла появится в окне. Повторное нажатие **Enter** переместит курсор в окно подтверждения («ОК»). Если Вы уверены, что все введено правильно, еще одно (последнее!) нажатие клавиши **Enter** заставит программу автоматически создать новый REC-файл по Вашему шаблону и перейти в режим ввода данных²⁰.

Ввод данных

После того, как на экране появилась созданная программой электронная форма (рис. 6), можно приступать к вводу данных. Для этого соответствующий текст вводится с клавиатуры в пустые поля экранной формы. Поля на экране выглядят как серые прямоугольники, слева от них расположены названия полей (вопросы), которые на экране высвечены желтым.

Рис. 6. Форма ввода данных в программе ENTER

Внизу слева, сразу под нижним краем формы ввода, выводится информация о названии и содержании поля (запись «AGE: Integers allowed» на рис. 6 означает, что поле называется AGE (возраст), а вводить в это поле можно только целые числа). В нижней строке, как всегда в Epi Info, имеется информация о назначении функциональных клавиш (подробнее об их назначении в ENTER чуть ниже). В крайнем правом углу

²⁰ Если Вы собираетесь вводить данные в уже существующий REC-файл, достаточно ввести его имя (можно выбрать из списка по нажатию F9) и нажать F4. Опция 1 всегда выбрана по умолчанию.

выводится номер записи: на рис. 6 база данных еще пуста, поэтому начинаем с самой первой записи, «**Rec = 1**».

Перемещение внутри поля осуществляется клавишами  и . Если вводится достаточно знаков, чтобы заполнить все поле (как, например, всегда бывает с полем «Дата»), курсор перейдет на следующее поле автоматически, в противном случае после завершения ввода в данном поле для перехода в следующее нужно нажать . Нажатие  или,  позволяет быстро перемещаться из одного поля в другое. Подробная информация о командах, выполняемых в программе ENTER, представлена в таблице:

Команды редактирования в программе ENTER:

	Однократное нажатие удаляет один знак слева от курсора
 и 	Перемещает курсор на один знак вправо или влево
 и 	Перемещает курсор на предыдущее или следующее поле
 и 	Перемещает курсор на страницу вверх или вниз, если вопросник не помещается на одном экране
	Перемещает курсор на первое поле вопросника
	Перемещает курсор на последнее поле вопросника
	Включает/выключает режим вставки/замены. Когда режим вставки выключен, вновь печатаемые символы вводятся поверх старых, заменяя их на новые. Когда режим вставки включен, вновь вводимые символы вставляются между старыми в позиции курсора
	Однократное нажатие удаляет один знак в позиции курсора
	Позволяет проводить поиск в файле данных и находить записи, отвечающие определенным критериям
	Создает новую пустую запись. Обычно это требуется после просмотра нескольких записей или выполнения операции поиска
	Выполняет поиск записи по ее номеру. После нажатия F2 следует ввести номер интересующей Вас записи
	Выполняет поиск первой записи, удовлетворяющей критериям, введенным по нажатию ^F
	Находит следующую запись, удовлетворяющую критериям, введенным по нажатию ^F
	Находит предыдущую запись, удовлетворяющую критериям, введенным по нажатию ^F
	Печать вопросника на принтере
	Помечает текущую запись как удаленную. Следует помнить, что сразу после нажатия F6 на экране ничего не изменится. Только после того как вы перейдете на следующую запись и снова вернетесь к «удаленной»,

можно заметить, что в нижней строке экрана после номера записи появилась звездочка. Это значит, что запись помечена для удаления и будет проигнорирована программой ANALYSIS. Физически запись будет удалена только в том случае, если вы перезапишете файл данных в программе ANALYSIS, пользуясь командами ROUTE и WRITE RECFILE. Для отмены удаления нужно повторно нажать F6, звездочка у номера записи при этом исчезнет



Переход на предыдущую запись



Переход на следующую запись



Закрывает файл вопросника и возвращает к главному экрану Epi Info

После завершения ввода первой записи (нажатия в последнем поле) внизу экрана (рис. 7) появляется вопрос «Write data to disk (Y/N/Esc)?» (Записать данные на диск?). Если ввести в ответ «Y»²¹, данные будут сохранены (т.е. к базе данных добавится новая запись) и появится незаполненная форма ввода данных для заполнения следующей записи. Нажатие «Esc» позволит вновь редактировать текущую запись, а ввод «N» приведет к потере только что введенной записи.

Рис. 7. Пример заполненной формы ввода данных

епиб - ENTERX

Auto

РАССЛЕДОВАНИЕ ВСПЫШКИ В Г.ОСВЕГО

=====

1 Возраст (полных лет) AGE 52 Пол SEX M

Время ужина TIMESUPPER)2000 (13 ч 24 мин вводится как 1324)

Заболел? ILL Y Дата возникновения заболевания ONSETDATE 19/04/2000

Время появления симптомов заболевания ONSETTIME 2230

Какие из перечисленных блюд употреблял(а)? (отметьте Y или N)

Ветчина BAKEDHAM Y Шпинат SPINACH Y Картофельное пюре MASHEDPOTA Y

Салат из капусты CABBAGESAL N Фруктовое желе JELLO N Рулет ROLLS Y

Черный хлеб BROWNBREAD N Молоко MILK N Кофе COFFEE Y

Вода WATER N Пирожные CAKES N Ванильное мороженое VANILLA Y

Шоколадное мороженое CHOCOLATE N Фруктовый салат FRUITSALAD N

Write data to disk (Y/N/<Esc>)?

<Ctrl-N>-New <Ctrl-F>-Find F5-Print F6-Delete F9-Choices F10-Done Rec= 1

Последствия изменения QES-файла

Каждый REC-файл содержит в себе копию вопросника. Если изменить файл вопросника (QES-файл) или даже нечаянно удалить его, REC-файл может остаться неизменным. Программа ENTER способна определить наличие изменений в QES-файле, если он называется так же, как REC-файл (например, ECOLI.QES и ECOLI.REC)

²¹ Если Вы вводите данные по-русски, не забудьте переключить раскладку клавиатуры на латинский шрифт, Epi Info не понимает русского языка.

и находится в той же папке, что и файл данных²². После того как программа ENTER обнаружит изменения в QES-файле, она спросит, хотите ли вы отразить эти изменения в структуре REC-файла. В случае отрицательного ответа данные будут вводиться в форму, построенную по шаблону предыдущей версии вопросника. Однако если Вы сообщите программе о своем желании изменить REC-файл, изменится не только форма ввода данных, но и структура файла данных, а также, что весьма вероятно, его содержание.

К процедуре изменения REC-файла по измененному вопроснику нужно относиться с крайней осторожностью, при этом нужно принимать во внимание несколько важных соображений:

- Если Вы переименуете поле, в которое уже вводились данные, эти данные будут потеряны навсегда.
- Если Вы уменьшите ширину поля или его тип (формат), введенные ранее данные могут быть изменены или потеряны.
- Если Вы создадите новое поле под тем же именем, что и уже существующее, поместив его перед ним, все данные из оригинального поля будут конвертированы во вновь созданное. Если имеется разница в форматах полей, последствия могут быть самые непредсказуемые.

Именно поэтому к проектированию базы данных нужно отнестись со всей ответственностью на самом начальном этапе. Изменить впоследствии неудачно составленный вопросник без потери данных может оказаться невозможным (*NB!*).

Анализ данных. Программа ANALYSIS

Общие сведения о программе ANALYSIS. Команда READ.

Для запуска программы ANALYSIS нужно выбрать из меню «Программы» пункт «ANALYSIS Анализ данных» и нажать клавишу .

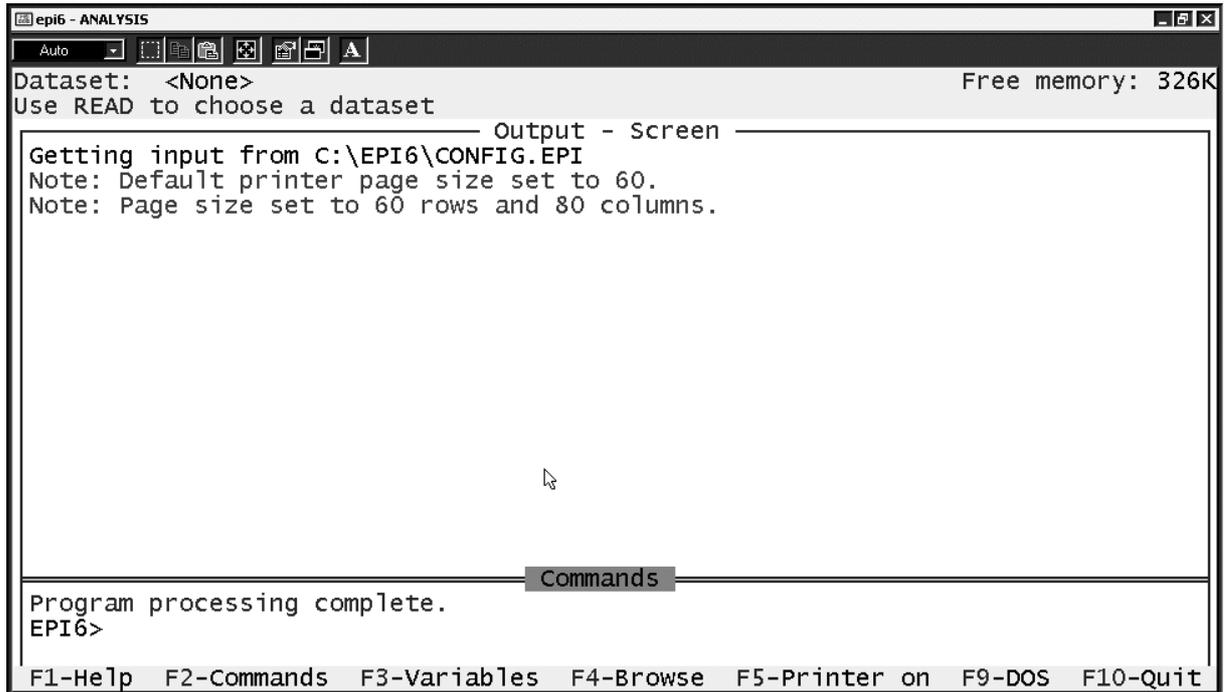
Появившийся экран программы (рис. 8) разделен на 4 зоны. В верхней части экрана выводится информация о текущем наборе данных и объеме свободной памяти. Большая часть экрана («Output») представляет собой окно вывода результатов работы программе. Ниже находится окно «Commands», в котором располагается командная строка. И, наконец, нижняя строка, как и в любом модуле Epi Info, содержит подсказки о назначении функциональных клавиш:

-  **Help** – вывод информации о назначении различных команд
-  **Commands** – список доступных команд
-  **Variables** – список переменных (полей) текущего файла данных
-  **Browse** – представление текущего файла данных в виде электронной таблицы
-  **Printer on/off** – включение/выключение прямого вывода на печать
-  **Update** – довольно опасный способ изменения файлов данных
-  **DOS** – выход в операционную систему DOS

²² Если этого не произошло, можно сделать это вручную, выбрав соответствующую опцию на стартовом экране программы ENTER («Revise structure of data file using revised .QES») и сообщив ей имя измененного QES-файла.

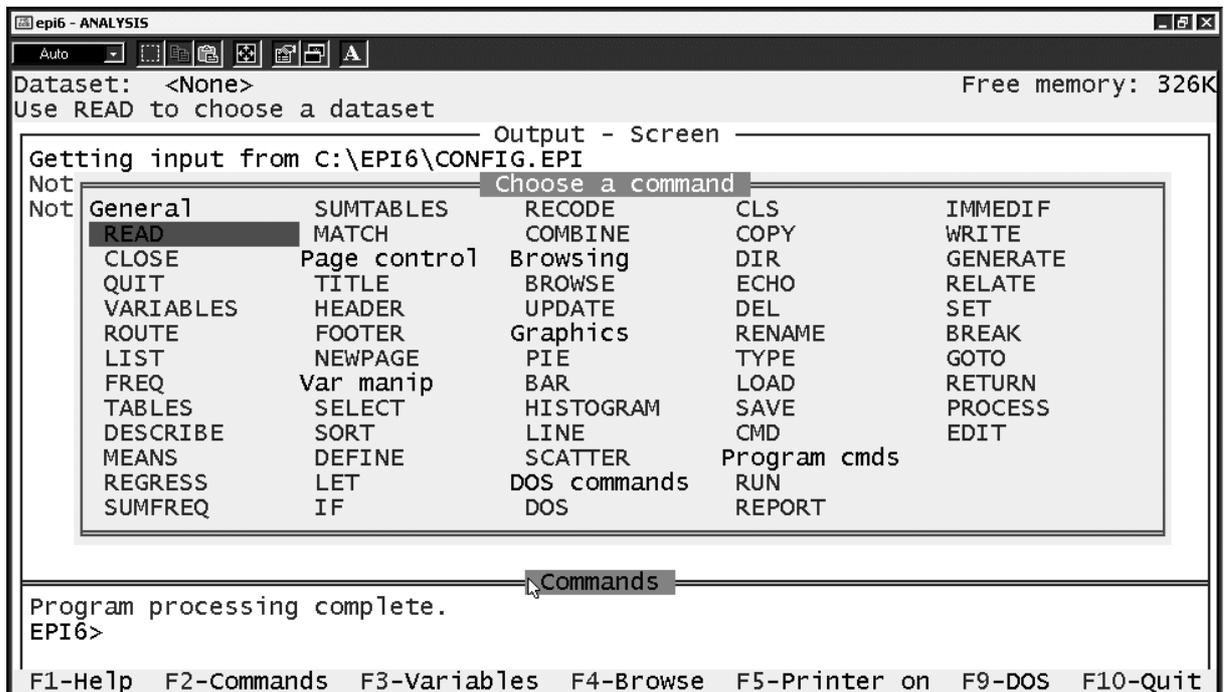
 **Quit** – завершение работы программы

Рис. 8. Стартовый экран программы ANALYSIS



Для того, чтобы начать работу с программой, нужно загрузить файл данных .REC. Для этого служит команда READ²³ (обратите внимание на подсказку в верхней строке: **Use READ to choose a dataset**). Эту (как и любую другую) команду в программе ANALYSIS можно ввести двумя способами: ввести непосредственно в командной строке или выбрать из списка команд по нажатию клавиши F2 (рис. 9).

Рис. 9. Список основных команд программы ANALYSIS



²³ Read по-английски означает «читать»

После появления команды READ в командной строке надо ввести название файла данных если требуется и нажать , например:

```
READ OSWEGO
```

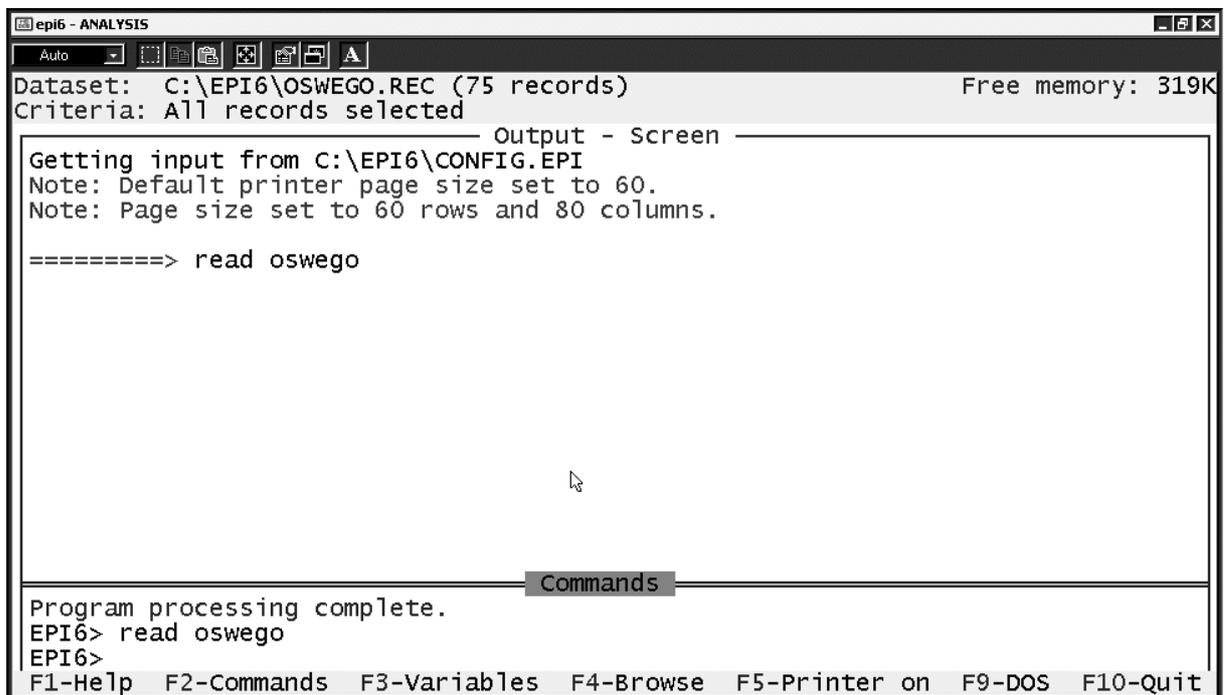
или

```
READ C:\DATA\OUTBREAK\OSWEGO
```

если файл данных находится не в текущей (с:\epi6 по умолчанию) директории.

После выполнения команды READ в верхней части экрана появляется сообщение о названии загруженного файла данных и количестве записей в этом файле (рис. 10). При этом следует обратить внимание, что количество записей, показываемое в верхней строке, включает также и записи, помеченные для удаления, хотя при обработке данных программа анализа не будет к ним обращаться.

Рис. 10. Экран программы ANALYSIS после выполнения команды READ



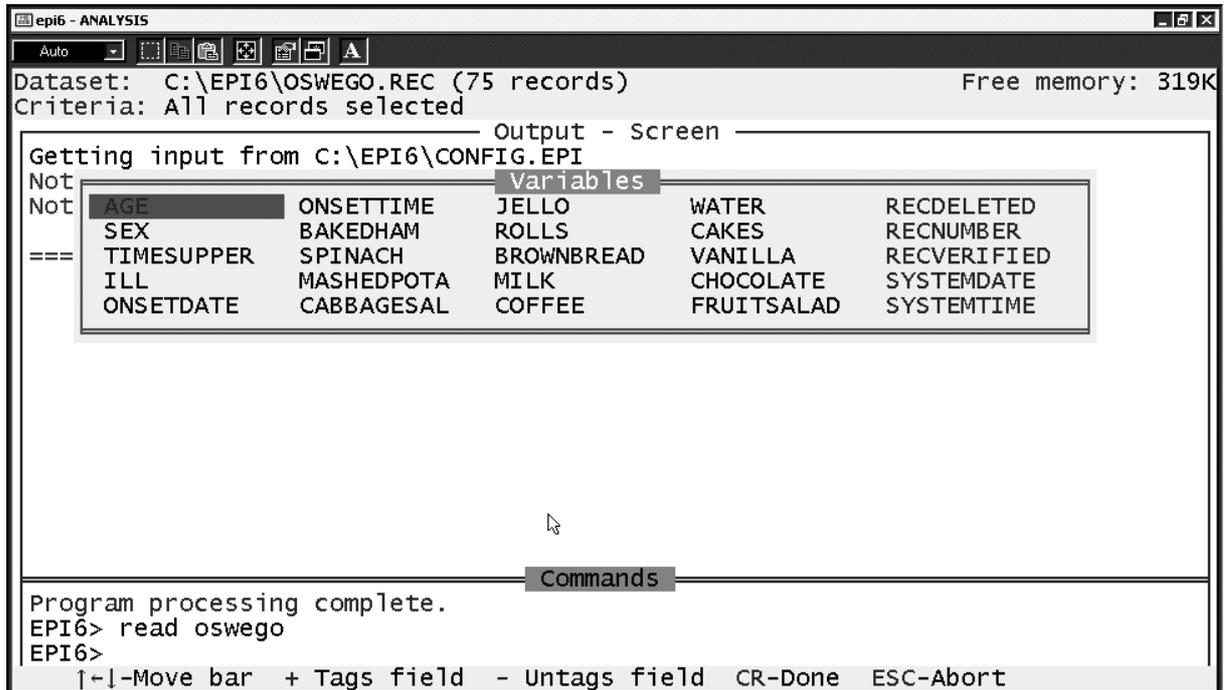
После того, как файл данных загружен, все команды программы ANALYSIS будут выполняться по отношению к данным, хранящимся в этом файле.

Основные команды анализа данных

Как уже было сказано, команду можно ввести из списка по нажатию F2 или ввести в командной строке. После имени команды анализа данных следует ввести имя переменной (нескольких переменных), для которой применяется команда, а также дополнительные опции выполнения команды, перечень которых можно посмотреть в файле справки, нажав F1. Если Вы не помните имена переменных (названия полей файла данных), можно, нажав клавишу F3, вызвать список переменных, хранящихся в открытом файле (см. рис. 11, на котором представлен список переменных, содержащихся в файле oswego.rec). Для ввода названия переменной в командную строку, нужно установить курсор на имя интересующей Вас переменной и нажать . Если необходимо ввести в командную строку несколько переменных, следует выделить их имена в списке клавишей  и нажать . Нужно помнить, что, после ввода

команды и всех необходимых аргументов, для ее выполнения следует снова нажать клавишу **↵** (**NB!**).

Рис. 11. Список переменных, содержащихся в файле `oswego.rec`: результат нажатия F3.



LIST и BROWSE

Для просмотра данных, содержащихся в открытом файле, можно использовать команду **BROWSE**²⁴ (клавиша F4). Нажатие этой клавиши вызывает появление электронной таблицы (см. содержимое файла `oswego.rec` на рис. 12), навигация по которой осуществляется клавишами управления курсором.

Рис. 12. Экран программы **ANALYSIS** после выполнения команды **BROWSE**.

REC	AGE	SEX	TIMESUPPER	ILL	ONSETDATE	ONSETTIME	BAKEDHAM	SPINACH	MASHEDPOTA	CABB
1	11	M		N			N	N	N	N
2	52	F	2000	Y	04/19	30	Y	Y	Y	N
3	65	M	1830	Y	04/19	30	Y	Y	Y	Y
4	59	F	1830	Y	04/19	30	Y	Y	N	N
5	13	F		N			N	N	N	N
6	63	F	1930	Y	04/18	2230	Y	Y	N	Y
7	70	M	1930	Y	04/18	2230	Y	Y	Y	N
8	40	F	1930	Y	04/19	200	N	N	N	N
9	15	F	2200	Y	04/19	100	N	N	N	N
10	33	F	1900	Y	04/18	2300	Y	Y	Y	N
11	65	M		N			Y	Y	Y	N
12	38	F		N			Y	Y	Y	N
13	62	F		N			Y	Y	N	Y
14	10	M	1930	Y	04/19	200	N	N	N	N
15	25	M		N			Y	Y	Y	Y
16	32	F		Y	04/19	1030	Y	Y	N	N
17	62	F		Y	04/19	30	N	N	N	N
18	36	M		Y	04/18	2215	Y	Y	N	Y
19	11	M		N			Y	Y	Y	Y
20	33	F		Y	04/18	2200	Y	Y	Y	Y
21	13	F	2200	Y	04/19	100	N	N	N	N

²⁴ *Browse* по-английски означает «просматривать, пролистывать»

Повторное нажатие F4 откроет экранную форму, как две капли воды похожее на экран программы ENTER, в которой будут представлены данные той записи, на которой стоял курсор.

Для возвращения в главное окно программы следует нажать клавишу F10.

Если нужно не просто просмотреть данные, но и вывести содержимое файла на печать или сохранить его в файле с результатами Вашей работы (подробнее об этом ниже), следует пользоваться командой LIST²⁵.

Ввод в командной строке команды LIST без аргументов выведет на экран²⁶ содержимое всего файла (точнее, той части файла, которая умещается на экране) в форме таблицы. Для просмотра всего файла (если он не поместился весь, в нижней строке экрана вывода результатов появится надпись «more») можно «пролистать» его содержимое экран за экраном, пользуясь клавишами  и . При необходимости построчного листания, следует нажимать  и .

Если Вам нужно просмотреть содержимое не всех, а лишь нескольких полей, нужно ввести их имена сразу после команды LIST, например

```
LIST AGE SEX ILL
```

В результате выполнения этой команды при работе с файлом oswego.rec на экране появится содержимое всех записей файла только в перечисленных выше полях базы данных.²⁷

```
REC  AGE  SEX  ILL
---  ---  ---  ---
  1   11   M   N
  2   52   F   Y
  3   65   M   Y
.....
 74   52   M   Y
 75   45   F   Y
```

Команда LIST **not** AGE SEX ILL выведет все записи базы данных для всех полей, кроме полей AGE, SEX и ILL. Точно так же можно задавать аргументы и для команды BROWSE.

FREQ

Команда FREQ позволяет вычислить частоту встречаемости каждого значения определенной переменной. В результате, например, применения команды FREQ к переменной SEX из файла oswego.rec, появится следующий результат:

```
SEX | Freq Percent Cum.
-----+-----
F   |  44  58.7%  58.7%
M   |  31  41.3%  100.0%
-----+-----
Total|  75 100.0%
```

По-русски это могло бы выглядеть следующим образом:

Пол	Частота (абс.)	Частота (%)	Накопленная частота (%)
Женский	44	58.7	58.7
Мужской	31	41.3	100
всего	75	100	

Команда FREQ применяется к одной переменной. Если ввести после команды FREQ имена нескольких переменных, программа ANALYSIS разместит результаты расчетов в отдельной таблице для каждой переменной, например:

²⁵ List по-английски означает «список, перечень»

²⁶ Или на принтер, или в файл

²⁷ Для экономии места приводим лишь 5 записей из 75.

FREQ ILL WATER

ILL	Freq	Percent	Cum.
+	46	61.3%	61.3%
-	29	38.7%	100.0%
Total	75	100.0%	

WATER	Freq	Percent	Cum.
+	24	32.0%	32.0%
-	51	68.0%	100.0%
Total	75	100.0%	

Приведенные выше результаты означают, что из 75 чел., данные опроса которых содержатся в файле `oswego.rec`, заболело 46 чел., т.е. 61.3%; 24 чел. (32.0%) из 75 опрошенных употребляли воду.

Команда `FREQ *` генерирует таблицы распределения по частоте встречаемости различных значений переменной для каждого поля базы данных. Команда `FREQ` с аргументом `/C` позволяет рассчитать 95% доверительные интервалы для относительных показателей, например:

FREQ SEX /C

SEX	Freq	Percent	Cum.	95% Conf Limit
F	44	58.7%	58.7%	46.7%-69.9%
M	31	41.3%	100.0%	30.1%-53.3%
Total	75	100.0%		

TABLES

Команда `TABLES` – это команда кросс-табуляции. Она позволяет генерировать таблицы сопряженности для категориальных переменных и выдает необходимые статистические показатели²⁸. Если после `TABLES` ввести в командную строку названия двух переменных, то значения первой переменной появятся в левом столбце получившейся таблицы, значения второй – в верхней строке таблицы. Если, например, попытаться оценить распределение участников вспышки в г.Освего по наличию заболевания (переменная `ILL`) и употреблению ванильного мороженого (переменная `VANILLA`), программа `ANALYSIS` построит соответствующую четырехпольную таблицу (обе переменные дихотомические) и произведет статистическую обработку данных, при этом результаты будут выглядеть так:

TABLES VANILLA ILL

VANILLA	ILL		Total
	+	-	
+	43	11	54
-	3	18	21
Total	46	29	75

Single Table Analysis

Odds ratio 23.45
 Cornfield 95% confidence limits for OR 5.07 < OR < 125.19*
 *May be inaccurate

²⁸ Описание возможностей команд анализа данных рассчитано на читателя, имеющего подготовку в области биостатистики. Подробное описание применяемых методов статистической обработки можно найти в соответствующих руководствах. Особенного внимания заслуживает прекрасное руководство по проведению статистического анализа именно средствами Epi Info: *Gerstman B.B. (2000). Data Analysis with Epi Info.* <http://www.sjsu.edu/faculty/gerstman/EpiInfo/>.

```

Maximum likelihood estimate of OR (MLE)                22.15
Exact 95% confidence limits for MLE                  5.22 < OR < 138.39
Exact 95% Mid-P limits for MLE                      5.93 < OR < 109.15
Probability of MLE >= 22.15 if population OR = 1.0   0.00000026

RISK RATIO (RR) (Outcome:ILL=+; Exposure:VANILLA=+)  5.57
95% confidence limits for RR                        1.94 < RR < 16.03

```

Ignore risk ratio if case control study

```

                Chi-Squares   P-values
                -----   -
Uncorrected:    27.22         0.00000018 <---
Mantel-Haenszel: 26.86         0.00000022 <---
Yates corrected: 24.54         0.00000073 <---

```

По-русски это выглядело бы следующим образом:

```

                ILL
VANILLA  |      +      - | Всего
-----+-----+-----
      + |     43     11 |    54
      - |      3     18 |    21
-----+-----+-----
      Всего |     46     29 |    75

```

Анализ одиночной таблицы

```

Отношение преобладаний29 (OR)                23.45
95% доверительные интервалы для OR по Cornfield    5.07 < OR < 125.19*
                *Могут оказаться неточными
Оценка максимального правдоподобия OR (MLE)        22.15
Точные 95% доверительные интервалы для MLE        5.22 < OR < 138.39
Точные 95% Mid-P интервалы для MLE                5.93 < OR < 109.15
Вероятность, что MLE >= 22.15 если OR в популяции = 1.0 0.00000026

ОТНОСИТЕЛЬНЫЙ РИСК (RR) (Исход:ILL=+; Экспозиция:VANILLA=+)  5.57
95% доверительные интервалы для RR                1.94 < RR < 16.03

```

Следует игнорировать относительный риск в исследовании случай-контроль

```

                Хи-квадрат   значение P
                -----   -
Без коррекции:    27.22         0.00000018 <---
По Mantel-Haenszel: 26.86         0.00000022 <---
С коррекцией по Yates    24.54         0.00000073 <---

```

Если ввести имя третьей переменной, она будет использована для стратификации. Например, TABLES VANILLA ILL SEX приведет к генерации двух таблиц, подобных приведенной в предыдущем примере, при этом одна из них будет содержать данные только для мужчин, другая – только для женщин.

Если нужно, чтобы таблица содержала, кроме абсолютных цифр, проценты, следует перед вводом команды анализа задать команду SET PERCENTS=ON:

```

SET PERCENTS=ON
TABLES VANILLA ILL

```

```

                ILL
VANILLA  |      +      - | Total
-----+-----+-----
      + |     43     11 |    54
      ► 79.6% 20.4% ► 72.0%
      | 93.5% 37.9% |
      - |      3     18 |    21
      ► 14.3% 85.7% ► 28.0%
      |  6.5% 62.1% |
-----+-----+-----
      Total |     46     29 |    75
      | 61.3% 38.7% |

```

²⁹ Иначе «отношение (соотношение) шансов»

MEANS и DESCRIBE

Команда MEANS предназначена для анализа непрерывных (интервальных) переменных. MEANS, выполняемая для одной переменной, позволяет получить целый ряд показателей описательной статистики, при этом вначале генерируется таблица, в которой выводятся все значения изучаемой переменной и частота встречаемости каждого значения.³⁰

```
MEANS AGE
```

AGE	Freq.	Percent	Cum.
3	1	1.3%	1.3%
7	2	2.7%	4.0%
8	2	2.7%	6.7%
.....			
70	1	1.3%	96.0%
72	1	1.3%	97.3%
74	1	1.3%	98.7%
77	1	1.3%	100.0%

Total	75	100.0%	

Total	Sum	Mean	Variance	Std Dev	Std Err
75	2761	36.813	460.181	21.452	2.477

Minimum	25%ile	Median	75%ile	Maximum	Mode
3.000	16.000	36.000	58.000	77.000	11.000

Student's "t", testing whether mean differs from zero.
T statistic = 14.862, df = 74 p-value = 0.00000

По-русски это выглядело бы так:

AGE	Частота	%	Накопл. частота
3	1	1.3%	1.3%
7	2	2.7%	4.0%
8	2	2.7%	6.7%
.....			
70	1	1.3%	96.0%
72	1	1.3%	97.3%
74	1	1.3%	98.7%
77	1	1.3%	100.0%

Всего	75	100.0%	

Всего	Сумма	Среднее	Дисперсия	Станд. отклонение	Станд. ошибка
75	2761	36.813	460.181	21.452	2.477

Минимум	25%иль ³¹	Медиана	75%иль ³²	Максимум	Мода
3.000	16.000	36.000	58.000	77.000	11.000

Коэффициент Student "t" для проверки отличия среднего от нуля.
Коэффициент T = 14.862, df³³ = 74 значение p = 0.00000

Чтобы избежать появления таблицы, следует вводить команду MEANS с опцией /n, например:

```
MEANS AGE /N
```

Сходные результаты дает команда DESCRIBE, выполнение которой выдает основные дескриптивные показатели для количественных переменных:

```
DESCRIBE AGE
```

³⁰ Для экономии места здесь приводятся не все значения переменной.

³¹ 25-ый перцентиль, нижний квартиль

³² 75-ый перцентиль, верхний квартиль

³³ Число степеней свободы (*degrees of freedom*)

Variable	Obs	Total	Mean	Variance	Std Dev
AGE	75	2761.000	36.813	460.181	21.452

По-русски это означает следующее:

Переменная	Кол-во наблюдений	Всего	Среднее	Дисперсия	Станд. отклонение
AGE	75	2761.000	36.813	460.181	21.452

Однако возможности команды MEANS не исчерпываются описательной статистикой. Если в командной строке после имени непрерывной переменной ввести название еще одной, категориальной переменной, выполнение команды MEANS приведет к появлению результатов однофакторного дисперсионного анализа. При этом первая переменная является зависимой, а вторая – независимой (фактором):³⁴

MEANS AGE ILL

AGE	ILL			Total
		+	-	
3		1	0	1
7		1	1	2
8		2	0	2
.....				
70		1	0	1
72		1	0	1
74		1	0	1
77		1	0	1

Total		46	29	75

ILL	Obs	Total	Mean	Variance	Std Dev
+	46	1806	39.261	477.264	21.846
-	29	955	32.931	423.709	20.584
Difference			6.330		

ILL	Minimum	25%ile	Median	75%ile	Maximum	Mode
+	3.000	17.000	38.500	59.000	77.000	15.000
-	7.000	14.000	35.000	50.000	69.000	11.000

ANOVA
(For normally distributed data only)

Variation	SS	df	MS	F statistic	p-value	t-value
Between	712.655	1	712.655	1.560	0.215602	1.249147
Within	33340.732	73	456.722			
Total	34053.387	74				

Bartlett's test for homogeneity of variance
Bartlett's chi square = 0.119 deg freedom = 1 p-value = 0.729786

The variances are homogeneous with 95% confidence.
If samples are also normally distributed, ANOVA results can be used.

Mann-Whitney or Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

Kruskal-Wallis H (equivalent to Chi square) = 1.161
Degrees of freedom = 1
p value = 0.

По-русски это могло бы выглядеть так:

AGE	ILL			Всего
		+	-	
3		1	0	1
7		1	1	2
8		2	0	2
.....				
72		1	0	1
74		1	0	1
77		1	0	1

Всего		46	29	75

³⁴ Для экономии места здесь приводятся не все значения переменной.

ILL	Кол-во наблюдений	Всего	Среднее	Дисперсия	Станд. отклонение
+	46	1806	39.261	477.264	21.846
-	29	955	32.931	423.709	20.584
Разность средних			6.330		

ILL	Минимум	25%иль	Медиана	75%иль	Максимум	Мода
+	3.000	17.000	38.500	59.000	77.000	15.000
-	7.000	14.000	35.000	50.000	69.000	11.000

Дисперсионный анализ
(Только для нормально распределенных данных)

Вариация	SS ³⁵	df ³⁶	MS ³⁷	критерий F	значение p	значение t
Между ³⁸	712.655	1	712.655	1.560	0.215602	1.249147
Внутри ³⁹	33340.732	73	456.722			
Общая	34053.387	74				

Тест однородности дисперсий *Bartlett*
Хи-квадрат по *Bartlett* = 0.119 степ. свободы = 1 значение p = 0.729786

Дисперсии однородны с достоверностью 95%
Если выборки также распределены нормально, результаты дисперсионного анализа использовать можно

Двухвыборочный тест по *Mann-Whitney* или *Wilcoxon* (тест *Kruskal-Wallis* для двух групп)

Kruskal-Wallis H (эквивалент хи-квадрат) = 1.161
Степени свободы = 1
значение p = 0.

REGRESS

Команда REGRESS служит для проведения линейного регрессионного анализа. Для ее выполнения следует ввести имена двух непрерывных переменных, первая из которых будет служить зависимой, а вторая – независимой. Выполнение команды позволяет получить коэффициенты регрессии и другие статистические показатели.

Графическое представление данных

Eri Info позволяет строить основные виды диаграмм, включая столбиковые (команда BAR), секторные (команда PIE) и линейные диаграммы (команда LINE), а также гистограммы (команда HISTOGRAM) и диаграммы рассеяния (команда SCATTER). Несмотря на то, что графика в Eri Info 6 довольно примитивна по сравнению со специализированными пакетами статистической графики⁴⁰, возможности программы позволяют получить по крайней мере общее представление о том, как выглядят анализируемые данные. При необходимости получения качественной графики можно экспортировать данные из Eri Info в другую программу (см. главу «Импорт и экспорт данных») и построить диаграммы там.

Для построения диаграммы следует ввести (или выбрать по F2) команду и имя переменной (двух переменных для диаграммы рассеяния) для графического представления. Так, например «PIE SEX» для данных из файла oswego.res приведет к построению секторной диаграммы распределения по полу (рис.13), а команда «HISTOGRAM AGE» позволит визуальнo оценить распределение участников вспышки по возрасту.

³⁵ Суммы квадратов отклонений

³⁶ Степени свободы

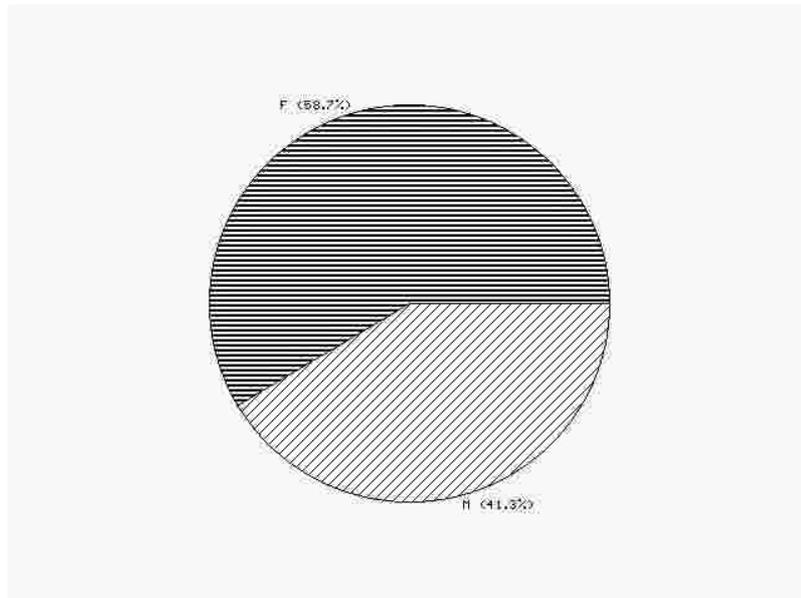
³⁷ SS/df

³⁸ По фактору

³⁹ Остаточная

⁴⁰ В Eri Info 2000 также имеется неплохой модуль графического анализа данных. Увы, эта версия еще не готова к широкому использованию...

Рис. 13. Результат выполнения команды PLE



Работа с переменными

Создание новых переменных. Команда DEFINE

Команда DEFINE позволяет создавать для использования в программе ANALYSIS новые переменные, которые не были созданы на этапе генерации файла данных .REC из вопросника. Для этого в командной строке нужно ввести команду DEFINE, название вновь создаваемой переменной и ее формат, например:

```
DEFINE NEW <A >
DEFINE NEWVAR <MM/DD/YYYY>
```

Вычисляемые переменные

В каких случаях может понадобиться новая переменная? Допустим, в Вашем вопроснике для ввода данных обследования пациентов есть пункт «Дата рождения»⁴¹ и соответствующее поле DOB⁴² в формате dd/mm/yyyy. Теперь Вам нужно знать возраст опрошенных, но такого поля в базе данных нет. Это значит, что переменную для записи возраста (назовем ее AGE) придется создавать в программе ANALYSIS. Для решения этой задачи придется ввести следующий набор команд:

```
DEFINE AGE ###
AGE = (DATEXAM - DOB) DIV 365.25
```

Команда в первой строке позволила создать совершенно новую переменную AGE, которой не было в базе данных. Это числовая переменная, значениями которой могут быть целые числа длиной не более 3 знаков. Следующая строка означает, что переменная AGE принимает значения, равные разности даты обследования (которая хранится в поле DATEXAM) и даты рождения пациента (поле DOB). Поскольку эта разница выражается в днях, для получения возраста, выраженного в годах, нужно разделить полученное число на 365.25. Оператор DIV означает целочисленное деление (частное от деления не содержит десятичных знаков после запятой)⁴³.

⁴¹ При создании вопросников всегда следует предпочесть дату рождения возрасту. При наличии даты рождения всегда можно рассчитать возраст на любую интересующую Вас дату.

⁴² Или, если хотите, DATAROZH.D. DOB значит *date of birth*.

⁴³ В обычных случаях в качестве оператора деления используется символ «/».

Еще один пример применения команды DEFINE. Допустим, в вопроснике есть поле HEIGHT⁴⁴ для хранения данных о росте пациента в сантиметрах, и по каким-то причинам Вам нужно иметь рост, измеренный в метрах (например, для расчета антропометрических показателей, когда этого требует формула). Это можно сделать следующим образом:

```
HEIGHT = HEIGHT / 100
```

При таком подходе могут возникнуть проблемы: если поле HEIGHT было задано в формате ###, который не допускает ввода десятичных знаков после запятой, 190см/100 превратится не в 1.9м, а в 1м, поскольку десятые доли негде будет разместить. Поэтому верным ходом будет использование команды DEFINE для создания нового поля HEIGHTM, допускающего хранение десятичных дробей:

```
DEFINE HEIGHTM #.##
HEIGHTM = HEIGHT / 100
```

Операторы IF-THEN

Рассмотрим применение логических операторов на нехитром примере:

```
DEFINE ТИМЕТОЕХАМ <Y>
IF AGE > 45 THEN ТИМЕТОЕХАМ = "Y" ELSE ТИМЕТОЕХАМ = "N"
```

В первой строке мы определили новую переменную ТИМЕТОЕХАМ⁴⁵ в формате «Yes/No» для того, чтобы выявить тех, кто нуждается в скрининговом обследовании. Следующая строка означает, что у всех, кто достиг 45-летнего возраста, в поле ТИМЕТОЕХАМ появится «Y» (то-есть «Да»), в противном случае появится «N» («Нет»). Если перевести содержимое второй строки на русский язык, получится что-то вроде:

```
ЕСЛИ AGE > 45 ТО ТИМЕТОЕХАМ = "Y" В ОСТАЛЬНЫХ СЛУЧАЯХ ТИМЕТОЕХАМ = "N"
```

Однако в любой базе данных могут оказаться пропущенные значения. Предположим, что запись о возрасте в поле AGE по каким-то причинам у одного или нескольких человек отсутствует. Отсутствующее значение не может быть больше 45. И хотя оно с таким же успехом не может быть меньше или равно 45, утверждение IF-THEN, приведенное выше, классифицирует человека «неопределенного возраста» как не нуждающегося в осмотре, и он выпадет из нашего поля зрения. Даже если на самом деле ему уже больше 45, он просто не попадет в списки для очередного осмотра. В этом случае хотелось бы, чтобы в поле ТИМЕТОЕХАМ какое-либо значение также отсутствовало, что позволило бы обратить внимание на пробелы в базе данных и предпринять соответствующие шаги. Проблему можно решить, используя два утверждения IF-THEN:

```
DEFINE ТИМЕТОЕХАМ <Y>
IF AGE > 45 THEN ТИМЕТОЕХАМ = "Y" ELSE ТИМЕТОЕХАМ = "N"
IF AGE = . THEN ТИМЕТОЕХАМ = .
```

Последняя строка означает «если возраст не указан, пусть в поле ТИМЕТОЕХАМ значение также отсутствует». Очень важно соблюсти приведенную последовательность логических утверждений. Если бы второе выражение IF-THEN (строка 3) шло первым, оно не имело бы никакого эффекта. IF-THEN-ELSE во второй строке просто перекрыло бы его.

Есть и другой способ добиться того же результата:

⁴⁴ С таким же успехом можно было назвать это поле ROST

⁴⁵ Можно назвать эту переменную «NUZHNO SMOTR» или «PORA OS MOTR» или еще какнибудь, все равно выглядит ужасно. Именно поэтому и нужен пояснительный текст в вопроснике.

```

DEFINE ТИМЕТОЕХАМ <Y>
IF AGE > 45 THEN ТИМЕТОЕХАМ = "Y"
IF (AGE >= 0) AND (AGE <= 45) THEN ТИМЕТОЕХАМ ="N"

```

Сразу после создания новой переменной (ТИМЕТОЕХАМ) значения в ней отсутствуют. Первое утверждение IF-THEN (строка 2) устанавливает значения переменной «Y», если возраст старше 45. Второе утверждение IF-THEN (строка 3) устанавливает значения ТИМЕТОЕХАМ "N", если возраст больше или равен 0 ⁴⁶ если он меньше или равен 45, т.е. возраст находится в пределах от 0 до 45. Все остальные варианты (возраст отсутствует) не удовлетворяют условиям ни одного из использованных выражений IF-THEN, благодаря чему в поле ТИМЕТОЕХАМ останутся пропущенные значения для всех, у кого возраст не был указан.

Выражение AGE >= 0 (больше или равно 0) не означает, что предполагается возможность отрицательного возраста. Дело в том, что отсутствующие значения в Epi Info трактуются как нижайшие из всех возможных в числовом отношении (даже «ниже» 0). И если бы мы скоментировали Epi Info

```
IF (AGE <= 45) THEN ТИМЕТОЕХАМ ="N" ELSE ТИМЕТОЕХАМ = "Y"
```

мы бы по-прежнему не решили проблемы отсутствующих значений⁴⁷, поскольку выражение «отсутствующее значение < 45» является истинным, и, значит, в поле ТИМЕТОЕХАМ появились бы "N". Следует очень осторожно обращаться с приемами формальной логики, чтобы не совершать ошибок при анализе данных.

Команда RECODE

Мы не случайно выше при обсуждении команд FREQ и TABLES делали оговорки по поводу того, что если привести таблицу целиком, все возможные значения возраста займут слишком много места. Часто группировка значений переменных (сокращение размерности признака) может быть необходима не только для облегчения восприятия данных, но и по формальным соображениям, в том числе для применения определенных методов статистического анализа. Для этих целей служит команда RECODE.

Самый простой способ группировки данных применяется в тех случаях, когда мы собираемся объединить значения переменной в группы известной длины. Для этого применяется оператор BY.

Предположим, что мы хотим разделить пациентов на возрастные группы с интервалом в 10 лет. Для этого сначала нужно создать новую текстовую переменную длиной не менее 10-12 знаков⁴⁸ (назовем ее, например, AGE10), а затем применить команду RECODE с оператором BY:

```

DEFINE AGE10 <A          >
RECODE AGE TO AGE10 BY 10

```

Вторая строка означает указание Epi Info перекодировать переменную AGE в переменную AGE10, разделив всех на возрастные группы по 10 лет.

Если применить эти команды к данным из уже набившего оскомину файла oswego.rec и затем выполнить команду «FREQ AGE10», получим следующий результат:

⁴⁶ Еще один логический оператор, AND (т.н. «логическое И»)

⁴⁷ Лучше, конечно, аккуратно заполнять базу данных, но практика показывает, что отсутствующие значения есть всегда

⁴⁸ Чтобы поместились результаты выполнения команды с оператором BY.

AGE10	Freq	Percent	Cum.
0 to 9	6	8.0%	8.0%
10 to 19	18	24.0%	32.0%
20 to 29	6	8.0%	40.0%
30 to 39	13	17.3%	57.3%
40 to 49	5	6.7%	64.0%
50 to 59	11	14.7%	78.7%
60 to 69	12	16.0%	94.7%
70 to 79	4	5.3%	100.0%
Total	75	100.0%	

Если Вас не устраивает автоматическая группировка, и хочется сгруппировать по другому (разделить на привычные группы: 0-2 года, 3-6 лет, 7-14 и т.д.), это придется делать следующим образом:

```
DEFINE AGEGR <A          >
RECODE AGE TO AGEGR 0-2="от 0 до 2" 3-6="от 3 до 6" \
7-14="от 7 до 14" 15-19="от 15 до 19" 20-29="от 20 до 29" \
30-39="от 30 до 39" 40-49="от 40 до 49" 50-59="от 50 до 59" \
60-69="от 60 до 69" 70-79="от 70 до 79" 80-НІ="СВЫШЕ 80" \
ELSE = .
```

Символ "\" означает, что команда продолжается на следующей строки (именно так нужно вводить команды в командной строке программы ANALYSIS, иначе нажатие клавиши  приведет к попытке выполнения неполностью введенной команды). В кавычках (все, что не является числами, командами, операторами и именами переменных, вводится в кавычках) можно вводить текст по-русски.⁴⁹

Выражение 80-НІ⁵⁰ означает то же, что «AGE >= 80», т.е. возраст 80 и более. Оператор НІ обозначает наибольшее из существующих в базе данных значений. Существует и оператор LO⁵¹, обозначающий минимальное из существующих (не отсутствующее!) значение.

Как уже говорилось, нужно всегда беспокоиться об отсутствующих значениях. Оператор ELSE, использованный выше, позволяет сохранить отсутствующие значения (".") и во вновь создаваемой переменной.

Команда RECODE без оператора BY позволяет перекодировать переменную одного типа в переменную любого другого типа.

Повторные RECODE имеют кумулятивный эффект. В отличие от IF-THEN, приоритет имеет первая введенная команда.

Команда SELECT

Одна из наиболее полезных и широко применяемых команд в программе ANALYSIS – команда SELECT⁵².

Эта команда позволяет выбирать из файла данные по определенным пользователем критериям (можно выбрать людей, имеющих определенные характеристики, или проанализировать данные за определенный период времени и т.п.). Например, если Вы

⁴⁹ Уважаемый читатель! Если у Вас хватило терпения добраться до этого места, сообщите об этом, пожалуйста, по адресу epi@beep.ru: авторам будет приятно.

⁵⁰ НІ – сокращение от англ. *high* (высокий)

⁵¹ LO – сокращение от англ. *low* (низкий)

⁵² *select* по-английски означает «выбирать»

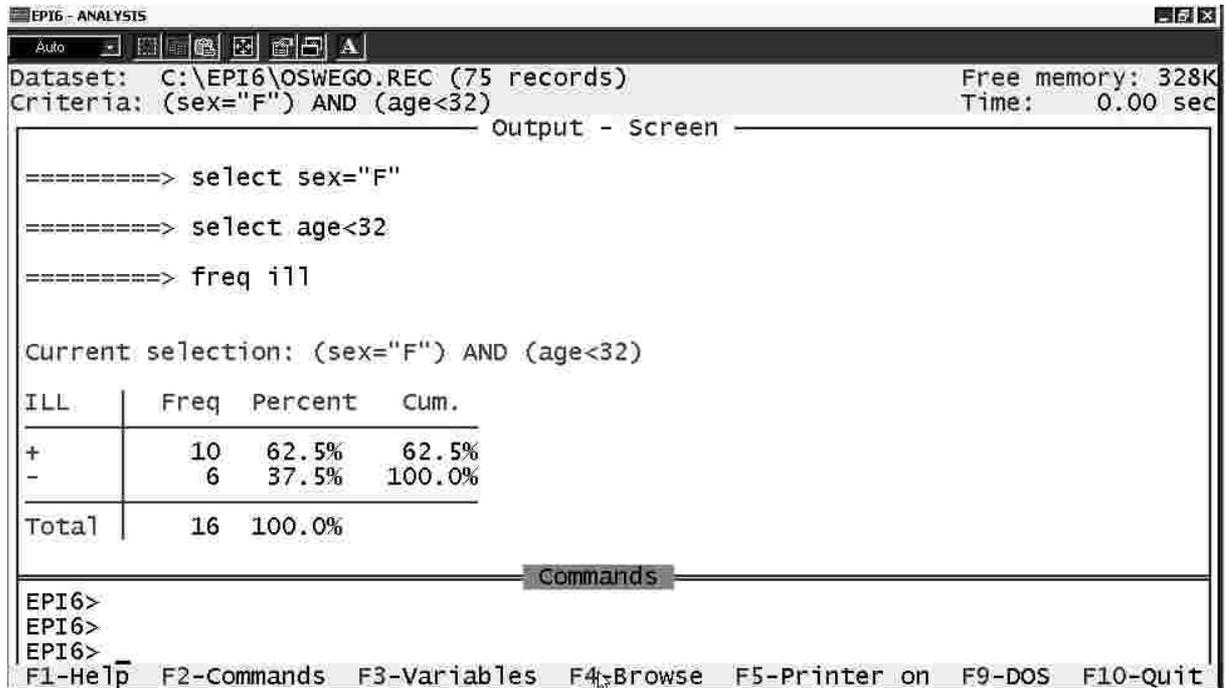
хотите выбрать из базы данных oswego.rec только женщин, в командной строке нужно ввести

```
SELECT (SEX="F")53
```

Теперь все операции будут проводиться только с данными о женщинах.

После выполнения команды SELECT в верхнем левом углу экрана появляется напоминание о действующих критериях отбора (рис. 14).

Рис. 14. Окно команды ANALYSIS после выполнения команды SELECT



Команда SELECT обладает кумулятивным действием: любой последующий отбор объединяется с предшествующим. Например, если Вам необходимо отобрать женщин моложе 32 лет, это можно сделать двумя способами:

```
SELECT (SEX="F")
SELECT AGE < 32
```

или сразу

```
SELECT (SEX = "F") AND (AGE < 32)
```

Представленный на рисунке результат соответствует выполнению любого из двух вариантов. Обратите внимание, что, кроме женщин моложе 32, будут отобраны женщины, не сообщившие свой возраст, поскольку «age <32» справедливо и для женщин, чей возраст = ., т.е. отсутствует.

Кроме оператора AND, могут применяться и другие арифметические и логические операторы, список которых приводится в конце этой главы.

Для того, чтобы вернуть возможность анализировать все данные (или воспользоваться другими критериями отбора), нужно отменить текущие критерии, набрав на клавиатуре (или выбрав из меню по F2) команду SELECT без аргументов:

```
SELECT
```

⁵³ Обратите внимание на кавычки, в которые заключено значение текстовой переменной

Сохранение данных в новом файле

Программа ANALYSIS позволяет сохранить Ваши данные частично или полностью, вместе со вновь созданными переменными⁵⁴, в новом .REC-файле. Для этого используется команда WRITE RECFILE:

```
ROUTE NEW.REC
WRITE RECFILE
```

В первой строке задано имя файла, во второй строке дана команда на его сохранение. Теперь все вновь созданные переменные, которые до выполнения команды считались временными, будут храниться на диске. В новом файле уже не будет записей, помеченных для удаления в программе ENTER (до сохранения нового файла они оставались на диске, хотя программа ANALYSIS к ним и не обращалась). Следует помнить, что все вычисляемые поля, которые автоматически обновлялись в ходе анализа, после их сохранения в новом файле обновляться больше не будут – теперь это обычные поля базы данных, содержащие результаты последнего вычисления, сделанного до сохранения файла.

Если файл данных содержит большое количество переменных, можно сохранить новый файл, который будет содержать только выбранные Вами поля. Команда

```
ROUTE OSWEGO2.REC
WRITE RECFILE AGE SEX ILL VANILLA
```

сохранит на диске новый файл, который будет содержать только переменные AGE, SEX, ILL и VANILLA.

При выполнении команды WRITE RECFILE можно задать критерии отбора. Например,

```
SELECT ILL="Y"
ROUTE OSWEGILL.REC
WRITE RECFILE
```

сохранит на диске файл, в котором будут содержаться только данные о заболевших.

Программы

Epi Info позволяет избежать утомительной необходимости повторного ввода одних и тех же команд⁵⁵, особенно если Вам приходится выполнять одни и те же процедуры анализа данных по мере их пополнения. Для этого можно сохранить список необходимых команд в обычном текстовом файле с расширением .PGM.

Самый простой способ создания собственных программ анализа данных – сохранить команды, которые вводились в командную строку в процессе работы с ANALYSIS и затем отредактировать получившийся файл в текстовом редакторе (например, EPED)

Для запуска таких программ достаточно ввести в командной строке программы ANALYSIS команду RUN и имя файла с программой:

```
RUN MYPROG.PGM
```

Например, чтобы автоматизировать утомительный ввод команды RECODE для распределения по возрастным группам пациентов, данные о которых хранятся в гипотетическом файле MYFILE.REC, можно создать небольшую программу AGEGROUP.PGM⁵⁶, которая может выглядеть следующим образом:

⁵⁴ Переменные, созданные командой DEFINE, являются временными, они хранятся только в оперативной памяти. Если они могут понадобиться при следующем сеансе работы, их нужно сохранить в новом файле.

⁵⁵ Достаточно вспомнить процедуру группировки по возрастным группам, описанную выше.

⁵⁶ Или, скажем, VOZRGRUP.PGM.

```

* AGEGROUP.PGM Дата создания 09/09/01 2:20 PM
* ПРОГРАММА ДЛЯ РАСПРЕДЕЛЕНИЯ ПО ВОЗРАСТНЫМ ГРУППАМ
* ОТКРЫТИЕ ФАЙЛА MYFILE.REC:
  READ MYFILE
* СОЗДАНИЕ ПЕРЕМЕННОЙ AGE (ВОЗРАСТ)
  DEFINE AGE ###
* ВЫЧИСЛЕНИЕ ВОЗРАСТА (КОЛ-ВО ПОЛНЫХ ЛЕТ) КАК РАЗНОСТИ МЕЖДУ
* ДАТОЙ ОБСЛЕДОВАНИЯ (DATEXAM) И ДАТОЙ РОЖДЕНИЯ (DOB):
  AGE = (DATEXAM - DOB) DIV 365.25
* СОЗДАНИЕ ПЕРЕМЕННОЙ AGEGR (ВОЗРАСТНАЯ ГРУППА):
  DEFINE AGEGR <A      >
* КОДИРОВАНИЕ ВОЗРАСТА (AGE) В ВОЗРАСТНЫЕ ГРУППЫ (AGEGR):
  RECODE AGE TO AGEGR 0-2="от 0 до 2" 3-6="от 3 до 6" \
  7-14="от 7 до 14" 15-19="от 15 до 19" 20-29="от 20 до 29" \
  30-39="от 30 до 39" 40-49="от 40 до 49" 50-59="от 50 до 59" \
  60-69="от 60 до 69" 70-79="от 70 до 79" 80-НІ="СВЫШЕ 80" \
  ELSE = .

```

Все, что начинается со звездочки (*), не является обязательным. Это просто произвольные комментарии, которые могут понадобиться, чтобы не забыть, что все это означает. Абзацные отступы тоже необязательны. На самом деле программа могла бы выглядеть и так:

```

* AGEGROUP.PGM
  READ MYFILE
  DEFINE AGE ###
  AGE = (DATEXAM - DOB) DIV 365.25
  DEFINE AGEGR <A      >
  RECODE AGE TO AGEGR 0-2="от 0 до 2" 3-6="от 3 до 6" \
  7-14="от 7 до 14" 15-19="от 15 до 19" 20-29="от 20 до 29" \
  30-39="от 30 до 39" 40-49="от 40 до 49" 50-59="от 50 до 59" \
  60-69="от 60 до 69" 70-79="от 70 до 79" 80-НІ="СВЫШЕ 80" \
  ELSE = .

```

Связанные базы данных. Команда RELATE

Предположим, Вам необходимо создать базу данных для сбора сведений о результатах микробиологического обследования пациентов. Вопросник (в упрощенном виде) может выглядеть следующим образом:

```

N истории болезни ISTVOL<A      > Фамилия И.О. FIO_____
Дата поступления DATAPOSTUP<dd/mm/yyyy> Дата выписки DATAVYPIS<dd/mm/yyyy>
Клинический диагноз KLINDS_____
N анализа ANALNUM<A      > Дата обследования DATANAL<dd/mm/yyyy>
Исслед. материал MATERIAL_____ Микроорганизм VOZB_____

```

Теперь, казалось бы, можно сохранить текст вопросника в файле с расширением .QES (например, micro.qes), затем с помощью программы ENTER создать файл данных (например, micro.rec) и приступить к их вводу. Однако не все так просто. Даже при однократном обследовании от пациента можно выделить несколько культур микроорганизмов различных видов из разных видов клинического материала. Более того, один и тот же пациент во время пребывания в больнице может обследоваться неоднократно (иногда каждый раз выделяя ассоциации микроорганизмов). Если воспользоваться приведенным выше вопросником, каждый раз при вводе нового результата обследования придется вводить все остальные данные об этом пациенте (в нашем примере их не так много, обычно гораздо больше). Попробуем решить проблему и переделаем вопросник так, чтобы учесть возможность выделения ассоциаций микроорганизмов и повторных обследований:

N истории болезни ISTVOL<A > Фамилия И.О. FIO _____
 Дата поступления DATAPOSTUP<dd/mm/yyyy> Дата выписки DATAVYPIS<dd/mm/yyyy>
 Клинический диагноз KLINDS _____

ПЕРВОЕ ОБСЛЕДОВАНИЕ:

Дата обследования DATANAL1<dd/mm/yyyy>
 N анализа ANALNUM11<A > Исслед. материал MATERIAL11 _____
 Микроорганизм 1 VOZB111 _____
 Микроорганизм 2 VOZB112 _____
 Микроорганизм 3 VOZB113 _____
 N анализа ANALNUM12<A > Исслед. материал MATERIAL12 _____
 Микроорганизм 1 VOZB121 _____
 Микроорганизм 2 VOZB122 _____
 Микроорганизм 3 VOZB123 _____

ВТОРОЕ ОБСЛЕДОВАНИЕ:

Дата обследования DATANAL2<dd/mm/yyyy>
 N анализа ANALNUM21<A > Исслед. материал MATERIAL21 _____
 Микроорганизм 1 VOZB211 _____
 Микроорганизм 2 VOZB212 _____
 Микроорганизм 3 VOZB213 _____
 N анализа ANALNUM22<A > Исслед. материал MATERIAL22 _____
 Микроорганизм 1 VOZB221 _____
 Микроорганизм 2 VOZB222 _____
 Микроорганизм 3 VOZB223 _____

Можно продолжать и дальше, но жаль бумаги. Мало того, что количество обследований и выделенных микроорганизмов может быть больше, чем мы предусмотрели (а может быть и меньше, тогда этот громоздкий вопросник останется полупустым). С точки зрения анализа данных это сущий кошмар.

Существует более простое и гораздо более эффективное решение. Нужно создать не один, а два вопросника (назовем их patient.qes и micro.qes).

Вопросник patient.qes должен выглядеть следующим образом:

N истории болезни ISTVOL<A > Фамилия И.О. FIO _____
 Дата поступления DATAPOSTUP<dd/mm/yyyy> Дата выписки DATAVYPIS<dd/mm/yyyy>
 Клинический диагноз KLINDS _____
 Микробиологическое обследование проводилось? OBSLED<Y>

В этом вопроснике содержится информация о пациенте, одна запись на каждого пациента. Поле ISTVOL (номер истории болезни) уникально.

Вопросник micro.qes должен выглядеть так:

N истории болезни ISTVOL<A > N анализа ANALNUM<A >
 Дата обследования DATANAL<dd/mm/yyyy>
 Исслед. материал MATERIAL _____ Микроорганизм VOZB _____

В этом вопроснике содержится информация о результатах обследования, одна запись на каждый выделенный микроорганизм. Этот вопросник тоже имеет поле ISTVOL (номер истории болезни), которое необходимо как для идентификации обследованного пациента, так и для связи двух вопросников между собой.

Процедура ввода данных в связанные базы обсуждается ниже, в главе, посвященной программе CHECK. Для проведения анализа данных, которые содержатся в нескольких связанных файлах, в программе ANALYSIS используется команда RELATE.

Чтобы проанализировать результаты микробиологического обследования, данные о котором содержатся в наших гипотетических файлах patient.rec и micro.rec, нужно выполнить следующие действия:

Сначала нужно открыть (READ) один из двух связанных файлов, а именно тот, который связан с другим соотношением «много к одному». В нашем случае это файл micro.rec, в котором множество записей о выделенных микроорганизмах относятся к одному пациенту (данные о котором хранятся в связываемом файле patient.rec):

```
READ MICRO57
```

Затем нужно выполнить команду RELATE, указав после имени команды имя поля, по которому связаны базы данных (оно является одинаковым для всех связываемых файлов, в рассматриваемом примере это поле ISTVOL) и название файла, который связан с уже открытым файлом соотношением «один ко многим»:

```
RELATE ISTVOL PATIENT
```

Теперь можно выполнять любые действия, обращаясь одновременно к переменным из обоих файлов. Команда RELATE действует все время до завершения текущего сеанса анализа. После выхода из программы все данные по-прежнему будут храниться в отдельных файлах.

Сохранение результатов анализа

Результаты анализа можно увидеть только на экране (этот режим включен по умолчанию), направить на печать (в этом случае у Вас наверняка возникнут проблемы с кириллицей) или сохранить в текстовом (ASCII) файле. Последняя опция наиболее удобна, так как получившийся текст можно редактировать в любом текстовом процессоре. Для управления выводом результатов служит команда ROUTE.

Команда «**ROUTE PRINTER**» позволит распечатать на принтере результаты анализа (то же самое можно сделать, нажав клавишу F5. Команда «**ROUTE SCREEN**» отменит предыдущую команду ROUTE и направит вывод результатов только на экран. А команда **ROUTE <имяфайла>**, где имя файла может быть любым, допустимым в DOS (кроме, конечно, «printer» и «screen»). Рекомендуется задать и расширение имени файла (.TXT) для того, чтобы можно было легко распознать тип файла в системе Windows. Если Вы собираетесь разместить файл с результатами анализа в директории (папке), отличной от установленной в Epi Info по умолчанию, необходимо указать и полный путь к файлу, например:

```
ROUTE C:\DATA\OUTBREAK\OSWEGO.TXT
```

Если файл с выбранным Вами именем уже существует, текущий вывод результатов будет просто добавлен в конец файла. При форматировании текста в редакторе (MS Word, Word Perfect или любом другом) следует выбрать пропорциональный шрифт, например, Courier: это позволит сохранить таблицы и другие элементы текста в том виде, в котором они были представлены на экране.

Следует помнить, что перед построением графика следует ввести команду «**ROUTE SCREEN**», в противном случае содержимое текстового файла со всеми уже сохраненными результатами будет безнадежно испорчено. Если же Вы хотите

⁵⁷ На всякий случай еще раз напоминаем, что после ввода каждой строки команд нужно нажимать 

сохранить в файле построенную диаграмму, для того чтобы в последствии вставить ее в файл MS Word (и т.п.), необходимо предварительно задать несколько команд инициализации. Например, если Вы хотите сохранить секторную диаграмму распределения участников вспышки в г. Освего по полу в графическом файле pie.bmp, следует ввести в командном окне следующий набор команд:

```
SET GRAPH = $BMP
ROUTE PIE.BMP
PIE SEX
```

Наиболее часто применяемые арифметические и логические операторы

+	Сложение
-	Вычитание
*	Умножение
/	Деление
DIV	Деление «нацело». Любой остаток меньше 1.0 отбрасывается.
^	Возведение в степень. X^2 - то же, что X^2 , а $16^{0.5}$ означает $\sqrt{16}$.
(Группировка выражений ⁵⁸
)	Группировка выражений
>	Больше чем
<	Меньше чем
>=	Больше или равно
<=	Меньше или равно
=	Равно
<>	Не равно
AND	Логическое И
OR	Логическое ИЛИ
NOT	Логическое НЕ

Перечисленные операторы (подробный список можно найти в «Руководстве пользователя») применяются для создания выражений, примеры которых уже приводились выше.

Управление вводом данных. Программа СНЕСК

Как бы внимателен ни был оператор, ошибки при вводе данных неизбежны⁵⁹. Bennett и соавт.[7] даже разработали классификацию таких ошибок:

- Транспозиции (напр., 19 при вводе данных превращается в 91)
- Ошибки копирования (напр., 0 (ноль) превращается в O (буква «O»))

⁵⁸ Для соблюдения последовательности действий

⁵⁹ Мы не обсуждаем здесь источники ошибок при сборе данных: с ними Epi Info ничего поделать не может.

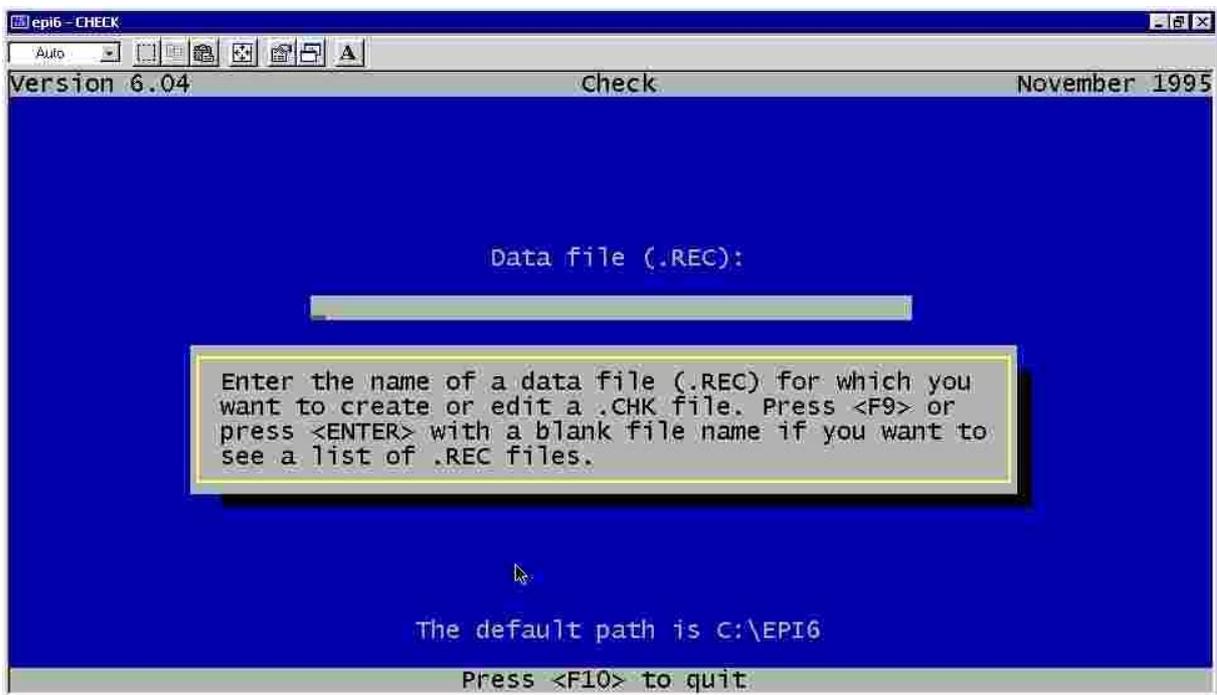
- Ошибки кодирования (напр., при изменении схемы кодирования)
- Ошибки маршрутизации, связанные с нарушением последовательности ответов (напр., вопросы задавались не те, или в последовательности, не соответствующей структуре вопросника)
- Ошибки, связанные с нарушением логической последовательности (противоречивые ответы, напр., если в поле «Аллергия на пенициллин» стоит «Да» после отрицательного ответа на вопрос об аллергических реакциях в анамнезе)
- Ошибки диапазона (ответы за рамками допустимого диапазона, напр. Возраст 900 лет вместо 90)

Даже после того, как бумажные формы сбора данных приведены в порядок (Вы убедились в полноте заполнения форм, разобрали почерк, проверили кодировку, выявили логические ошибки), сохраняется вероятность ошибок при вводе данных, а, значит, и необходимость проверки правильности ввода. Для этого в пакете Epi Info имеется специальная программа СНЕСК⁶⁰. Программа СНЕСК способна не только регулировать правильность ввода данных, но и облегчить его.

Для запуска программы СНЕСК нужно выбрать из меню «Программы» пункт «СНЕСК Управление Вводом Данных» и нажать клавишу .

Посреди стартового экрана (см. рис. 15) под заголовком «Data file (.REC)» расположено окно, в которое нужно ввести имя файла данных, для которого создается файл управления вводом (после сохранения он будет иметь то же имя, что и файл данных, но с расширением .CHK).

Рис. 15. Стартовый экран программы СНЕСК



После ввода имени файла данных (можно воспользоваться клавишей F9, чтобы вызвать список всех файлов с расширением .REC) появляется главный экран программы СНЕСК, напоминающий, на первый взгляд, окно программы ENTER (рис.16)

⁶⁰ Есть еще модуль VALIDATE для верификации ввода путем сравнения двух файлов данных, введенных разными операторами из одного и того же источника. В данном руководстве не обсуждается.

Рис. 16 Экран программы CHECK при работе с открытым файлом данных

ерпб - CHECK

Auto

МАССОВОЕ ПИЩЕВОЕ ОТРАВЛЕНИЕ В ЛИЦЕЕ

N ФАМИЛИЯ ВОЗРАСТ ГРУППА ОБЩЕЖИТИЕ

ЗАБОЛЕЛ?

КЛИНИЧЕСКИЕ ПРОЯВЛЕНИЯ:

Температура Боль в животе Жидкий стул Кратность стула

Тошнота Рвота Кратность рвоты Головная боль Слабость

ЛАБОРАТОРНЫЕ ДАННЫЕ:

Патогенная микрофлора Дата обследования ПМФ

Условно-патогенная флора Дата обследования УПФ

Вирусы Дата обследования Вир

ДАТА ЗАБОЛЕВАНИЯ ВРЕМЯ ДАТА ОБРАЩЕНИЯ

ДАТА РЕГИСТРАЦИИ ГОСПИТАЛИЗИРОВАН ДАТА ГОСПИТАЛИЗАЦИИ

ДИАГНОЗ: Первоначальный Заключительный

ПИЩЕВОЙ АНАМНЕЗ:

N: No entries allowed

F1/F2-Min/Max F5-Link fields Legal: F6-Add Shift-F6-Display Ctrl-F6-Delete

F3-Repeat F9-Edit field Jump: F7-Add Shift-F7-Display Ctrl-F7-Delete

F4-MustEnter F10-Quit Codes: F8-Add Shift-F8-Display Ctrl-F8-Delete

Внизу экрана расположен список команд, выполняемых программой CHECK и соответствующих им функциональных клавиш:

- F1 F2 **Min/Max** – задание диапазона допустимых значений в поле
- F3 **Repeat** – автоматический ввод значения предыдущей записи
- F4 **MustEnter** – установка правила обязательного ввода данных в поле
- F5 **Link fields** – связывание полей
- F6 **Legal: Add** – добавление допустимых значений
- Shift F6 **Legal: Display** – показ допустимых значений
- Ctrl F6 **Legal: Delete** – удаление допустимых значений
- F7 **Jump: Add** – добавление переходов
- Shift F6 **Jump: Display** – показ переходов
- Ctrl F6 **Jump: Delete** – удаление переходов
- F8 **Codes: Add** – добавление кодов
- Shift F6 **Codes: Display** – показ кодов
- Ctrl F6 **Codes: Delete** – удаление кодов
- F9 **Edit fields** – редактирование полей
- F10 **Quit** – завершение работы программы

Основные команды программы СHECK

Задание диапазона допустимых значений

Для задания нижней границы диапазона нужно ввести минимально допустимое значение в поле вопросника и нажать F1. После этого в это же поле нужно ввести максимально допустимое значение (верхняя граница диапазона) и нажать F2. Если Вы проделали эту процедуру для задания диапазона возможных значений возраста (напр., 0–100) в поле AGE, в нижнем левом углу экрана появится сообщение

```
AGE: VALID VALUES: 0 TO 100
```

Теперь при вводе данных в программе ENTER оператору при всем желании не удастся ввести возраст 900 лет или даже 101 год.

Автоматический ввод значения предыдущей записи

Программа СHECK дает возможность автоматического повторения предыдущей записи. Если, например, приходится вводить данные о сотнях людей, которые находятся в зоне обслуживания одной поликлиники, совсем необязательно каждый раз вводить название (или код) этого учреждения. Достаточно установить курсор в поле «Поликлиника» и нажать F3. Никакое сообщение о результатах выполнения команды в левом нижнем углу не появится, но при вводе данных в программе ENTER каждый раз при переходе в это поле будет появляться предыдущая запись, до тех пор, пока Вы не перейдете к другой поликлинике и не введете новое значение⁶¹.

Обязательный ввод данных

В некоторых случаях нужно быть уверенным, что оператор не забудет ввести данные в поле, без записи в котором дальнейший ввод данных и анализ становится бессмысленным. Для этого нужно установить курсор в такое поле и нажать F4. В нижней строке появится сообщение «You must enter data». Теперь при вводе данных в программе ENTER оператор не сможет продолжать ввод, не введя данные в поле, где ввод обязателен.

Переходы

Иногда вопросники могут содержать поля, которые при вводе можно (и нужно) пропустить. Например, при отсутствии заболевания следует пропустить поля, в которых содержится информация о дате и симптомах заболевания. Для того, чтобы избежать при этом утомительного перемещения курсора вручную, можно задать условия, при которых курсор автоматически перемещается на следующее (в порядке логической последовательности) поле. Если, например, в поле ILL (файл oswego.rec, см. рис.6) ввести «N», и нажать F7, в нижней строке экрана появится сообщение «Put cursor on field to jump to on entry of N and press F7»⁶². После этого курсор следует переместить в поле BAKEDHAM и снова нажать F7. Теперь, если при вводе данных на вопрос «Заболел?» (поле ILL) дать отрицательный ответ, курсор сразу переместится в область данных об эпидемиологическом анализе, минуя поля ONSETDATE («Дата заболевания») и ONSETTIME («Время заболевания»)

Кроме того, команда JUMP может понадобиться для перехода в специальные области вопросника, где содержатся вопросы, имеющие значение только в определенных случаях (например, ввод данных в поле «Желтуха» понадобится, если в поле «Диагноз» ввели «гепатит»).

⁶¹ Оно теперь тоже будет повторяться.

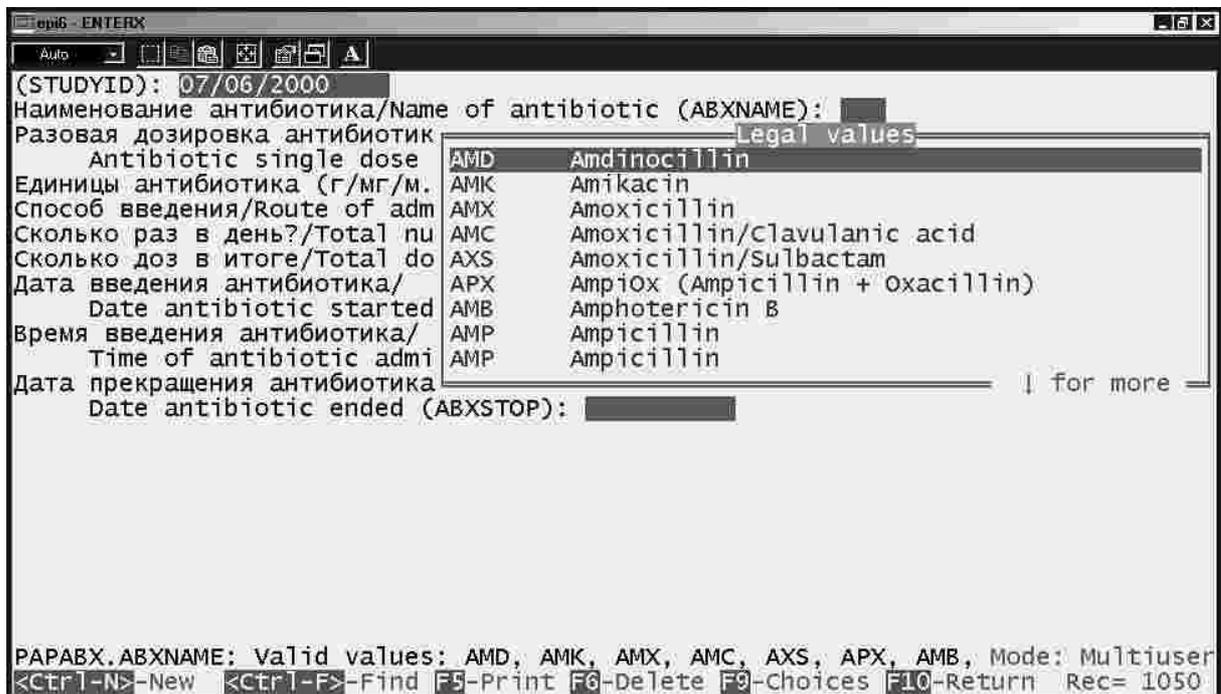
⁶² «Поставьте курсор на поле, на которое следует перепрыгнуть при вводе N и нажмите F7»

Допустимые значения

Предположим, что в Вашем вопроснике, кроме всех прочих, имеется, например, поле «Диагноз». Если при вводе данных с клавиатуры оператор ошибется, и сделает опечатку, в поле «Диагноз», вместо записи «Шигеллез Зонне» может появиться, скажем, «Шигиллез Зонне». При анализе данных компьютер будет считать, что имеет дело с двумя разными диагнозами. Кроме того, довольно утомительно каждый раз вводить диагнозы с клавиатуры.

Программа СНЕСК позволяет создавать справочники, содержащие перечень допустимых значений. Для этого нужно в поле, для которого создается справочник, последовательно ввести все допустимые значения, каждый раз нажимая клавишу F6. В результате настройки, при вводе данных в программе ENTER достаточно нажать в нужном поле клавишу F9, выбрать из разворачивающегося списка необходимое значение и нажать  (рис. 17).

Рис. 17. Экран программы ENTER со списком допустимых значений поля.



СНК-файлы: сохранение, редактирование

После завершения всех операций в программе СНЕСК следует выйти из программы обычным способом (F10) и утвердительно (Y) ответить на запрос программы о сохранении файла. После этого будет автоматически создан файл со всеми только что созданными настройками с расширением .СНК. При необходимости этот файл можно редактировать в программе EPED или любом другом текстовом процессоре.

Приведенные выше примеры использования программы СНЕСК далеко не исчерпывают всех возможностей управления вводом данных. С помощью СНК-файлов можно создавать вычисляемые поля, обеспечивать вывод контекстной помощи и многое другое (подробно об этом можно прочитать в Руководстве пользователя).

В частности, как уже говорилось выше, с помощью файлов .СНК можно организовать ввод данных в связанные файлы. Для этого общие поля, по которым связываются файлы, должны быть проиндексированы (в СНК-файлах для каждого связываемого файла после названия поля должно стоять слово KEY). В СНК-файле для главного файла данных (файл с соотношением «один ко многим») после названия общего поля

должны следовать слова KEY UNIQUE. После названия того поля в главном файле, при переходе к которому происходит связывание, должна следовать команда RELATE, название общего поля и название связываемого файла. Если переход не является обязательным, а осуществляется по определенным условиям, команде RELATE может предшествовать соответствующее выражение IF-THEN. Для примера можно снова воспользоваться приведенной выше базой данных для ввода результатов микробиологического обследования (файлы patient.rec и micro.rec, см. с.36).

В файле patient.chk после названия общего поля (ISTBOL) должна быть запись KEY UNIQUE:

```
ISTBOL
      KEY UNIQUE
```

В файле micro.chk после ISTBOL должна быть запись KEY:

```
ISTBOL
      KEY
```

Кроме того, в файле patient.chk после названия поля OBSLED должны следовать следующие команды:

```
IF OBSLED = "Y"
THEN
      RELATE ISTBOL MICRO
ENDIF
```

Теперь при вводе данных в программе ENTER достаточно открыть файл patient.rec. При утвердительном ответе на вопрос о проведении микробиологического обследования откроется форма для ввода данных в micro.rec. После завершения ввода данных о выделенных возбудителях можно вернуться в patient.rec, нажав F10.

Статистический калькулятор STATCALC

Рис. 18 Главный экран программы STATCALC



Для запуска программы STATCALC нужно выбрать из меню «Программы» пункт «STATCALC Стат. калькулятор» и нажать клавишу .

Программа STATCALC может оказаться полезной, если Ваши данные уже сгруппированы в таблицу 2x2 (2xN).

После выбора из меню пункта «Tables <2 x 2, 2 x n>» (нужно нажать клавишу ) , появляется окно ввода табличных данных.

Рис. 19. Окно ввода табличных данных программы STATCALC



Экранная форма представляет собой классическую четырехпольную таблицу сопряженности:

		Заболевание	
		+	-
Экспозиция	+	a	b
	-	c	d

Теперь достаточно ввести в эту форму соответствующие данные и нажать F4 для проведения вычислений. STATCALC вычисляет отношение пробладаний и относительный риск с доверительными интервалами, хи-квадрат (тремя способами!), а также точный критерий *Fisher*:

```

+ Disease -
+-----+-----+
+| 5 | 30 | 35
+-----+-----+
-| 2 | 57 | 59
+-----+-----+
E 7 87 94
x
p
o
s
u
r
e

Analysis of Single Table
Odds ratio = 4.75 (0.75 <OR< 37.91*)
Cornfield 95% confidence limits for OR
*Cornfield not accurate. Exact limits preferred.
Relative risk = 4.21 (0.86 <RR< 20.58)
Taylor Series 95% confidence limits for RR
Ignore relative risk if case control study.

Chi-Squares P-values
-----
Uncorrected : 3.78 0.0517436
Mantel-Haenszel: 3.74 0.0530043
Yates corrected: 2.37 0.1238245
Fisher exact: 1-tailed P-value: 0.0642775
                2-tailed P-value: 0.0976064

An expected cell value is less than 5.
Fisher exact results recommended.

F2 More Strata; <Enter> No More Strata; F10 Quit

```

По-русски это выглядело бы так:

		+ Влезнь -			
+	5	30		35	
-	2	57		59	
Э	7	87		94	

Анализ одиночной таблицы
 Отношение преобладаний (OR) = 4.75 (0.75 <OR< 37.91*)
 95% доверительные интервалы для OR по Cornfield
 *по Cornfield не корректно. Препочтительны точные интервалы.
 Относительный риск = 4.21 (0.86 <RR< 20.58)
 95% доверительные интервалы для RR по Taylor
 Игнорируйте относительный риск в исследовании случай-контроль

	Хи-квадраты	P-значения
Без коррекции:	3.78	0.0517436
Mantel-Haenszel:	3.74	0.0530043
С коррекцией по Yates:	2.37	0.1238245
Точный критерий Fisher:	1-сторонний P-value: 0.0642775	
	2-сторонний P-value: 0.0976064	

Ожидаемое значение в ячейке таблицы менее 5.
 Рекомендуются точные результаты по Fisher.

Нажатие клавиши F2 позволяет вводить очередные страты при проведении стратификационного анализа с вычислением после ввода последней страты соответствующих статистических показателей. Выход из программы осуществляется, как всегда, по F10.

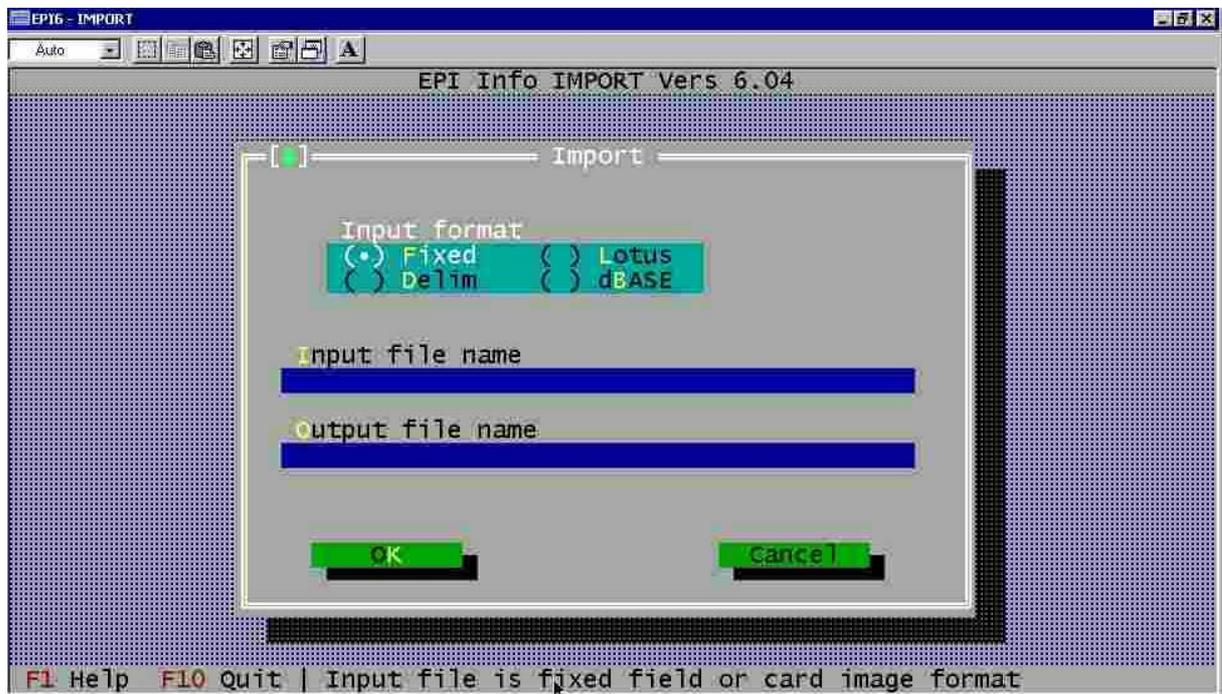
STATCALC позволяет выполнять и другие статистические процедуры, которые в данном руководстве не обсуждаются

Импорт и экспорт данных

Epi Info имеет достаточно широкие возможности для экспорта (программа EXPORT) и импорта (программа IMPORT) данных.

Для запуска программы IMPORT нужно выбрать из меню «Программы» пункт «IMPORT Импорт файлов» и нажать клавишу .

Рис. 20. Главный экран программы IMPORT

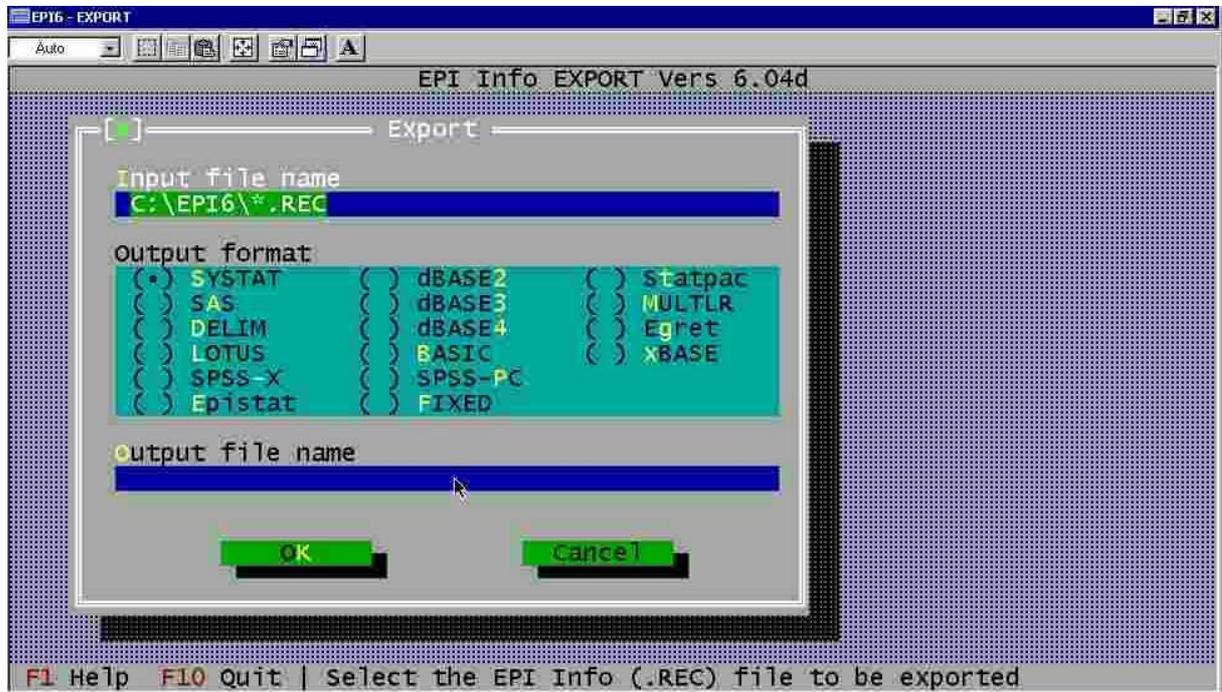


Epi Info способна импортировать файлы четырех форматов: 2 текстовых формата, Lotus (.WKS) и dBase (.DBF). Практически любая компьютерная программа включает в себя средства сохранения своих файлов в этих форматах.

Для импорта в окне с заголовком «**Input file name**» нужно ввести имя импортируемого файла, а в окне «**Output file name**» - желаемое имя файла .REC. Для завершения процедуры следует нажать экранную клавишу с надписью «**ОК**».

Для запуска программы EXPORT нужно выбрать из меню «**Программы**» пункт «**EXPORT Экспорт файлов**» и нажать клавишу .

Рис. 21. Главный экран программы EXPORT



Epi Info может экспортировать данные в форматы наиболее популярных пакетов статистической обработки (Systat, SAS, SPSS, Epistat, Statpac, Egret), формат dBase II, III, и IV (.DBF), формат Lotus 1-2-3 (.WKS) и некоторые другие. Этого вполне достаточно, чтобы экспортировать данные Epi Info практически в любой пакет, где, в случае необходимости, можно продолжить их обработку.

Для экспорта данных в окне с заголовком «**Input file name**» нужно ввести имя экспортируемого файла .REC, а в окне «**Output file name**» - желаемое имя файла. Для завершения процедуры следует нажать «**ОК**».

Литература:

- 1 Прикладная медицинская статистика / Зайцев В.М., Зуева Л.П., Лифляндский В.Г., Маринкин В.И., Самусь А.А., Еремин С.Р.; под ред. В.М.Зайцева, В.Г.Лифляндского. – СПб.: СПбГМА им И.И.Мечникова, 2000. – 300 с.
- 2 Harbage B, Dean AG. Distribution of epi info software: an evaluation using the Internet. Am J Prev Med. 1999 May;16(4):314-7
- 3 Dean AG, Dean JA, Coulombier D, Brendel KA, Smith DC, Burton AH, Dicker RC, Sullivan K, Fagan RF, Arner TG. Epi Info Version 6: a word processing, database, and statistics program for public health on IBM-compatible microcomputers. Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A., 1995.
- 4 Dean AG, Arner TG, Sangam S, Sanki GG, Friedman R, Latinga M, Zubieta JC, Sullivan KM, Smith DC. Epi Info 2000, a database and statistics program for public health

professionals for use on Windows 95, 98, NT, and 2000 computers. Centers for Disease Control and Prevention, Atlanta, Georgia, U. S. A., 2000.

5 Lauritsen JM, Bruus M., Myatt MA, EpiData, version 2.0. An extended tool for validated entry and documentation of data. The EpiData Association, Odense Denmark, 2001

6 Brown SM, Eremin SR, Zueva LP, O'Rourke EJ, eds. Hospital Epidemiologist's Workstation, version 2.0. American International Health Alliance, Washington, D.C., USA, January 2001

7 Bennett, S., Myatt, M., Jolley, D., & Radalowicz. A. (1996). Data Management for Surveys and Trials. A Practical Primer Using Epi Info. Llanidloes, Powys, Great Britain: Brixton Books.