

В серии:

Библиотека ALT Linux

Gnumeric

Электронная таблица для всех

И. А. Хахаев

Москва
ALT Linux
2011

УДК 004.91
ББК 32.97
Х27

Gnumeric: Электронная таблица для всех / И. А. Хахаев
Х27 — М. : ALT Linux, 2011. — 192 с. : ил. — (Библиотека ALT Linux).

ISBN 978-5-905167-05-8

Электронная таблица Gnumeric часто воспринимается как «облегченный» вариант офисной электронной таблицы OpenOffice.org/LibreOffice Calc. Однако более тесное знакомство с Gnumeric открывает её особенности и возможности, которыми не обладают другие пакеты электронных таблиц. Предлагаемая книга как раз и является попыткой показать эти возможности и особенности для читателей, в принципе знакомых с программами такого класса. Однако можно использовать эту книгу в качестве введения в электронные таблицы для школьников, студентов и специалистов в различных областях.

Сайт книги: <http://www.altlinux.org/Books:Gnumeric>

Книга адресована широкому кругу читателей, использующих электронные таблицы в своей учебной, научной или профессиональной деятельности.

УДК 004.91
ББК 32.97

Материалы, составляющие данную книгу, распространяются на условиях лицензии GNU FDL. Книга содержит следующий текст, помещаемый на первую страницу обложки: «В серии “Библиотека ALT Linux”». Название: «Gnumeric: Электронная таблица для всех». Книга не содержит неизменяемых разделов. Авторы разделов указаны в заголовках соответствующих разделов. ALT Linux — торговая марка компании ALT Linux. Linux — торговая марка Линуса Торвальдса. Прочие встречающиеся названия могут являться торговыми марками соответствующих владельцев.

ISBN 978-5-905167-05-8

© Хахаев И. А., 2011
© ALT Linux, 2011

Оглавление

Оглавление	3
Глава 1. Введение	5
1.1 Предмет описания, структура и особенности книги .	5
1.2 Основные достоинства Gnumeric	6
1.3 Справочная система и источники информации . . .	7
1.4 Условия использования и распространения материала	7
Глава 2. Основы работы в Gnumeric	9
2.1 Управление файлами	9
2.2 Управление листами	15
2.3 Управление ячейками	20
2.4 Автозаполнение: генерация рядов данных	33
2.5 Формулы. Абсолютная и относительная адресация .	39
2.6 Функции	44
Глава 3. Работа со списками	67
3.1 Сортировка	67
3.2 Автофильтр	71
3.3 Расширенный фильтр	76
3.4 Функции базы данных	81
Глава 4. Диаграммы в Gnumeric	84
4.1 Принципы создания и форматирования диаграмм .	84
4.2 Несколько диаграмм на графике	90
4.3 Диаграммы для отображения коммерческих данных	94
4.4 Отображение научных данных	101
4.5 Статистические диаграммы	107
Глава 5. Инструменты Gnumeric для статистиков	112
5.1 Основные принципы создания и форматирования диаграмм	112
5.2 Описательные статистики	112
5.3 Прогнозирование	116
5.4 Корреляция	120

5.5	Ковариация	120
5.6	Регрессия	122
5.7	Анализ Фурье	124
5.8	Гистограмма	125
5.9	Выборка	128
5.10	Ранги и проценты	131
5.11	Дисперсионный анализ	132
5.12	Два средних	137
5.13	Две дисперсии: F-тест	143
5.14	Оценка выживаемости	144
Глава 6. Регрессионный анализ в Gnumeric		151
6.1	Небольшое теоретическое введение	151
6.2	Реализация вычислений на модели	152
Глава 7. Линейная оптимизация		159
7.1	Оптимизация как задача линейного программирования	159
7.2	Пример задачи линейной оптимизации	160
7.3	Решение в Gnumeric	161
Глава 8. Подбор параметра		165
Глава 9. Моделирование рисков методом Монте-Карло		170
9.1	Общее описание задачи	170
9.2	Построение модели	171
9.3	Использование модели	175
Список литературы		183
Список иллюстраций		184
Список таблиц		190

Глава 1

Введение

1.1 Предмет описания, структура и особенности книги

Gnumeric — это электронная таблица (ЭТ), которая разрабатывается как проект с открытым исходным кодом (Free and Open Source Software — FOSS в англоязычной терминологии, свободное программное обеспечение — СПО — в российской терминологии). Gnumeric разрабатывается сообществом программистов как часть проекта GNOME Office, однако она существует (и может быть установлена) независимо от других компонентов GNOME Office.

Пакет является кросс-платформенным, основан на интерфейсной библиотеке GTK+ и распространяется по лицензии GPL (GNU General Public License). Версии для Linux и xBSD-систем входят в пакетную базу практически всех современных дистрибутивов. Версии для Windows — отдельные проекты, установочные файлы можно найти на сайте проекта (<http://projects.gnome.org/gnumeric/>) или с помощью поисковых систем Интернета (Yandex, Google и пр.).

Обычное применение электронных таблиц — офисная деятельность. По сравнению с широко распространёнными офисными пакетами ЭТ в Gnumeric отсутствует представление табличных данных с вычислением промежуточных итогов, минимизирована система интерактивных фильтров, а для получения сводной таблицы требуется использовать встроенную функцию. Однако Gnumeric имеет богатые возможности обработки и отображения результатов обработки данных, вследствие чего этот пакет можно позиционировать скорее не как офисный, а как инструмент начального уровня для обработки числовых данных в инженерных задачах и в статистике.

Первые три главы описывают базовые возможности Gnumeric как электронной таблицы — работу с объектами таблицы, построение

формул, использование различных встроенных функций, возможности поиска информации в списках и построение диаграмм.

В отдельной главе рассматриваются некоторые возможности Gnuplot как инструмента для статистического анализа данных, в частности, такие специальные возможности как проверка гипотез и оценка выживаемости по Каплану-Майеру.

Затем рассматриваются отдельные частные задачи — применение Gnuplot для регрессионного анализа (нелинейной подгонки), решение задач линейной оптимизации и поиска корней уравнений (подбор параметра).

В последней главе рассматривается применение Gnuplot для построения экономико-математической модели с использованием метода Монте-Карло.

Проект Gnuplot быстро развивается усилиями международного сообщества, новые версии появляются несколько раз в год. За время подготовки этой книги также произошли некоторые изменения, поэтому редакторский коллектив взял на себя труд заменить некоторые иллюстрации для обеспечения соответствия версии, актуальной на момент предпечатной подготовки.

1.2 Основные достоинства Gnuplot

Сравнение различных программных средств — бесполезное и неблагодарное занятие. Любая программа создаётся как инструмент для решения определённых задач, и в тех случаях, когда задачи и условия их решения соответствуют замыслу создателей программы, она и является лучшей. Поэтому ниже приведены достоинства пакета Gnuplot с позиции автора в порядке убывания важности. Понятно, что у каждого пользователя может существовать своя система ценностей.

1. Свободная лицензия, что для пользователей означает возможность использовать пакет в любых целях на законных основаниях без всяких ограничений и отчислений.
2. Очень высокое качество диаграмм, большие возможности их форматирования и экспорта в различные графические форматы.
3. Кросс-платформенность, которая обеспечивает совместимость на уровне пользовательских файлов.
4. Поддержка стандартных (в смысле стандартов ISO) и распространённых форматов файлов.
5. Отсутствие ограничений количества листов в файле электронной таблицы (Нужно заметить, что нет никаких данных по

ограничению количества листов. Опытным путём удалось сделать 365 — по количеству дней в году.)

6. Наличие большого набора встроенных инженерных и статистических функций.
7. Возможность изменения цвета «ярлычков» для листов ЭТ и цвета текста для имён листов ЭТ.
8. Возможность форматирования отдельных элементов текста в ячейках ЭТ (отдельные слова и символы в ячейке ЭТ можно выделять цветом и шрифтом).
9. Возможность создания пользовательских функций (расширений) на языках Perl и Python.

1.3 Справочная система и источники информации

В качестве справочной системы можно использовать сетевую документацию (<http://projects.gnome.org/gnumeric/doc/gnumeric.shtml>) и подсказки по встроенным функциям. Однако и то и другое, скорее всего, будет на английском языке. Несмотря на достаточно «почтенный» возраст (проект Gnumeric старше проекта OpenOffice.org), информационных ресурсов, кроме сайта проекта, практически нет. Статья в Википедии на английском языке очень короткая, а на русском — ещё короче.

Отдельные короткие заметки в русском сегменте Интернета (сайты citkit.ru, orepnet.ru, linuxcenter.ru) не содержат сколько-нибудь полезной информации. Таким образом получается, что данная книга является первой попыткой создания пользовательской документации по Gnumeric на русском языке.

Часть примеров, рассмотренных в этой книге, взята из книги «OpenOffice.org: Теория и практика», частично использованы материалы из сетевой документации, некоторые примеры придуманы автором.

1.4 Условия использования и распространения материала

Материал, изложенный в этой книге, не является переводом оригинальной документации, не заменяет её и не претендует на полноту изложения. Автор рассматривает те возможности и особенности пакета, которые ему интересны или на которые, по мнению автора, нужно

обратить внимание. Возможно, что у другого автора получилась бы другая книга.

Содержание этой книги распространяется на условиях лицензии GNU Free Documentation License (FDL), что означает возможность его копирования и распространения в любых целях неограниченно.

Файлы примеров опубликованы на сайте книги (<http://www.altlinux.org/Books:Gnumeric>) и также могут свободно и неограниченно использоваться в личных, учебных и прочих целях. Автор надеется, что эта книга будет полезна как людям, уже имеющим опыт работы с электронными таблицами, так и начинающим пользователям.

Глава 2

Основы работы в Gnumeric

2.1 Управление файлами

При запуске Gnumeric получаем достаточно стандартный вид приложения электронной таблицы (ЭТ), рис. 2.1.

Документ ЭТ состоит из листов, каждый лист электронной таблицы может иметь переменное число строк и столбцов (см. далее главу «Управление листами»), а количество листов может быть более 256

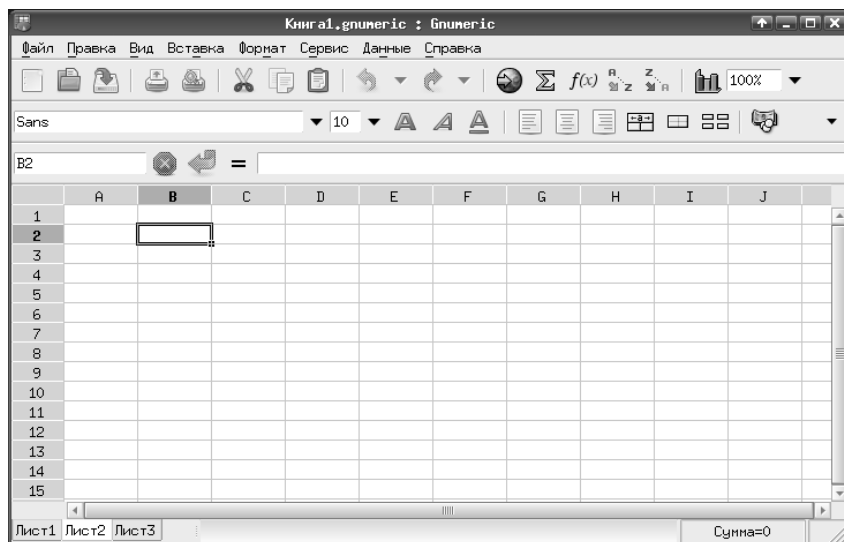


Рис. 2.1. Общий вид окна ЭТ Gnumeric

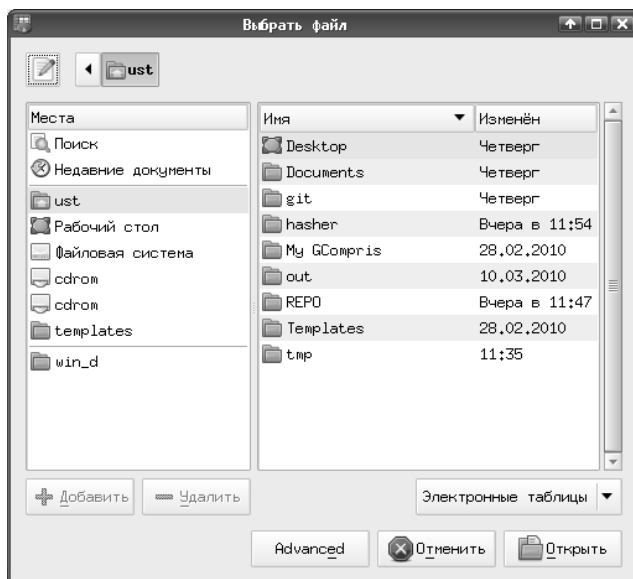


Рис. 2.2. Простой вид диалога открытия файла

(как уже упоминалось во «Введении», сведений об ограничении количества листов ЭТ автору обнаружить не удалось).

Для открытия файла можно использовать пиктограмму «Открыть файл» в верхней панели инструментов (с изображением «папки»), команду главного меню «Файл/Открыть» или комбинацию клавиш `CTRL+O` (буква `O` — от слова «Open»). В результате появится GTK-диалог открытия файла (рис. 2.2). Поскольку подобные диалоги присутствуют во многих кросс-платформенных GTK-приложениях (в частности, в GIMP и в Inkscape), рассмотрим некоторые особенности этого диалога.

Самая левая кнопка в верхней части диалогового окна «отвечает» за ввод или отображение имени открываемого файла. Если она нажата, в диалоге появляется текстовое поле ввода «Расположение:» (см. рис. 2.3), которое даёт возможность сразу ввести полное имя нужного файла.

Также в верхней части диалога расположена строка указания пути к нужному файлу, причём каждому каталогу в этом пути соответствует кнопка. Если первая кнопка в пути — кнопка со стрелкой, это означает, что путь строится относительно «домашних» каталогов (точка монтирования `/home` в POSIX-системах). Если на эту кнопку со стрелкой нажать, увидим абсолютный путь от начала дерева каталогов («Файловая система» или точка монтирования `/`).

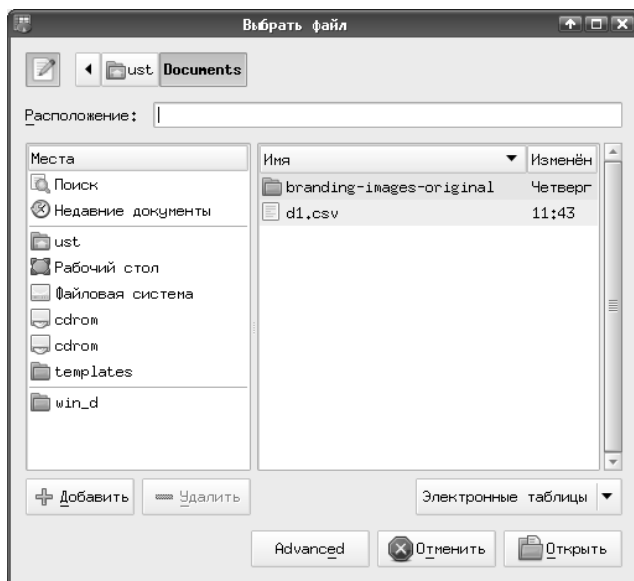


Рис. 2.3. Диалог открытия файла со строкой ввода имени и просмотр содержимого каталога

Основная часть диалогового окна состоит из двух панелей. Левая панель — «Места» — состоит из трёх секций. Верхняя секция позволяет выбрать для открытия один из документов, с которым недавно работали («Недавние документы»).

Средняя секция показывает стандартный набор каталогов, в которых могут находиться пользовательские файлы. В этот стандартный набор входят домашний каталог пользователя, «Рабочий стол» пользователя, а также начало дерева каталогов («Файловая система» или точка монтирования /).

В нижнюю секцию пользователь может добавлять свои часто используемые каталоги, выбрав их в правой панели и нажав кнопку «Добавить». Это работает для каталогов, отмеченных курсором (подсветкой), за исключением каталога **Desktop** («Рабочий стол»).

Важно заметить, что для перехода в новое «место» (в левой панели диалога) требуется одиночный щелчок левой кнопкой мыши, а для открытия каталога в правой панели диалога нужен двойной щелчок левой кнопки мыши.

По мере продвижения «вглубь» иерархии каталогов файловой системы в верхней части диалога увеличивается количество кнопок, показывающих путь к открываемому файлу.

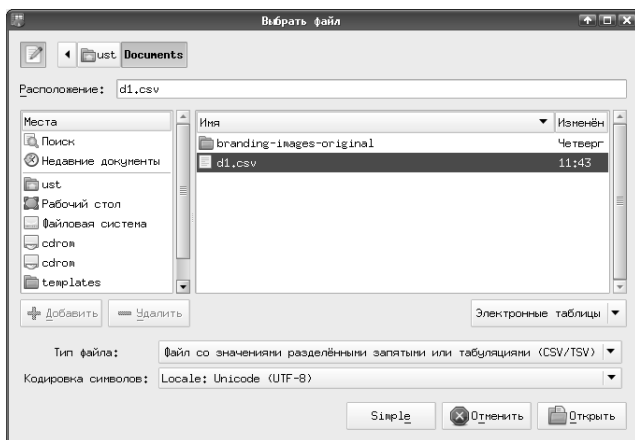


Рис. 2.4. Диалог открытия файла с дополнительными возможностями выбора типа файла и кодировки

Нажатие на кнопку "Advanced" («Расширенный») открывает ещё две настройки в этом диалоге (рис. 2.4) — возможность выбора типа файла и возможность выбора кодовой страницы (кодировки) для текстовых файлов.

Пусть в выбранном каталоге находится текстовый файл формата CSV (Comma Separated Values — текст, разделённый запятыми). Тогда после выбора соответствующего типа и кодовой страницы и нажатия на кнопку «Открыть» файл окажется импортированным в лист ЭТ, причём Gnumeric правильно распознает текст и числа (рис. 2.5). Дело в том, что ЭТ автоматически выравнивает текст по левому краю ячеек, а числа — по правому. Более подробно форматирование данных в ячейках будет рассматриваться в главе «Управление ячейками».

Для сохранения результатов работы используются кнопка с пиктограммой «диска» или «дискеты» («Сохранить текущую книгу»), либо команда главного меню «Файл/Сохранить», либо комбинация клавиш CTRL+S (буква S — от "Save"). Для изменения имени (включая расположение) или типа файла используется команда «Файл/Сохранить как...» или комбинация клавиш SHIFT+CTRL+S. При любом варианте вызова команды сохранения в первый раз открывается диалог сохранения файла (рис. 2.6). При последующих сохранениях он не открывается, если не выбрана команда «Сохранить как...».

О не сохранённых изменениях в файле свидетельствует символ «звёздочка» (*) в строке заголовка окна перед именем файла (в нашем примере получилось бы *d1.csv).

Щелчок левой кнопкой мыши по треугольничку слева от надписи «Просмотреть другие папки» переключает полный и краткий вари-

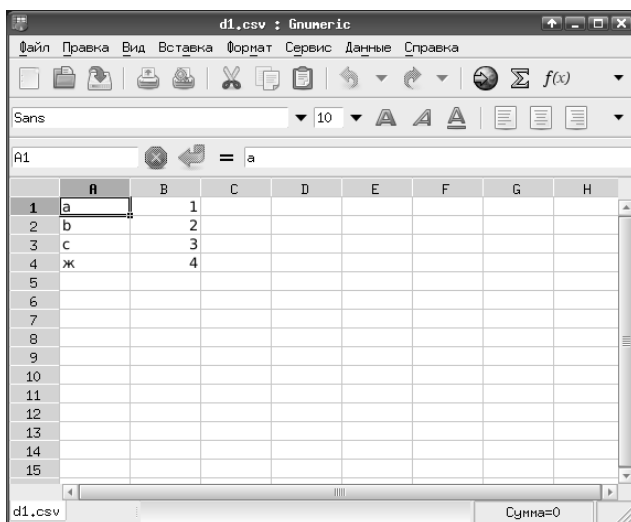


Рис. 2.5. Результат импорта текстового файла

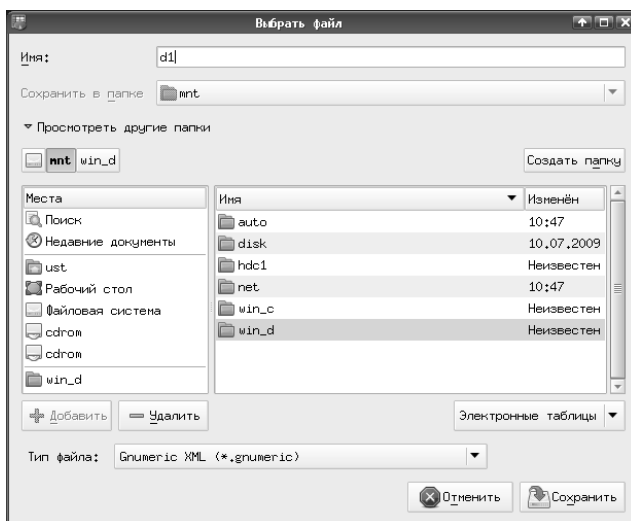


Рис. 2.6. Диалог сохранения файла (полный вариант)

анты диалога сохранения файла. В кратком варианте можно задать или изменить имя и тип файла, а в полном — ещё и выбрать каталог для сохранения.

Импорт файлов (открытие, загрузка) возможен из следующих известных форматов:

- Applix (*.as)
- Data Interchange Format (.dif)
- GNU Oleo (*.oleo)
- Gnumeric XML (*.gnumeric)
- HTML (*.html, *.htm)
- Lotus 123 (*.wk1, *.wks, *.123)
- MS Excel (tm) (*.xls)
- MS Excel (tm) 2003 SpreadsheetML
- MS Excel (tm) 2007
- MultiPlan (SYLK)
- Quattro Pro (*.wb1, *.wb2, *.wb3)
- SC/xspread
- Импорт текстового файла (настраиваемый)
- Импорт файлов в формате Plain Perfect (PLN)
- Файл базы данных или первичного индекса Paradox (*.db, *.px)
- Файл в формате "Linear and integer program" (*.mps)
- Файл со значениями разделёнными запятыми или табуляциями (CSV/TSV)
- Формат Open Document (*.sxc, *.ods)
- Формат файла XBase (*.dbf)

Экспорт файлов (сохранение, выгрузка) возможен в следующие распространённые форматы:

- Data Interchange Format (.dif)
- GLPK Linear Program Solver
- Gnumeric XML (*.gnumeric)
- HTML 3.2 (*.html)

- HTML 4.0 (*.html)
- LaTeX 2e (*.tex)
- LaTeX 2e (*.tex) фрагмент таблицы
- PLSolve Linear Program Solver
- MS Excel (tm) 2007
- MS Excel (tm) 5.0/95
- MS Excel (tm) 97/2000/XP
- MS Excel (tm) 97/2000/XP и 5.0/95
- MultiPlan (SYLK)
- ODF/OpenDocument без дополнительных элементов (*.ods)
- ODF/OpenDocument с дополнительными элементами (*.ods)
- TROFF (*.me)
- XHTML (*.html)
- База данных Paradox (*.db)
- Значения разделённые запятыми (CSV)
- Текст (настраиваемый)
- Фрагмент HTML (*.html)
- Экспорт в PDF

При экспорте в текстовый файл также возможно указание кодировки выходного файла, а также символа-разделителя и варианта окончания строки (в стиле UNIX, Mac OS X или Windows).

При экспорте в PDF экспортируются все листы с колонтитулами, форматированием и диаграммами (если они есть). Таким образом, экспорт в PDF равносителен печати всего документа в файл.

2.2 Управление листами

Для изменения названий (имён) листов ЭТ можно использовать контекстное меню (рис. 2.7), вызываемое щелчком правой кнопкой мыши по «ярлычку» листа. Выбор пункта «Управление листами...» вызывает диалог управления свойствами листов (рис. 2.8).

В принципе, все операции с листами, которые можно делать с помощью контекстного меню, выполняются в этом диалоге. Лист выбирается щелчком левой кнопкой мыши по соответствующей строчке

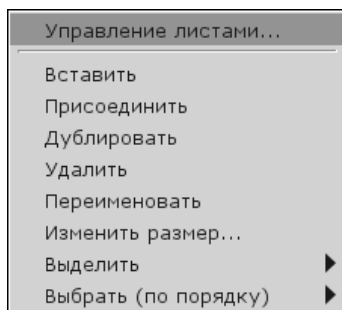


Рис. 2.7. Контекстное меню листа ЭТ

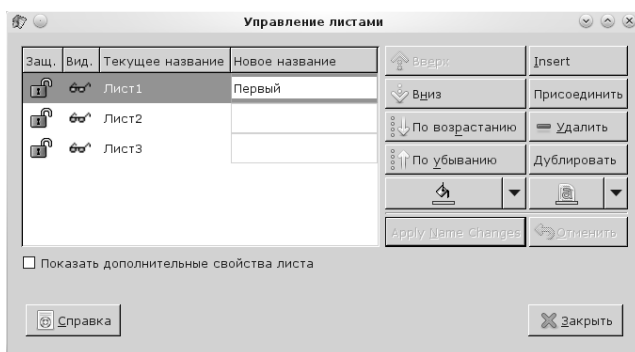


Рис. 2.8. Диалог управления листами в Gnumeric

в списке листов. Для изменения имени листа щёлкаем левой кнопкой мыши в столбце «Новое название» для выбранного листа, пишем нужный текст и нажимаем кнопку «Применить изменение названий». После этой операции новое имя листа переместится в столбец «Текущее название». Для изменения порядка следования листов перемещаем выбранный лист вверх или вниз по списку соответствующими кнопками справа. Кнопка «Вставить» вставляет новый лист перед выбранным, а кнопка «Присоединить» вставляет новый лист после последнего имеющегося. Кнопка «Удалить» позволяет удалить выбранный лист с возможностью восстановления случайно удалённого.

Для изменения цвета «ярлычка» листа и названия листа служат кнопки «Цвет заливки» и «Цвет символов». Нажатие на часть кнопки со «стрелочкой» справа от пиктограммы открывает палитру цветов (рис. 2.9), из которой можно выбрать цвет из типового набора. Клеточки в нижнем ряду заполняются «пользовательскими» цветами. Для установки цвета, не входящего в типовой набор (пользовательского) нужно щёлкнуть по кнопке «Другой цвет...» в нижней



Рис. 2.9. Типовой набор цветов в Gnumeric

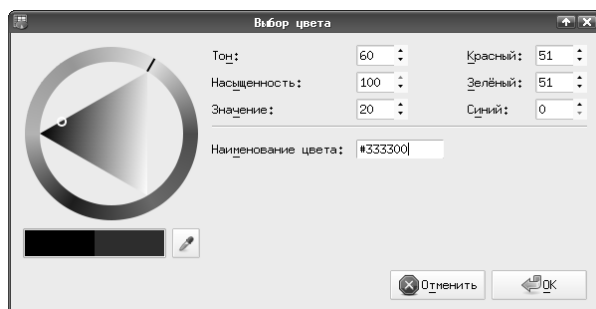


Рис. 2.10. GTK-диалог выбора цвета

части палитры и выбрать цвет с помощью GTK-диалога выбора цвета (рис. 2.10).

GTK-диалог выбора цвета предоставляет несколько возможностей формирования цвета объекта. Во-первых, можно установить точное значение цвета в координатах HSV (Hue-Saturation-Value или «Тон-Насыщенность-Яркость», диапазон изменения значений компонента «Тон» от 0 до 360, остальных — от 0 до 100) или RGB (Red-Green-Blue или «Красный-Зелёный-Синий», диапазон изменения компонентов от 0 до 255). Во-вторых, можно установить HTML-эквивалент цвета в шестнадцатеричном выражении (поле «Наименование цвета»). В-третьих, цвет можно выбрать «на глаз» (визуально), вращая цветной треугольник «протягиванием» чёрного отрезка по цветному кольцу, а затем «перетаскивая» белый кружок внутри треугольника. Наконец, кнопка с изображением пипетки под цветным кольцом даёт возможность выбрать цвет произвольной точки экрана. При любом способе выбора все варианты «цветовых координат» изменяются согласованно, а под цветным кругом показываются образцы текущего и нового цвета объекта.

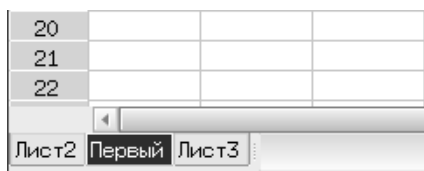


Рис. 2.11. Результат управления листами

При изменении цвета «ярлычка» листа ЭТ разумно также изменить цвет названия листа, чтобы фон и текст были достаточно контрастны.

На рис. 2.11 показаны результаты изменения цвета фона, имени и порядка следования листов ЭТ.

Менять порядок листов можно простым «перетаскиванием» ярлычка, при этом появится вертикальная стрелка, указывающая на новую позицию перемещаемого листа.

Листы ЭТ в Gnumeric имеют дополнительные атрибуты защиты, видимости и направления (в диалоге управления листами на рис. 2.8 — столбцы слева от имени листа). Изменяются эти атрибуты простым щелчком левой кнопкой мыши в соответствующем столбце диалога.

Установка атрибута защиты предотвращает изменение данных в ячейках листа, снятие атрибута видимости позволяет скрыть лист в окне ЭТ и убрать его «ярлычок» (при этом в диалоге управления листами отображаются все листы). Изменение атрибута «Направление» меняет порядок столбцов таблицы с направления «слева направо» на направление «справа налево» (первый столбец оказывается справа), что может быть полезным при использовании соответствующих систем письменности.

Дополнительные возможности управления внешним видом листов ЭТ можно получить с помощью вложенного меню «Формат/Лист» через пункт «Формат» главного меню (рис. 2.12).

Назначение режимов, которые включаются и отключаются в этом вложенном меню, достаточно очевидно. Нужно только заметить, что режим «Использовать нотацию R1C1» приводит к изменению порядка формирования адреса. Если в стандартном варианте сначала указывается имя столбца (например, А), а потом номер строки (например, 3) и получается адрес типа А3, то в режиме адресации R1C1 сначала указывается номер строки (Row), а затем — номер столбца (Column) и тот же адрес будет выглядеть как R3C1.

Вернёмся к диалогу управления листами (рис. 2.8) и включим режим показа дополнительных свойств («Показать дополнительный

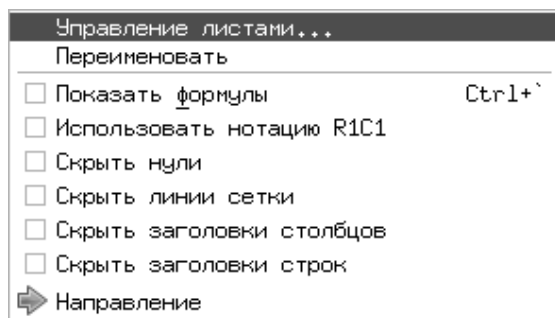


Рис. 2.12. Вложенное меню «Формат/Лист»

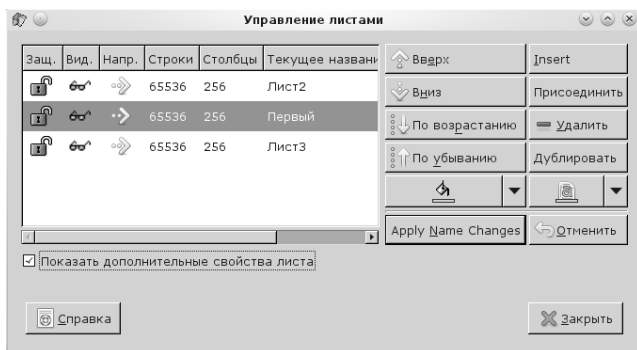


Рис. 2.13. Показ дополнительных свойств в диалоге управления листами ЭТ

свойства листа»). В поле свойств листа диалога появятся новые столбцы «Напр.», «Строки» и «Столбцы» (рис. 2.13).

Дело в том, что Gnumeric позволяет изменять максимальные значения количества строк и столбцов для листов ЭТ, в том числе и отдельно для каждого листа. Для этого используется вызов диалога «Изменить размер...» из контекстного меню листа ЭТ (рис. 2.14).

Перемещая ползунки, можно изменять количество строк и столбцов на листах ЭТ. При выключенном режиме «Применить изменение ко всем листам» новые настройки будут применены только к выбранному листу (выбирается тот лист, на ярлычке которого было вызвано контекстное меню).

Минимальное количество строк и столбцов, как видно из рис. 2.14 — 128, максимальное количество столбцов на листе — 16386, а строк — более 16 миллионов. Установим для одного из листов зна-

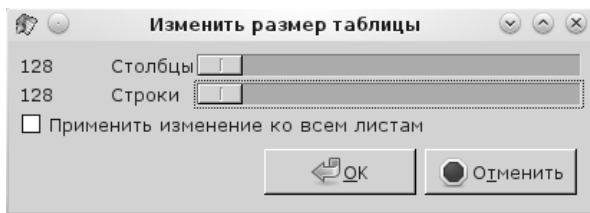


Рис. 2.14. Диалог изменения количества строк и столбцов листа ЭТ

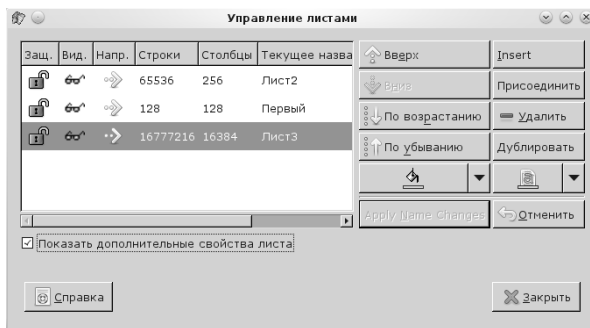


Рис. 2.15. Листы с различным количеством строк и столбцов в Gnumeric

чения по минимуму, для другого — по максимуму и посмотрим на результат в диалоге управления листами (рис. 2.15).

2.3 Управление ячейками

Для адресации ячеек может использоваться как нотация A1, так и нотация R1C1. Адрес текущей (активной) ячейки отображается над левым верхним углом таблицы.

Адреса ячеек используются как операнды в формулах. Использование в формулах конкретных значений является нежелательным (за редким исключением).

В ячейках могут содержаться числа, текстовые данные или календарные даты. Числа и даты автоматически выравниваются по правой границе столбца, а текст — по левой. Таким образом, можно всегда распознать ошибку ввода.




Для чисел десятичным разделителем является точка. Даты в «европейском» варианте лучше вводить с использованием символа / (например, 12/05/2007).

Если необходимо, чтобы числовые данные интерпретировались как текст (например, почтовые индексы или ИНН), то начинать ввод следует с символа «апостроф» — '. Например, чтобы ввести почтовый индекс 143921, нужно набрать '143921.

Ввод данных в ячейки листа таблицы осуществляется путем набора символов (букв) и цифр с последующим нажатием ENTER. При этом указатель активной ячейки («рамка») смещается вниз на одну строку. Для перехода от ячейки к ячейке в любом «направлении» по листу ЭТ можно использовать клавиши управления курсором, а также щелчок левой кнопкой мыши по целевой ячейке. Переход на другую ячейку автоматически означает окончание ввода данных в текущей ячейке.

Курсор в ЭТ Gnumeric может быть четырёх видов. Варианты курсора и ситуации, при которых они используются, описаны в таблице ниже.

Таблица 2.1: Варианты курсора в Gnumeric

Вид курсора	Ситуация, назначение
	Стандартный курсор (режим позиционирования). В таком режиме происходит указание ячеек (щелчком левой кнопкой мыши) или выделение диапазона (блока) ячеек (протаскиванием мыши с нажатой левой кнопкой).
	Курсор перемещения (при подведении курсора к рамке активной ячейки или выделенного диапазона). В таком режиме протаскивание мыши перемещает выбранный объект по листу ЭТ.
	Курсор заполнения (может иметь вид «косого» белого крестика), появляется при позиционировании мыши в нижнем правом углу активной ячейки или выделенного диапазона (блока). При выделении диапазона ячеек в таком режиме происходит копирование содержимого активной ячейки (данных или формулы) на выделенный диапазон.

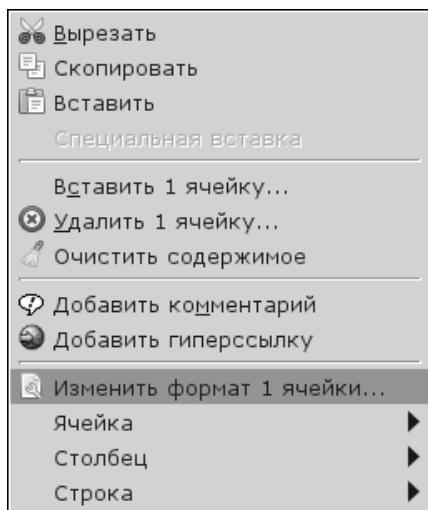


Рис. 2.16. Контекстное меню ячейки ЭТ

Таблица 2.1 — продолжение

Вид курсора	Ситуация, назначение
I	Курсор редактирования. Появляется при переходе в режим редактирования содержимого ячейки.

Для редактирования содержимого ячейки следует нажать клавишу F2 или сделать двойной щелчок мышки по ячейке, после чего редактируемая ячейка изменяет цвет фона. При редактировании возможны стандартные операции редактирования текста, а также изменение начертания и/или цвета отдельных символов текста в ячейке (при этом символы нужно выделять в строке ввода, а не в ячейке, см. пример рис. 2.22). Завершается редактирование нажатием клавиши ENTER или щелчком мыши в какой-либо другой ячейке. Для отмены изменений и возврата к предыдущему состоянию следует нажать клавишу ESC.

Управление представлением данных в ячейках осуществляется настройками форматов ячеек («Формат/Ячейки...» в главном меню или вызов диалога «Изменить формат 1 ячейки...» через контекстное меню ячейки (рис. 2.16) с помощью правой кнопки мыши).

Изменения формата касаются активной ячейки или выделенного блока ячеек (включая строки и столбцы полностью). Диалог «Формат

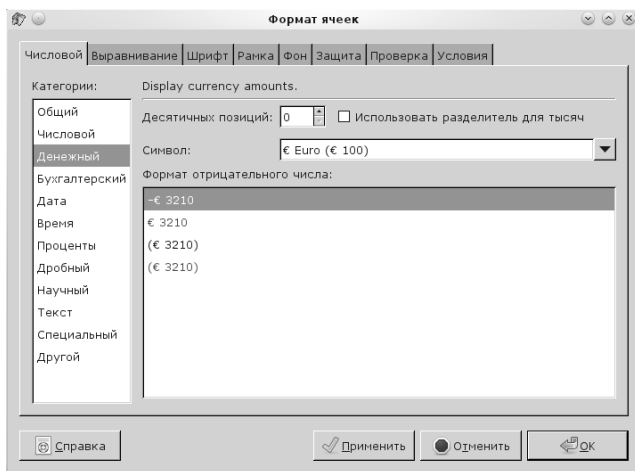


Рис. 2.17. Настройка представления чисел в ячейках ЭТ

ячеек» (рис. 2.17) имеет несколько вкладок, обеспечивающих различные настройки параметров ячейки или диапазона ячеек.

- Вкладка «Числовой» позволяет установить вид чисел в ячейке (количество десятичных знаков, вид представления валют и дат, форму отображения дробей как десятичных или обыкновенных), а также преобразовать числа в текст. Для понимания каждого варианта представления данных полезно с ними поэкспериментировать.
- Вкладка «Выравнивание» позволяет определить расположение содержимого в ячейке, задать горизонтальное и вертикальное выравнивание или угол поворота. Если нужно, чтобы текст в ячейке размещался в несколько строк, на этой вкладке включается режим «Переносить текст».
- Вкладка «Шрифт» позволяет задать гарнитуру (начертание) и кегль (размер) шрифта, а также цвет и другие атрибуты.
- Вкладка «Рамка» позволяет задать вид границ ячеек. Для границ настраиваются наличие, стиль и цвет линии. Можно также «перечеркивать» ячейки, делая диагональные штрихи в прямом или обратном направлении.
- Вкладка «Фон» позволяет задать цвет фона, вид штриховки и цвет штриховки ячейки или блока ячеек.
- Вкладка «Защита» позволяет установить защиту от изменений для ячейки или блока ячеек.

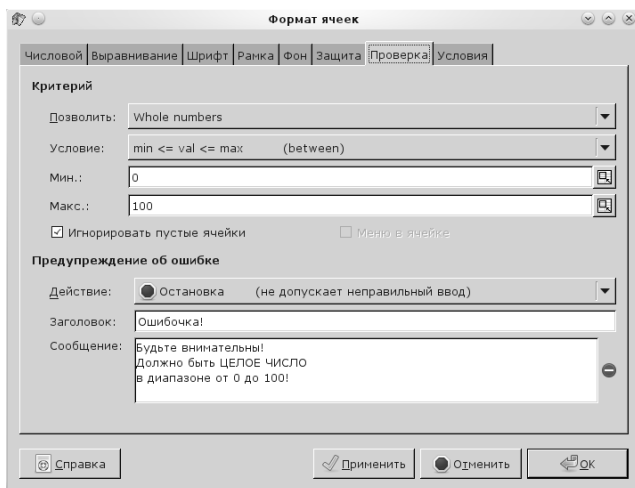


Рис. 2.18. Настройка проверки правильности ввода данных в ячейку ЭТ

- Вкладка «Проверка» позволяет настроить проверку соответствия данных при их вводе, так что при ошибочном вводе может быть выдано предупреждение или ошибочный ввод прямо запрещается.

Возможность проверки правильности данных при вводе — очень полезная возможность. Пусть, например, в некотором диапазоне ячеек необходим ввод только целых чисел, причем пустая ячейка (отсутствие данных) не является ошибкой. С помощью настройки параметров проверки («Формат ячеек: Проверка») обеспечим блокировку ошибочного ввода и появление предупреждения об ошибке (рис. 2.18).

Теперь, если в какую-то ячейку из этого диапазона попытаться ввести что-либо неправильное, появится соответствующее предупреждение (рис. 2.19).

Контекстное меню ячейки (рис. 2.16) позволяет установить для ячейки комментарий. Комментарий — это какой-то текст, поясняющий назначение или особенности содержимого ячейки. При вызове команды «Добавить комментарий» из контекстного меню появляется диалог редактирования комментария (рис. 2.20). В области ввода можно писать произвольный текст, который станет комментарием к ячейке после нажатия на кнопку ОК.

При создании комментария можно управлять атрибутами текста. Включение/выключение режима «Переносить окно свойств» в описываемой версии не даёт какого-либо эффекта.

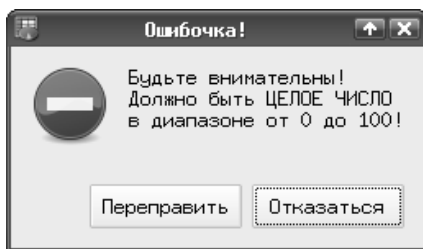


Рис. 2.19. Сообщение об ошибке при неправильном вводе

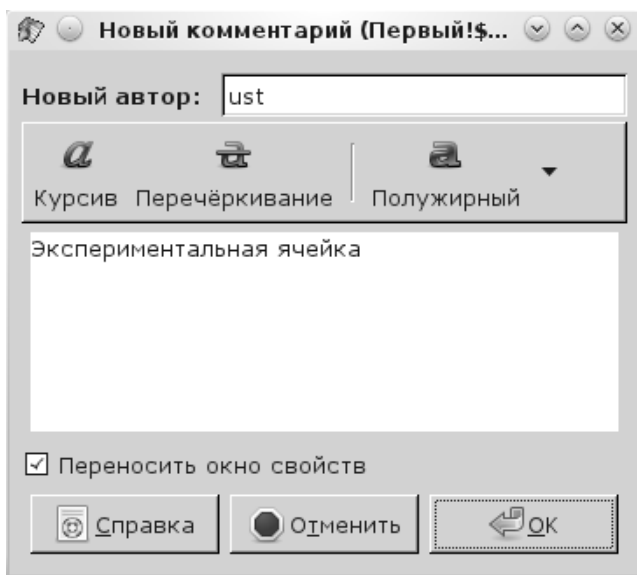


Рис. 2.20. Создание комментария для ячейки ЭТ

Важно, что для каждого комментария устанавливается автор (текущий пользователь). При изменении комментария другим пользователем сохраняется история изменений.

Ячейки с комментарием обозначаются красным треугольничком в верхнем правом углу ячейки (рис. 2.21). Комментарий можно увидеть, если навести курсор на этот треугольничек. Для удаления комментария следует вызвать диалог «Удалить 1 комментарий» с помощью контекстного меню ячейки.

Интересной особенностью ЭТ Gnumeric является возможность управления атрибутами отдельных символов текста в ячейке. Для этого в режиме редактирования (по нажатию F2 или двойному клику по ячейке) следует выделить нужные символы и использовать кнопки

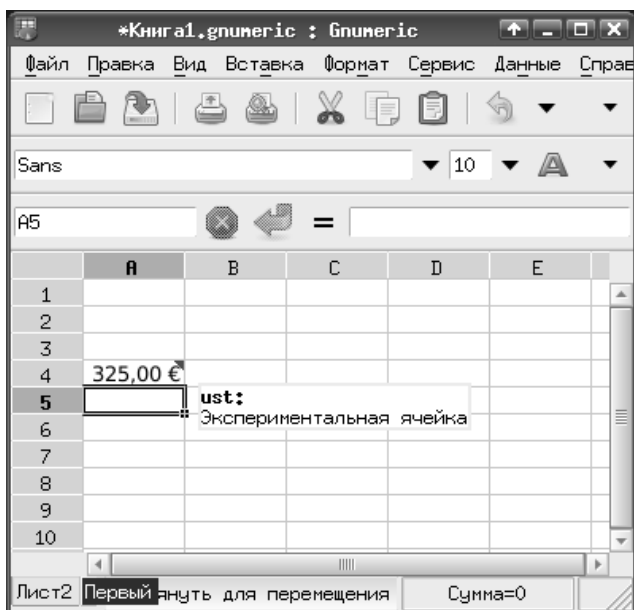


Рис. 2.21. Просмотр комментария ячейки

управления атрибутами («Полужирный», «Наклонный», «Подчеркнутый» и «Передний план») в панели инструментов Gnumeric. Пример такой модификации текста показан на рис. 2.22.

Операции изменения формата данных, цвета переднего плана, фона и обрамления, а также копирования, перемещения и вставки можно проводить не только с одиночными ячейками, но и с группами (блоками) ячеек. Для этого требуется выделить группу ячеек. Для выделения нескольких соседних ячеек используется «протаскивание» мыши с нажатой левой кнопкой от верхнего левого до правого нижнего угла выделяемого блока или используются клавиши управления курсором («стрелки») при нажатой клавише SHIFT. На рис. 2.23 показан вид области листа ЭТ с выделенным блоком ячеек.

Весь выделенный блок оказывается в общей «рамке», а фон всех выделенных ячеек меняет цвет, за исключением ячейки, которая была активной до выделения блока. Для всех выделенных ячеек с помощью контекстного меню и кнопок панели инструментов можно установить одинаковый формат данных, обрамление и цвет фона и текста, шрифт и другие параметры, за исключением комментария. Комментарий всё равно будет устанавливаться только для активной ячейки.

Для снятия выделения достаточно перейти в любую ячейку ЭТ щелчком левой кнопки мыши или с помощью клавиш-«стрелок».

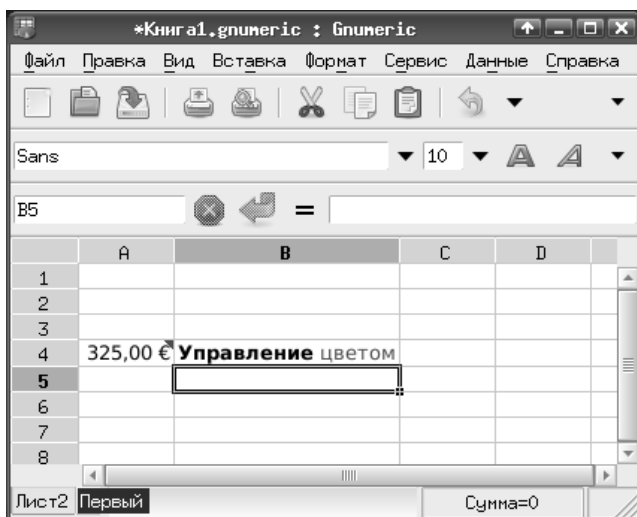


Рис. 2.22. Пример выделения отдельных символов в ячейке B4

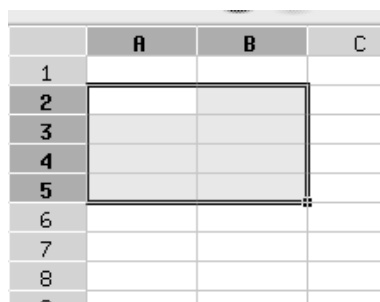


Рис. 2.23. Пример выделения непрерывного блока ячеек

Если необходимо выделить несколько ячеек, расположенных в произвольных позициях листа, используется щелчок левой кнопкой мыши по нужным ячейкам при нажатой клавише **CTRL**. Пример такого выделения показан на рис. 2.24.

Возможно также проводить операции со всеми ячейками в строке (строках) или столбце (столбцах) листа ЭТ. Для выделения всей строки нужно щелкнуть левой кнопкой мыши по номеру строки (рис. 2.25).

Контекстное меню строки, получаемое после щелчка правой кнопкой мыши по номеру строки (рис. 2.26) отличается от контекстного меню ячейки. Для строк появляются операции «Вставить 1 строку» и «Удалить 1 строку».

	А	В	С	
1				
2				
3				
4				
5				
6				
7				
8				

Рис. 2.24. Выделение нескольких произвольных ячеек

	А	В	С	Д	
1					
2					
3					
4					
5					
6					
7					
8					

Рис. 2.25. Результат выделения строки

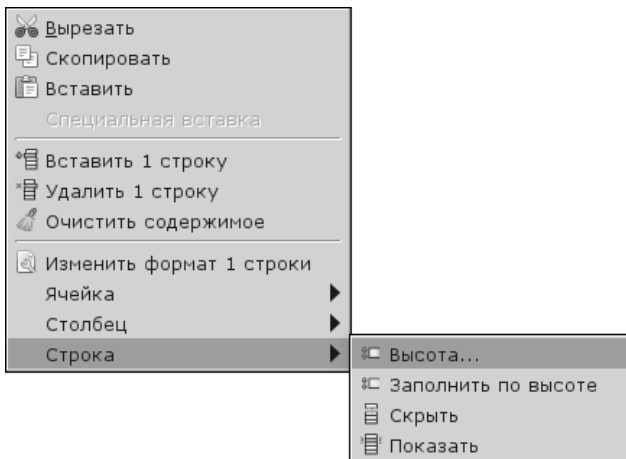


Рис. 2.26. Контекстное меню строки

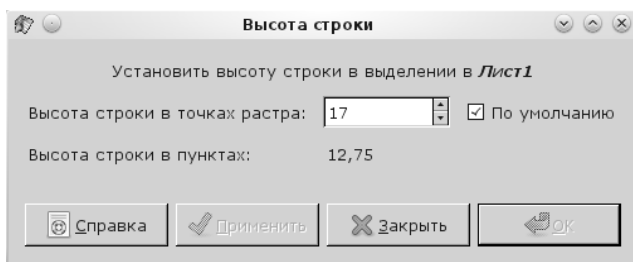


Рис. 2.27. Диалог настройки высоты строки

	А	В	С	Д
1				
2				
3				
4				
5				
6				
7				
8				

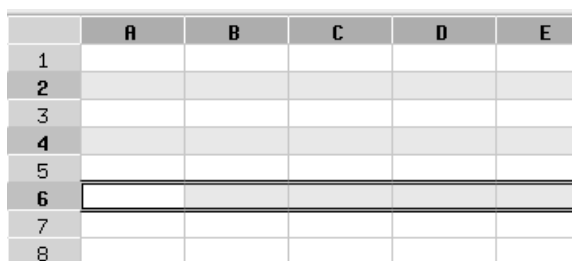
Рис. 2.28. Выделение соседних строк

Диалог «Строка/высота...» (рис. 2.27) позволяет задавать высоту строки в точках экрана (pixels), причём она автоматически пересчитывается в типографские пунктах (пт, 1 пт = 1/72 дюйма) с учётом разрешения экрана.

Явное указание высоты строки может потребоваться для точного форматирования документа, однако в большинстве случаев высота устанавливается автоматически исходя из размера шрифта и параметров форматирования.

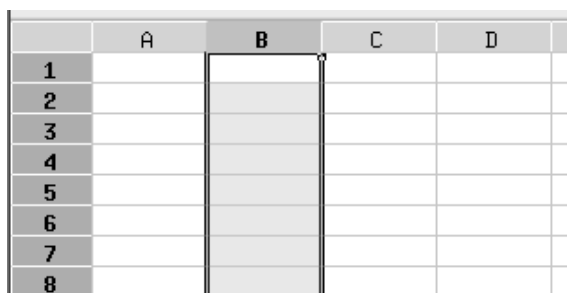
При выборе операции «Вставить 1 строку» новая строка появляется перед выделенной (выделенная строка смещается «вниз», её номер увеличивается на 1). При выборе операции «Скрыть» строка становится невидимой и её номер пропадает (нарушается непрерывность номеров). Чтобы снова увидеть скрытую строку, нужно выделить соседние строки и из контекстного меню выбрать операцию «Показать» (предлагается поэкспериментировать самостоятельно).

Для выделения нескольких соседних строк можно «протащить» мышью по номерам строк или выделить одну строку, а затем использовать «стрелки» при нажатой клавише SHIFT (рис. 2.28).



	А	В	С	Д	Е
1					
2					
3					
4					
5					
6					
7					
8					

Рис. 2.29. Выделение нескольких произвольных строк



	А	В	С	Д
1				
2				
3				
4				
5				
6				
7				
8				

Рис. 2.30. Выделение столбца

Тем же самым способом, каким выделяются произвольные ячейки, выделяются и произвольные строки (рис. 2.29), только щелкать мышью нужно по номерам строк.

Аналогично строкам, столбцы выделяются целиком при щелчке левой кнопкой по имени столбца (букве), как показано на рис. 2.30.

Контекстное меню столбца показано на рис. 2.31.

Диалог настройки ширины столбца (рис. 2.32) позволяет точно устанавливать значение этого параметра. Однако, в отличие от высоты строк, ширина столбцов не изменяется автоматически.

Пусть в ячейки ЭТ введён текст, как показано на рис. 2.33. Содержимое ячейки В2 «не помещается» в видимую ширину столбца (на самом деле текст никуда не пропадает, в чём легко убедиться в режиме редактирования).

Для подбора нужной ширины столбца можно воспользоваться диалогом настройки ширины столбца (рис. 2.32), можно «растянуть» столбец за правую границу области имени столбца (прямоугольник, в котором написана буква), а можно по этой правой границе дважды щелкнуть левой кнопкой мыши, вызвав таким образом операцию «Автоподбор ширины». Результат автоподбора ширины для столбца В показан на рис. 2.34.

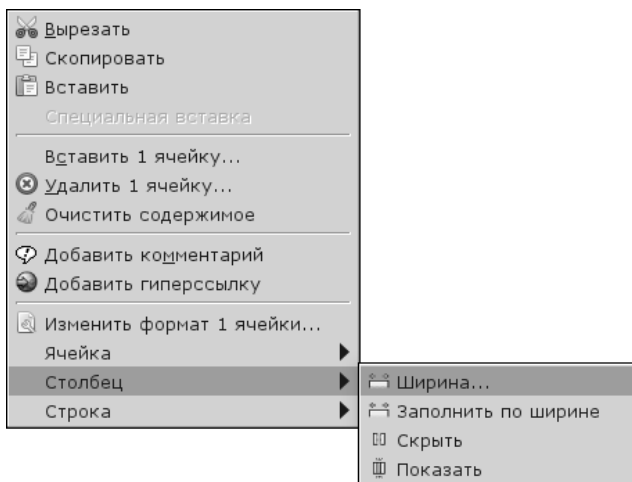


Рис. 2.31. Контекстное меню столбца

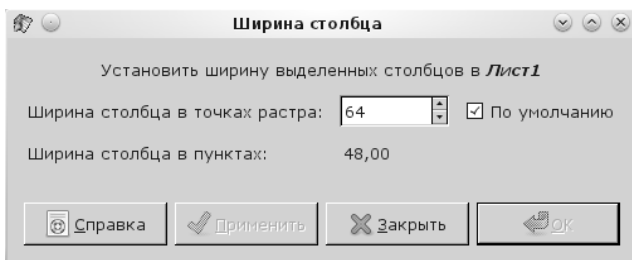


Рис. 2.32. Диалог настройки ширины столбца

	А	В	С	Д
1				
2		Автопод ширины		
3				
4				
5				
6				

Рис. 2.33. Пример недостаточной ширины столбца

	А	В	С	Д
1				
2		Автоподбор	ширины	
3				
4				
5				
6				

Рис. 2.34. Результат автоподбора ширины

	А	В	С	Д
1				
2		Автоподбор	ширины	
3				
4				
5				
6				
7				
8				

Рис. 2.35. Автоподбор ширины соседних столбцов

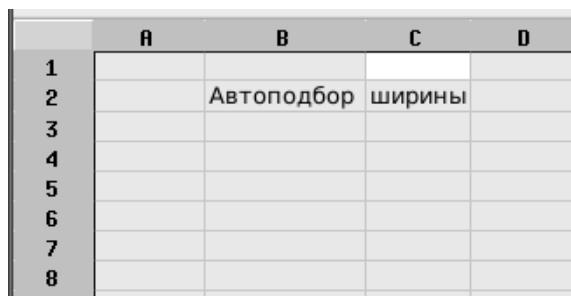
	А	В	С	Д	Е
1					
2		Автоподбор	ширины		
3					
4					
5					
6					
7					
8					

Рис. 2.36. Выделение нескольких произвольных столбцов

Если выделить соседние столбцы, то двойной щелчок по правой границе последнего выделенного столбца приведёт к выполнению автоподбора ширины для всех выделенных столбцов (рис. 2.35, сравните с рис. 2.34).

При выделении нескольких произвольных столбцов с помощью клавиши **CTRL** (рис. 2.36) для «пустых» столбцов автоподбор ширины не действует.

После упражнений с выделением отдельных ячеек, строк и столбцов логично задаться вопросом: а нет ли возможности выделить сразу



	А	В	С	Д
1				
2		Автоподбор	ширины	
3				
4				
5				
6				
7				
8				

Рис. 2.37. Выделение всех ячеек листа

все ячейки листа? Оказывается, такая возможность тоже существует. Для выделения всех ячеек листа нужно щёлкнуть левой кнопкой мыши по «кнопке» над столбиком с номерами строк (самый верхний левый угол таблицы, над номером 1 и левее буквы А). Это равносильно выбору команды главного меню «Правка/Выделение/Все».

Основное применение всех электронных таблиц — «программирование без программирования», то есть создание таких связей между ячейками, чтобы на основе исходных данных в одних ячейках получался конечный результат в других ячейках. Поэтому для дальнейшей работы с таблицей будут нужны исходные данные. Они могут быть введены вручную, экспортированы из внешнего файла (например, из текстового) или смоделированы (сгенерированы) с помощью соответствующих инструментов пакета.

2.4 Автозаполнение: генерация рядов данных

Рассмотрим возможности *Glumeric* по генерации исходных данных. Исходные данные, как правило, располагаются в непрерывном блоке ячеек (столбце или строке) и затем используются для каких-либо вычислений. Для определенности будем располагать исходные данные в столбце.

Для создания последовательностей исходных данных (векторов, таблиц) можно использовать возможности заполнения блоков данных, используя пункт главного меню «Данные» (рис. 2.38).

Вложенное меню «Заполнить» как раз и предоставляет различные способы создания блоков ячеек с исходными данными (рис. 2.39).

Здесь мы рассмотрим только три варианта — генерация последовательностей («Автозаполнение»), создание прогрессий («Серии...») и создание выборок последовательностей случайных чи-

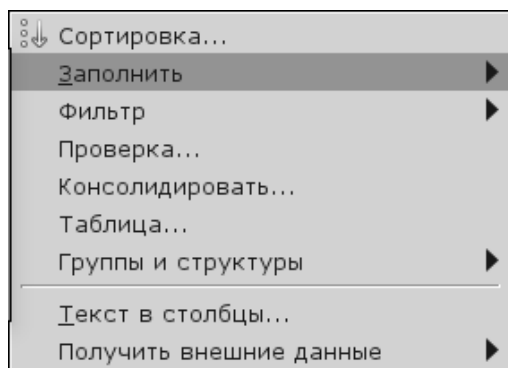


Рис. 2.38. Пункт «Данные» главного меню Gnumeric

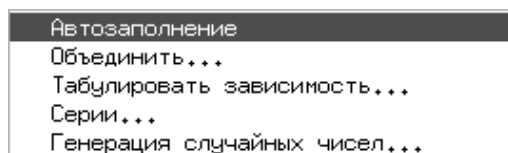


Рис. 2.39. Вложенное меню «Заполнить»

сел с различными видами функций распределения («Случайные числа.../Некоррелированные...»).

Для начала сформируем последовательность натуральных чисел от 0 до 25. Для этого введем в ячейку A3 значение «0», а в A4 — значение «1» (без кавычек!), затем выделим ячейки от A3 до A28 (всего 25 ячеек) путем «протаскивания» мыши с нажатой левой кнопкой. Далее вызовем функцию автозаполнения ряда: «Данные/Заполнить/Автозаполнение» и наблюдаем результат.

Далее сформируем последовательность из 25 чисел, кратных 7, начиная с 0. По аналогии с предыдущим случаем, введем в ячейку B3 значение «0», а в ячейку B4 — значение «7», после чего повторим операции выделения и вызова функции автозаполнения и посмотрим на результат.

Такую же операцию можно провести с датами. Пусть известно, что 21 мая 2007 года и 28 мая этого же года были понедельниками. На какие даты будут приходиться понедельники в течение последующих 23 недель? (Аналогично предыдущему случаю, только вместо чисел — даты). Соответственно, вводим в ячейку C3 значение «21/05/07», в C4 — «28/05/07», выделяем снова 25 ячеек и вызываем автозаполнение. Затем устанавливаем формат дат как «dd/mm/yy» с помощью диалога «Формат ячеек». Результат показан на рис. 2.40.

	А	В	С	Д
1	Использование Автозаполнения			
2	через 1	через 7	даты через 7	
3	0	0	21/05/07	
4	1	7	28/05/07	
5	2	14	04/06/07	
6	3	21	11/06/07	
7	4	28	18/06/07	
8	5	35	25/06/07	
9	6	42	02/07/07	
10	7	49	09/07/07	
11	8	56	16/07/07	
12	9	63	23/07/07	
13	10	70	30/07/07	
14	11	77	06/08/07	
15	12	84	13/08/07	
16	13	91	20/08/07	
17	14	98	27/08/07	
18	15	105	03/09/07	
19	16	112	10/09/07	
20	17	119	17/09/07	
21	18	126	24/09/07	
22	19	133	01/10/07	
23	20	140	08/10/07	
24	21	147	15/10/07	
25	22	154	22/10/07	
26	23	161	29/10/07	
27	24	168	05/11/07	
28	25	175	12/11/07	

Рис. 2.40. Результаты использования автозаполнения

Нужно отметить, что с другими (не числовыми) данными функция «Автозаполнение» не работает. Таким образом, функция «Автозаполнение» продолжает зависимость, определенную по первым двум ячейкам диапазона, на все ячейки выделенного диапазона.

Следующая возможность автоматической генерации исходных данных — создание серий. Она может потребоваться, если заранее неизвестно количество ячеек для автозаполнения, или если это количество велико, а также если зависимость значения следующей ячейки от предыдущей не является линейной.

Рассмотрим примеры генерации серий чисел со значениями от 1 до 100. Диалог заполнения ячеек сериями значений («Данные/Заполнить/Серии...») содержит три вкладки (рис. 2.41).

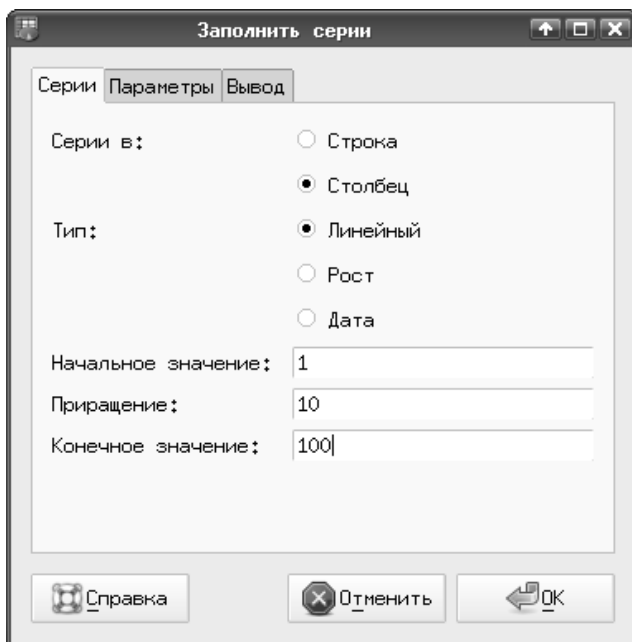


Рис. 2.41. Настройка серий для заполнения ячеек данными

- Вкладка «Серии» позволяет установить основные параметры серии. Здесь определяется направления заполнения (по строке или по столбцу), вид заполнения (линейный или рост с заданным коэффициентом), а также начальное значение, приращение и конечной значение. Не все эти параметры обязательно указывать, однако начальное значение указывать обязательно, а конечное значение и приращение указываются в зависимости от того, известны или нет количество ячеек в серии и максимальное значение элемента серии.
- Вкладка «Параметры» активируется, если для формирования серии выбраны даты. В этом случае указывается единица приращения дат (календарный день, рабочий день, месяц или год).
- Вкладка «Вывод» позволяет указать диапазон вывода полученной серии значений. По умолчанию вывод происходит в выделенный диапазон ячеек (если он есть) или начиная с активной ячейки. Однако можно сформировать серию на новом листе или открыть новый документ Gnumeric (книгу) и создать серию значений там.

	А	В	С	Д
1	Заполнение сериями			
2	Больше на 10	Рост в 2 раза	Рост в 3 раза	
3	1	1	1	
4	11	2	3	
5	21	4	9	
6	31	8	27	
7	41	16	81	
8	51	32		
9	61	64		
10	71			
11	81			
12	91			
13				

Рис. 2.42. Результат заполнения сериями

Сформируем серию значений, линейно изменяющихся по столбцу (арифметическую прогрессию). Устанавливаем соответствующие параметры на вкладке «Серии», указываем в качестве начального значения число 1, приращение 10 и конечное значение 100 (рис. 2.41). Больше никаких дополнительных настроек не требуется. После нажатия на ENTER (или использования кнопки «ОК» в диалоге) наблюдаем результат (рис. 2.42). Видно, что последнее значение серии меньше указанного максимального значения. Однако если провести эксперимент с начальным значением, равным 0, то последним значением в серии будет 100. Таким образом, можно сделать очевидный вывод, что конечное значение в серии никогда не превышает указанного максимального значения.

Теперь в соседних столбцах создадим серии с ростом в 2 и в 3 раза при тех же начальных и конечных значениях (именно поэтому в качестве начального значения выбрана 1, а не 0). Коэффициент роста указывается как приращение. Результаты показаны на рис. 2.42).

С заполнением диапазона ячеек сериями дат читателям предлагается разобраться самостоятельно.

Следующая интересная возможность генерации исходных данных — заполнение диапазона ячеек случайными числами с различными функциями плотности распределения. Всего предлагается около 30 вариантов. Для использования этой возможности полезно иметь выделенный диапазон ячеек (блок). Можно реализовать двумерное поле случайных чисел, если выделить не часть строки или столбца, а прямоугольный диапазон и сделать соответствующие настройки в диалоге «Генерация случайных чисел» («Данные/Заполнить/Случайные числа.../Некоррелированные...», рис. 2.43).

Диалог «Генерация случайных чисел» также имеет три вкладки.

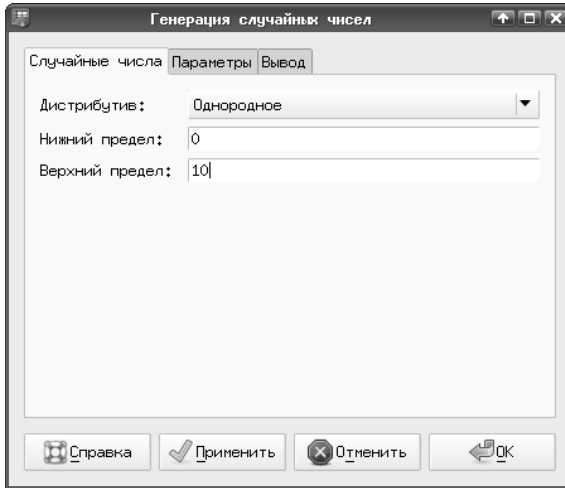


Рис. 2.43. Диалог генерации случайных чисел (однородное распределение)

- Вкладка «Случайные числа» позволяет определить вид функции плотности распределения (названный словом «дистрибутив»), а также параметры этой функции. Для разных распределений количество и значения параметров будут разными.
- Вкладка «Параметры» позволяет определить размер выборки (сколько чисел нужно) и количество переменных (1-мерная последовательность или 2-мерное поле). Если предварительно выделен диапазон ячеек, размер выборки устанавливается автоматически.
- Вкладка «Вывод» позволяет (как и в предыдущем случае) указать диапазон вывода полученной серии значений. По умолчанию вывод происходит в выделенный диапазон ячеек (если он есть) или начиная с активной ячейки. Однако можно сформировать серию на новом листе или открыть новый документ Gnumeric (книгу) и создать серию значений там.

В качестве примера создадим три вектора из 100 случайных чисел с разными распределениями: равномерным в диапазоне от 0 до 10 (как на рис. 2.43), нормальным со средним значением 5 и стандартным отклонением 2, а также с распределением x^2 со значением параметра $v = 2$. Для полученных результатов установим формат с четырьмя десятичными знаками (рис. 2.44)

Рассмотренные возможности по генерации исходных данных будут полезны при дальнейшей работе с пакетом Gnumeric.

	А	В	С	Д
1	Генерация случайных чисел			
2	Равномерное	Нормальное	Хи-квадрат	
3	3,5215	4,1070	1,0284	
4	2,2043	2,9831	0,8274	
5	5,4872	6,9444	1,1866	
6	9,6423	5,1187	0,1518	
7	1,6817	6,7977	0,6176	
8	7,0084	8,4337	2,0553	
9	7,2888	9,3311	0,9406	
10	0,7312	1,9206	0,2733	
11	5,7347	3,9104	0,7215	
12	2,2740	4,8930	1,4979	
13	7,4336	3,2952	0,6764	
14	4,5522	5,7443	0,1311	
15	4,2461	2,8635	2,5090	
16	8,8224	6,4100	2,9347	
17	0,2124	4,9169	1,1188	
18	4,4424	4,7765	0,4385	
19	7,4021	4,5417	2,7815	
20	5,3035	6,1519	0,6666	
21	7,9344	2,3908	2,6562	
22	4,1669	4,9635	8,2444	

Рис. 2.44. Три вектора случайных чисел

2.5 Формулы. Абсолютная и относительная адресация

Формулы в электронных таблицах предназначены для вычислений значений в ячейках таблицы на основе данных, записанных в другие ячейки. Результатом работы формулы может быть число, дата, текст или отсутствие данных (т.е. после вычислений можно получать пустые ячейки). Формула записывается в той ячейке, в которой должен быть результат. Ввод формулы начинается с нажатия на символ = на клавиатуре, после чего вводятся числа, адреса ячеек и знаки операций. Например, формула =3+5 даст всегда результат 8, а формула =A3/12 будет давать различные результаты при изменении значения в ячейке A3.

Таким образом, при изменении содержимого ячеек, адреса которых используются в формулах, результаты пересчитываются автоматически. Эта особенность является ключевой для всех электронных таблиц. Из-за этого свойства крайне не рекомендуется использовать в формулах конкретные значения, если только они не являются неотъемлемой частью правил вычисления (например, если площадь круга вычисляется в евклидовой геометрии как πR^2 , то двойка может ис-

	A	B	C	D	E
1					
2					
3	Фамилия, инициалы	Должность	Оклад за день	Дней работы	Заработок
4	Сиднов К.С.	Бухгалтер	50	22	
5	Чернышевская С.П.	Швея	43	25	
6	Чкалов В.П.	Плотник	45	15	
7	Романов П.А.	Мореплаватель	70	31	
8	Штек Е.А.	Уборщица	22	25	
9	Зубарева О.А.	Ветеринар	34	8	
10	Жеребцова Н.Н.	Швея	40	25	
11	Зубов П.Р.	Директор	85	30	

Рис. 2.45. Исходные данные для задачи о заработках

пользоваться в формуле вычисления площади круга, а вот конкретное значение радиуса — нет).

Электронная таблица с хорошо составленными формулами представляет из себя мощное средство решения вычислительных задач и, по сути, является программой, выдающей результаты на основе содержимого ячеек с исходными данными.

Теперь рассмотрим некоторые примеры использования формул и адресации данных в таблицах Gnumeric.

Начнем вычисления в ЭТ с простой задачи. Пусть имеется список из семи человек, для каждого из которых известны фамилия, инициалы, должность, оклад за день работы и число отработанных дней. Требуется вычислить заработок каждого лица.

Формируем таблицу, начиная с ячейки A3, в соответствии с рис. 2.45. Для исправления ошибок в ячейках электронной таблицы используется режим редактирования строки ввода, который включается клавишей F2. Завершение редактирования обеспечивается клавишами ENTER (с сохранением изменений) или ESC (без сохранения изменений).

Если при вводе информации ширина ячейки представляется недостаточной, ее можно скорректировать после завершения ввода всех данных. Как уже упоминалось выше, всегда есть возможность подобрать ширину столбца автоматически двойным щелчком левой кнопки мыши на правой границе столбца в строке имен столбцов.

Для вычисления заработка нужно просто перемножить попарно числа из третьей (столбец C) и четвертой (столбец D) колонок. Результаты вычислений должны быть в пятой колонке (столбец E). С учетом возможностей ЭТ, формулу (т.е. правила) для вычислений можно написать один раз, а потом скопировать. Формулу надо писать там, где должен появиться первый результат (в нашем примере — в ячейке E4, под заголовком «Заработок»). Переводим указатель активной ячейки в клетку E4 и нажимаем клавишу «=» (указание на начало ввода

Фамилия, инициалы	Должность	Оклад за день	Дней работы	Зарплата
Сионов К.С.	Бухгалтер	50	22	1100
Чернышевская С.П.	Швея	43	25	1075
Чкалов В.П.	Плотник	45	15	675
Романов П.А.	Мореплаватель	70	31	2170
Штек Е.А.	Уборщица	22	25	550
Зубарева О.А.	Ветеринар	34	8	272
Жеребцова Н.Н.	Швея	40	25	1000
Зубов П.Р.	Директор	85	30	2550

Рис. 2.46. Результат расчёта заработка

формулы). После этого щелкаем левой кнопкой мыши по ячейке, в которой записан оклад за день (С4), нажимаем на клавиатуре знак операции (умножение — «*») и щелкаем левой кнопкой мыши по ячейке с количеством отработанных дней (D4, после чего нажимаем ENTER. В ячейке E4 появляется результат (число 1100), а переместив указатель активной ячейки на E4, в строке ввода увидим формулу =C4*D4.

Теперь скопируем эту формулу в оставшиеся ячейки. Поместив указатель активной ячейки на E4, наведем мышь на нижний правый угол указателя активной ячейки (там есть маленький черный квадратик). Двойной щелчок левой кнопкой мыши в этой позиции приведет к автоматическому копированию формулы до конца блока исходных данных. Результаты показаны на рис. 2.46.

Если изменить какие-то числа в столбцах С и D, то числа в столбце E будут автоматически пересчитываться.

Перемещая указатель активной ячейки по столбцу E заметим, что адреса ячеек в расчетной формуле изменяются. Это происходит потому, что в нашей формуле использованы относительные адреса ячеек. Формула просто перемножает содержимое ячеек, находящихся слева от ячейки с результатом.

Таким образом, формула «запомнила» взаимное расположение ячеек с данными и с результатом и при копировании это взаимное расположение сохраняется. Это очень полезное свойство ЭТ, избавляющее от необходимости писать одну и ту же формулу много раз.

Если в какой-либо ячейке расчетного столбца (столбца «Зарплата») перейти в режим редактирования (F2), то можно увидеть формулу. При перемещении текстового курсора по формуле будут подсвечиваться ячейки, содержащие данные для формулы (рис. 2.47).

На следующем этапе посчитаем налог на доходы физических лиц, который будет начислен на рассчитанные ранее значения заработка. Пусть ставка налога фиксирована и составляет 13%. Тогда наша таблица дополняется в соответствии с рис. 2.48 (здесь и в следующих иллюстрациях к этому примеру первый столбец «обрезан»).

Фамилия, инициалы	Должность	Оклад за день	Дней работы	Зарплата
Сионов К.С.	Бухгалтер	50	22	1100
Чернышевская С.П.	Швея	43	25	=C5*D5
Чкалов В.П.	Плотник	45	15	675
Романов П.А.	Мореплавателъ	70	31	2170
Штек Е.А.	Уборщица	22	25	550
Зубарева О.А.	Ветеринар	34	8	272
Жеребцова Н.Н.	Швея	40	25	1000
Зубов П.Р.	Директор	85	30	2550

Рис. 2.47. Просмотр формулы в режиме редактирования

В	С	Д	Е	Ф
	Ставка налога	13%		
Должность	Оклад за день	Дней работы	Зарплата	Сумма налога
Бухгалтер	50	22	1100	
Швея	43	25	1075	
Плотник	45	15	675	
Мореплавателъ	70	31	2170	
Уборщица	22	25	550	
Ветеринар	34	8	272	
Швея	40	25	1000	
Директор	85	30	2550	

Рис. 2.48. Таблица для вычислений с параметром

Сумму налога легко сосчитать по правилу $\text{Сумма налога} = \text{зарплата} * \text{ставка_налога}$. Указав соответствующие адреса ячеек, в ячейке F4 записываем формулу `=E4*D1` и копируем ее во все оставшиеся ячейки. При этом получается неожиданный результат (рис. 2.49).

В этом случае использование относительной адресации привело к ошибке — запомнив взаимное расположение ячеек результата и исходных данных (первого заработка в списке и ставки налога) программа ЭТ повторяет это взаимное расположение для остальных строк списка (в чем можно убедиться, войдя в режим редактирования, как показано на рис. 2.49). Чтобы не создавать дополнительный столбец с одним и тем же значением ставки налога, в соответствующей формуле надо использовать *абсолютный* адрес ячейки, содержащей параметр (в данном случае — значение ставки налога).

Для указания абсолютного адреса к букве столбца или номеру строки добавляется префикс `$` и формула для расчета суммы налога приобретает вид `=E4*D1` (для добавления символов `$` при редактировании формулы можно использовать клавишу `F4`). Отредактировав

В	С	Д	Е	Ф
	Ставка налога	13%		
Должность	Оклад за день	Дней работы	Заработок	Сумма налога
Бухгалтер	50	22	1100	143
Швея	43	25	1075	0
Плотник	45	15	675	=E6*D3
Мореплавателъ	70	31	2170	47740
Уборщица	22	25	550	13750
Ветеринар	34	8	272	4080
Швея	40	25	1000	31000
Директор	85	30	2550	63750

Рис. 2.49. Неправильный результат вычислений с параметром

В	С	Д	Е	Ф
	Ставка налога	13%		
Должность	Оклад за день	Дней работы	Заработок	Сумма налога
Бухгалтер	50	22	1100	143
Швея	43	25	1075	139,75
Плотник	45	15	675	=E6*\$D\$1
Мореплавателъ	70	31	2170	282,1
Уборщица	22	25	550	71,5
Ветеринар	34	8	272	35,36
Швея	40	25	1000	130
Директор	85	30	2550	331,5

Рис. 2.50. Правильные вычисления с параметром

формулу в ячейке F4, копируем ее снова в оставшиеся ячейки и получаем правильный результат (рис. 2.50).

В режиме редактирования теперь видно, что во всех ячейках при вычислении суммы налога происходит обращение к ячейке, содержащей ставку налога, независимо от строки таблицы (вспомним, что одинаковое количество знаков после запятой во всех ячейках можно получить, изменив формат представления чисел).

Итак, абсолютный адрес указывает программе ЭТ, что нужно всегда обращаться к одной и той же ячейке (если поставлено два префикса \$), строке (если \$ поставлен перед номером строки) или столбцу (если \$ — перед буквой столбца). Использование абсолютных адресов позволяет работать с условно-постоянными величинами (ставка налога, курс валюты, текущая дата и пр.), причем их значения заносятся в таблицу только один раз, что экономит время и место.

	Заработок	Сумма налога
2	1100	143
5	1075	139,75
5	675	87,75
1	2170	282,1
5	550	71,5
8	272	35,36
5	1000	130
0	2550	331,5
	=sum(E4:E12)	

Рис. 2.51. Автосуммирование по столбцу

Нужно заметить, что формулы позволяют связывать между собой не только ячейки в пределах одного листа, но и ячейки на разных листах документа. В результате можно строить «трехмерные» электронные таблицы.

Полезная и часто используемая возможность электронных таблиц — автосуммирование. Для использования этой возможности нужно установить указатель активной ячейки в позицию, в которой нужно получить результат и нажать на панели инструментов Gnumeric кнопку Σ (знак суммы). Программа автоматически определит непрерывный блок ячеек выше или слева от целевой и предложит вариант функции для вычисления результата. Обратите внимание, что диапазон ячеек указывается с использованием двоеточия, как показано на рис. 2.51.

В качестве упражнения посчитайте аналогичным образом сумму всех налогов.

2.6 Функции

2.6.1 Селектор функций и помощник по формулам

Для проведения математических, тригонометрических, экономических или инженерных вычислений четырех действий арифметики явно недостаточно. Поэтому все электронные таблицы имеют большое число встроенных функций, которые могут включаться в состав формул. Каждая функция имеет имя и список аргументов, которые помещаются в скобки сразу за именем функции. Могут быть функции с пустым множеством аргументов, такие как `pi()` или `today()`, а могут быть функции с неограниченным количеством аргументов (такие

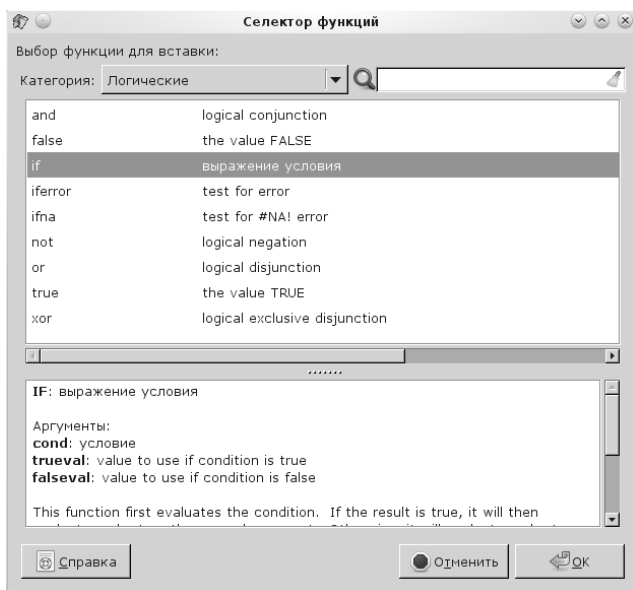


Рис. 2.52. Диалог выбора функции (селектор функций)

как `sum(...)` или `average(...)`), однако для большинства функций количество аргументов фиксировано.

В зависимости от назначения и синтаксиса функции, аргументом функции может быть число, текст, дата или логическое (булево) выражение. Чаще всего в качестве аргумента (или в составе аргумента) используются адреса ячеек. В случае нескольких аргументов разделителем аргументов является точка с запятой (символ `;`). Если в качестве аргумента используется текст, то он должен быть заключен в кавычки (символ `”`). Кроме того, в аргументах функций могут использоваться другие функции и арифметические выражения (например, возможны формулы типа `=2*sin(3*A3*pi()/4)`).

Для вызова селектора встроенных функций используется либо главное меню («Вставка/Функция»), либо кнопка `f(x)` в панели инструментов. После этого появляется диалог выбора функции. На рис. 2.52 показан пример для логической функции `if()`, которая подробнее будет рассмотрена ниже.

При выборе функции сначала выбирается категория функций (в верхней части диалога), а затем — конкретная функция (в средней части диалога). В нижней части диалога приводится объяснение назначения и структуры функции. После нажатия на кнопку «ОК» появляется диалог определения аргументов выбранной функции (помощник

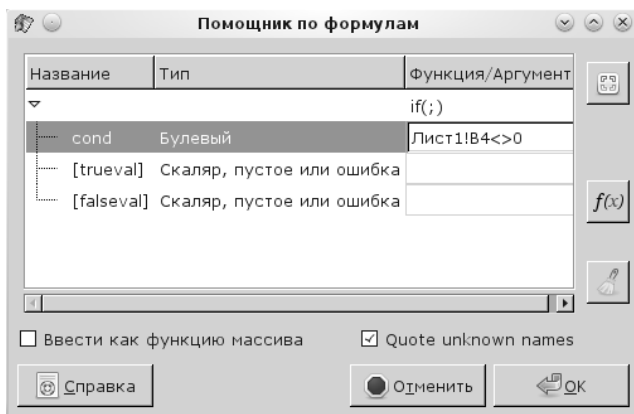


Рис. 2.53. Составление формулы)

по формулам, рис. 2.53). Для приведенной в примере функции `if()` нужно определить три аргумента.

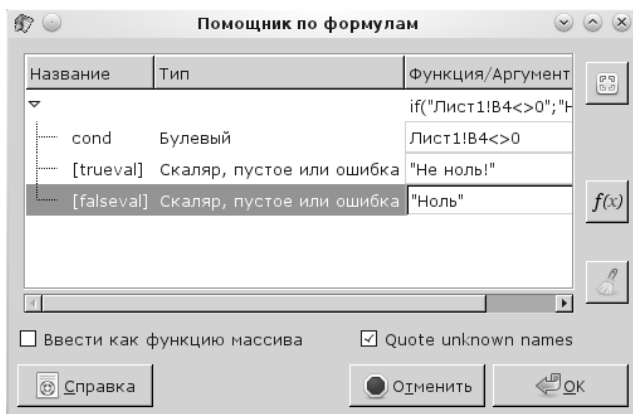
Для определения аргументов нужно щелкнуть мышкой в поле ввода (в столбце «Функция/Аргумент», указать ячейку со значением аргумента (можно щелкнуть мышкой по нужной ячейке) и при необходимости дописать остальное. Завершается определение аргумента нажатием на **ENTER**, после чего с помощью щелчка мышкой переходим к определению следующего аргумента. По мере определения аргументов функция дописывается.

Окончательный результат построения функции `if()` показан на рис. 2.54. В данном случае проверяется ячейка **C3**. Если в ней содержится ненулевое значение, должно быть выведено сообщение «Не ноль!», в противном случай — сообщение «Ноль»

Кнопка `f(x)` в этом диалоге позволяет вставить функцию в качестве аргумента формируемой функции. При необходимости можно редактировать ссылки типа «Лист1!B4», превращая их в ссылки типа «B4».

После окончательного определения всех аргументов и нажатия на кнопку **OK** можно наблюдать результат работы функции. При необходимости можно редактировать функцию либо в строке ввода по нажатию клавиши **F2**, либо снова вызвать диалог определения аргументов функции, используя кнопку `f(x)`, когда ячейка с результатом функции является активной.

Далее кратко рассмотрим несколько основных групп функций, а потом перейдем к конкретными примерам.

Рис. 2.54. Функция `if()` с аргументами

2.6.2 Математические функции

`Gnumeric` содержит около 80 встроенных математических функций. Описывать все нет необходимости (использование и особенности функций типа `abs()`, `sin()`, `tan()` или `pi()` достаточно очевидны). Приведем здесь краткое описание некоторых более редких математических функций.

Таблица 2.2: Некоторые неочевидные математические функции

Название, аргументы	Назначение
<code>atan2(b1;b2)</code>	Вычисляет арктангенс отношения $b2/b1$ без учета знаков аргументов.
<code>beta(a;b)</code>	Вычисляет бета-функцию (интеграл Эйлера I рода) для положительных целых аргументов.
<code>betaln(a;b)</code>	Вычисляет натуральный логарифм бета-функции для положительных целых аргументов.
<code>combin(n;k)</code>	Вычисляет количество уникальных комбинаций из n по k .
<code>expm1(x)</code>	Вычисляет $\exp(x)-1$ с высокой точностью.
<code>factdouble(m)</code>	Вычисляет двойной факториал положительного числа m , т.е. $m!!$. Если m не целое, дробная часть обрезается.
<code>hypot(a;b;...)</code>	Вычисляет квадратный корень из суммы квадратов аргументов.

Таблица 2.2 — продолжение

Название, аргументы	Назначение
multinomial (x1;x2;...)	Вычисляет отношение факториала суммы аргументов к произведению их факториалов.
quotient(a;b)	Вычисляет целую часть результата деления a на b.
roman(m;тип)	Преобразует натуральное число в «римскую» форму записи. Необязательный аргумент «тип» определяет вариант нотации (классический или краткий).
seriessum(X;n;m; коэффициенты)	Вычисляет сумму степенного ряда вида $aX^n + bX^{n+m} + cX^{n+2m} + dX^{n+3m} \dots$, т.е. X — основание степени, n — начальное значение показателя степени, m — инкремент показателя степени, a, коэффициенты при каждом члене ряда, записанные в блоке ячеек таблицы.
sqrtpi(m)	Вычисляет квадратный корень из $m \cdot \pi()$ (<code>sqrt(m*pi())</code>).
sumsq (x1;x2;...)	Вычисляет сумму квадратов аргументов.
sumx2my2 (вектор1;вектор2)	Вычисляет сумму разностей квадратов соответствующих элементов векторов. Вектор1 и вектор2 — блоки ячеек одинаковой длины.
sumx2py2 (вектор1;вектор2)	Вычисляет сумму сумм квадратов соответствующих элементов векторов.
sumxmy2 (вектор1;вектор2)	Вычисляет сумму квадратов разностей соответствующих элементов векторов.

Логических функций вообще немного, поэтому рассмотрим их все. Каждая логическая функция может быть аргументом другой, если это позволено синтаксисом.

Таблица 2.3: Логические функции в Gnumeric

Название, аргументы	Назначение
false()	Аргументов не имеет. Всегда выдает логическое значение «ЛОЖЬ» (FALSE).
true()	Аргументов не имеет. Всегда выдает логическое значение «ИСТИНА» (TRUE).

Таблица 2.3 — продолжение

Название, аргументы	Назначение
and(условие1; условие2;...)	Имеет неограниченное количество аргументов (более одного). Выдает логическое значение «ИСТИНА» (TRUE), если выполняются все условия, приведенные в аргументах. Пример синтаксиса: and(c1>0,b4=3,d2<b4), где c1, b4 и d2 — адреса ячеек.
or(условие1; условие2;...)	Имеет неограниченное количество аргументов (более одного). Выдает логическое значение «ИСТИНА» (TRUE), если выполняются хотя бы одно из условий, приведенных в аргументах. Синтаксис аналогичен функции and().
not(условие)	Аргументом является условие или результат работы логической функции (логическое TRUE или FALSE). Выдает логическое значение «ИСТИНА» (TRUE), если условие не выполняется или аргумент установлен в FALSE.
xor(условие1; условие2;...)	Имеет неограниченное количество аргументов (более одного). Выдает логическое значение «ИСТИНА» (TRUE), если выполняется нечетное количество условий (исключающее ИЛИ).
if(условие; действие1; действие2)	Проверяет условие, и если оно выполняется, производится действие1 (возможна проверка еще какого-то условия, выполнение арифметических операций или вычисление по формуле с функциями, а также вывод текста). В противном случае выполняется действие2 (с теми же особенностями).

Важным свойством функции if () является возможность использования результата работы одной функции if () в качестве аргумента другой функции if (). Однако количество таких вложений ограничивается максимально допустимой длиной текста в ячейке, а также здравым смыслом (формулы с количеством проверок условий более 7-8 крайне трудно анализировать в случае ошибок). Также в качестве условий в if () могут использоваться любые синтаксически допустимые комбинации логических функций.

2.6.3 Функции комплексного переменного

Функций работы с комплексными числами в Gnumeric насчитывается более сорока, поэтому здесь рассмотрим только основные (и наиболее интересные с точки зрения автора).

Таблица 2.4: Некоторые неочевидные математические функции

Название, аргументы	Назначение
<code>complex(x1;x2;символ)</code>	Формирует комплексное число из двух вещественных. Третий необязательный аргумент (символ) позволяет изменить обозначение мнимой единицы. Если он не указан, будет сформировано комплексное число вида $x1+x2i$.
<code>imabs(complex)</code>	Вычисляет модуль комплексного числа. Например, если в ячейке E3 записано комплексное число $5+3i$ (как результат функции <code>complex()</code>), то <code>imabs(E3)</code> выдаст 5,83095.
<code>imargument(complex)</code>	Вычисляет аргумент комплексного числа (показатель степени при экспоненциальном представлении). Для примера $5+3i$ выдаст значение 0,54042.
<code>imreal(complex)</code>	Выдает вещественную часть комплексного числа.
<code>imaginary(complex)</code>	Выдает мнимую часть комплексного числа.
<code>imconjugate(complex)</code>	Вычисляет комплексно сопряженное число.
<code>imdiv(complex1; complex2)</code>	Вычисляет целую часть результата деления двух комплексных чисел.
<code>iminv(complex)</code>	Выполняет преобразование $1/z$.
<code>impower(complex;power)</code>	Возводит комплексное число в степень, которая тоже может быть комплексным числом.
<code>improduct(complex1; complex2;...)</code>	Вычисляет произведение комплексных чисел (обычная операция умножения не работает!).

Далее рассмотрим использование логических и математических функций, а также функций комплексного переменного на классическом примере вычисления корней квадратного уравнения с произвольными коэффициентами.

	А	В	С
1	Решатель квадратных уравнений		
2			
3	Коэффициенты: А		1
4		В	2
5		С	3
6	Дискриминант: D		-8
7	Корни:	X1	-1+2,82842712474619i
8		X2	-1-2,82842712474619i
9			

Рис. 2.55. Решение квадратного уравнения

Итак, заданы три коэффициента А, В и С квадратного уравнения вида

$$A \cdot x^2 + B \cdot x + C = 0 \quad (2.1)$$

Требуется вычислить корни x_1 и x_2 , которые в общем случае могут быть комплексными. Перед вычислением корней вычислим дискриминант D:

$$D = B^2 - 4 \cdot A \cdot C \quad (2.2)$$

И затем воспользуемся формулой для вычисления корней:

$$\frac{x_{1,2} = -B + -\sqrt{D}}{2 \cdot A} \quad (2.3)$$

Однако при отрицательном дискриминанте корни будут комплексными. Такое комплексное число будет иметь вещественную часть $(-B/2A)$ и мнимую часть (с точностью до знака)

$$\Im(x_{1,2}) = + - \frac{\sqrt{D}}{2 \cdot A} \quad (2.4)$$

В то же время для неотрицательных значений дискриминанта будут работать обычные правила вычисления корней (в соответствии с формулой (3)). Таким образом, в формуле для вычисления корня должна присутствовать проверка дискриминанта на отрицательность, и при отрицательном дискриминанте должно быть сформировано комплексное число. При неотрицательном дискриминанте используются обычные функции и арифметические действия.

После столь долгих рассуждений пора показать таблицу и формулы для вычислений (рис. 2.55).

Формулы для вычислений приведены ниже:

Таблица 2.5: Формулы для решения квадратного уравнения

Адрес ячейки, назначение	Формула
C6: Дискриминант	=C4^2-4*C3*C5
C7: Корень X1	=if(C6<0;complex(-C4/(2*C3);sqrt(abs(C6)));(-C4+sqrt(C6))/(2*C3))
C8: Корень X2	=if(C6<0;complex(-C4/(2*C3),-sqrt(abs(C6)));(-C4-sqrt(C6))/(2*C3))

2.6.4 Календарные функции

Календарные функции в Gnumeric находятся в категории «Дата/Время», всего таких функций более 30. Основными являются функции «разложения» даты на составляющие — выделения из даты номера дня в месяце (функция `day()`), номера месяца в году (`month()`) и года (`year()`) — и функция обратного преобразования `date(year(), month(), day())`, которая конструирует данные типа «дата» из номера года, месяца и дня.

Также часто используются функции автоматического определения дня недели (`weekday()`), текущей даты (`today()`) и текущего момента времени (`now()`).

Имеются также интересные (и полезные) функции преобразования дат — `date2unix()` и обратная ей `unix2date()`. Они выполняют преобразование даты в количество секунд, прошедших с начала «эры UNIX», и наоборот («эра UNIX» началась в полночь 1 января 1970 года).

Для знакомства с основными календарными функциями рассмотрим следующую задачу. Дан список лиц, для которых известны фамилии, пол и даты рождения. Определить:

- День недели, на который приходится день рождения каждого человека в текущем году. Если день рождения приходится на выходные дни, вывести текст «УРА!», в остальных случаях вывести текст «УВЫ...».
- Возраст на настоящий момент.
- Дату выхода на пенсию для каждого человека.

Для этой задачи воспользуемся исходными данными (списком фамилий) из задачи про доходы и налоги. Пол установим в соответствии с фамилиями, даты рождения введем произвольно (как уже упоминалось, удобно вводить дату в виде 23/11/88, а программа приведет ее в нужный вид).

Для получения решения по пункту «а» необходимо проделать некоторые промежуточные вычисления. Сначала нужно для каждого лица сформировать дату рождения в текущем году на основании дня и месяца рождения, а также номера текущего года. Тогда первая формула (в ячейке D4 на рис. 2.56) будет иметь вид:

```
=DATE(YEAR(TODAY());MONTH(C4);DAY(C4))
```

Очевидно, что функция `today()` просто выдает текущую дату, а функция `date()` формирует дату из номера года, номера месяца и номера дня в месяце, определяемых соответственно, с помощью функций `year()`, `month()` и `day()`. Естественно, при желании вместо результатов работы функций можно использовать адреса ячеек, содержащих числа или просто числа. Соответственно, формула для окончательного результата по пункту «а» (в ячейке E4) будет иметь вид:

```
=IF(OR(WEEKDAY(D4;2)=6;WEEKDAY(D4;2)=7);"УРА!";"УВЫ...")
```

где функция `weekday()` вычисляет день недели для даты, указанной в первом аргументе. Вторым аргумент функции `weekday()` определяется правила нумерации дней недели. В данном случае первый день недели соответствует понедельнику.

Для обработки всего списка просто копируем эти формулы вниз. Для получения решения по пункту b воспользуемся функцией вычисления разности дат — `datedif()`. Эта функция имеет три аргумента — начальную дату, конечную дату и строку-параметр, задающую единицы измерения разности. Значения параметра `y`, `m` или `d` позволяют найти разность дат соответственно, в полных годах, полных месяцах и в днях. С учетом того, что возраст определяется в полных годах, запишем формулу для возраста в следующем виде:

```
=datedif(C4;TODAY();"y")
```

Для получения решения по пункту c снова нужно формировать даты, используя значения параметров возраста выхода на пенсию. Эти возрасты на момент написания книги составляют 60 лет для мужчин и 55 для женщин, однако они могут в любой момент измениться, поэтому конкретные числа в формулу записывать не будем. Дата выхода на пенсию формируется с использованием условия проверки пола. Итак, получаем формулу:

```
=IF(B4="М";DATE(YEAR(C4)+$G$1,MONTH(C4);DAY(C4));  
DATE(YEAR(C4)+$G$2;MONTH(C4);DAY(C4)))
```

Результирующая таблица показана на рис. 2.56.

Последний столбец (дата выхода на пенсию) приведен в формате `dd/mm/уууу` для иллюстрации корректного выполнения вычислений.

А	В	С	Д	Е	Ф	Г
	Сегодня:	03/04/10			Мужчины	60
					Женщины	55
Фамилия, инициалы	Пол	Дата рожд.	Д/р в этом году		Возраст	На пенсию
Сионов К.С.	М	11/03/56	11/03/10	УВЫ...	54	11/03/2016
Чернышевская С.П.	Ж	03/11/65	03/11/10	УВЫ...	44	03/11/2020
Чкалов В.П.	М	22/08/99	22/08/10	УРА!	10	22/08/2059
Романов П.А.	М	08/01/88	08/01/10	УВЫ...	22	08/01/2048
Штек Е.А.	Ж	06/06/66	06/06/10	УРА!	43	06/06/2021
Зубарева О.А.	Ж	05/05/55	05/05/10	УВЫ...	54	05/05/2010
Жеребцова Н.Н.	Ж	29/09/77	29/09/10	УВЫ...	32	29/09/2032
Зубов П.Р.	М	16/01/68	16/01/10	УРА!	42	16/01/2028

Рис. 2.56. Пример вычисления с датами

2.6.5 Функции поиска соответствий

В тех случаях, когда использование функции `if()` становится неудобным по причине большого количества вложений или большой длины формулы, а также для повышения эффективности вычислений, в электронных таблицах используются функции поиска соответствий. К таким функциям относятся `lookup()`, `vlookup()`, `hlookup()`, `match()` и `index()`, находящиеся в категории «Поиск».

Понимание работы этих функций может сначала вызывать некоторые затруднения, однако они (вместе с `if()`) обеспечивают мощные средства обработки данных, поэтому их освоение стоит затраченного времени.

Функции `lookup()`, `vlookup()` и `hlookup()` в качестве одного из аргументов используют так называемые «ассоциативные массивы» (справочники). Ассоциативный массив — это структура данных, оформленная в виде таблицы, первый столбик которой содержит «ключи» — данные, которые участвуют в формировании условий. В следующих столбиках ассоциативного массива содержатся значения, соответствующие ключам. Таким образом, по значению ключа можно однозначно получить какие-то другие данные (в языках программирования такие структуры называются «хэш-таблицы»).

В программах ЭТ ассоциативные массивы реализуются как блоки ячеек (справочные таблицы), содержащие минимум два столбца. Первый столбец содержит ключи, второй — значения, соответствующие ключам.

Для примера рассмотрим следующую задачу. В ралли участвуют гонщики на автомобилях, для которых известны марки и расход топлива в литрах на 100 км. Дан список, в котором указаны фамилии гонщиков и марки автомобилей. Известна протяженность трас-

E	F	G	
	Марка а/м	Расход, л/100км	
	BMW	13	
	FIAT	10,5	
	Ford	15	
	Honda	11,4	
	Nissan	14,3	
	Toyota	12,5	
	Volvo	13,3	

Рис. 2.57. Справочник для задачи о ралли

A	B	C	D
	Длина трассы L	500	
Фамилия	Марка а/м	Расход на трассу	
Шумахер	Toyota		
Бельмондо	FIAT		
Валигора	Nissan		
Пороханов	Volvo		
Коннери	BMW		
Сигал	Ford		
Кэмпбелл	Honda		

Рис. 2.58. Исходные данные задачи о ралли

сы гонок в километрах L. Определить расход топлива для каждого участника гонок.

Для решения задачи составим справочную таблицу для семи разных марок а/м, и расположим ее в ячейках F3:G10 (рис. 2.57). Полезно соблюдать алфавитный порядок текстовых значений в первом столбике справочной таблицы.

Данные (список участников и марки их а/м) запишем в ячейки A3:B10 (рис. 2.58). Длину трассы L, которая является параметром, запишем в ячейку C1. Для вычисления полного расхода топлива используем формулу с функцией lookup().

Формула в ячейке C4 будет выглядеть следующим образом.

А	В	С	Д
	Длина трассы L		500
Фамилия	Марка а/м	Расход на трассу	
Шумахер	Toyota	62,5	
Бельмондо	FIAT	52,5	
Валигора	Nissan	71,5	
Пороханов	Volvo	66,5	
Коннери	BMW	65	
Сигал	Ford	75	
Кэмпбелл	Honda	57	

Рис. 2.59. Решение задача о ралли

=LOOKUP(B4;\$F\$4:\$F\$10;\$G\$4:\$G\$10)*\$C\$1/100

Функция `lookup()` считывает содержание ячейки, указанной в первом аргументе, ищет это значение в диапазоне ячеек (столбце), указанном во втором аргументе и выдаёт соответствие этому значению из диапазона ячеек (столбца), указанном в третьем аргументе. Таким образом, для получения результата по названию а/м находим расход топлива на 100 км и умножаем это значение на количество сотен километров. Принципиально важно указывать абсолютные адреса блоков ячеек справочной таблицы.

Итоговая таблица показана на рис. 2.59.

Ограничение функции `lookup()` — только один столбец соответствий. Более «мощной» является функция `vlookup()`, в которой второй аргумент определяет весь блок ячеек, содержащих ассоциативный массив (справочник), а третий аргумент указывает, в каком столбце ассоциативного массива нужно искать соответствие ключу. Четвёртый (необязательный) аргумент определяет порядок сортировки первого («ключевого») столбца справочника. Если он не указан или равен 1 (логическая ИСТИНА), то первый столбец ассоциативного массива для функции `vlookup()` должен содержать числа, отсортированные по возрастанию, или текст, отсортированный в алфавитном порядке. Если значения в первом столбце не отсортированы, то четвёртый аргумент должен быть установлен в 0. Ещё одним большим достоинством функции `vlookup()` является возможность работы с диапазонами значений ключа.

Для примера рассмотрим вычисление суммы годового налога при прогрессивной налоговой шкале. Пусть при годовом доходе до 10000 у.е. ставка налога составляет 12%, до 30000 у.е. — 20%, до

А	В	С
Фамилия	Годовой доход	Сумма налога
Шумахер	78000	
Бельмондо	46000	
Валигора	29000	
Пороханов	30000	
Коннери	10000	
Сигал	18400	
Кэмпбелл	5670	

Рис. 2.60. Исходные данные задачи о налогах

Е	Ф
Сумма дохода	Ставка налога
0	0,12
10000	0,2
30000	0,25
50000	0,35

Рис. 2.61. Справочник к задаче о налогах

50000 у.е. — 25%, и при большем доходе — 35%. Для создания таблицы данных используем фамилии из предыдущего примера, а суммы годового дохода запишем такие, чтобы можно было реализовать все варианты ставок налога (рис. 2.60). Таблицу данных разместим в диапазоне А3:В10.

Справочную таблицу размещаем в диапазоне Е3:Ф7 (рис. 2.61), формула в ячейке С4 с использованием функции `vlookup()` выглядит следующим образом:

$$=VLOOKUP(B4; \$E\$4: \$F\$7; 2) * B4$$

Во втором столбце ассоциативного массива находим ставку налога, соответствующую доходу, а затем получаем сумму налога, умножая ставку налога на величину дохода (рис 2.62).

А	В	С
Фамилия	Годовой доход	Сумма налога
Шумахер	78000	27300,00
Бельмондо	46000	11500,00
Валигора	29000	5800,00
Пороханов	30000	7500,00
Коннери	10000	2000,00
Сигал	18400	3680,00
Кэмпбелл	5670	680,40

Рис. 2.62. Решение задачи о налогах

Важно, что `vlookup()` выбирает значения, соответствующие нижней границе диапазона значений ключей, и используется минимальное значение ключа в интервале.

Четвертый аргумент функции `vlookup()` также влияет и на возможность интервального просмотра. Если этот аргумент имеет значение ИСТИНА (1) или опущен, то интервальный просмотр работает, как описано выше. Если этот аргумент имеет значение ЛОЖЬ (0), то функция `vlookup()` ищет точное соответствие. Если таковое не найдено, то возвращается значение ошибки N/A (Н/Д — «нет данных»). Таким образом, для использования возможности работы с интервалами значений в `vlookup()` первый столбец справочной таблицы (ассоциативного массива) обязательно должен быть отсортирован по возрастанию.

Функция `hlookup()` работает аналогично `vlookup()`, только порядок следования «ключей» — не сверху вниз, а слева направо.

Теперь рассмотрим формат функций `match()` и `index()`.

Функция `match(искомое_значение; искомый_массив; тип_сопоставления)` находит позицию (порядковый номер) искомого значения в одномерном массиве. Значение аргумента «тип сопоставления» — (-1, 0 или 1) — зависит от того, упорядочен ли массив (-1 — массив упорядочен по убыванию, находится место наименьшего значения, которое больше или равно искомому, 0 — массив может быть неупорядоченным, находится место первого значения, равного исходному, 1 — массив упорядочен по возрастанию, находится место наибольшего значения, которое меньше или равно искомому).

Таблица 2.6. Входные данные

Входная плата, у.е.	1	1,5	2	2,5	3	3,5	5
Кол-во посетителей	200	175	160	140	124	110	70

А	В	С	Д	Е	Ф	Г	Н
Входная плата, ч.е.	1	1,5	2	2,5	3	3,5	5
Кол-во посетителей	200	175	160	140	124	110	70
Выручка	200	262,5	320	350	372	385	350
Макс. выручка	385	max(B4:H4)					
Позиция	6	match(B6:B4:H4)					
Оптимальная плата	3,5	index(B2:H4;1:B7)					

Рис. 2.63. Решение задачи о дискотеке

Функция `index(массив; номер_строки; номер_столбца)` находит значение элемента, находящегося в заданном массиве на пересечении заданных строки и столбца.

Далее рассмотрим пример. Пусть на основе эксперимента получена следующая зависимость посещаемости дискотеки от входной платы:

Необходимо определить оптимальную входную плату.

Очевидно, что оптимуму соответствует максимальная выручка. Соответственно, решение задачи состоит в том, чтобы подсчитать выручку в каждом случае, найти максимальную и написать формулу, показывающую входную плату, соответствующую максимальной выручке. То есть нужно определить номер столбца, в котором получается максимальная выручка и вывести значение входной платы.

Определяем выручку в каждом случае (умножив входную плату на количество посетителей), затем функцией `max()` находим наибольшую выручку. После этого функцией `match()` определяем, на каком месте в массиве она находится и функцией `index()` смотрим, какая входная плата находится на этом месте (рис. 2.63). Рядом с ячейками с результатами приведены формулы для получения этих результатов (используется функция `expression()`).

В этом примере участвует функция `max()`, которая относится к категории статистических функций, к знакомству с которыми теперь и перейдем.

2.6.6 Статистические функции

Основные статистические функции — это `min()`, `max()` и `average()`, вычисляющие, соответственно, минимальное, максимальное и среднее значение для набора аргументов. Аргументы могут быть либо перечислены через точку с запятой (если данные находятся в несмежных ячейках), либо в качестве аргумента может быть использован диапазон ячеек. Также к статистическим функциям можно отнести функции `sumif()` и `countif()`, которые формально находятся в Gnumeric среди математических функций, а также функцию `rank()`, определяющее «рейтинг» какого-то значения в списке аналогичных значений.

В качестве примера рассмотрим следующую задачу.

Дан список участников соревнования среди студентов по бегу на 100 метров и метанию мяча. В таблице (рис. 2.64) указаны пол (юноша или девушка) и результаты. Определить места каждого участника в каждом виде соревнований, минимальный, максимальный и средний результаты в каждом виде соревнований, на сколько юноши (в среднем) метают мяч дальше, чем девушки.

Сразу под списком в соответствующих столбцах подсчитываем максимальный, минимальный и средний результаты (по столбцу C =`Max(C2:C16)`, =`Min(C2:C16)`, =`AVERAGE(C2:C16)`) и аналогично по столбцу D).

Затем подсчитываем количество юношей и количество девушек. В ячейку B22 вносим формулу

```
=COUNTIF(B2:V16;"юноша"),
```

а в ячейку B23 формулу

```
=COUNTIF(B2:V16;"девушка")
```

В ячейки C22 и C23 записываем формулы для подсчета суммы результатов по метанию для юношей и девушек соответственно

```
=SUMIF(B2:V16;"юноша",D2:D16)
```

```
=SUMIF(B2:V16;"девушка",D2:D16)
```

После чего подсчитываем среднее значение в ячейках D22 и D23, разделив сумму результатов на количество участников в каждой группе. Затем подсчитываем разницу средних значений (рис. 2.64).

Для иллюстрации использования функции `rank()` рассмотрим следующую задачу: На основе данных из задачи о соревнованиях распределить места в соревнованиях по бегу среди юношей, а также определить, сколько человек показали результаты ниже среднего (т.е. время больше среднего).

Для распределения участников по местам как раз и потребуется функция `rank()`, а кроме того потребуется `countif()` для вычисления

А	В	С	Д
Фамилия	Пол	БЕГ	МЕТАНИЕ
Беляева	девушка	17,6	35
Васильева	девушка	18,2	55
Виноградов	юноша	14,3	38
Воробьева	девушка	16,8	46
Герасимов	юноша	15,2	55
Дмитриев	юноша	12,3	57
Егоров	юноша	14,5	56
Иванова	девушка	15,9	47
Карпов	юноша	15,3	50
Крылова	девушка	15,9	38
Матвеева	девушка	18,4	42
Новикова	девушка	16,3	53
Осипова	девушка	17,5	51
Павлова	девушка	16,0	54
Петров	юноша	13,4	57
наибольшее значение		18,4	57
наименьшее значение		12,3	35
среднее значение		15,8	48,93
	количество	сумма метров	среднее значение
юноши	6	313	52,17
девушки	9	421	46,78
	разница		5,39

Рис. 2.64. Решение задачи о соревнованиях

с условием. Однако условие для `countif()` обязательно должно быть текстом, поэтому при формировании условия с вычисляемыми данными целесообразно использовать текстовую функцию `concatenate()`. Результат использования этих функций показан на рис. 2.65.

Формула для определения места участника соревнований (ячейка С2) будет выглядеть следующим образом:

```
=rank(B2;$B$2:$B$7;1)
```

Абсолютные адреса диапазона использованы для обеспечения возможности копирования формулы. Третий аргумент установлен в 1, что обеспечивает обратный порядок распределения мест, т.е. чем меньше значение, тем лучше место (первое место — минимальный результат). Для обеспечения прямого порядка распределения мест

А	В	С
Фамилия	БЕГ	Места
Виноградов	14,3	3
Герасимов	15,2	5
Дмитриев	12,3	1
Егоров	14,5	4
Карпов	15,3	6
Петров	13,4	2
Среднее время	14,1666666666667	
Условие	>14,1666666666667	
Хуже среднего	4	

Рис. 2.65. Задача о распределении мест

(первое место — максимальный результат) третий аргумент нужно установить в 0 или не указывать.

Среднее время определяется с помощью функции `average()`, а условие для подсчета аутсайдеров формируется с помощью текстовой функции `concatenate()`, которая «сцепляет» строки для формирования одного значения строкового типа. В данном случае в ячейке B10 записана формула

```
=concatenate(">";B9).
```

Подсчет аутсайдеров выполняется по формуле (ячейка B11)

```
=countif(B2:B7;B10).
```

К статистическим функциям относятся также функции `count()` (подсчет количества числовых значений в диапазоне ячеек) и `counta()` (подсчет количества не пустых ячеек в диапазоне). Кроме того, большое число статистических функций вычисляют статистические параметры различных распределений и обеспечивают генерацию случайных чисел с заданными параметрами распределений.

2.6.7 Текстовые (строковые) функции

Одна из типовых задач в офисной работе — изменения формы представления списков людей или организаций. Избежать трудоёмкой работы по переписыванию текста из одного вида в другой помогают функции работы с текстом (строковые функции).

Почему-то подобные задачи кажутся очень сложными, поэтому покажем пример решения и рассмотрим минимальный набор функций, необходимых для получения результата.

	A	D	H	J	L
1					
2	Фамилия, имя, отчество	Фамилия	Имя	Отчество	Результат
3	Виноградов Василий Семёнович	Виноградов	Василий	Семёнович	Виноградов В.С.
4	Герасинов Иван Сергеевич	Герасинов	Иван	Сергеевич	Герасинов И.С.
5	Дмитриев Сергей Никитович	Дмитриев	Сергей	Никитович	Дмитриев С.Н.
6	Егоров Василий Иванович	Егоров	Василий	Иванович	Егоров В.И.
7	Карпов Владимир Яковлевич	Карпов	Владимир	Яковлевич	Карпов В.Я.
8	Петров Борис Андреевич	Петров	Борис	Андреевич	Петров Б.А.

Рис. 2.66. Пример преобразования списка

Пусть имеется список лиц с полными фамилиями, именами и отчествами (например, список группы слушателей каких-нибудь курсов). Для составления журнала полные имена и отчества не требуются, а требуются только инициалы. Задача заключается в преобразовании имён и отчеств в буквы инициалов.

На рис. 2.66 показан пример решения такой задачи. Следует обратить внимание, что имеются скрытые столбцы, в которых содержатся результаты промежуточных вычислений.

Алгоритм решения задачи может быть таким:

1. Определяем длину строки «Фамилия, имя, отчество»
2. Определяем длину фамилии (количество букв до первого пробела)
3. Делим строку «Фамилия, имя, отчество» на фамилию и всё что осталось (получаются строки «Фамилия» и «Имя, отчество»)
4. Определяем длину строки «Имя, отчество»
5. Определяем длину имени (количество букв до первого пробела в строке «Имя, отчество»)
6. Делим строку «Имя, отчество» на имя и всё что осталось (получаются строки «Имя» и «Отчество»)
7. Выделяем первую букву имени
8. Выделяем первую букву отчества
9. Создаём итоговую строку из строки «Фамилия» и первых букв имени и отчества

Строковые функции, которые понадобятся для решения данной задачи, описаны в таблице ниже.

Таблица 2.7: Некоторые строковые функции

Название, аргументы	Назначение
len(str)	Вычисляет длину (количество символов) для строки str.
find(str1;str2;start)	Определяет позицию (номер символа), с которой начинается подстрока str1 в строке str2, начиная с позиции start. Если аргумент start не указан, поиск идёт с начала строки.
left(str;n)	Выделяет n символов с начала строки str. Если аргумент n не указан, функция возвращает первый символ строки.
right(str;n)	Выделяет n символов с конца строки str. Если аргумент n не указан, функция возвращает последний символ строки.
concatenate (str1;str2;. . . ;strN)	Формирует одну строку из «фрагментов» — строк str1, str2, . . . , strN.

В следующей таблице приведены формулы, использованные при решении задачи.

Таблица 2.8: Формулы для решения квадратного уравнения

Формула	Адрес ячейки, назначение
=len(A3)	B3: Длина исходной строки
=find(" ";A3)	C3: Позиция первого пробела (количество букв в фамилии)
=left(A3;C3)	D3: Строка «Фамилия»
=right(A3;B3-C3)	E3: Строка «Имя, отчество»
=len(E3)	F3: Длина строки «Имя, отчество»
=find(" ";E3)	G3: Позиция первого пробела в строке «Имя, отчество» (количество букв в имени)
=left(E3;G3)	H3: Строка «Имя»
=left(H3)	I3: Первая буква имени
=right(E3;F3-G3)	J3: Строка «Отчество»
=left(J3)	K3: Первая буква отчества
=concatenate (D3;" "; I3;" ";K3;" ")	L3: Результат

Символьные (строковые) значения в формулах (в аргументах функций) следует указывать в кавычках.

В качестве других полезных строковых функций (по мнению автора) нужно отметить функции преобразования регистров `lower()` (переводит все символы строки в нижний регистр, т.е. в строчные буквы) и `upper()` (переводит все символы строки в верхний регистр, т.е. в прописные буквы), функцию `mid()`, которая позволяет вывести заданное количество символов строки, начиная с заданного символа, а также функцию `value()`, превращающую строку из символов-цифр в число.

2.6.8 Обработка матриц

`Gnumeric` обеспечивает базовые возможности по работе с матрицами. Список функций приведён в таблице ниже. Все эти функции в качестве аргументов используют диапазоны ячеек, в которые записаны элементы матриц.

Таблица 2.9: Функции для обработки матриц

Название, аргументы	Назначение
<code>transpose(matrix)</code>	Выполняет транспонирование матрицы <code>matrix</code> (строки становятся столбцами и наоборот). Функция находится в категории «Поиск».
<code>mdeterm(matrix)</code>	Вычисляет определитель квадратной матрицы.
<code>minverse(matrix)</code>	Вычисляет матрицу, обратную по отношению к исходной, при условии, что определитель не равен нулю.
<code>mmult (matrix1;matrix2)</code>	Вычисляет матричное произведение. Результирующая матрица имеет количество строк как в <code>matrix1</code> и столбцов как в <code>matrix2</code> .

Чтобы в результате операций с матрицами получить тоже матрицу (если это нужно), в «Помощнике по формулам» нужно установить режим «Ввести как функцию массива» (рис. 2.67).

Пример вычислений с матрицами показан на рис. 2.68.

В `Gnumeric` отсутствует функция для создания единичной матрицы, но такая матрица легко может быть получена умножением произвольной матрицы на обратную ей, что и проиллюстрировано на рис. 2.68 (здесь -0 указывает на то, что в результате погрешностей вычислений в связи с ограниченной точностью представления получилось вещественное отрицательное число с очень маленьким модулем).

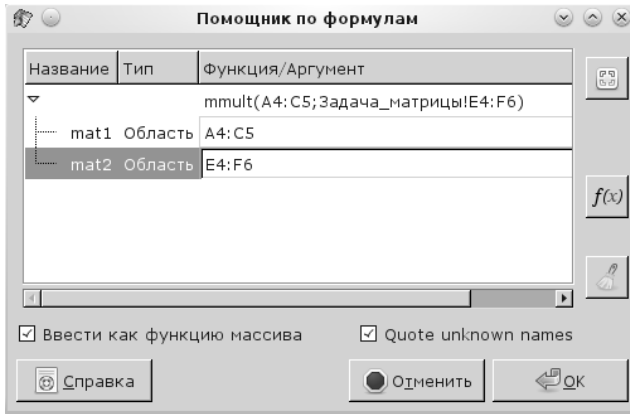


Рис. 2.67. Создание формулы в режиме работы с матрицами

	A	B	C	D	E	F	G	H	I	J	
1											
2	Матрица1				Матрица2			Матрица3			
3											
4		1	2	3		7	8		3	4	5
5		4	5	6		9	1		7	8	9
6						2	3				
7											
8	Трансп. M1				M1xM2			M2xM3			
9		1	4			31	19		77	92	107
10		2	5			85	55		34	44	54
11		3	6						27	32	37
12											
13	Матрица4				Обр. M4			Единичная			
14		3	5	6		-2,5	2	1	1	-0	-0
15		1	3	4		9,5	-8	-3	0	1	0
16		7	8	9		-6,5	6	2	0	-0	1
17											
18	Det. M4	2									

Рис. 2.68. Пример матриц и результатов операций с ними

Глава 3

Работа со списками

В этой главе рассмотрим набор возможностей Gnumeric по работе с простейшей базой данных, а именно — списком. При обработке списков в электронных таблицах используются операции сортировки по одному или нескольким признакам, а также выборки данных по условию (фильтры).

Пусть имеется некоторая модельная база данных по сотрудникам мифического магазина, состоящая из 10 полей и 78 строк (1-я строка — имена полей). Поля: «ФИО» — текстовое, «Дата рожд» — дата, «Нач.стажа» — дата, «Пол» — текстовое (одна буква), «С/п» — текстовое (одна буква), «Детей» — число (целое), «Секц» — текстовое, «Образ» — текстовое, «Должность» — текстовое, «Оклад» — число. Оклады указываются в некоторых условных единицах (у.е.).

На рис. 3.1 показано начало списка.

Для выполнения сортировки списка и использования фильтров используются элементы пункта «Данные» главного меню (рис. 3.2).

3.1 Сортировка

Сначала посмотрим, как в Gnumeric производится сортировка списка. Перед началом операции необходимо выделить весь диапазон ячеек, занимаемый списком, включая строку с именами полей (выделение производится либо «протаскиванием» мыши с нажатой левой кнопкой, либо перемещением указателя активной ячейки с помощью клавиш управления курсором — «стрелок» — при нажатой клавише SHIFT). Кнопки сортировки в панели инструментов программы обеспечивают сортировку диапазона по первому столбцу соответственно по возрастанию или по убыванию. Для детального управления условиями сортировки надо использовать вызов диалога «Данные/Сортировка...» из главного меню программы (рис. 3.3).

А	В	С	Д	Е	F	G	Н	I	J
ФИО	Дата рожд	Нач.ст.стаж	Пол	С/п	Детей	Сенц	Образ	Должность	Оклад
Агафонов Ю.А.	06/08/80	08/03/97	м	х	0	ОФ	ср	продавец-2	1050,00
Агеев С.П.	22/10/54	02/04/74	м	ж	3	ВР	ср	механик	900,00
Андреева О.П.	02/12/55	13/03/72	ж	э	1	МР	ср/сп	продавец-1	1200,00
Аникина И.Г.	09/10/63	02/09/91	ж	э	1	АУР	в	экономист	1450,00
Анисимов Г.С.	11/07/66	14/08/87	м	ж	2	МО	в	товаровед	1250,00
Аннцшин С.П.	10/10/80	04/09/98	м	ж	0	ОФ	ср	продавец-2	1050,00
Антоненко И.А.	14/01/78	18/12/97	ж	э	1	МР	ср	продавец-3	980,00
Атоманюк Т.Ф.	12/02/44	24/03/61	ж	э	3	БХ	в	товаровед	1250,00
Багирова Д.К.	03/11/60	13/02/81	ж	э	1	Б	ср/сп	продавец-2	1030,00
Базина Л.П.	23/11/79	12/12/97	ж	э	1	Б	ср	продавец-3	900,00
Баранов П.П.	22/06/43	02/03/63	м	ж	2	МР	ср	продавец-2	1100,00
Белов Н.Н.	22/06/53	02/03/75	м	ж	2	МР	в	зав.сенц	1300,00
Березнина А.С.	06/10/34	03/09/55	ж	э	1	МР	в	зав.сенц	1400,00
Бирюкова А.А.	08/05/49	28/09/72	ж	р	1	АУР	в	директор	2000,00
Богданов И.И.	07/02/68	09/11/85	м	х	0	КС	в	коммерч.дир	1800,00
Болотов Т.Г.	02/08/75	02/03/98	м	х	0	АУР	в	зам.дир	1950,00
Бондарчук С.Г.	02/08/60	02/03/79	м	х	0	ВР	ср	грузчик	600,00

Рис. 3.1. Фрагмент исходного списка

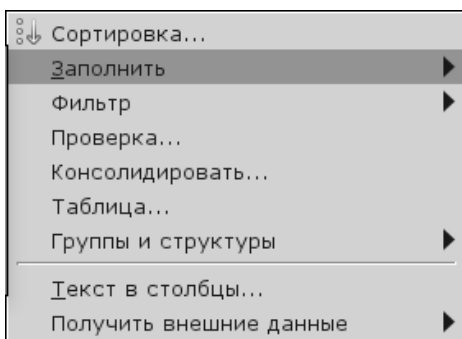


Рис. 3.2. Меню «Данные»

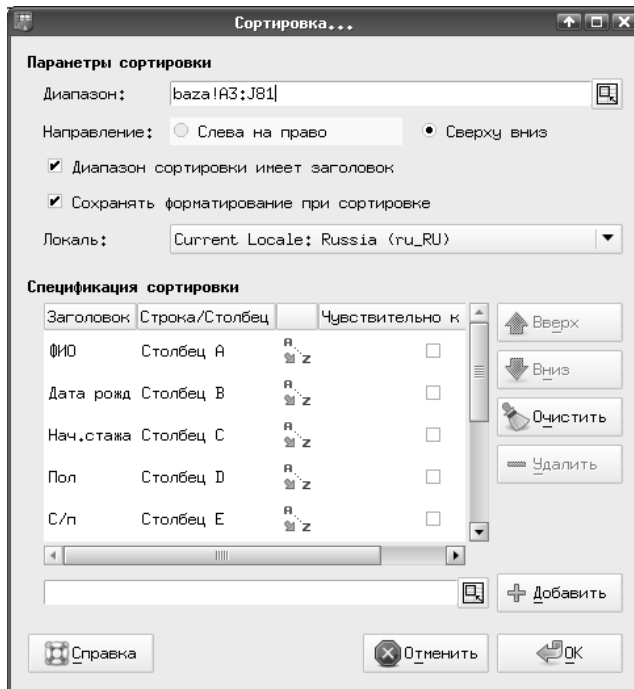


Рис. 3.3. Диалог настройки параметров сортировки

Выключение режима «Диапазон сортировки имеет заголовок» позволяет вместо номеров столбцов использовать имена полей. Щелчок мышкой по стрелке ($A \rightarrow Z$) позволяет изменять направление сортировки (вместо сортировки по возрастанию устанавливать порядок сортировки по убыванию). Назначение переключателя «Чувствительно к регистру» очевидно, и он действует для текстовых полей. Кнопки «Вверх» и «Вниз» позволяют изменять порядок критериев сортировки, а кнопки «Удалить» и «Добавить» — соответственно, удалять и добавлять критерии сортировки.

Отсортируем список по следующим критериям — сначала женщины, потом мужчины, и для каждой группы — по убыванию количества детей. В диалоге сортировки уберём лишние поля, а для оставшихся установим нужный порядок сортировки. Тогда условия сортировки будут выглядеть в соответствии с рис. 3.4.

Фрагмент результатов сортировки показан на рис. 3.5.

Далее рассмотрим возможности выбора данных из списка по заданным критериям, то есть применение фильтров.

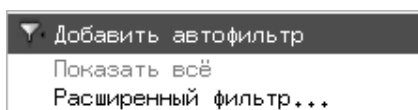


Рис. 3.6. Вложенное меню «Фильтр»

ФИО	Дата рожд	Нач. стажа	П	С	Дети	Сем	Обра	Должность	Оклад
Агафонов Ю.А.	06/08/80	08/03/97	м	х	0	0Ф	ср	продавец-2	1050,00
Агеев С.П.	22/10/54	02/04/74	м	ж	3	ВР	ср	механик	900,00
Андреева О.П.	02/12/55	13/03/72	ж	з	1	МР	ср/сп	продавец-1	1200,00
Аникина И.Г.	09/10/63	02/09/91	ж	з	1	АУР	в	экономист	1450,00

Рис. 3.7. Список с включенным автофильтром

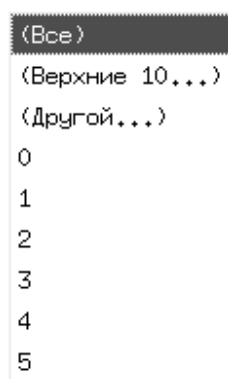


Рис. 3.8. Варианты выбора для поля «Детей»

3.2 Автофильтр

Автофильтр включается выбором команды «Данные/Фильтр/Добавить автофильтр» (рис. 3.6). При этом указатель активной ячейки должен находиться в одной из ячеек диапазона, занимаемого списком, например, в строке с именами полей. После включения Автофильтра в каждой ячейке строки с именами полей появляется значок раскрывающегося списка (рис. 3.7).

При раскрытии списка показываются все варианты значений в выбранном поле, а также варианты (Все), (Верхние 10...) и (Другой...). Список вариантов критериев выбора для поля «Детей» показан на рис. 3.8.

Самый простой вариант использования Автофильтра — выбор данных по точному соответствию значений. При этом критерии могут устанавливаться по нескольким полям одновременно, что соответ-

	А	В	С	Д	Е	Ф	Г	Н	И	Ж
1	Задание: мужчины со средне-специальным образованием									
2	Автофильтр									
3										
4	ФИО	Дата рожд	Нач.стажа	П	С	Дети	Сем	Образ	Должность	Оклад
37	Киреев М.П.	14/07/60	06/06/77	н	р	1	МР	ср/сп	продавец-2	1080,00
73	Тарасов Т.Ю.	28/10/52	26/04/74	н	х	0	ВР	ср/сп	повар	1150,00
83										

Рис. 3.9. Результат поиска мужчин со средним специальным образованием

ствует логической операции «И», т.е. все выбранные условия должны одновременно выполняться. Например, выберем мужчин, имеющих среднее специальное (ср/сп) образование. Результат показан на рис. 3.9.

О том, что данные подверглись фильтрации, можно узнать по двум признакам. Во-первых, некоторые строки таблицы оказываются скрытыми. Во-вторых, кнопки раскрывающихся списков для полей, по которым установлен фильтр, также меняют свой вид — маленькие черные треугольнички оказываются повернуты набок (и при определенных настройках интерфейса тоже меняют цвет, что заметно на приведённом рисунке).

Чтобы вернуть список в первоначальный вид, можно выполнить команду «Данные/Фильтр/Удалить автофильтр» из главного меню, а можно для «отфильтрованных» полей установить автофильтр в вариант (Все).

В списке критериев Автофильтра вариант (Другой...) позволяет установить для выбранного поля два условия, связав их логическим выражением «И» или «ИЛИ». Сначала рассмотрим возможности формирования условий поиска для текстовых полей.

Пусть требуется выбрать из общего списка людей, фамилии которых начинаются на «Ми» или «Ни». Тогда диалог настройки автофильтра по полю «ФИО» будет выглядеть следующим образом (рис. 3.10).

В данном случае варианты сравнения выбираются из списка слева. Список вариантов показан на рис. 3.11.

В рассмотренном примере для формирования условий логичным является использование логической функции «ИЛИ», поскольку фамилия не может начинаться одновременно с разных букв. Результаты поиска показаны на рис. 3.12.

Для полей текстового типа можно использовать варианты условий "equals", "does not equal", "begins with", "does not begin with", "ends with", "does not end with", "contains" и "does not contain".

Пусть теперь надо найти людей, фамилии которых состоят из пяти или шести букв. Для этого нужно указать, что в поле «ФИО»

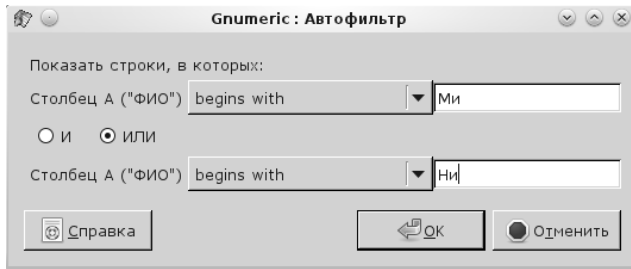


Рис. 3.10. Поиск по началу фамилии

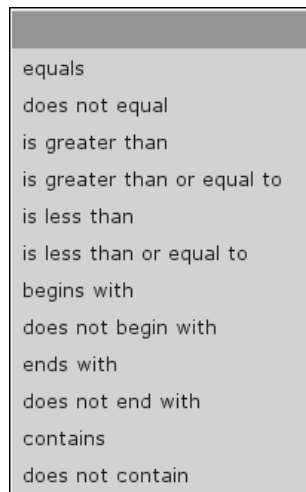


Рис. 3.11. Варианты условий в автофильтре

А	В	С	Д	Е	Ф	Г	Н	И	Ж
Задание: фамилии, начинающиеся на "Ми" или "Ни"									
Автофильтр по образцу									
ФИО	Дата ро	Нач.ста	П	С	Дет	Се	Обр	Должнос	Окла
Милашевич Н.М.	11/04/64	04/05/83	ж	з	0	ВР	н/ср	уборщица	750,00
Минина М.Р.	22/08/73	29/07/90	ж	з	1	ТОП	н/ср	фасовщица	800,00
Михайлова К.Н.	27/10/51	08/10/72	ж	з	2	ТОП	ср	ст.кассир	1250,00
Никодимова О.Е.	02/06/72	22/01/90	ж	з	2	Б	ср	продавец-3	900,00
Николаева С.И.	16/02/67	03/03/83	ж	р	2	МР	ср	продавец-1	1200,00
Николаева В.Л.	02/01/53	23/12/70	ж	н	1	Х	ср	продавец-3	900,00

Рис. 3.12. Результат поиска по началу фамилии



Рис. 3.13. Поиск слов с заданным количеством символов

А	В	С	Д	Е	Ф	Г	Н	И	Ж
Задание: фамилии из 5 или 6 букв									
Автофильтр с символами подстановки									
ФИО	Дата ро	Нач.ста	П	С	Дет	Се	Обр	Должность	Окла
Агеев С.П.	22/10/54	02/04/74	н	ж	3	ВР	ср	механик	900,00
Базина Л.П.	23/11/79	12/12/97	ж	э	1	Б	ср	продавец-3	900,00
Белов Н.Н.	22/06/53	02/03/75	н	ж	2	МР	в	зав.секц	1300,00
Бурова Д.Д.	18/05/43	13/01/59	ж	н	4	ОФ	ср/сп	продавец-1	1200,00
Гарин Ю.П.	07/12/50	06/11/73	н	ж	2	АУР	в	зам.дир	1750,00
Гранин Ю.В.	17/12/65	29/04/83	н	ж	1	Б	в	продавец-2	1030,00
Дрозд И.И.	22/09/35	06/09/60	ж	э	1	ОФ	ср	продавец-1	1200,00
Ильина Ю.В.	15/07/65	04/09/85	ж	э	1	Б	в	продавец-2	1030,00
Киреев М.П.	14/07/60	06/06/77	н	р	1	МР	ср/сп	продавец-2	1080,00
Минина М.Р.	22/08/73	29/07/90	ж	э	1	ТОП	н/ср	фасовщица	800,00
Мухина Л.Л.	13/04/64	04/12/85	ж	э	3	Б	ср	продавец-3	900,00
Озеров К.П.	13/12/49	31/01/69	н	ж	3	ОФ	в	зав.секц	1400,00
Пятова М.В.	17/03/56	09/06/75	ж	э	2	ВР	н/ср	приемщикпосуды	800,00
Тришко Р.П.	12/07/74	12/08/92	н	э	0	Х	ср	продавец-3	950,00

Рис. 3.14. Результат поиска по длине фамилий

содержатся пять (или шесть) любых символов, после которых обязательно стоит пробел. Тогда для указания любого одиночного символа используется символ подстановки «?». В этом случае условия поиска по полю «ФИО» будут выглядеть следующим образом (рис. 3.13).

Соответствующий результат показан на рис. 3.14.

Для полей числового типа (в том числе дат) при формировании условий поиска используются варианты "equals", "does not equal", "is less than", "is greater than", "is less than or equal to" и "is greater than or equal to".

Сначала рассмотрим работу с данными числового типа (включая даты) на примере поиска сотрудников, родившихся в 1975 году. Год, как известно, начинается 1 января, а заканчивается 31 декабря. По-

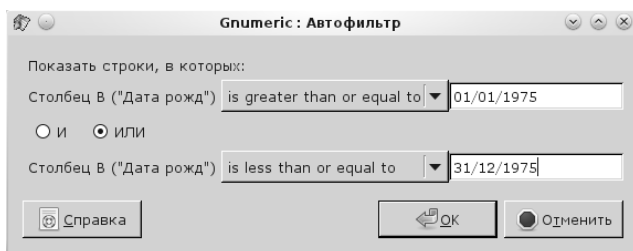


Рис. 3.15. Условие поиска в диапазоне дат

А	В	С	Д	Е	Ф	Г	Н	И	Ж
Задание: сотрудники, родившиеся в 1975г.									
Автофильтр с условиями по датам									
ФИО	Дата ро	Нач.ста	П	С	Дет	Сл	Обр	Должнос	Оклад
Болотов Т.Г.	02/08/75	02/03/98	м	х	0	АУР	в	зам.дир	1950.00
Дорофеева Л.К.	20/07/75	02/06/92	ж	н	0	Б	ср	продавец-3	900.00
Лаврентьева Ю.В.	02/01/75	25/07/91	ж	н	0	ОФ	ср	продавец-3	950.00

Рис. 3.16. Результаты поиска в диапазоне дат

этому для поля «Дата рождения» сформируем условия в соответствии с рис. 3.15.

Результат применения фильтра показан на рис. 3.16.

Наконец, рассмотрим ситуацию, когда для формирования условий нет возможности напрямую указать значения, но можно получить эти значения после некоторых расчётов.

Пусть теперь нужно получить список сотрудников, начавших трудовую деятельность летом. Поскольку в поле «Нач.стажа» нет возможности выбрать конкретный месяц и для числовых полей нельзя воспользоваться символами подстановки, воспользуемся базовыми возможностями электронной таблицы и создадим новое (расчетное) поле «Месяц» с помощью функции `month()` (перед созданием нового поля нужно отключить автофильтр).

Условие для поиска по расчётному полю показано на рис. 3.17, а результат — на рис. 3.18.

Таким образом в Gnumeric с помощью Автофильтра можно эффективно проводить поиск данных, задавая критерии для нескольких полей по очереди, используя либо точное совпадение значений, либо условия, связанные отношениями «И» или «ИЛИ». В принципе, практически для любых выборок можно создавать расчетные поля (одно или несколько), используя текстовые, математические, логические и любые другие функции электронных таблиц, однако есть более эффективные приемы работы, которые описываются ниже.

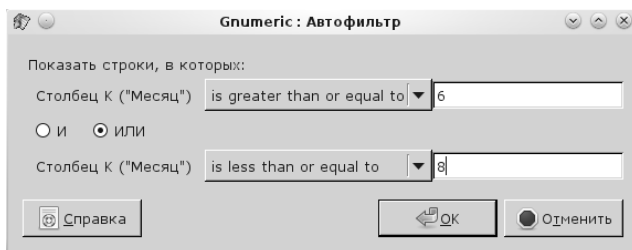


Рис. 3.17. Условия выбора летних месяцев

А	В	С	Д	Е	Ф	Г	Н	И	Ж	К
Задание: сотрудники, начавшие трудовую деятельность летом										
Автофильтр по расчетному полю										
ФИО	Дата ро	Нач.ста	П	С	Дет	Се	Обр	Должнос	Окл	Ме
Анисимов Г.С.	11/07/66	14/08/87	м	ж	2	МО	в	товаровед	1250	8
Борисов А.Г.	23/11/37	25/07/61	м	ж	1	ВР	ср	грузчик	600	7
Григорьева Ж.Д.	25/01/71	04/07/88	ж	э	2	ТОП	ср/сп	касс.контр	1200	7
Дорофеева Л.К.	20/07/75	02/06/92	ж	н	0	Б	ср	продавец-3	900	6
Киреев М.П.	14/07/60	06/06/77	м	р	1	МР	ср/сп	продавец-2	1080	6
Китуничев Д.М.	02/02/67	30/07/91	ж	э	1	МР	ср	продавец-3	980	7
Константинова Н.	20/03/58	04/08/75	ж	э	4	МР	в	зан.зав.се	1300	8
Лаврентьева Ю.В.	02/01/75	25/07/91	ж	н	0	ОФ	ср	продавец-3	950	7
Лазаренко П.Г.	02/07/48	02/06/66	ж	н	0	Б	н/ср	продавец-3	900	6
Минина М.Р.	22/08/73	29/07/90	ж	э	1	ТОП	н/ср	фасовщица	800	7
Озерова П.П.	03/09/54	03/07/71	ж	э	0	МР	ср/сп	продавец-2	1050	7
Пуговкина Н.В.	31/05/63	14/07/79	ж	э	1	ОФ	ср	продавец-3	950	7
Пятова М.В.	17/03/56	09/06/75	ж	э	2	ВР	н/ср	приемщикпо	800	6
Трушко Р.П.	12/07/74	12/08/92	м	э	0	Х	ср	продавец-3	950	8
Янковская А.Л.	02/04/74	30/06/91	ж	н	0	ТОП	ср/сп	кассир	1150	6

Рис. 3.18. Результат поиска по расчётному полю

Для ситуаций, когда по одному полю необходимо указать более двух условий, или условия являются противоречивыми, используется расширенный фильтр.

3.3 Расширенный фильтр

Расширенный фильтр позволяет реализовать подобие запросов QBE (query by example — запрос по образцу), используемых в настоящих базах данных. Для использования расширенного фильтра необходимо для каждого запроса формировать блок критериев. Блок критериев должен состоять минимум из двух ячеек — имени поля и условия поиска по этому полю. Условие должно быть либо числом, либо текстом (аналогично условиям в функциях `sumif()` и `countif()`),

J	K
Оклад	Год
	1960
	1983
	1990

Рис. 3.19. Создание блока критериев

которые были рассмотрены в предыдущей главе). Блок критериев целесообразно располагать над списком с данными. Обязательно наличие пустой строки перед блоком критериев и после него. Таким образом, в нашем примере перед диапазоном исходных данных нужно вставить несколько строк.

Пусть из списка сотрудников требуется выбрать лиц, начавших трудовую деятельность в 1960, 1983 и в 1990 годах. Здесь потребуется создать расчетное поле «Год», аналогично тому, что делалось при рассмотрении автофильтра. По этому полю нужно удовлетворить одновременно трем условиям, поэтому автофильтр не годится.

Для использования расширенного фильтра в первую очередь необходимо сформировать блок критериев. Он формируется путем копирования строки с именами полей в пустую строку над таблицей данных, а затем под именем поля «Год» в блоке критериев записываются одно под другим три условия (искомые значения годов), как показано на рис. 3.19. Условия, записанные одно под другим, обеспечивают выполнение логической операции «ИЛИ».

После этого выделяется диапазон исходных данных, включая строку с именами полей и вызывается диалог настройки расширенного фильтра «Данные/Фильтр/Расширенный фильтр...». Этот диалог имеет две вкладки. Первая вкладка (Ввод) позволяет определить диапазоны исходных данных и блока критериев (поэтому блок критериев уже должен существовать) (рис. 3.20). В нашем случае данные занимают диапазон $\$A\$9:\$K\83 , а критерии — диапазон K4:K7 (в блок критериев входит имя поля «Год» и три значения под ним). Вторая вкладка («Вывод») позволяет определить, куда будут записываться результаты работы фильтра (рис. 3.22).

При задании диапазона списка критериев нет возможности выделить эти диапазоны в ЭТ, если диалог настройки расширенного фильтра полностью открыт, как показано на рис. 3.20. Для указания диапазонов путём выделения блоков ЭТ мышью нужно свернуть

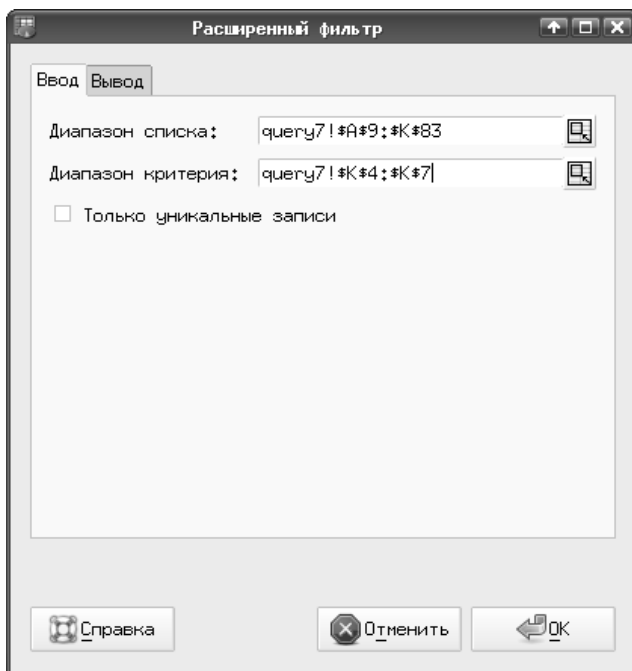


Рис. 3.20. Определение диапазона исходных данных и блока критериев для расширенного фильтра

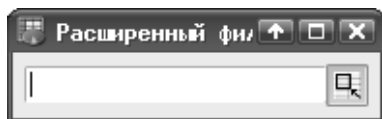


Рис. 3.21. Окно для указания диапазона ячеек

диалог, нажав на «кнопку» справа от соответствующего поля ввода. Диалог примет вид окна указания диапазона (рис. 3.21), после чего уже можно выделять нужный блок ячеек. По окончании выделения нажатием на ту же «кнопку» следует вернуть диалоговое окно в первоначальный вид.

В данном случае выбран вариант «Фильтровать на месте», однако результаты работы фильтра могут быть скопированы в другой диапазон на том же листе, на другой лист или даже в другой документ. Блок критериев также может быть размещен на другом листе, но это уже дело вкуса и привычки.

Теперь рассмотрим использование расширенного фильтра в случае противоречивых условий. Предположим, что нужно выбрать жен-

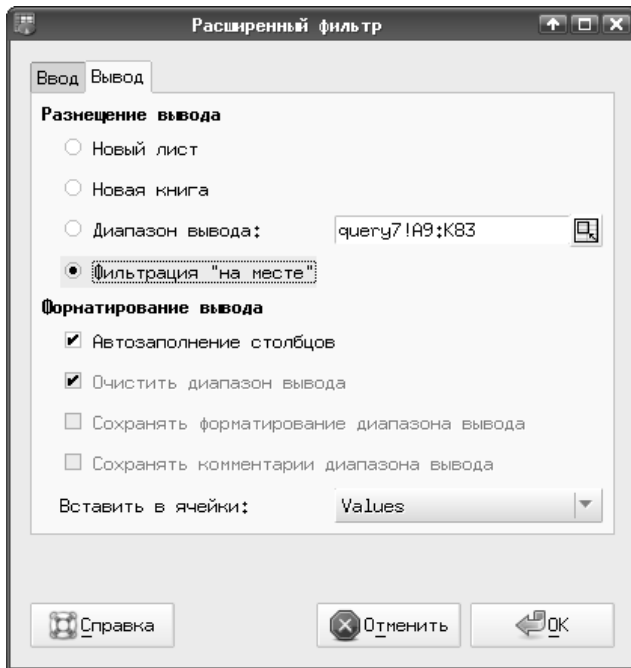


Рис. 3.22. Определение расположения результатов работы расширенного фильтра

ФИО	Дата рожд	Нач. стажа	Пол	С/п	Детей	Секц	Образ	Должность	Оклад	Год
Гранин Ю.В.	17/12/65	29/04/83	м	ж	1	Б	в	продавец-2	1030,00	1983
Дрозд И.И.	22/09/35	06/09/60	ж	з	1	ОФ	ср	продавец-1	1200,00	1960
Милашевич Н.М.	11/04/64	04/05/83	ж	з	0	ВР	н/ср	цборщица	750,00	1983
Минина М.Р.	22/08/73	29/07/90	ж	з	1	ТОП	н/ср	фасовщица	800,00	1990
Никодимова О.Е.	02/06/72	22/01/90	ж	з	2	Б	ср	продавец-3	900,00	1990
Николаева С.И.	16/02/67	03/03/83	ж	р	2	МР	ср	продавец-1	1200,00	1983
Петрушевич М.Т.	03/09/70	10/05/90	ж	н	3	Б	ср	продавец-3	900,00	1990
Федоренко Г.Ш.	14/05/67	15/03/90	ж	з	1	ТОП	ср	кассир	1150,00	1990

Рис. 3.23. Результат работы расширенного фильтра

D	E	F	G	H
Пол	С/п	Детей	Секц	Образ
#		>0		в
м		>0		ср

Рис. 3.24. Сложные условия для расширенного фильтра

J	K
Оклад	
<1112	<
	1111,6667

Рис. 3.25. Условия с вычисляемым критерием поиска

щин с высшим образованием, имеющих детей, и мужчин со средним образованием, также имеющих детей. Критерии для фильтра показаны на рис. 3.24.

Из приведенного рисунка видно, что в условиях расширенного фильтра можно использовать как точное соответствие, так и операции сравнения для числовых полей. Расположение условий в одной строке означает одновременное выполнение условий (отношение «И»), а расположение условий друг под другом означает требование выполнения хотя бы одного из условий (отношение «ИЛИ»).

В условиях расширенного фильтра можно также использовать результаты работы формул. Например, нужно найти сотрудников с окладом ниже среднего по предприятию. Тогда сначала подсчитываем средний оклад с помощью функции `average()` по столбцу с окладами (например, в ячейке K6), рядом в какой-то ячейке (например, в K5) записываем знак сравнения (в ячейке, а не в условии, потому что критерий поиска может измениться), а в условии пишем формулу `=concatenate(K5;round(K6;0))`. Функция `round()` используется для округления среднего значения до целого.

Блок критериев, полученный с использованием формулы, показан на рис. 3.25.

Формулы, использованные при формировании условия, обеспечивают динамическое изменение условия при изменении исходных данных.

3.4 Функции базы данных

Отдельная группа функций электронной таблицы (категория «База данных») позволяет проводить вычисления на основе данных из списка с условиями, определяемыми блоками критериев (как в расширенном фильтре). При использовании этих функций для диапазона данных, занимаемого списком, автоматически производится отбор значений в указанном столбце по указанным критериям, и производятся соответствующие вычисления.

Некоторые часто используемые функции этой категории приведены в таблице ниже.

Таблица 3.1: Некоторые функции категории «База данных»

Название, аргументы	Описание
daverage(диапазон;поле; критерии)	Вычисляет среднее значение по указанному полю для данных, соответствующих критерию (пример: средний оклад у мужчин).
dcount(диапазон;поле; критерии)	Вычисляет количество значений по указанному числовому полю (пример: количество незамужних женщин, имеющих детей).
dcounta(диапазон;поле; критерии)	Вычисляет количество значений по указанному полю (пример: количество мужчин).
dmin(диапазон;поле; критерии)	Вычисляет минимальное значение по указанному числовому полю (пример: минимальный оклад у мужчин).
dmax(диапазон;поле; критерии)	Вычисляет максимальное значение по указанному числовому полю (пример: максимальное имеющееся у сотрудника количество детей).
dsum(диапазон;поле; критерии)	Вычисляет сумму значений по указанному числовому полю (пример: общее количество детей у холостых мужчин).

L	M	
Блоки критериев:		
Пол	Образ	
ж	н/ср	
		9
Пол	Образ	
ж	ср/сп	
		10
Пол	Образ	
ж	ср	
		27
Пол	Образ	
ж	в	
		14

Рис. 3.26. Критерии и результаты вычислений для функции `dcounta()`

Остальные функции баз данных в Gnumeric, при необходимости, легко изучить самостоятельно.

Рассмотрим пример использования функций базы данных для вычисления количества женщин с различными уровнями образования. Внимательно посмотрев описание функций категории «База данных», можно понять, что для такой задачи потребуется использовать функцию `dcounta()`, которая подсчитывает количество значений (непустых ячеек) в указанном столбце при указанных условиях. Условия и результаты работы функции показаны на рис. 3.26.

Блок критериев должен состоять как минимум из двух ячеек — имени поля и условия поиска по этому полю. Поскольку все поля текстовые, то условием является полное соответствие текста. В данном случае (для первого результата в ячейке M6) имеем формулу `=dcounta(A7:J85;H7;L4:M5)`, где:

- `A7:J85` — диапазон ячеек, занимаемый списком (базой данных) в абсолютных адресах;
- `H7` — абсолютный адрес ячейки с именем поля, по которому производится подсчет (в данном примере — поле «Образ»);
- `L4:M5` — диапазон ячеек блока критериев.

Результаты подсчета находятся соответственно в ячейках M6, M9, M12 и M15.

Работа со списками не является сильной стороной Gnumeric, однако выполнение часто требуемых операций всё-таки обеспечивается.

В следующей главе мы рассмотрим возможности Gnumeric по построению диаграмм, которые являются действительно серьёзными.

Глава 4

Диаграммы в Gnumeric

4.1 Принципы создания и форматирования диаграмм

Изучение диаграмм в Gnumeric и особенностей работы с ними начнём с построения круговой диаграммы по данным, полученным в предыдущей главе при вычислении количества женщин с различными уровнями образования.

Результаты вычислений оформим в виде таблицы в ячейках M17:N21, например, как показано на рис. 4.1.

Для построения диаграммы прежде всего нужно выделить диапазон данных (включая заголовки). Затем вызывается диалог вставки диаграммы (либо кнопкой в панели инструментов, либо командой главного меню «Вставка/Диаграмма...», рис. 4.2). Этот диалог позволяет сначала определить тип диаграммы, а затем установить все нужные параметры для выбранного типа. Для нашей диаграммы выберем тип «Круговая диаграмма». На рис. 4.2 показано, что для этого типа диаграммы существует четыре варианта.

Образование	Женщин
н/ср	9
ср/сп	10
ср	27
в	14

Рис. 4.1. Исходные данные для диаграммы

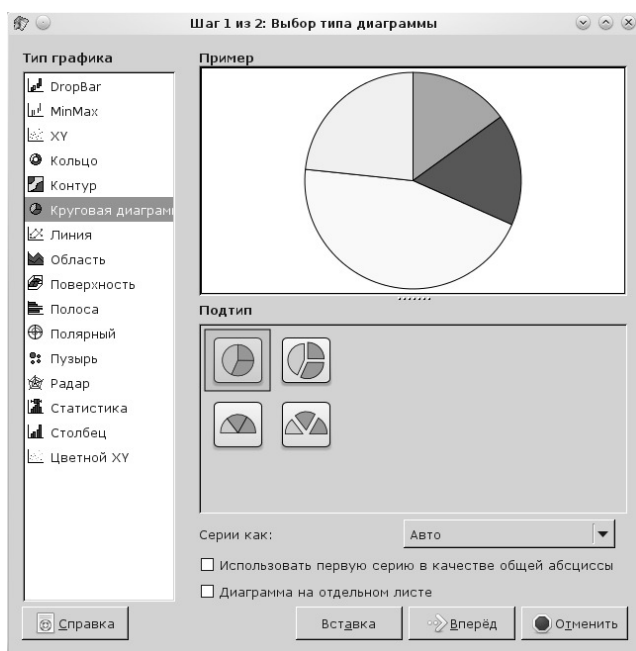


Рис. 4.2. Выбор типа диаграммы

После выбора нужного типа диаграммы нажатием на кнопку «Вперёд» переходим к этапу настройки диаграммы (рис. 4.3).

Область построения содержит различные объекты. Основой (объектом верхнего уровня) является график, это собственно и есть область построения. На графике может располагаться одна или несколько диаграмм. В нашем примере имеется одна диаграмма типа PlotPie, отображающая один набор данных (названный по заголовку блока ячеек, содержащего эти данные).

Кнопка «Добавить» под списком объектов позволяет добавлять на график объекты, соответствующие свойствам текущего выделенного объекта. Кнопка «←» служит для удаления лишних объектов, а кнопки-стрелки позволяют перемещать выбранный объект по иерархии объектов графика (если это в принципе возможно).

Для области графика настраиваются стиль и тема оформления. Сначала изменим тему. На вкладке «Тема» из списка тем выберем вариант Gurr1 и вернёмся на вкладку настройки стиля.

При смене темы изменяется цветовая палитра диаграммы. На вкладке «Стиль» изменим параметры оформления графика «Контур» и «Заливка». Для контура оформления установим стиль «сплошная линия», чёрный цвет и размер в две точки. Для заливки выберем ва-

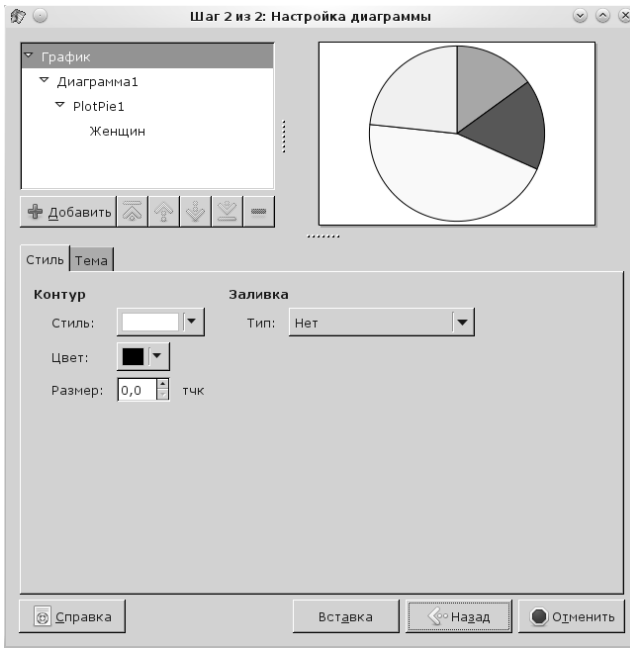


Рис. 4.3. Настройка объектов диаграммы

риант «Двухцветный градиент», начальный цвет ярко-синий, конечный цвет — чёрный, направление — снизу вверх (синий внизу, чёрный наверху). Результаты изменения темы оформления и свойств графика сразу видны в области просмотра в верхней правой части диалога (рис. 4.4).

Следующим действием будет создание заголовка графика. Для этого нажмём кнопку «Добавить», когда выделен объект Graph (График) и из списка вариантов выберем «Заголовок to График».

В списке объектов появится новый объект «Заголовок1». Настроим заголовок в соответствии с рис. 4.5. Текст (например, «Уровни образования у женщин») пишется в поле ввода на вкладке «Данные». На вкладке «Шрифт» выбирается шрифт надписи и его атрибуты. Цвет шрифта разумно выбирать контрастным относительно фона области графика в той области, где располагается заголовок. На вкладках «Текст» и «Позиция» устанавливаются ориентация текста (горизонтально, вертикально или под произвольным углом) и расположение заголовка на графике. Пока оставим все эти настройки без изменений.

На рис. 4.6 показаны настройки ряда данных для диаграммы. На вкладке «Данные» можно определить или переопределить ячейки, содержащие данные (для указания ячеек и блоков ячеек путём выде-

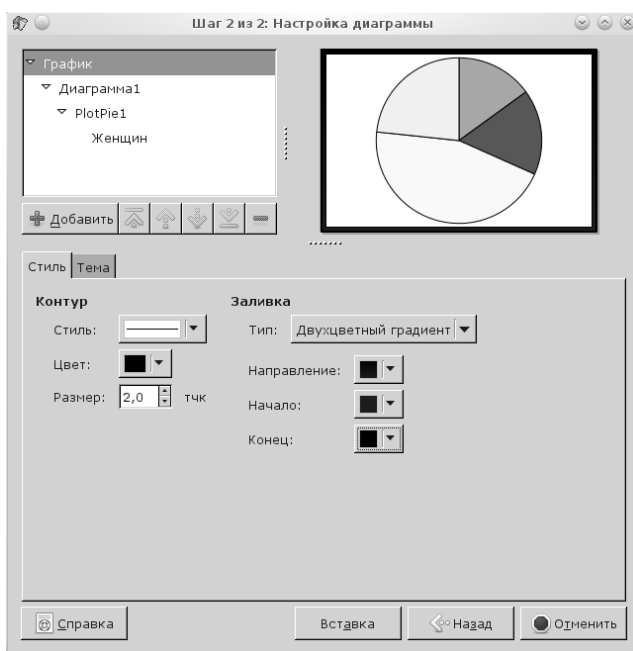


Рис. 4.4. Настройка оформления области графика

ления мышью используются кнопки «Свернуть/развернуть диалог» справа от полей ввода).

Название ряда данных берётся из ячейки, указанной в поле «(Название)». В это поле можно просто вписать нужный текст. Режим "Show in Legend" определяет, будет ли выбранный ряд данных отображаться в пояснении к диаграмме («легенде»), когда она будет помещена на диаграмму.

Для добавления легенды нужно выделить объект «Диаграмма1», нажать на кнопку «Добавить» и из списка вариантов выбрать «Подпись to Диаграмма1». Затем определяются стиль, шрифт для текста легенды и расположение легенды. Расположим легенду справа сверху области диаграммы, как показано на рис. 4.7.

Для придания выразительности круговой диаграмме или выделения какого-нибудь элемента данных часто используется вариант «вырезанного» сектора, когда один из секторов круговой диаграммы отделяется от остальных небольшим промежутком. На рис. 4.8 показан диалог настройки такого вырезанного сектора (объект «Точка» на диаграмме).

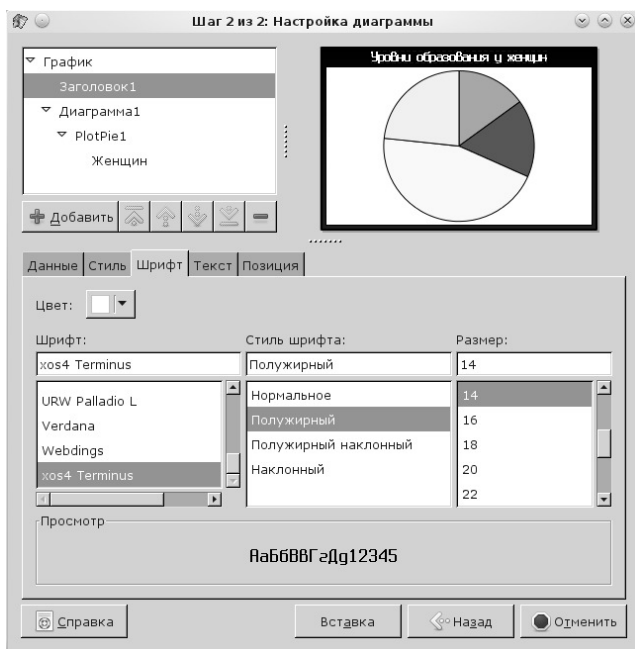


Рис. 4.5. Настройка заголовка графика

Сектора нумеруются по часовой стрелке от вертикальной линии, поэтому выделенный сектор имеет номер 0. Для примера, показанного на рис. 4.8 также изменено название ряда данных.

После завершения всех настроек и нажатия на кнопку «Вставка» в диалоге настроек графика нужно выделить на листе ЭТ область графика, протаскивая мышью с нажатой левой кнопкой. При этом указывается размер области как в точках экрана, так и в типографских пунктах (производится пересчёт точек в пункты на основании значения разрешения экрана).

После того, как будет выделена область желаемого размера, кнопку мыши нужно отпустить и график появится на листе электронной таблицы, при этом он останется в режиме редактирования (рис. 4.9).

Маркеры по углам и в серединах сторон области графика позволяют менять размеры при перемещении маркеров мышью.

Для перехода в диалог настроек графика нужно дважды щёлкнуть по нему мышью.

Однако у круговых диаграмм в Gnumeric есть и существенный недостаток — отсутствие возможности выводить значения данных в числовом виде (в абсолютных единицах или в процентах).

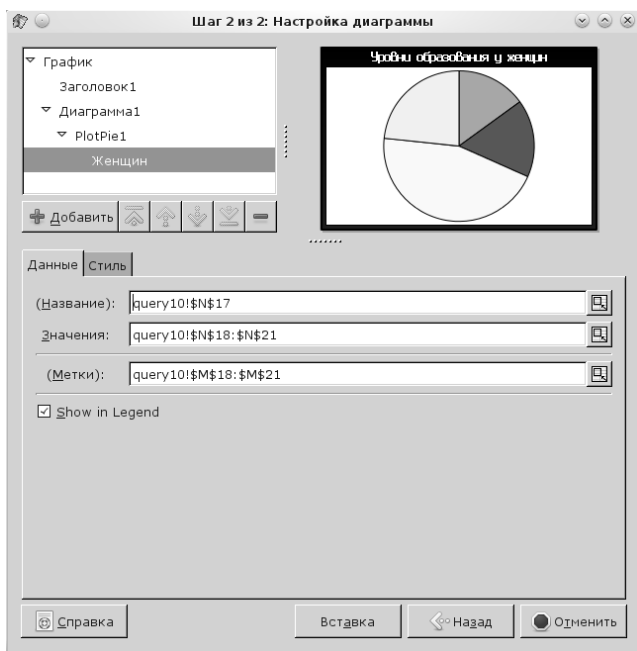


Рис. 4.6. Настройка исходных данных для диаграммы

4.1.1 Экспорт диаграмм в графические файлы

Gnumeric сильно облегчает подготовку иллюстраций для различного рода изданий благодаря возможности прямого экспорта диаграмм в графические файлы различных форматов.

Для использования этой возможности достаточно открыть контекстное меню диаграммы (рис. 4.10) щелчком правой кнопки мыши на диаграмме.

Нужно заметить, что выбор пункта «Свойства» в этом контекстном меню вызовет диалог настройки графика.

Выбор пункта «Сохранить как изображение» откроет диалог экспорта диаграммы (рис. 4.11).

Как видно, это такой же GTK-диалог управления файлами, как и диалог открытия/сохранения файлов, разобранный в главе «Управление файлами».

Для некоторых графических форматов в дополнительных настройках ("Export settings") появляется возможность изменить разрешение для изображения.

Список форматов, в которых возможен экспорт диаграммы как изображения, показан на рис. 4.12.

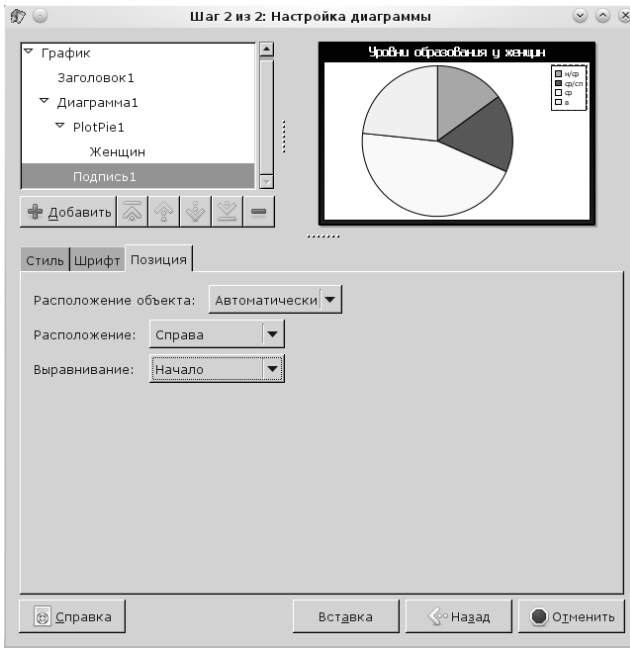


Рис. 4.7. Настройка подписи к диаграмме (легенды)

При экспорте в форматы PNG, JPEG и PS можно изменять разрешение для экспортируемого изображения. Таким образом можно выполнить часто встречающееся требование издательств предоставлять иллюстрации с разрешением 300 точек на дюйм (300 dpi).

4.2 Несколько диаграмм на графике (на примере круговой диаграммы)

Пусть нужно проследить за изменением статей расходов семьи по сезонам. Для каждого сезона будем рисовать круговую диаграмму по статьям расходов, а для сравнения диаграмм их целесообразно иметь рядом на одном графике. Такие диаграммы в пакетах научной графики называются Multiply Charts («множественные диаграммы»). Таблица исходных данных показана на рис. 4.13.

Как обычно при создании диаграммы, выделяем весь диапазон данных с подписями, вызываем диалог создания диаграммы и выбираем вариант круговой диаграммы.

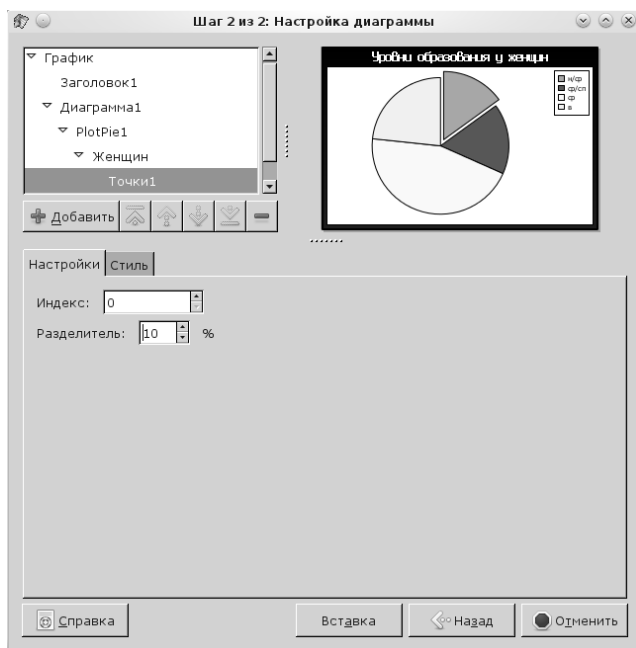


Рис. 4.8. Настройка «вырезанного» сектора



Рис. 4.9. Готовый график на листе ЭТ в режиме редактирования

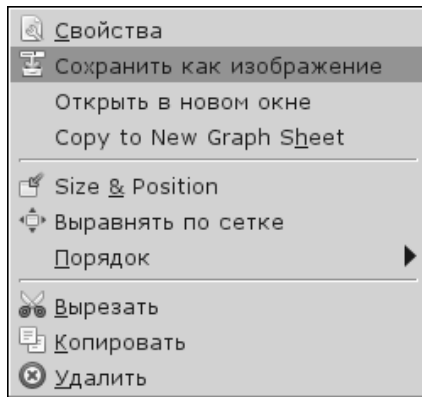


Рис. 4.10. Контекстное меню диаграммы

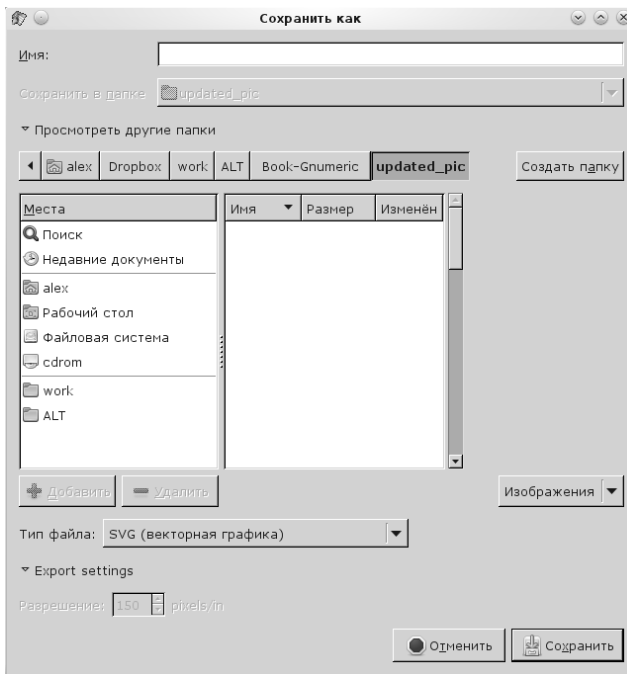


Рис. 4.11. Диалог экспорта диаграммы в Gnumeric

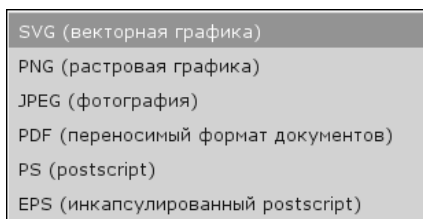


Рис. 4.12. Форматы для экспорта диаграммы

	А	В	С	Д
1	Расходы по сезонам			
2		Пища	Одежда	Квартира
3	Зима	15000	23000	1900
4	Весна	17000	21000	2000
5	Лето	14000	15000	2500
6	Осень	18000	18000	3000
7				

Рис. 4.13. Исходные данные для множественной диаграммы

Для графика в целом установим тему *Gurri* без всякого обрамления с белым фоном, определим заголовок «Структура расходов по сезонам», настроим исходные данные для ряда данных (рис. 4.14).

азвание ряда данных берётся из названия строки, данные — из соответствующей строки таблицы, а заголовки данных — из названий столбцов таблицы исходных данных.

Поскольку планируется на графике разместить четыре диаграммы, рассмотрим настройки размещения для первой диаграммы, а остальные можно будет разместить по аналогии.

Размещение диаграммы на графике настраивается на вкладке «Позиция» в диалоге настроек объекта «Диаграмма1» (рис. 4.15).

Некоторое пространство сверху требуется для отображения заголовка графика (в рассматриваемом примере — 10%), поэтому для двух диаграмм остаётся пространство в 90% высоты области графика.

Добавим к диаграмме название и легенду, а на график — оставшиеся варианты диаграмм (для объекта «График» — «Добавить/Диаграмма to График»).

Результат размещения нескольких диаграмм на графике (множественная диаграмма) показан на рис. 4.16.

В дальнейшем при рассмотрении примеров диаграмм возможность размещения нескольких диаграмм на одном графике будет использоваться там, где это целесообразно.

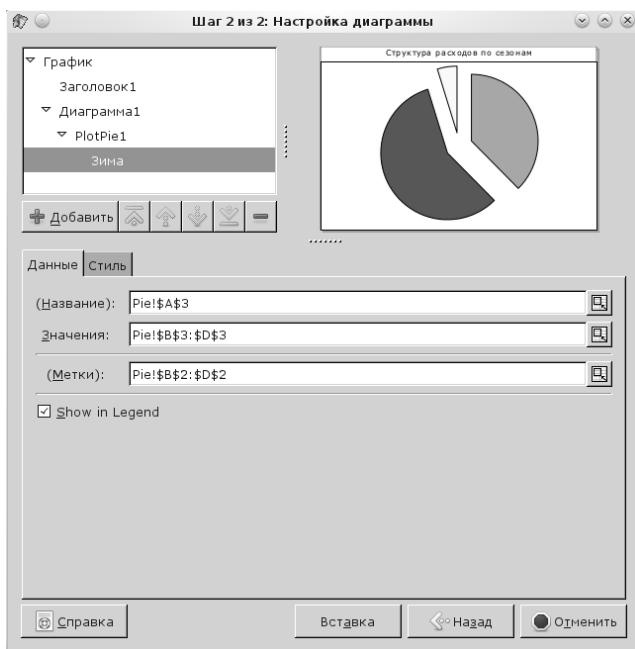


Рис. 4.14. Исходные данные ряда данных

4.3 Диаграммы для отображения коммерческих данных

В эту группу вариантов диаграмм выделены виды, которые не отображают зависимости в строгом смысле этого слова, а отображают изменение каких-то показателей для пронумерованных объектов. Для этих видов диаграмм по горизонтальной оси откладывается не значение переменной, а номер некоторого объекта (периода, предприятия, человека). Важно, что в таких диаграммах горизонтальная ось (абсцисса, ось X) воспринимается как равномерная шкала последовательных значений.

Попытки использовать диаграммы этой группы для отображения зависимостей при произвольно распределённых значениях абсциссы приводят к досадным ошибкам.

Один из вариантов диаграмм такого типа — круговая диаграмма уже был рассмотрен во всех подробностях.

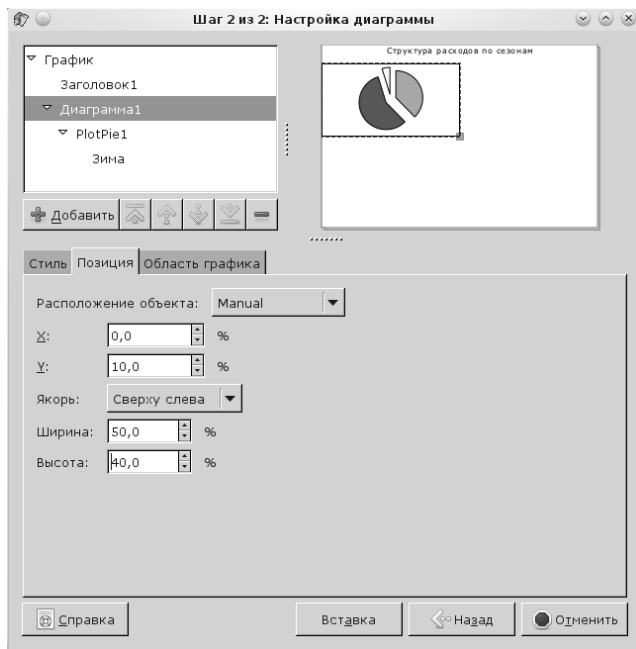


Рис. 4.15. Размещение диаграммы на графике

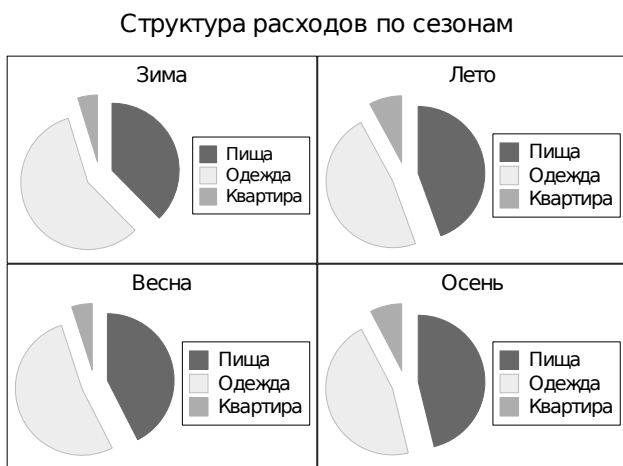


Рис. 4.16. Несколько диаграмм на графике

	А	В	С	Д
1		Группа1	Группа2	Отдел
2	январь	7	1	12
3	февраль	8	2	13
4	март	9	3	14
5	апрель	6	5	15
6	май	5	7	15
7	июнь	4	9	17
8	июль	5	11	15

Рис. 4.17. Исходные коммерческие данные

4.3.1 Линейчатая диаграмма

Для использования этого типа диаграммы необходимо минимум два вектора данных одинаковой длины. Диаграмма имеет три подтипа — обычная «Линии», линий с накоплением значений («Stacked Line») и «Процентные линии». Кроме того, для каждого подтипа возможен вариант без маркеров по умолчанию и с маркерами. Рассмотрим использование диаграмм всех трех подтипов (варианты с маркерами) на примере данных по условным баллам оценки работы отдела продаж и двух групп менеджеров этого отдела.

Исходные данные показаны на рис. 4.17, а варианты диаграмм на одном графике — на рис. 4.18.

Если в диапазон ячеек, выделенный перед построением диаграммы, включить текстовые ячейки (на рис. 4.17 выделены цветом), то они будут использованы для меток по горизонтальной оси и для имен рядов данных.

В случае диаграммы с накоплением первый ряд данных (в данном случае — самый левый) используется как «подставка», и к значениям данных первого ряда прибавляются значения второго, а к их сумме — значения третьего и т.д.

В случае «процентной» диаграммы последний ряд (в данном случае — самый правый) принимается за 100%, а первый образует «подставку», но его значения вычисляются как доли от значений последнего ряда. Значения второго и последующих рядов данных формируются по тому же принципу, что и для «стопки», но опять-таки в долях от значений последнего ряда.

4.3.2 Диаграмма областей

Диаграммы такого типа по сути очень похожи на линейчатые, только пространство под линией заливается цветом или штриховкой. Имеется те же самые три подтипа — обычная диаграмма областей, об-

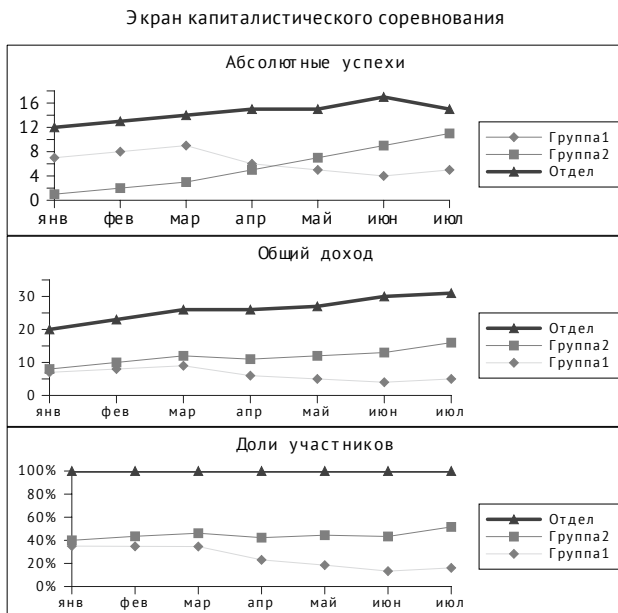


Рис. 4.18. Варианты линейчатой диаграммы

ласти с накоплением значений ("Stacked Area") и «процентная» диаграмма областей. Поэтому воспользуемся данными из предыдущего примера и построим те же диаграммы в виде диаграмм областей (рис. 4.19).

В первом (верхнем) варианте диаграммы область для серии «Отдел» сделана прозрачной, иначе она бы полностью перекрыла остальные области.

4.3.3 Полосковая диаграмма

Для тех же исходных данных все три подтипа диаграммы этого типа приведены на рис. 4.20.

4.3.4 Столбиковая диаграмма

Этот тип диаграммы, называемый также "Column Chart", во всем похож на предыдущий тип, только значения изображаются не горизонтальными полосками, а вертикальными столбиками. Подтипы здесь те же самые, поэтому рисунок достаточно просто вообразить.

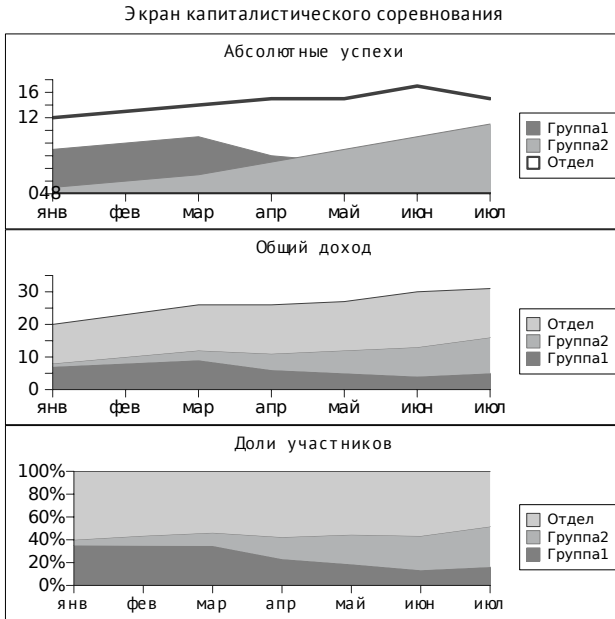


Рис. 4.19. Диаграммы областей

4.3.5 Кольцевая диаграмма

Кольцевая диаграмма предназначена для отображения соотношения величин при дискретном изменении значения параметра. Каждому значению параметра соответствует «сектор», а значения переменных отображаются по кольцам. Для этой диаграммы используем данные из примера множественной круговой диаграммы (рис. 4.13).

Соответствующая кольцевая диаграмма показана на рис. 4.21. Как видно, она достаточно мало информативна и может служить скорее для украшения презентации, чем для реального анализа данных.

4.3.6 Диаграмма радар

Этот вид диаграммы предназначен для отображения серий данных по категориям, так же как и линейчатая и столбиковая диаграммы. Однако серии данных должны располагаться исключительно по столбцу. Рассмотрим пример такой диаграммы (рис. 4.22) на примере тех же данных, что и для кольцевой диаграммы.

Этот тип диаграммы в Gnumeric имеет три подтипа. Первый подтип — линии без точек, второй (показан на рис. 4.22) — линии с точ-

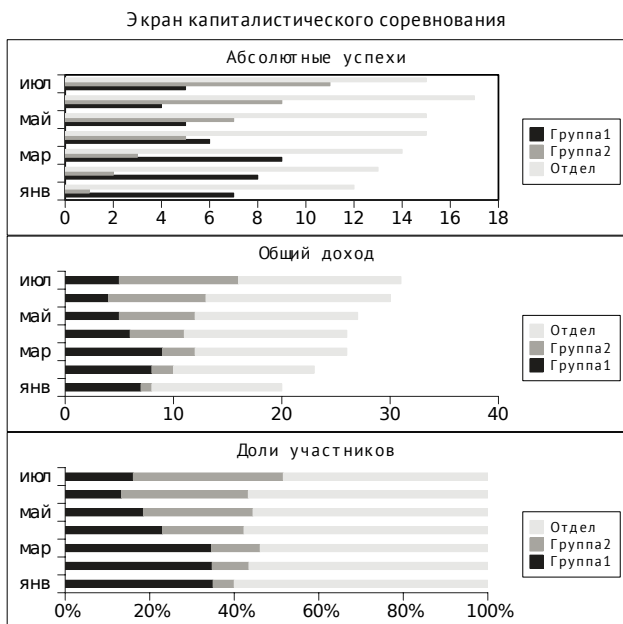


Рис. 4.20. Полосковые диаграммы

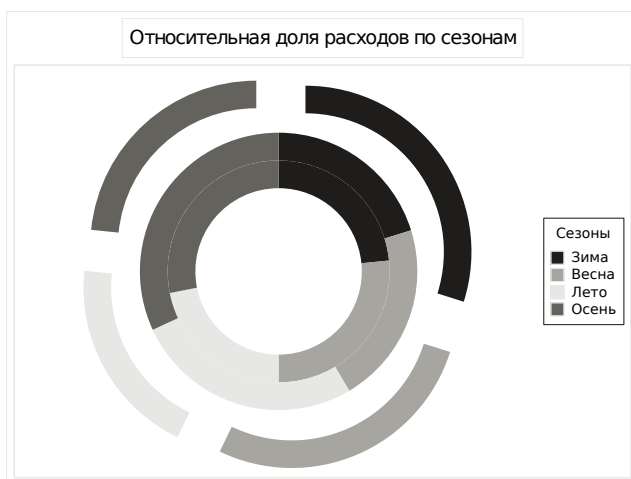


Рис. 4.21. Кольцевая диаграмма

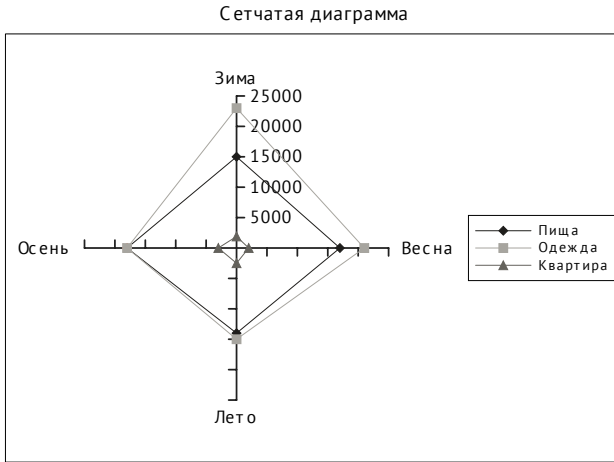


Рис. 4.22. Диаграмма радар

ками данных ("Dotted radar"), а третий — линии, ограничивающие области ("Area radar").

4.3.7 Диаграмма отклонений (DropBar)

Серия данных для этой диаграммы образуется двумя векторами одинаковой длины, первый из которых (по умолчанию левый или верхний) задает опорные точки, а значения второго вектора показывают отклонения от опорных точек. В качестве значений по горизонтальной оси (абсцисс) используются номера точек, начиная с 1. На рис. 4.23 показаны отклонения функции $\cos()$ от функции $\sin()$. Цвет для положительных отклонений используется в легенде. Отрицательные отклонения показаны другим цветом. Можно для опорных значений и для отклонений добавить огибающие линии с маркерами или без. На вкладке «Данные» при настройке серии можно переопределить опорный вектор (Начало) и вектор отклонений (Конец).

4.3.8 Диаграмма пределов (MinMax)

Этот вид диаграммы похож на диаграмму отклонений, только отклонения показываются не разноцветными «палочками», а отрезками. Можно выбрать вертикальную или горизонтальную ориентацию отрезков, а также наличие или отсутствие маркеров на концах отрезков. Вид маркеров, их размер, цвет, толщину и цвет линии можно настраивать при настройке серии. Также можно переопределять век-

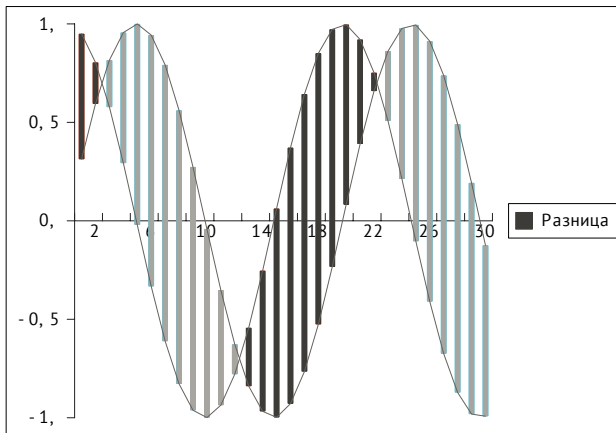
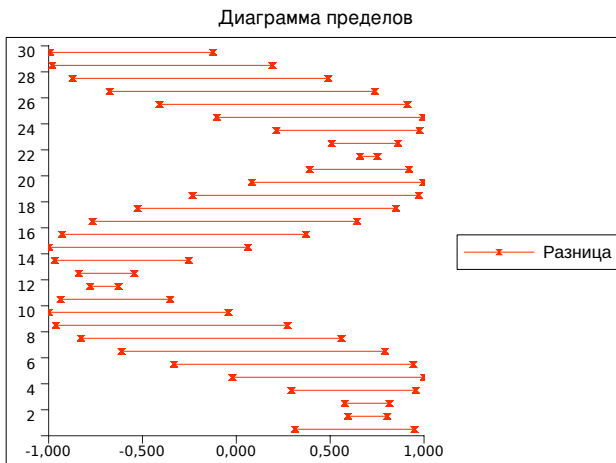
Рис. 4.23. Отклонения значений $\cos()$ от значений $\sin()$ 

Рис. 4.24. Отклонения функций в виде диаграммы пределов

тора для нижних и верхних значений. На рис. 4.24 показан пример диаграммы пределов в «горизонтальной» ориентации.

4.4 Отображение научных данных

К этой группе можно отнести виды диаграмм, предназначенных для отображения зависимостей вида $y = f(x)$ или $z = f(x, y)$ при

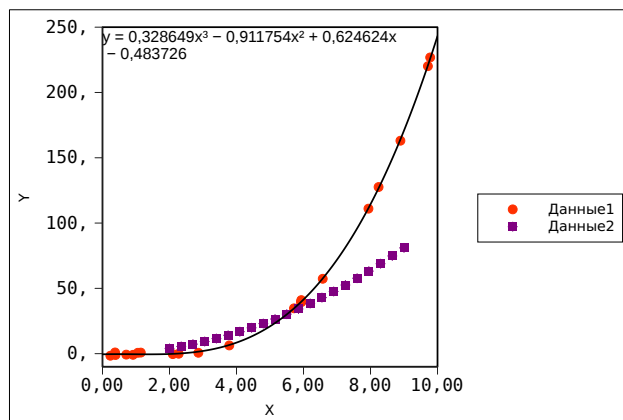


Рис. 4.25. Диаграмма XY с двумя независимыми наборами данных

произвольных значениях аргументов (данные могут быть не отсортированы и распределены неравномерно).

4.4.1 Диаграмма XY

Диаграмма XY (она же точечная, она же ScatterPlot) — один из самых часто встречающихся вариантов визуализации данных. В Gnumeric с помощью этого вида диаграммы можно отобразить несколько независимых наборов данных (разной длины и при разных значениях абсцисс) и для каждого набора при желании построить несколько уравнений регрессии. На рис. 4.25 показаны графики для двух независимых наборов, один из которых «искажен» аддитивной случайной добавкой, и для него построено уравнение регрессии (полинома третьего порядка). Более подробно построение линий и уравнений регрессии рассматривается в главе «Регрессионный анализ в Gnumeric».

Как видно на рис. 4.25, наборы данных имеют разную длину и точки имеют совершенно разные значения по оси X. В офисных электронных таблицах такую диаграмму построить довольно сложно, а в некоторых случаях — невозможно.

Следует обратить внимание на подтипы диаграммы XY (рис. 4.26).

Варианты со ступенчато изменяющимися значениями могут быть полезны при изображении диаграмм дискретных процессов. Пример такого изображения, который иллюстрирует процесс дискретизации электрического сигнала, показан на рис. 4.27.

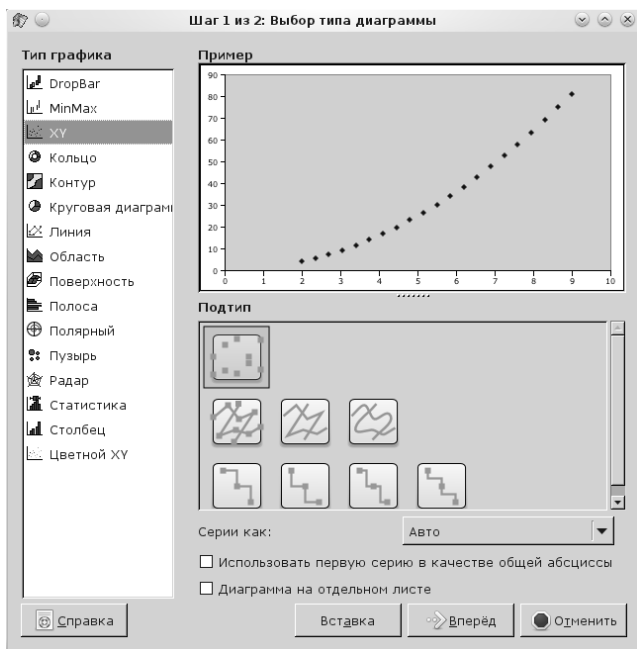


Рис. 4.26. Подтипы XY-диаграмм

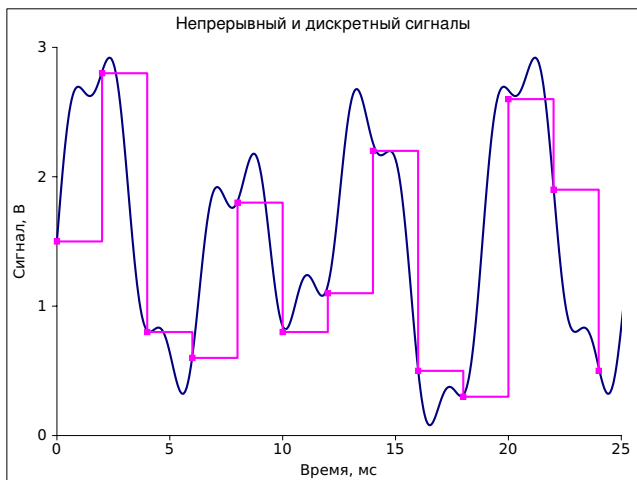


Рис. 4.27. Иллюстрация процедуры дискретизации с использованием диаграммы XY

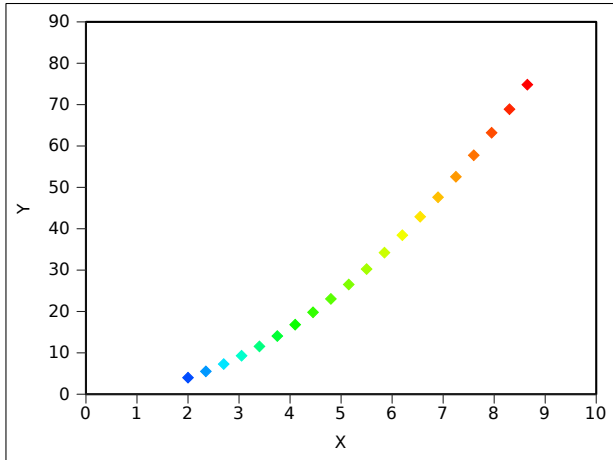


Рис. 4.28. Диаграмма Цветной XY

4.4.2 Цветной XY

Этот вариант является модификацией обычной точечной диаграммы, но для точек можно задать вектор цветов. Таким образом можно выделять цветом отдельные значения. При этом линии, соединяющие точки, будут плавно менять цвет от одной точки к другой. Желаящие могут самостоятельно определить «палитру» цветов точек для такой диаграммы.

Пример такой диаграммы показан на рис. 4.28.

4.4.3 Пузырьковая диаграмма

Этот тип диаграммы является вариантом точечной диаграммы (XY), но значения данных отображаются кругами («пузырями»), причем площадь круга может быть задана отдельным вектором данных. На рис. 4.29 приведен пример данных для построения такой диаграммы, а на рис. 4.30 — два варианта диаграммы. В первом (верхнем) варианте размеры «пузырей» заданы вектором R (см. рис. 4.29), а во втором (нижнем) размеры «пузырей» прямо определяются значениями Y (т.е. вектора для Y и размеров «пузырей» совпадают). Из сравнения верхнего и нижнего вариантов видно, что несмотря на наличие отдельного вектора размеров, значение Y также играет роль для отображения «пузыря».

	A	B	C
1	X	Y	R
2	1	1	9
3	3	9	8
4	2	4	7
5	5	25	6
6	7	49	5
7	4	16	4
8	9	81	3
9	6	36	2
10	8	64	1

Рис. 4.29. Данные для пузырьковой диаграммы

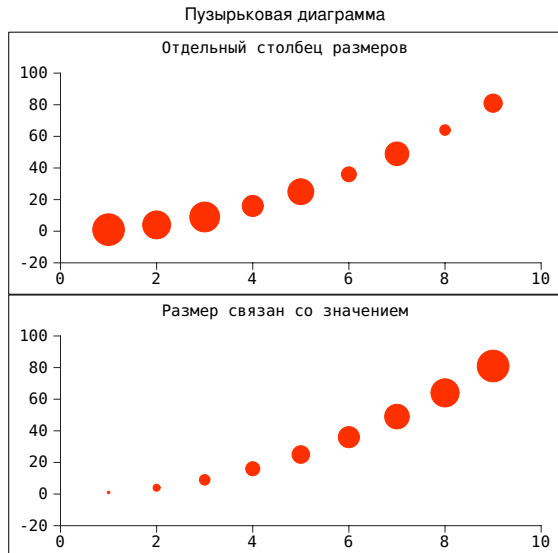


Рис. 4.30. Варианты пузырьковой диаграммы

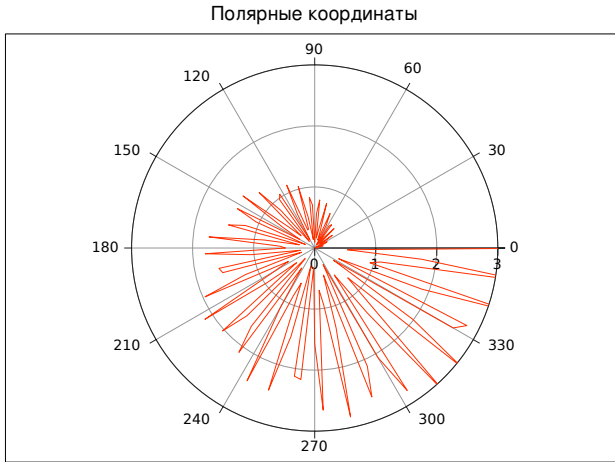


Рис. 4.31. График в полярных координатах

4.4.4 График в полярных координатах

Этот тип диаграммы удобен для отображения всякого рода угловых зависимостей в полярных координатах (например, распределения энергии для цилиндрических или сферических волн). На рис. 4.31 показан вид зависимости

$$A(\alpha) = (\sin^2 \alpha + \cos^2 \alpha) / \pi \quad (4.1)$$

4.4.5 Диаграмма высот (Контур)

Этот тип диаграммы является одним из вариантов отображения 3D-диаграмм, имеющих в Gnumeric. Для построения такой диаграммы требуется матрица (прямоугольная таблица) значений. Изображается «географическая» карта поверхности, в виде областей, ограниченных линиями равной высоты по Z . Интервал высот между линиями заливается цветом. К сожалению, используются контрастные цвета, которые невозможно настроить, поэтому картина получается яркой, но малопонятной. На рис. 4.32 показан пример диаграммы такого типа для функции $Z = \sin^2(x) - \cos^2(y)$ при $x \in [0; 3]$, $y \in [-1, 5; 1, 5]$.

4.4.6 Диаграмма поверхности

Для этого вида диаграммы в Gnumeric существует два варианта — вариант, использующий «матрицу» значений «Поверхность» и вариант, для которого требуются «тройки» значений «XYZ поверхность».

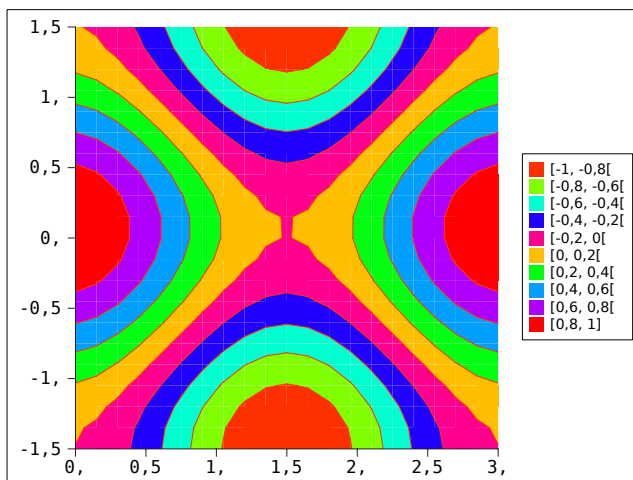


Рис. 4.32. Диаграмма высот

Для обоих вариантов в диалоге свойств графика помимо форматирования осей, отображения данных, области графика и диаграммы добавляется настройка вида 3D-диаграммы. Соответствующий диалог «Вращение» показан на рис. 4.33.

Область диаграммы можно вращать за выделенную точку на круге, куб также можно вращать или «качать» вперёд-назад, а ползунок позволяет изменять «угол зрения» на трёхмерный объект.

Диаграмма, построенная по матрице значений для функции $Z = \sin^2(x) - \cos^2(y)$ при $x \in [0; 3]$, $y \in [-1, 5; 1, 5]$. 4.34.

Для поверхности, которая строится по «тройкам» значений XYZ, данные нужно формировать следующим образом. Сначала фиксируется начальное значение X, при постоянном X изменяется Y и добавляются соответствующие значения Z. Потом устанавливается следующее значение X и т.д. Пример поверхности, построенной по точкам XYZ и использующий те же значения, что и предыдущий вариант, показан на рис. 4.35.

По сравнению с поверхностью, построенной по сетке, эта поверхность показывается в другом ракурсе (развёрнуты оси X и Y).

4.5 Статистические диаграммы

Все эти диаграммы сведены в одну группу «Статистика» и разделены по подтипам. Для изучения особенностей этих диаграмм удобно пользоваться возможностью генерации выборок случайных величин.

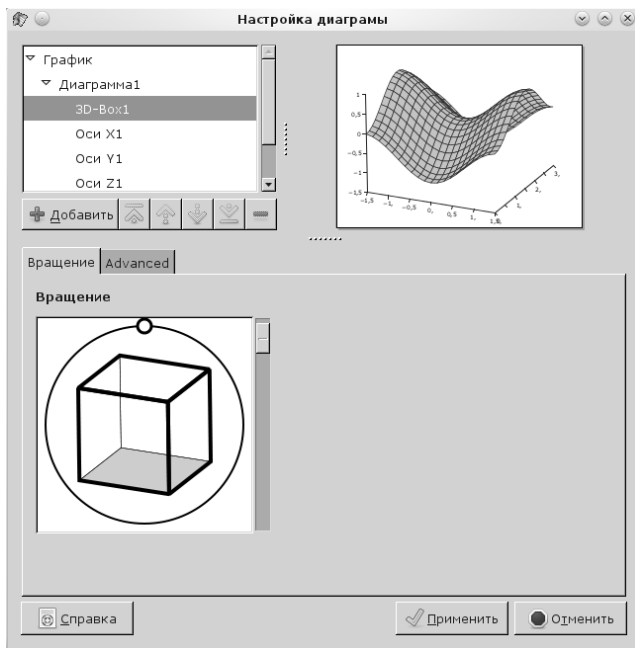


Рис. 4.33. Настройка трёхмерного вида

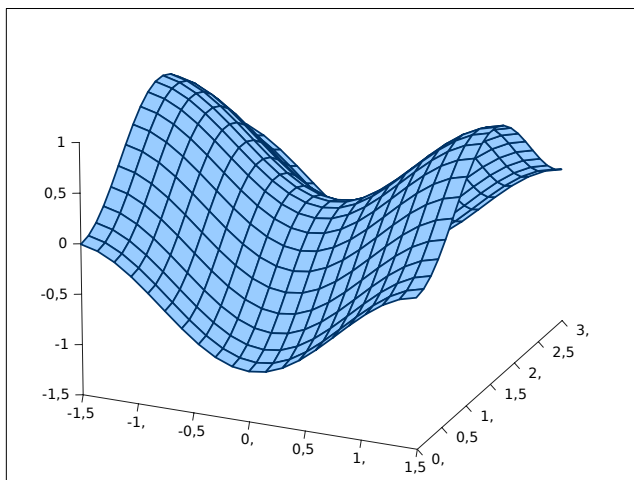


Рис. 4.34. Поверхность, построенная по сетке значений

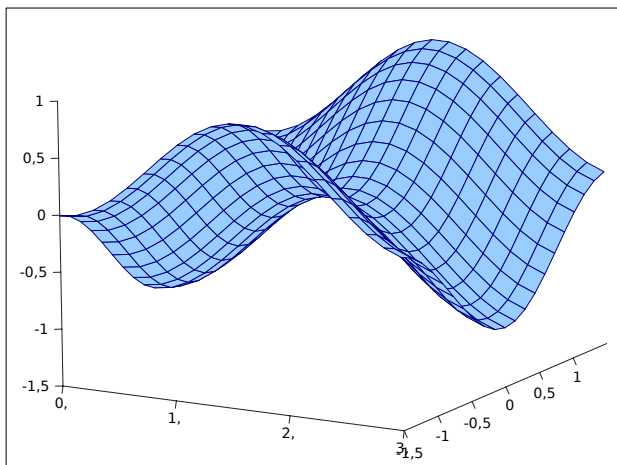


Рис. 4.35. Поверхность, построенная по точкам XYZ

Все рассмотренные ниже диаграммы построены на основе выборок нормально распределённых случайных чисел.

4.5.1 Ящичковая (коробчатая) диаграмма (Box Plot)

Вид диаграммы в варианте «показывать выбросы» (Horizontal Box Plot showing outliers) для двух независимых выборок приведён на рис. 4.36.

Линия поперёк «ящика» — медиана выборки, левая и правая границы «ящика» — соответственно, нижний и верхний квартили. «Усы» показывают минимальное и максимальное значение выборки при нормальной распределении. Выбросы — точки, которые выходят за пределы нормального распределения и «портят статистику».

Цвета линий и «ящичков», а также размеры точек выбросов настраиваются в диалогах для серий данных и для диаграммы в целом.

4.5.2 График распределения вероятности (Probability Plot)

Вариант такого графика показан на рис. 4.37.

По этому графику можно определить, насколько параметры выборки соответствуют заданному распределению (в данном случае — нормальному). В случае полного соответствия все точки располага-

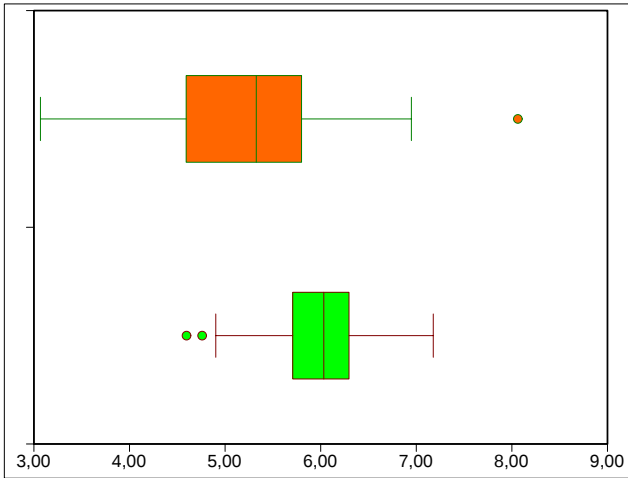


Рис. 4.36. Ящичковая диаграмма с выбросами для двух выборок

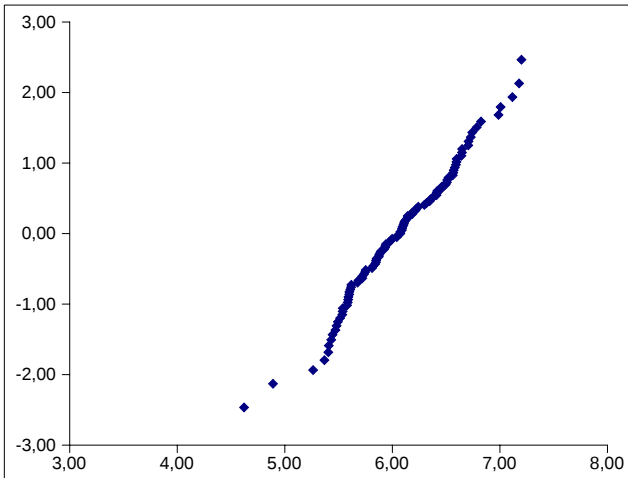


Рис. 4.37. График распределения вероятности

ются на одной прямой. Чем больше отклонения от прямой линии, тем хуже соответствие.

Для быстрой проверки соответствия выборки заданному распределению можно добавить к серии данных линию регрессии (см. главу «Регрессионный анализ в Gnumeric»).

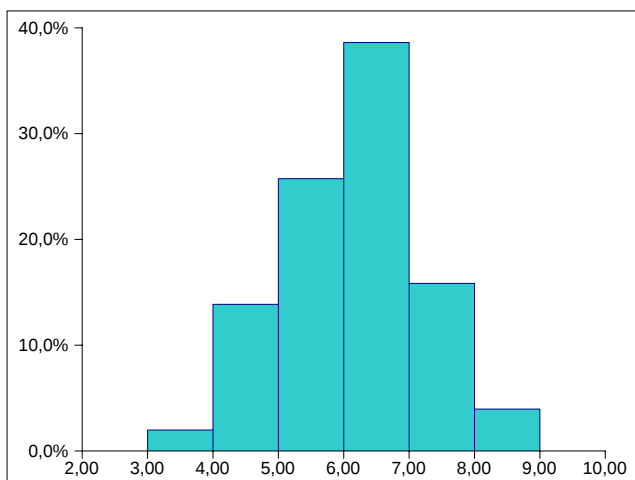


Рис. 4.38. Гистограмма

4.5.3 Гистограмма

Гистограмма показывает количество (или долю) значений выборки, попадающих в тот или иной интервал значений, поэтому для построения гистограммы нужно задать блок ячеек (вектор), содержащий значения границ интервалов.

Пример гистограммы показан на рис. 4.38.

Нужно отметить, что в версиях Gnupercis, существующих во время написания книги (весна 2011 года) использование гистограммы как подтипа статистической диаграммы давало достаточно непонятные результаты. Поэтому лучше использовать построение гистограммы как элемента статистического анализа («Статистика/Описательные статистики/Частотная таблица/Гистограмма...»).

Более подробно процесс построения гистограммы будет рассматриваться в главе «Инструменты Gnupercis для статистиков».

Глава 5

Инструменты Gnumeric для статистиков

Инструменты статистической обработки данных находятся в пункте главного меню «Статистика» (рис. 5.1). В этой главе рассмотрим принципы работы большинства из них, поскольку от версии к версии добавляются новые инструменты и возможности.

5.1 Основные принципы создания и форматирования диаграмм

5.2 Описательные статистики

Исследование возможностей Gnumeric по статистической обработке данных начнем с простейшей задачи — получения основных ста-

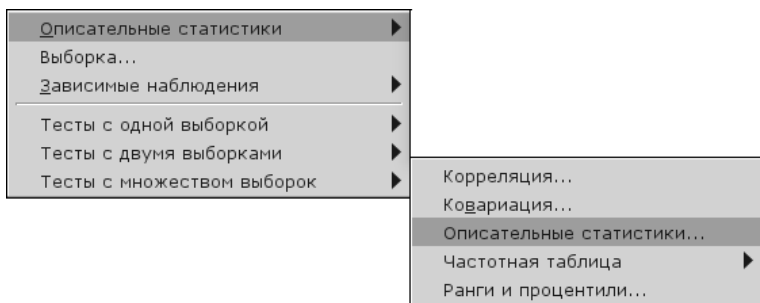


Рис. 5.1. Средства статистического анализа

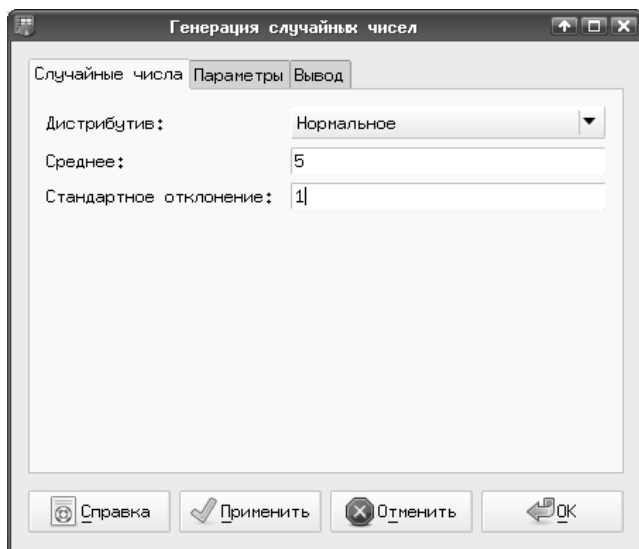


Рис. 5.2. Выбор распределения для создания исходных данных

статистических характеристик выборки. В качестве исходных данных будем использовать диапазоны ячеек, заполненные последовательно случайных чисел. Примеры данных здесь приводить не имеет смысла, поэтому будет описываться вид исходного модельного распределения и его параметры, а на рисунках будут приводиться диалоги формирования исходных данных и результаты.

Для начала сформируем выборку с нормальным распределением, задав среднее значение 5 и стандартное отклонение 1. Это делается в помощью диалога «Генерация случайных чисел» (рис. 5.2), вызываемого из главного меню («Правка/Заполнить/Генерация случайных чисел...») или «Данные/Заполнить/Генерация случайных чисел...»).

На вкладке «Случайные числа» устанавливаем вид распределения (здесь названный «Дистрибутив») — Нормальное, Среднее значение — 5 и Стандартное отклонение — 1. На вкладке «Параметры» устанавливаем Число переменных — 1 и Размер выборки — 25 (рис. 5.3).

Наконец, на вкладке «Вывод» устанавливаем диапазон вывода — диапазон ячеек, начиная, например, с A4 на текущем листе (рис. 5.4). После нажатия на кнопки «Применить» и «ОК» будет получено 25 случайных чисел с заданным законом распределения.

Теперь получим базовые статистические характеристики этой выборки как будто мы про неё ничего не знаем. Для этого выделит наши данные и вызовем диалог «Описательные статисти-

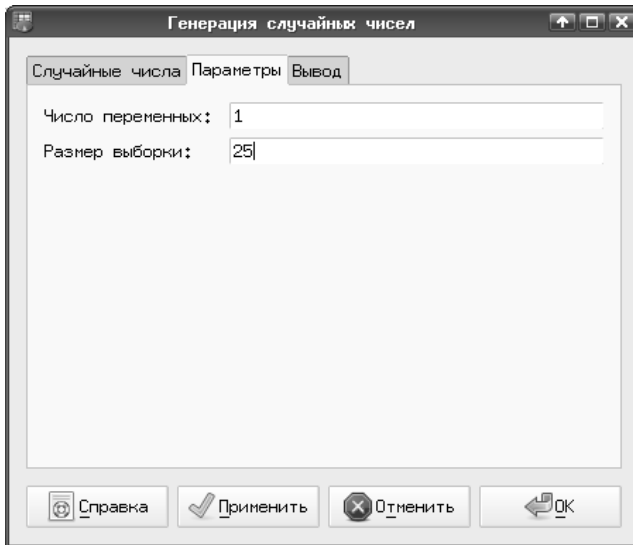


Рис. 5.3. Определение параметров выборки

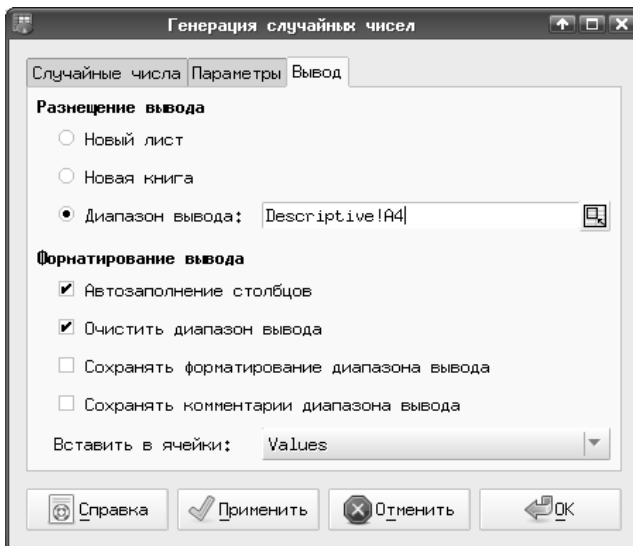


Рис. 5.4. Настройка размещения результатов

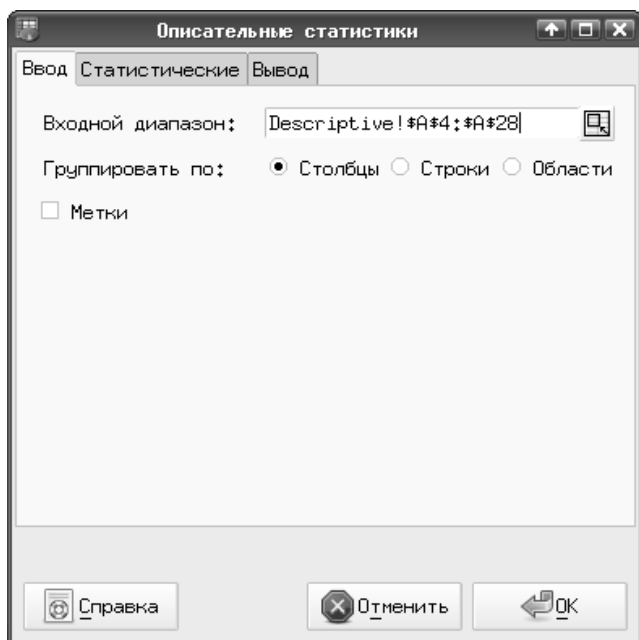


Рис. 5.5. Определение диапазона данных для обработки

ки» («Статистика/Описательные статистики/Описательные статистики...») (рис. 5.5).

На вкладке «Ввод» проверяем правильность диапазона ввода, на вкладке «Статистические» при необходимости уточняем доверительный интервал и другие параметры (можно все оставить по умолчанию, как на рис. 5.6). Наконец, на вкладке «Вывод» опять-таки задаем ячейку текущего листа, с которой начнется вывод результатов (рис. 5.7).

Очевидно, что раз 25 точек данных начинаются с A4, то имеет смысл выводить результаты статистического анализа после окончания данных. Хотя ничто не мешает вывести результаты на отдельный лист, это не очень удобно, если приходится сравнивать характеристики нескольких выборок.

В дальнейшем без особой необходимости все эти однотипные диалоги приводиться не будут. Теперь посмотрим на результаты обработки исходных данных — те самые описательные статистики для нормального распределения (рис. 5.8).

Из приведённых результатов видно, что сгенерированы были действительно случайные числа. Вычисленные по выборке значения близки к параметрам, по которым формировалась эта выборка, но совпадение не идеальное, т.е. фактор «случайности» действительно

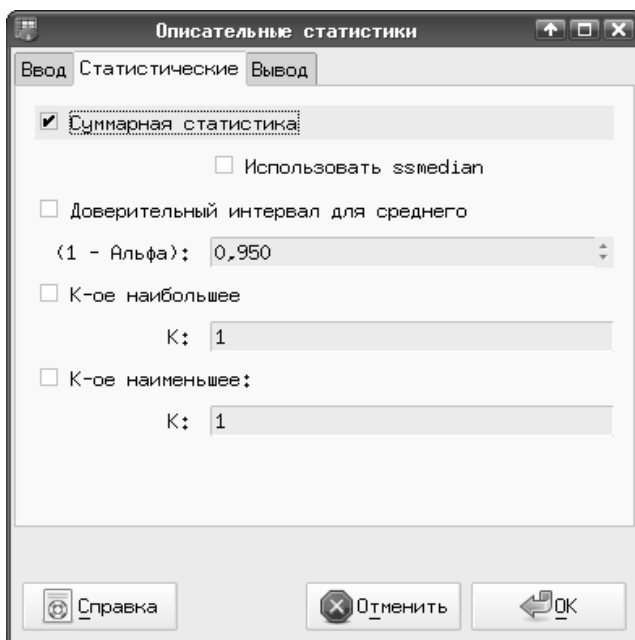


Рис. 5.6. Уточнение параметров обработки данных

имеет место. Отсутствие значения для моды, вероятно, связано с тем, что исходная выборка воспринимается как вариативный ряд, в котором нет варианты с максимальной частотой, поскольку значения не повторяются.

Теперь проделаем те же операции для однородного распределения в диапазоне

$[-2;2]$ и посмотрим на результаты (рис. 5.9).

Здесь стандартное отклонение очень велико по сравнению с диапазоном от минимума до максимума, что неудивительно для однородного распределения.

Таким образом, инструмент «Описательные статистики» позволяет получить практически все необходимые статистические характеристики имеющейся выборки.

5.3 Прогнозирование

Статистическое прогнозирование (в англоязычных статистических программах — forecasting) является на самом деле сглаживанием, которое применяется для выделения тенденции при сильном разбросе точек исходных данных. В Gnumeric эта процедура может

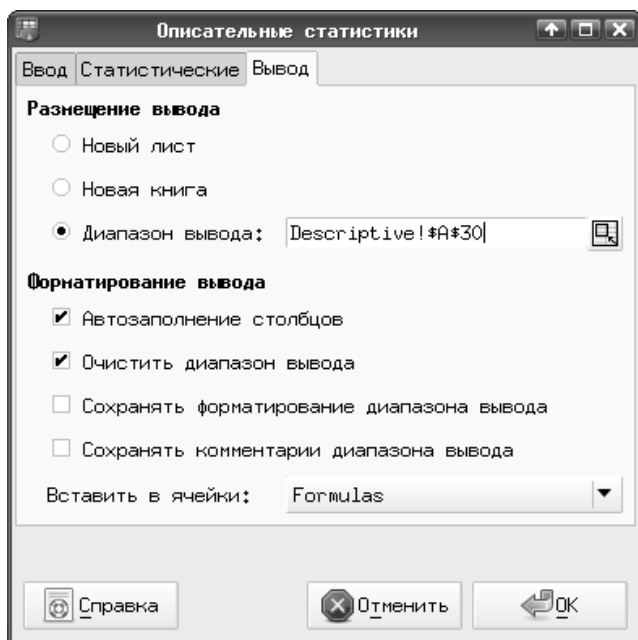


Рис. 5.7. Определение размещения результатов обработки

		Столбец 1
30		
31	Среднее	5,04464909892853
32	Стандартная ошибка	0,18236526945984
33	Медиана	5,01330411151615
34	Мода	#N/A
35	Стандартное отклонение	0,91182634729922
36	Выборочная дисперсия	0,83142728762905
37	Экссесс	1,11813424036683
38	Ассиметрия	0,65839749421218
39	Диапазон	4,06723579731569
40	Минимум	3,45842165522619
41	Максимум	7,52565745254188
42	Сумма	126,116227473213
43	Количество	25

Рис. 5.8. Описательные статистики для нормального распределения

	Столбец 1
Среднее	0,15410167465159
Стандартная ошибка	0,18644560472611
Медиана	0,15616328055933
Мода	#N/A
Стандартное отклонение	0,93222802363054
Выборочная дисперсия	0,86904908804211
Эксцесс	-0,66484799688632
Ассиметрия	0,10377506362021
Диапазон	3,56023754139636
Минимум	-1,63299821817276
Максимум	1,9272393232236
Сумма	3,8525418662898
Количество	25

Рис. 5.9. Описательные статистики для однородного распределения

проводиться двумя способами — методом экспоненциального сглаживания и методом скользящего среднего (соответственно, команды главного меню «Статистика/Зависимые наблюдения/Прогнозирование/Экспоненциальное сглаживание...» и «Статистика/Зависимые наблюдения/Прогнозирование/Скользящее среднее...»). При выборе сглаживания методом скользящего среднего можно указать количество точек, по которым будет проводиться усреднение.

Рассмотрим пример с некоторыми экспериментальными данными (рис. 5.10). Вектор X представляет собой некоторую независимую переменную, вектор Y — измеренные значения. Правее приведены результаты экспоненциального сглаживания и сглаживания методом скользящего среднего по трем точкам. Поскольку при сглаживании для данного значения Y оказываются задействованы предыдущие и последующие значения, то количество «сглаженных» точек меньше, чем количество исходных. Это видно как по отсутствию последнего значения в обоих случаях сглаживания, так и из сообщения «N/A (нет данных)» в начале последовательности. Для скользящего среднего по трем точкам результат вообще начинается только с третьей точки последовательности.

График исходных данных и результатов сглаживания показан на рис. 5.11. Нужно заметить, что сами операции сглаживания («прогнозирования») дают только числовые значения.

Из графика видно, что скользящее среднее (сплошная линия) в данном примере дает лучший результат, но чем больше точек участу-

2				
3			Эксп. сглаживание	Скольз. среднее(3)
4			Столбец 1	Столбец 1
5	X	Y	#N/A	#N/A
6	1	80,54	80,540000	#N/A
7	2	54,21	59,476000	61,920000
8	3	51,01	52,703200	43,493333
9	4	25,26	30,748640	31,566667
10	5	18,43	20,893728	18,933333
11	6	13,11	14,666746	14,763333
12	7	12,75	13,133349	11,643333
13	8	9,07	9,882670	9,406667
14	9	6,4	7,096534	6,633333
15	10	4,43	4,963307	4,740000
16	11	3,39	3,704661	3,326667
17	12	2,16	2,468932	2,416667
18	13	1,7	1,853786	1,666667
19	14	1,14	1,282757	1,163333
20	15	0,65		

Рис. 5.10. Исходные данные и результаты сглаживания

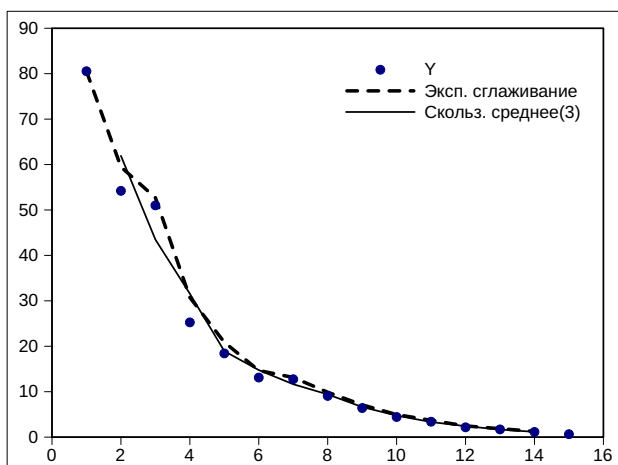


Рис. 5.11. Графическое представление исходных данных и результатов сглаживания

ют в усреднении (и чем более гладкой получается кривая), тем больше точек в начале и в конце теряются.

5.4 Корреляция

Корреляционный анализ, как известно, позволяет выявить взаимосвязь двух случайных величин. Коэффициент корреляции (корреляция по Пирсону) может принимать значения от -1 до 1. Чем ближе к 1 абсолютная величина коэффициента корреляции, тем сильнее связаны исследуемые случайные величины.

Для примера рассмотрим несколько выборок. Первая выборка (назовем ее «Выборка1») является нормально распределенной случайной величиной со средним значением 5 и стандартным отклонением 1, вторая («Выборка2») — также нормально распределенная случайная величина со средним значением 4 и стандартным отклонением 2, третья («Выборка3») — случайная величина с однородным распределением в интервале [-2;2]. Четвертая выборка («Выборка4») получена путем удвоения значений Выборки1 и добавления к результату десятой доли значений Выборки3 (т.е. $\text{Выборка4} = 2 * \text{Выборка1} + \text{Выборка3} / 10$) в каждой точке.

Пусть в каждой выборке будет по 25 значений, и начинать генерацию исходных данных будем со столбца А. После создания выборок вызовем диалог «Корреляция» («Статистика/Описательные статистики/Корреляция...», рис. 5.12).

Присвоение имён столбцам исходных данных и использование режима «Метки» в диалоге «Корреляция» позволяет на выходе получить именованные результаты (рис. 5.13).

Видно, что каждая выборка сама с собой прекрасно коррелирует (коэффициент равен 1), а четвертая выборка с первой дает коэффициент почти равный 1 (почти, но не совсем, поскольку Выборка4 искажена дополнительным влиянием Выборки3).

Для упражнения полезно вычислить коэффициент корреляции двух независимых выборок случайных величин с одинаковыми параметрами распределения.

5.5 Ковариация

Коэффициент ковариации также позволяет определить взаимосвязи случайных величин, но, в отличие от коэффициента корреляции, этот параметр не является нормированным, поэтому его значение не несет никакой очевидной информации. Более информативным является вычисление коэффициента корреляции. Посмотрим на результаты вычисления ковариации («Статистика/Описательные статисти-

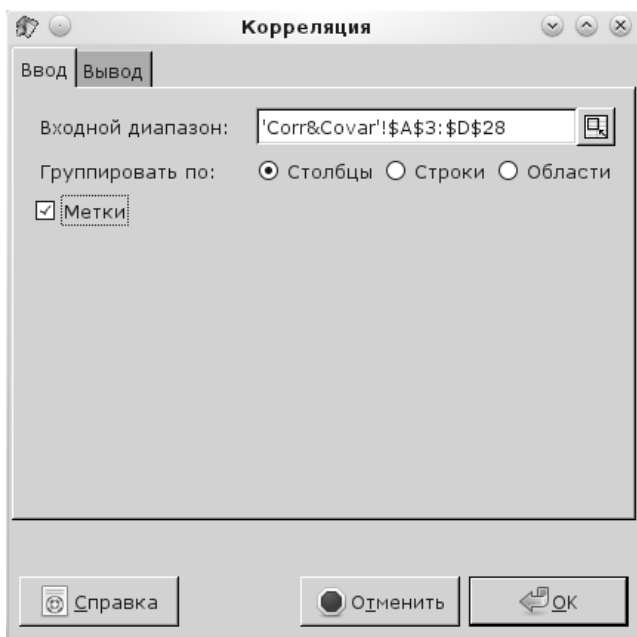


Рис. 5.12. Определение исходных данных для вычисления корреляции

30					
31	<i>Корреляции</i>	<i>Выборка1</i>	<i>Выборка2</i>	<i>Выборка3</i>	<i>Выборка4</i>
32	<i>Выборка1</i>	1,000000			
33	<i>Выборка2</i>	-0,330514	1,000000		
34	<i>Выборка3</i>	-0,282875	0,276372	1,000000	
35	<i>Выборка4</i>	0,999637	-0,325260	-0,256946	1,000000

Рис. 5.13. Результаты вычисления корреляции выборок

37	Ковариации	Выборка1	Выборка2	Выборка3	Выборка4
38	Выборка1	3,573914			
39	Выборка2	-1,428732	5,22851		
40	Выборка3	-0,563336	0,66571	1,109692	
41	Выборка4	7,091494	-2,79089	-1,015704	14,081418

Рис. 5.14. Результат вычисления ковариации выборок

ки/Ковариация...») для тех же исходных данных, что и в примере вычисления корреляции (рис. 5.14).

Видно, что без дополнительных усилий как-то интерпретировать результаты вычисления ковариации достаточно сложно.

5.6 Регрессия

Регрессия как элемент статистического анализа в Gnumeric проводится по линейной модели, а отклонения рассматриваются как нормально распределённые случайные ошибки. В результате проведения такого регрессионного анализа («Статистика/Зависимые наблюдения/Регрессия...») вычисляется множество параметров.

Рассмотрим результаты регрессии для примера исходных данных, использованных ранее для сглаживания (см. рис. 5.10).

При создании иллюстрации (рис. 5.15) данные из длинных строчек (параметры "F" и "P") были перенесены вниз, т.е. в реальной таблице ячейки "F" и «Значимость F» находятся в той же строке, что и параметр «степень свободы», а значения параметров "P", «Ниже 95%» и «Выше 95%» — в тех же строках, что и значения коэффициентов регрессии.

По результатам выполнения регрессии можно получить следующее уравнение $Y = -4.59 * X + 55.677$. Параметр «Столбец1» дает коэффициент наклона прямой, а параметр «Пересечение» — точку пересечения прямой с осью Y.

Интересно сравнить результаты регрессионного анализа, проведенного таким образом, с уравнениями регрессии, которые можно получить на диаграмме XY (рис. 5.16).

На рис. 5.16 точками показаны исходные данные, пунктирной линией — линейная регрессия (верхнее уравнение), параметры которой в точности совпадают с вычисленными с помощью «статистического» регрессионного анализа. Сплошная линия и нижнее уравнение соответствуют экспоненциальной модели регрессии, которая дает гораздо лучший коэффициент определенности (критерий Пирсона).

Итоговый вывод			
Регрессионные статистики			
Множественная регрессия		-0,852668	
Коэффициент определенности		0,727043	
Подобранный коэффициент определенности		0,706046	
Стандартная ошибка		13,054635	
Наблюдения		15	
Дисперсионный анализ			
	степень свободы	сумма квадратов	Квадрат среднего
Регрессия	1	5901,17958892857	5901,17958892857
Остатки	13	2215,50541107143	170,423493159341
Всего	14	8116,685	
	F	Значимость F	
	34,626561629101	5,37181850557115E-05	
	Коэффициенты	Стандартная ошибка	t Stat
Пересечение	55,676571428571	7,09334247757616	7,8491305903499
Столбец 1	-4,590821428571	0,78016365764622	-5,88443384099957
	Значение P	Ниже 95%	Выше 95%
	2,752699921E-06	40,3523366704577	71,0008061866851
	5,371818506E-05	-6,27626254146192	-2,90538031568094

Рис. 5.15. Результаты регрессионного анализа экспериментальных данных

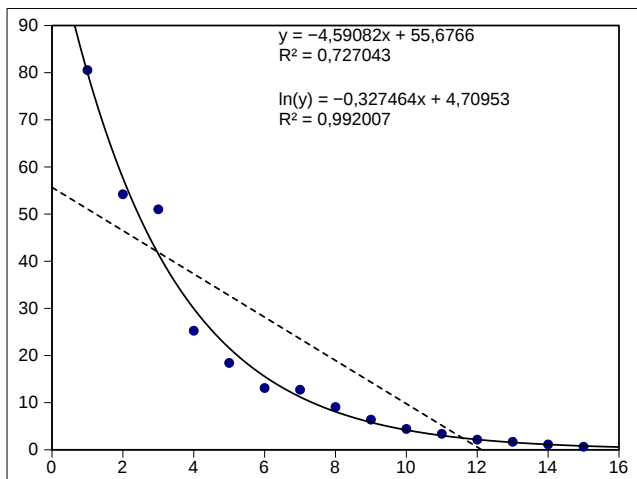


Рис. 5.16. Исходные данные и кривые регрессии на диаграмме XY

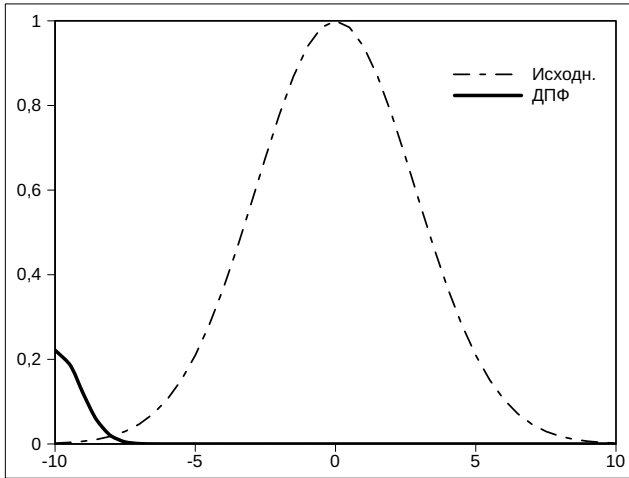


Рис. 5.17. Гауссова кривая и её Фурье-образ

Более подробно построение кривых регрессии будет рассматриваться в главе «Регрессионный анализ в Gnumeric».

5.7 Анализ Фурье

Модуль преобразования Фурье («Статистика/Зависимые наблюдения/Анализ Фурье...») позволяет вычислять дискретное преобразование Фурье (ДПФ) для заданного ряда данных (режим «Инверсия» на вкладке «Параметры» диалога «Анализ Фурье» позволяет, по-видимому, вычислять обратное преобразование Фурье).

Рассмотрим два примера. Первый пример — преобразование Фурье для гауссовой кривой, задаваемой формулой (5.1).

$$y = e^{\left(-\frac{(x-x_0)^2}{\sigma^2}\right)} \quad (5.1)$$

Известно, что для такой функции Фурье-образ будет иметь вид такой же функции (с точностью до нормировки). Пусть для простоты x_0 имеет значение 0, а дисперсия пусть будет равна 4. Графики исходной функции (пунктир) и её Фурье-образа (сплошная линия) показаны на рис. 5.17.

В Gnumeric вычисляются действительная и мнимая части Фурье-образа, так что для получения окончательного результата можно воспользоваться математической функцией `hypot()`, вычисляющей квадратный корень из суммы квадратов аргументов.

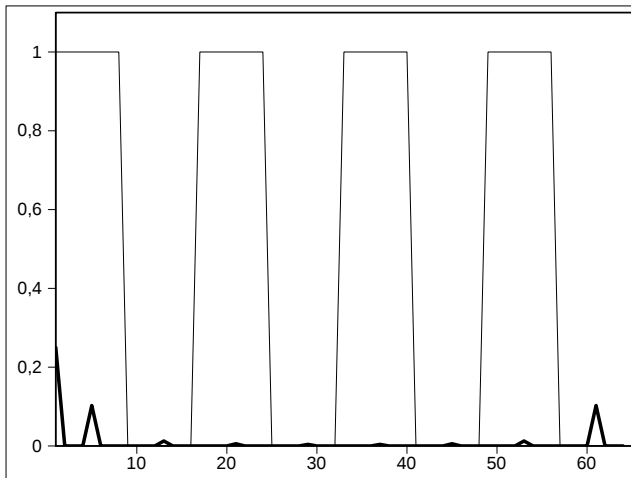


Рис. 5.18. «Импульсная» функция и её Фурье-образ

Теперь попробуем получить Фурье-образ для периодической «прямоугольной» функции (прямоугольных импульсов), причем состояние «0» и состояние «1» делятся по половине периода (рис. 5.18).

По горизонтальной оси отложены некие условные единицы. Период исходной функции составляет 16 единиц (всего взято 64 точки). Фурье-образ показан более жирной линией. Видно, что Фурье-образ симметричен относительно середины горизонтальной оси. Чем больше точек по горизонтали взято и чем больше период исходной функции, тем больше компонентов ряда Фурье можно различить.

Специально для читателей с инженерным и техническим образованием нужно заметить, что на рис. 5.18 показан квадрат модуля Фурье-образа, что соответствует, например, распределению энергии при дифракции электромагнитных волн на периодической решетке. В полном соответствии с теорией дифракции для заданных параметров решетки «энергия» в первом порядке (первый слева пик) составляет около 10% абсолютной величины.

5.8 Гистограмма

Инструмент «Гистограмма» («Статистика/Описательные статистики/Частотная таблица/Гистограмма...») вычисляет количество значений в выборке, попадающих в заданный интервал значений. Границы интервалов (отрезки, "cutoffs") могут быть заданы заранее или вычислены исходя их максимального и минимального значений и желаемого количества интервалов (рис. 5.19).



Рис. 5.19. Определение диапазонов значений для создания гистограммы

В качестве тестовых значений сформируем выборку из 39 нормально распределённых случайных величин со средним значением 5 и стандартным отклонением 2.

(На вкладке «Ввод» обычным образом задаётся диапазон ячеек с исходными данными, поэтому эту вкладку диалога настройки гистограммы не обсуждаем).

На вкладке «Двоичные» определяется способ учёта значений, на границах отрезков (рис. 5.20). Если какое-то значение точно (с учётом «машинного нуля») попадает на границу интервала (отрезка), то для границы, отмеченной квадратной скобкой («[» или «]»), оно учитывается в этом интервале (отрезке), а для границы, отмеченной круглой скобкой — в соседнем (предыдущем или следующем).

На вкладке «Графики и параметры» нужно определить вид диаграммы, которая будет сформирована и формат вывода результатов. Достаточно разумно заказать вывод гистограммы и представления результатов в процентах, как показано на рис. 5.21.

Наконец, на вкладке «Вывод», как обычно в Gnumeric, определяется лист и диапазон ячеек на листе, в который будут выводиться результаты (рис. 5.22).

После нажатия на кнопку ОК строится гистограмма и вычисляются частоты попадания значений выборки в заданные отрезки. Однако

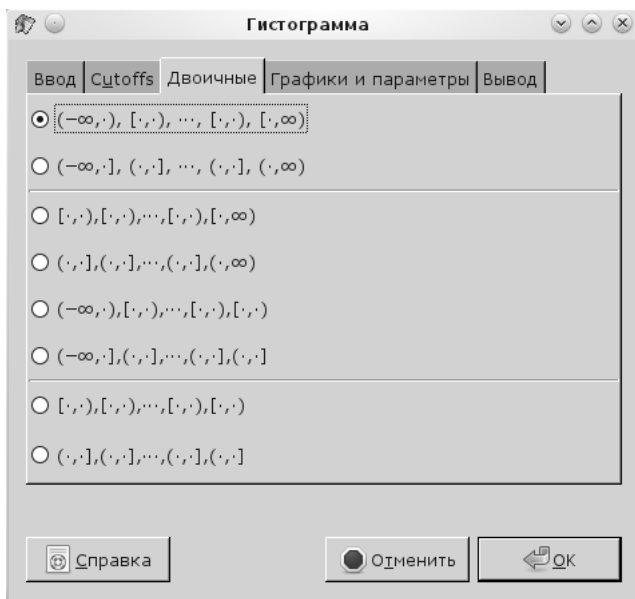


Рис. 5.20. Определение правил учёта границ отрезков

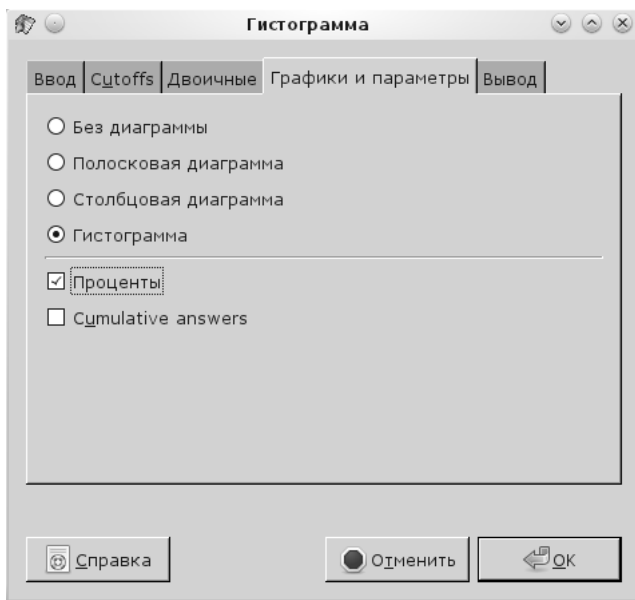


Рис. 5.21. Определение вида получаемой диаграммы и формата вывода результатов

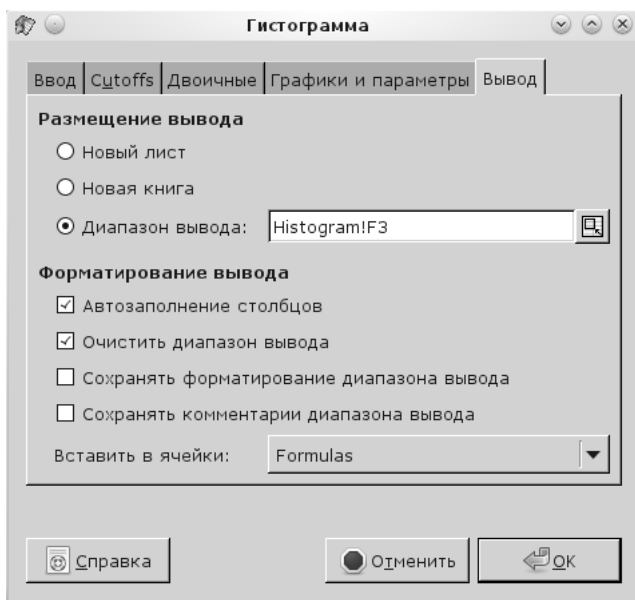


Рис. 5.22. Указание диапазона для вывода результатов

позиция графика гистограммы и диапазона ячеек с результатами совпадают, поэтому график нужно отодвинуть, чтобы увидеть числа (см. рис. 5.23).

График гистограммы можно теперь настраивать обычным образом, изменяя размер, форматы вывода по осям, цвет для серии данных и т.д.

5.9 Выборка

С помощью диалога «Выборка» можно выбрать несколько серий данных из некоторого вектора данных. Для примера в качестве исходного вектора рассмотрим столбец чисел от 0 до 25 (назовем столбец словом «Данные»). В диалоге «Выборка» («Статистика/Выборка...») на вкладке «Ввод», как обычно, определяется диапазон ячеек, содержащих исходные данные, на вкладке «Вывод» — расположение результатов.

Вкладка «Параметры» (рис. 5.24) заслуживает отдельного обсуждения.

Есть возможность задать периодическую выборку и случайную. Пример параметров для периодической выборки показан на рис. 5.24. Параметр «Смещение» определяет начальную позицию в векторе исходных данных. Для вектора, имеющего имя, его можно оставить рав-

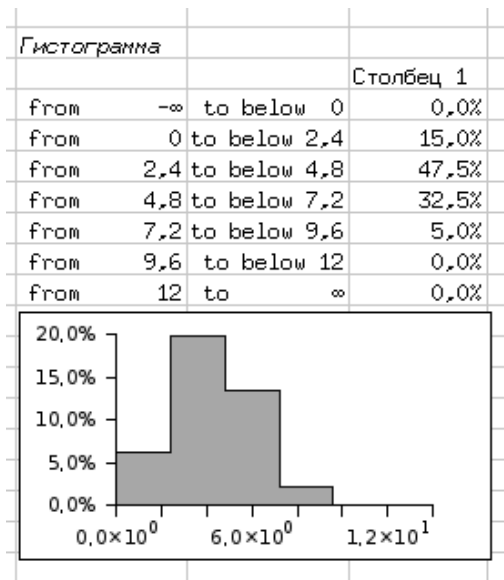


Рис. 5.23. Результаты вычислений и график гистограммы

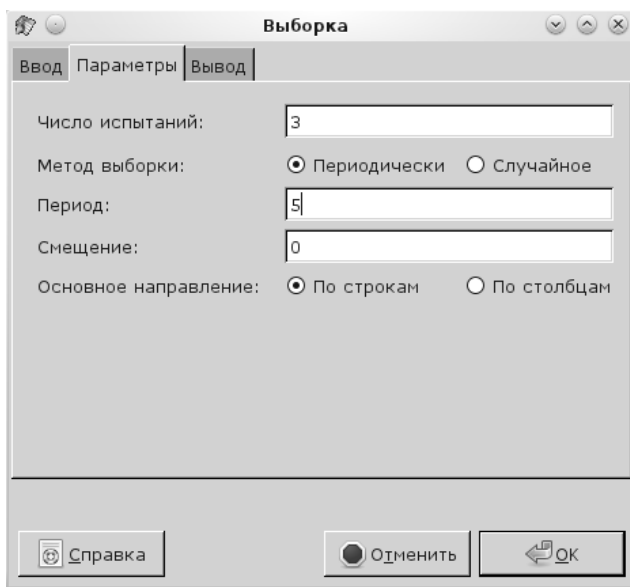


Рис. 5.24. Определение параметров периодической выборки

Данные	Данные	Данные
4	4	4
9	9	9
14	14	14
19	19	19
24	24	24

Рис. 5.25. Результат трёх испытаний для периодической выборки

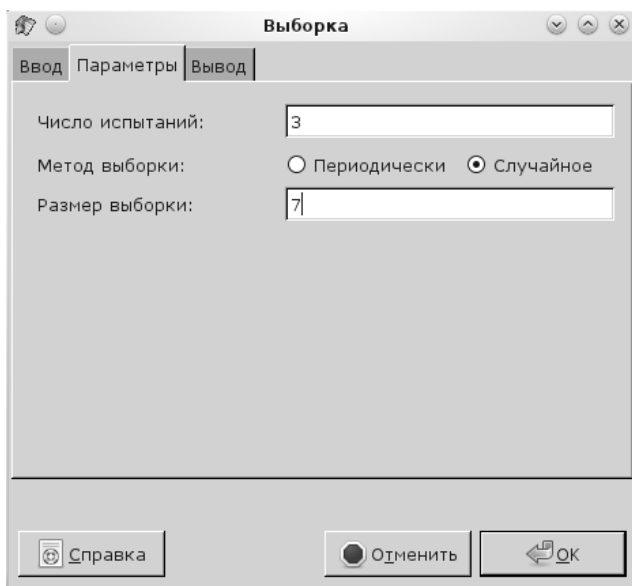


Рис. 5.26. Определение параметров случайной выборки

ным 0, если на вкладке «Ввод» установлен режим «Метки» (всё становится понятным, если провести несколько экспериментов). Результат такой выборки показан на рис. 5.25.

Видно, что при периодической выборке размер периода задаёт интервал между выбираемыми значениями (в данном примере выбирается каждое пятое значение, начиная с номера 0), и в различных сериях (испытаниях) значения повторяются.

Пример параметров случайной выборки показан на рис. 5.26, а результат — на рис. 5.27.

Данные	Данные	Данные
8	24	6
24	7	4
19	9	1
17	17	13
1	21	17
3	4	20
0	10	14

Рис. 5.27. Результат трёх испытаний при случайной выборке

Видно, что для случайной выборки количество точек определяется параметром «Размер выборки» и в различных сериях (испытаниях) значения не повторяются.

Инструмент «Выборка» может быть использован для превращения вектора данных в матрицу, а также для «прореживания» больших наборов исходных данных.

5.10 Ранги и проценти

Инструмент «Ранги и проценти» выполняет сортировку выборки по убыванию, определяет максимальное значение в выборке, присваивая ему первый ранг и значение в 100%, остальные значения выстраивает относительно первого и определяет, какой процент составляет текущая величина в выборке относительно максимального значения, а также указывает позицию каждого значения в выборке.

Рассмотрим, например, гипотетические результаты единого госэкзамена по ботанике среди группы в 25 участников (фамилии не указаны, поскольку они не используются для формирования данных). Таблица исходных данных приведена на рис. 5.28 (столбец «Баллы»).

На вкладке «Параметры» диалога «Ранг и процентиль» (рис. 5.29) можно определить способ вычисления ранга (варианты «Средний ранг» и «Верхний ранг»).

Поскольку требуется распределить весь набор значений на количество мест, равных количеству значений, то для повторяющихся значений получается деление мест (два одинаковых значения занимают два уровня рангов). Поэтому при использовании режима «Средний ранг» получаем дробные значения рангов (например, от 13 до 14 ранга имеется два значения, итого в среднем ранг получается 13,5). Результаты вычислений в таком режиме иллюстрируются группой «Средний ранг» на рис. 5.28.

Баллы	Средний ранг				Верхний ранг			
	Точки	Баллы	Ранг	Процентиль	Точки	Баллы	Ранг	Процентиль
12	7	93	1,5	100,00%	7	93	1	100,00%
34	10	93	1,5	100,00%	10	93	1	100,00%
28	6	88	3	91,67%	6	88	3	91,67%
47	22	78	4	87,50%	22	78	4	87,50%
67	14	77	5,5	83,33%	14	77	5	83,33%
88	23	77	5,5	83,33%	23	77	5	83,33%
93	15	68	7	75,00%	15	68	7	75,00%
45	5	67	8,5	70,83%	5	67	8	70,83%
48	20	67	8,5	70,83%	20	67	8	70,83%
93	11	65	10	62,50%	11	65	10	62,50%
65	16	58	11	58,33%	16	58	11	58,33%
38	9	48	12	54,17%	9	48	12	54,17%
19	4	47	13,5	50,00%	4	47	13	50,00%
77	25	47	13,5	50,00%	25	47	13	50,00%
68	17	46	15,5	41,67%	17	46	15	41,67%
58	24	46	15,5	41,67%	24	46	15	41,67%
46	8	45	17	33,33%	8	45	17	33,33%
43	18	43	18	29,17%	18	43	18	29,17%
34	12	38	19	25,00%	12	38	19	25,00%
67	2	34	21	20,83%	2	34	20	20,83%
34	19	34	21	20,83%	19	34	20	20,83%
78	21	34	21	20,83%	21	34	20	20,83%
77	3	28	23	8,33%	3	28	23	8,33%
46	13	19	24	4,17%	13	19	24	4,17%
47	1	12	25	0,00%	1	12	25	0,00%

Рис. 5.28. Исходные данные и расчёты рангов и процентилей

При использовании режима «Верхний ранг» используется минимальное значение ранга для повторяющихся значений, что иллюстрируется группой «Верхний ранг» на рис. 5.28.

5.11 Дисперсионный анализ

Согласно определению, данному в классической книге В.Е.Гмурмана «Теория вероятностей и математическая статистика», «дисперсионный анализ (ДА) применяют, чтобы установить, оказывает ли существенное влияние некоторый качественный фактор F , который имеет p уровней $F_1, F_2 \dots F_p$ на изучаемую величину X . Основная идея дисперсионного анализа состоит в сравнении «факторной дисперсии», порождаемой воздействием фактора, и «остаточной дисперсии», обусловленной случайными причинами. Если различие этих дисперсий значимо, то фактор оказывает существенное влияние на X .

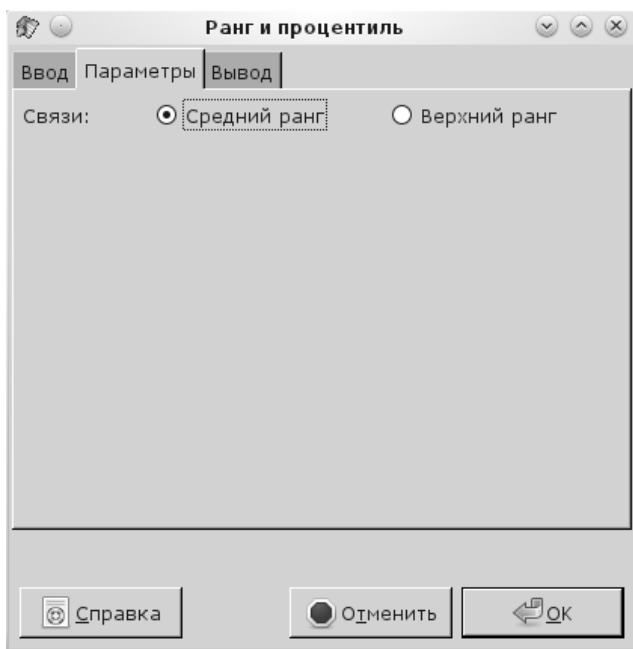


Рис. 5.29. Определения способа вычисления рангов

В более сложных случаях исследуют воздействие нескольких факторов на нескольких постоянных или случайных уровнях и выясняют влияние отдельных уровней и их комбинаций (многофакторный анализ)».

5.11.1 Однофакторный дисперсионный анализ

Рассмотрим пример однофакторного ДА на основе данных, взятых из учебного пособия «Статистические методы в инженерных исследованиях (лабораторный практикум)» (авторы Бородюк В.П., Воцинин А.П., Иванов А.З и др., см. список литературы).

Исследуется зависимость долговечности у электрических лампочек (в часах) от технологии изготовления (фактор x). В качестве исходных данных используется отклонение долговечности от «стандартного» значения в 1500 часов для четырёх неравночисленных серий образцов из разных партий (см. рис. 5.30).

В таблице приведены отклонения для различных образцов (y — номера образцов).

	y1	y2	y3	y4	y5	y6	y7	y8
серия 1	100	110	150	180	200	200	300	
серия 2	80	140	140	200	250			
серия 3	-40	50	100	120	140	160	240	320
серия 4	10	20	30	70	100	180		

Рис. 5.30. Исходные данные для однофакторного ДА

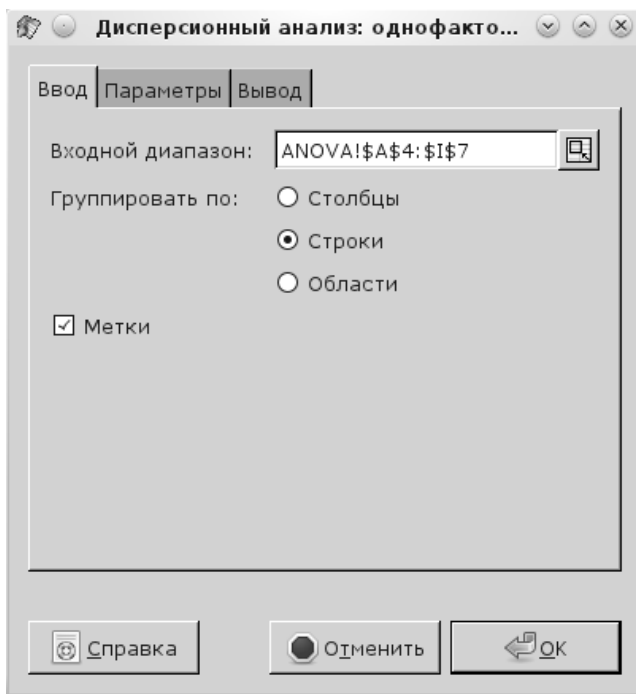


Рис. 5.31. Определение исходных данных для однофакторного ДА

Диалог определения исходных данных для однофакторного ДА «Статистика/Тесты с множеством выборок/Дисперсионный анализ/Однофакторный...» показан на рис. 5.31.

Важно не включать в диапазон данных лишние ячейки, в данном случае — ячейки с названиями образцов, что обеспечивается включением режима «Метки».

На вкладке «Параметры» устанавливается уровень значимости «Альфа» (по умолчанию — общепринятое значение 5%).

Дисперсионный анализ: однофакторный				
ИТОГО				
Группы	Количество	Сумма	Среднее	Дисперсия
серия 1	7	1240	177,142857142857	4557,1428571429
серия 2	5	810	162	4220
серия 3	8	1090	136,25	12169,642857143
серия 4	6	410	68,3333333333333	4136,6666666667
Дисперсионный анализ				
Источник дисперсии	Сумма квадратов	степень свободы	Квадрат среднего	F
Между группами	42694,771062271	3	14231,5903540904	2,0859969982526
В группах	150093,69047619	22	6822,44047619048	
Всего	192788,46153846	25		
			Значение P	F критическое
			0,13120350077972	3,0491249886521

Рис. 5.32. Результаты однофакторного ДА

Вкладка «Вывод» — стандартная для всех диалогов статистического анализа, на ней определяется местоположение результатов вычислений.

Результаты показаны на рис. 5.32. Для повышения компактности вывода некоторые ячейки перенесены.

Какой же из этого всего следует вывод? А вывод такой: поскольку вычисленное значение результата «F» (F-критерий) меньше, чем «F критическое» для данного уровня значимости, влияние фактора (технологии изготовления) на исследуемый параметр (долговечность лампочек) является несущественным.

5.11.2 Двухфакторный дисперсионный анализ

В качестве примера применения Gnumeric для двухфакторного дисперсионного анализа рассмотрим задачу, приведенную в статье «Практикум по статистике с пакетами StatGraphics, Statistica, SPSS» на <http://www.exponenta.ru/educat/systemat/goritskii/part2/LR7/2.asp> (см. список литературы).

Исследуется урожайность четырех сортов пшеницы, ц/га (фактор А, 4 уровня) от используемого вида удобрений (5 уровней фактора В). Данные получены с 20 участков равной площади и одинакового почвенного состава. Требуется выяснить влияние сорта пшеницы и типа удобрений на урожайность.

Таблицы исходных данных приведена на рис. 5.33.

На рис. 5.34 показана вкладка «Ввод» диалога «Дисперсионный анализ: двухфакторный». Вкладки «Параметры» и «Вывод» являются стандартными, поэтому показывать их нет особого смысла.

	A1	A2	A3	A4
B1	19	25	17	21
B2	22	19	19	18
B3	26	23	22	25
B4	18	26	20	23
B5	21	22	21	24

Рис. 5.33. Исходные данные для двухфакторного ДА

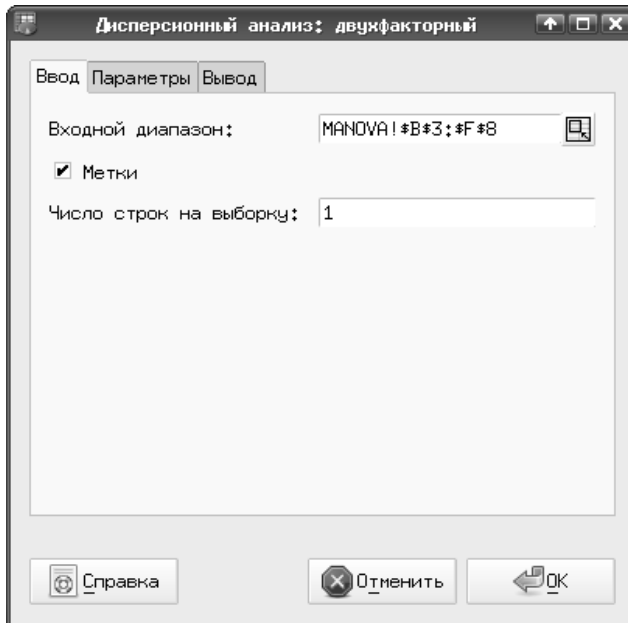


Рис. 5.34. Определение исходных данных для двухфакторного ДА

<i>Дисперсионный анализ: двухфакторный без воспроизведения</i>					
<i>Итого</i>	<i>Количество</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>	
<i>B1</i>	4	82	20,5	11,6666667	
<i>B2</i>	4	78	19,5	3	
<i>B3</i>	4	96	24	3,33333333	
<i>B4</i>	4	87	21,75	12,25	
<i>B5</i>	4	88	22	2	
<i>A1</i>	5	106	21,2	9,7	
<i>A2</i>	5	115	23	7,5	
<i>A3</i>	5	99	19,8	3,7	
<i>A4</i>	5	111	22,2	7,7	
<i>Дисперсионный анализ</i>					
<i>Источник дисперсии</i>	<i>Сумма квадратов</i>	<i>степень свободы</i>	<i>Квадрат среднего</i>		
<i>Строки</i>	46,2	4	11,55		
<i>Столбцы</i>	28,550000000000	3	9,51666666666679		
<i>Ошибка</i>	68,200000000000	12	5,68333333333333		
<i>Всего</i>	142,95	19			
	<i>F</i>	<i>Значение P</i>	<i>F критическое</i>		
	2,0322580645161	0,1536620984991	3,25916672690125		
	1,6744868035191	0,2251061113715	3,49029481949754		

Рис. 5.35. Результаты двухфакторного ДА

Результаты вычислений показаны на рис. 5.35. Для уменьшения размера рисунка итоговые значения ("F", "P" и «F критическое») перенесены на другую строку.

При использовании Gnumeric получены те же значения уровней значимости P (0,153 и 0,225), что и в примере первоисточника. Соответственно, делается вывод о том, что в результате дисперсионного анализа не обнаружено влияние сорта пшеницы и типа удобрения на урожайность, что также видно из того, что при вычисленных уровнях значимости значения критерия F получаются меньше, чем соответствующие значения параметра «F критическое».

Нужно заметить, что использование свободно распространяемого пакета Gnumeric для решения подобных задач выглядит значительно привлекательнее использования пакета Statistica ценой около \$700 USD, не говоря уже о StatGraphics или SPSS, легально приобрести которые весьма затруднительно.

5.12 Два средних

Вложенное меню «Два средних» («Статистика/Тесты с двумя выборками/Два средних») предоставляет набор инструментов для про-

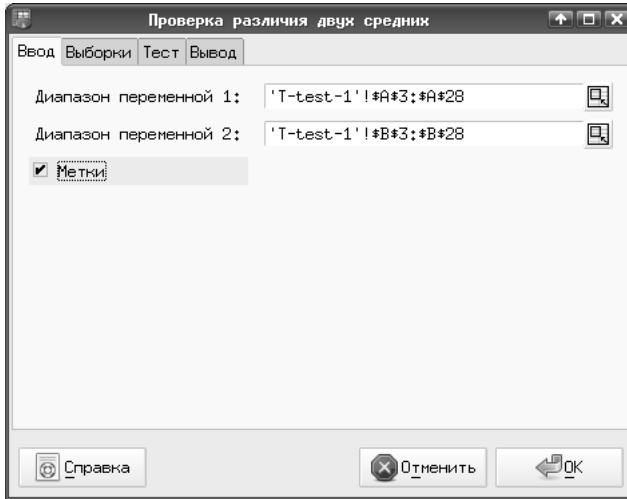


Рис. 5.36. Настройка исходных данных для проверки гипотезы

верки гипотезы о равенстве (или неравенстве) средних значений выборок (генеральных совокупностей). В качестве исходных данных будем средствами Gnumeric генерировать выборки с нормальным распределением.

5.12.1 Равные выборки: Т-тест

Пусть известно, что в двух выборках имеется равное количество значений случайных величин, и известно, что у этих выборок равные дисперсии. В данном случае Т-тест дает возможность определить дисперсии и средние значения для этих выборок и посчитать разницу между средними.

Для примера рассмотрим 25 нормально распределенных случайных значений со средним значением 5 и стандартным отклонением 1 (Выборка1) и 25 нормально распределенных случайных значений со средним значением 7 и стандартным отклонением 1 (Выборка2).

На вкладке «Ввод» диалога «Проверка различия двух средних» («Статистика/Тесты с двумя выборками/Два средних/Равные выборки: Т-тест...») указываем диапазоны для исходных данных, причем нужно проследить, чтобы все адреса были абсолютными (см. рис. 5.36). Режим «Метки», как всегда, позволяет использовать в выводе названия векторов данных.

На вкладке «Выборки» (рис. 5.37) указываем, что данные непарные, дисперсии неизвестны, но равны.

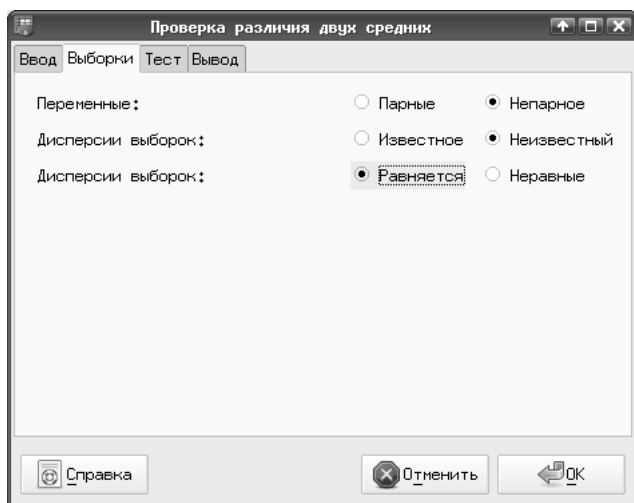


Рис. 5.37. Настройка режимов анализа

	Выборка1	Выборка2
<i>Среднее</i>	4,9856949420708	6,9219356509208
<i>Дисперсия</i>	1,00959256662225	1,1100760960711
<i>Наблюдения</i>		25
<i>Накопленная дисперсия</i>	1,05983433134665	
<i>Гипотетическое среднее отклонение</i>		0
<i>Наблюдаемое среднее отклонение</i>	-1,93624070885001	
<i>df</i>		48
<i>t Stat</i>	-6,64959754973821	
<i>P (T<=t) одностороннее</i>	1,2675474797465E-08	
<i>t критическое одностороннее</i>	1,67722419612551	
<i>P (T<=t) двухстороннее</i>	2,5350949594930E-08	
<i>t критическое двухстороннее</i>	2,01063475762452	

Рис. 5.38. Результаты проверки гипотезы

На рис. 5.38 показаны результаты вычислений, причем заметно прекрасное согласование результатов с заранее заданными параметрами выборок.

Проделав всё это с любыми модельными данными, в ячейках блока результатов можно увидеть формулы, по которым производятся расчеты.

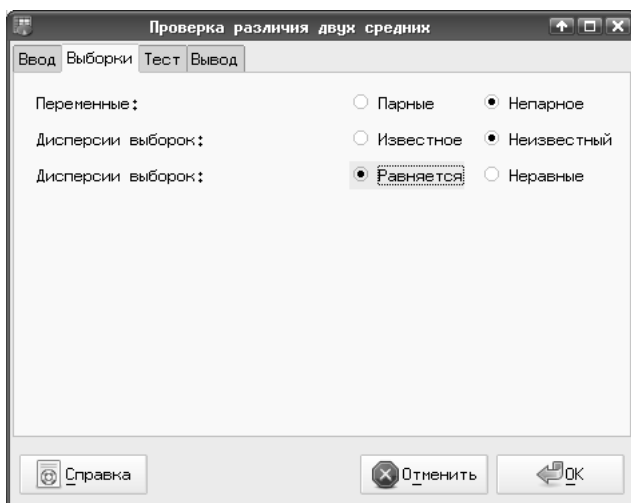


Рис. 5.39. Настройка режимов вычислений

5.12.2 Неравные выборки, равные дисперсии: Т-тест

Теперь рассмотрим ситуацию, когда выборки имеют разное количество точек. Пусть параметры Выборки1 остаются прежними (25 точек, нормальное распределение, среднее значение 5 стандартное отклонение 1), а для Выборки2 установим следующие параметры — нормальное распределение со средним 7, стандартным отклонением 1 и количеством точек 20.

В этом случае на вкладке «Выборки» диалога «Проверка различия двух средних» все оставляем по умолчанию (рис. 5.39), и наблюдаем результаты (рис. 5.40).

Настройки режимов вычислений в этом случае совпадают с предыдущим случаем.

Опять-таки, наблюдается соответствие результатов расчетов с ранее определенными параметрами выборок.

5.12.3 Неравные выборки, неравные дисперсии: Т-тест

Используя те же параметры распределения для Выборки1, что и в предыдущих случаях, для Выборки2 возьмем 20 точек, среднее значение 7 и стандартное отклонение 2.

На вкладке «Выборки» установим режимы в соответствии с рис. 5.41 и посмотрим результат (рис. 5.42).

	Выборка1	Выборка2
Среднее	4,9290079315592	7,1298748506069
Дисперсия	0,88875584053372	0,9422275270696
Наблюдения	25	20
Накопленная дисперсия	0,91238286481703	
Гипотетическое среднее отклонение	0	
Наблюдаемое среднее отклонение	-2,20086691904767	
df	43	
t Stat	-7,68040227794376	
P (T<=t) одностороннее	6,752644595038E-10	
t критическое одностороннее	1,68107070320362	
P (T<=t) двухстороннее	1,350528919008E-09	
t критическое двухстороннее	2,01669219922857	

Рис. 5.40. Результаты проверки гипотезы

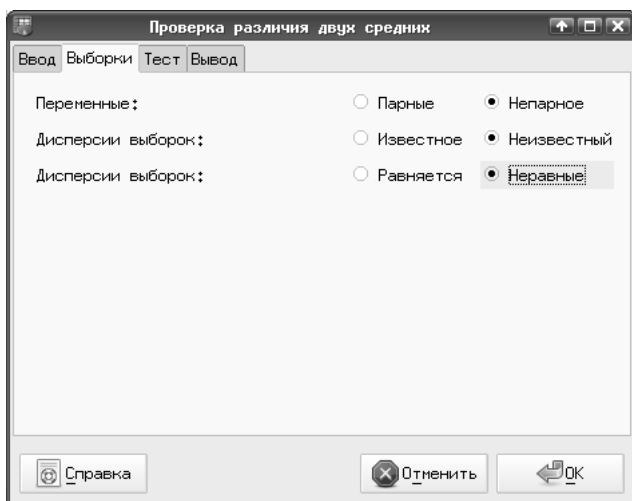


Рис. 5.41. Настройка режимов вычислений

	Выборка1	Выборка2
<i>Среднее</i>	5,22363347199752	6,9627418308079
<i>Дисперсия</i>	1,08564555028485	3,8771835295732
<i>Наблюдения</i>	25	20
<i>Гипотетическое среднее отклонение</i>	0	
<i>Наблюдаемое среднее отклонение</i>	-1,7391083588104	
<i>df</i>	27,3780733293963	
<i>t Stat</i>	-3,5701914147587	
<i>P (T<=t) одностороннее</i>	0,00067176549533	
<i>t критическое одностороннее</i>	1,7024535832589	
<i>P (T<=t) двухстороннее</i>	0,00134353099067	
<i>t критическое двухстороннее</i>	2,0505055360291	

Рис. 5.42. Результаты проверки гипотезы

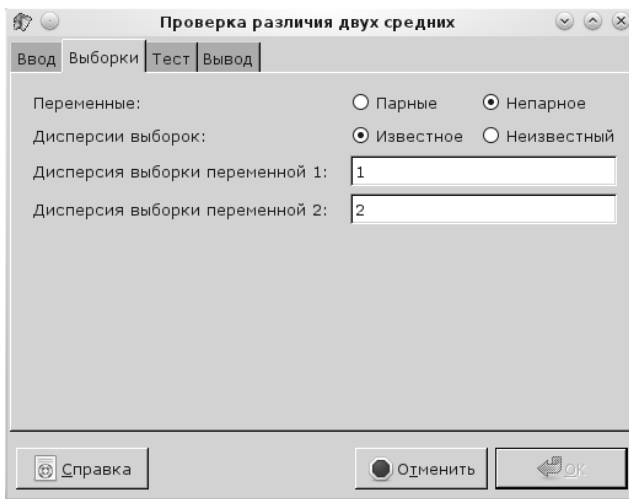


Рис. 5.43. Настройка режимов вычислений

5.12.4 Известные дисперсии: Z-тест

Применим эту процедуру к слегка измененным модельным данным. Пусть Выборка1 остается с прежними параметрами, а для Выборки2 (среднее 7, 20 точек) установим дисперсию 2, для чего стандартное отклонение должно быть установлено как 1,41.

На вкладке «Выборка» устанавливаем режимы и значения дисперсии в соответствии с рис. 5.43 и получаем результат, показанный на рис. 5.44.

	Выборка1	Выборка2
<i>Среднее</i>	4,81474951148839	6,523494868105
<i>Известная дисперсия</i>		1 2
<i>Наблюдения</i>		25 20
<i>Гипотетическое среднее отклонение</i>		0
<i>Наблюдаемое среднее отклонение</i>	-1,70874535661631	
<i>z</i>	-4,56681406121364	
<i>P (Z<=z) одностороннее</i>	2,47596446779E-06	
<i>z критическое одностороннее</i>	1,64485362695147	
<i>P (Z<=z) двухстороннее</i>	4,95192893550E-06	
<i>z критическое двухстороннее</i>	1,95996398454005	

Рис. 5.44. Результаты проверки гипотезы

F-Тест	Выборка1	Выборка2
<i>Среднее</i>	4,78087892427695	7,27042414268456
<i>Дисперсия</i>	1,37528345243102	1,44920309531595
<i>Наблюдения</i>		25 25
<i>df</i>		24 24
<i>F</i>	0,94899290297968	
<i>P (F<=F) правостороннее</i>	0,55049212345448	
<i>F критическое правостороннее</i>	1,98375956848961	
<i>P (F<=F) левостороннее</i>	0,44950787654552	
<i>F критическое левостороннее</i>	0,50409334673626	
<i>P двухстороннее</i>	0,89901575309104	
<i>F критическое двухстороннее</i>	0,44066893449361	2,26927727762143

Рис. 5.45. F-тест. Равные дисперсии

5.13 Две дисперсии: F-тест

Этот инструмент позволяет проверить гипотезу о равенстве (или неравенстве) двух дисперсий. В качестве исходных данных будем использовать те же модельные выборки, что и в случае проверки гипотезы о равенстве двух средних.

В качестве первого примера рассмотрим 25 нормально распределенных случайных значений со средним значением 5 и стандартным отклонением 1 (Выборка1) и 25 нормально распределенных случайных значений со средним значением 7 и стандартным отклонением 1 (Выборка2). В этом случае дисперсии однозначно равны.

Результат проведения теста показан на рис. 5.45.

F-Тест	Выборка3	Выборка4
Среднее	5,1093984261493	4,62256516995347
Дисперсия	0,70692477100268	3,83214446689538
Наблюдения		25
df		24
F	0,18447236974221	
P < F(<=f) правостороннее	0,99994957126822	
F критическое правостороннее	1,98375956848961	
P < F(<=F) левостороннее	5,0428731778185E-05	
F критическое левостороннее	0,50409334673626	
P двухстороннее	0,00010085746356	
F критическое двухстороннее	0,44066893449361	2,269272727762143

Рис. 5.46. F-тест. Неравные дисперсии

Теперь рассмотрим вариант, при котором дисперсии отличаются в два раза (Выборка3 и Выборка4). Результат теста показан на рис. 5.46.

Вывод получается следующий: чем сильнее отличаются дисперсии выборок, тем меньше значение F.

5.14 Оценка выживаемости (оценка Каплана-Майера)

Общие сведения о задаче анализа выживаемости можно получить в статье А.Б. Меркова «Об анализе выживаемости» или в руководствах по коммерческим статистическим пакетам (см. список литературы). Суть задачи заключается в том, чтобы по набору признаков (характеристик) определить время сохранения объектом этих характеристик («время жизни») или распределение вероятностей сохранения характеристик в заданных пределах. Соответственно, можно строить прогнозы (предсказывать) среднее «время жизни» (время сохранения характеристик) таких объектов. Объектами могут быть вещества, устройства (приборы), сооружения и конструкции, а также живые существа. Чаще всего оценка выживаемости упоминается в связи с медицинской практикой.

В тех случаях, когда время наблюдения (продолжительность испытаний) меньше, чем «время жизни» конкретного объекта, получается, что «время жизни» точно не меньше времени наблюдения, а вот какое оно конкретно — узнать уже нельзя. Такие данные называются «цензурированными» (censored). Для группы объектов, участвующих в испытаниях возможны одновременно цензурированные и нецензу-

	А	В	С
1	Длительность	Группа	Цензурир.
2	1	1	1
3	1	2	1
4	2	1	1
5	2	2	1
6	3	2	0
7	3	1	1
8	4	1	1
9	4	1	0
10	4	2	1
11	4	2	1
12	5	1	0
13	6	1	0
14	7	1	1
15	9	2	0
16	11	2	1
17	12	1	0
18	12	2	0
19	13	2	1
20	18	1	0
21	19	2	0

Рис. 5.47. Пример исходных данных для анализа выживаемости

рированные данные для различных экземпляров (например, при исследованиях срока службы энергосберегающих ламп в течение 10000 часов часть ламп вышла из строя в течение испытаний, а часть — так и не испортилась).

Пример использования Gnumeric для оценки выживаемости по Каплану-Майеру взят из справки по Gnumeric (Gnumeric 1.10.x).

Заготовим исходные данные в соответствии с рис. 5.47.

Первый столбец («Длительность») означает время испытаний (наблюдений) для каждого исследуемого экземпляра. В столбце «Группа» задаётся принадлежность объекта к группе объектов (группы могут отличаться местоположением, периодом времени наблюдений и другими признаками и обстоятельствами). В данном примере имеется только две группы. Наконец, в третьем столбце указывается признак «цензурированности» данных (если в ячейке 1 — данные цензурированы).

Все данные носят дискретный характер («время жизни» изменяется дискретно).

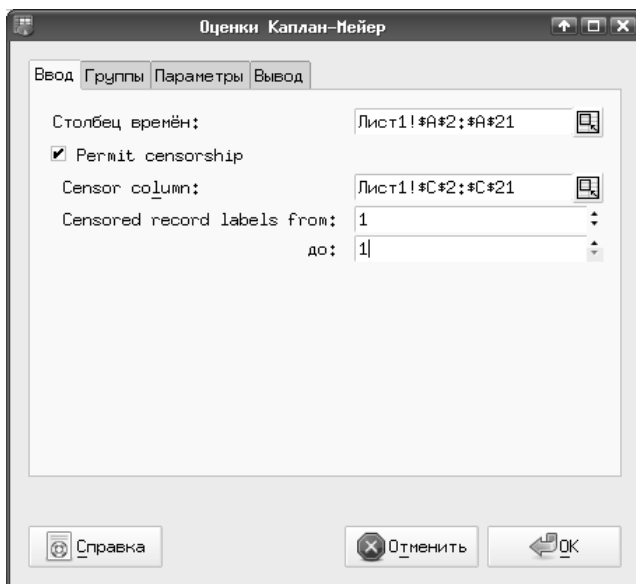


Рис. 5.48. Настройка исходных данных для анализа

Диалог настройки анализа вызывается через вложенное меню «Статистика/Зависимые наблюдения». Сначала определяется набор исходных данных и их цензурированность (вкладка «Ввод» диалогов, рис. 5.48). Использование цензурированных данных разрешается включением соответствующего режима (Permit censorship).

На вкладке «Группы» задаётся количество групп и номера, которые их определяют. Теоретически можно объединять несколько групп в одну, указав диапазон номеров «от» и «до» (рис. 5.49). Для установки номера группы используются поля со счётчиками (для редактирования поля нужно дважды щёлкнуть в нём левой кнопкой мыши).

В этом примере (и по умолчанию) используется две группы, но с помощью кнопок «Добавить» и «Удалить» количество групп можно изменять так, как требуется.

На вкладке «Параметры» (рис. 5.50) определяется объём итоговой информации. Различные виды результатов можно включать и выключать. Пусть в рассматриваемом примере будет выводиться максимально полный набор результатов.

Наконец, на вкладке «Вывод» (рис. 5.51) имеет смысл выбрать вариант создания нового листа, поскольку количество выводимых результатов достаточно велико.

В результате получается график, на котором отмечены точки с цензурированными данными для обеих групп (рис. 5.52), а также вы-

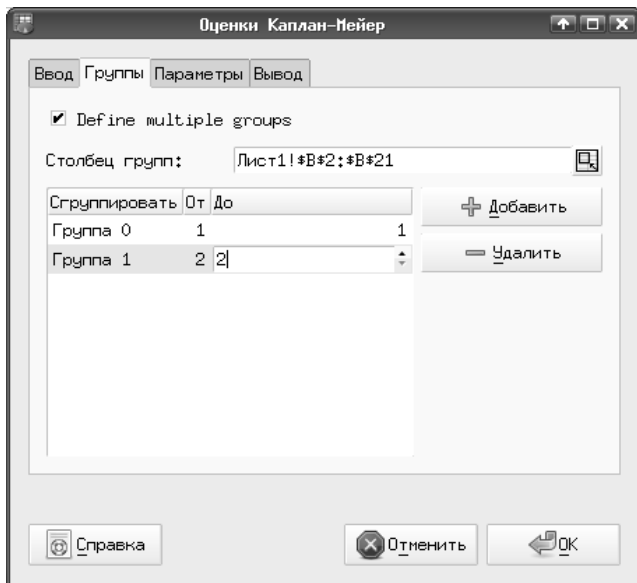


Рис. 5.49. Настройка групп исследуемых объектов

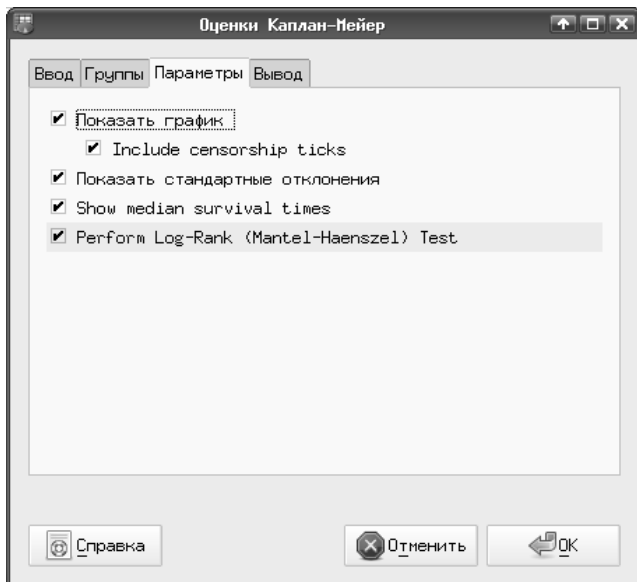


Рис. 5.50. Настройка результатов анализа

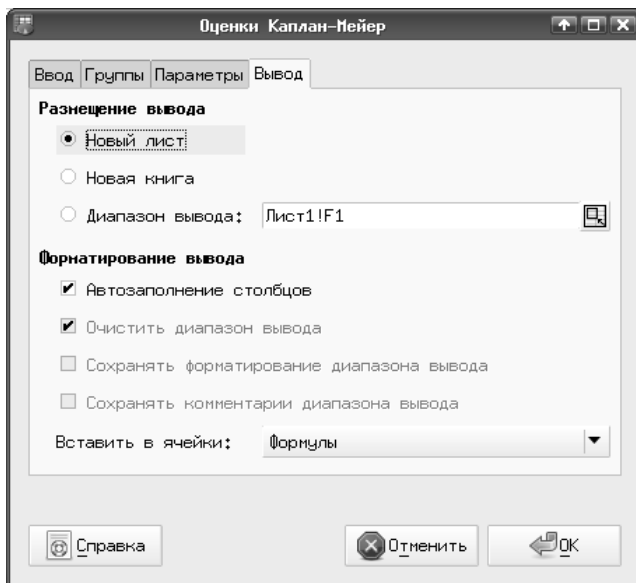


Рис. 5.51. Определение расположения результатов анализа

даются численные результаты. На рис. 5.52 результаты для первой группы показаны сплошной линией, цензурированные точки — треугольниками, а результаты для второй группы — «точечной» линией, цензурированные точки — ромбы.

Численные результаты для первой группы показаны на рис. 5.53. Наличие деления на 0 при времени в 19 единиц, видимо, связано с тем, что для первой группы («Группа0») нет цензурированных данных для такого «времени жизни».

В следующих столбцах располагаются результаты для второй группы («Группа1»). Для получения иллюстрации столбцы таблицы от В до F были скрыты (рис. 5.54).

Наконец, общее сравнение среднего времени выживаемости в группах обеспечивается тестом Log-Rank (рис. 5.55).

Значение p позволяет оценить различие среднего времени жизни по группам. На основании полученной в рассматриваемом примере величины p делается вывод, что эти значения статистически различимы.

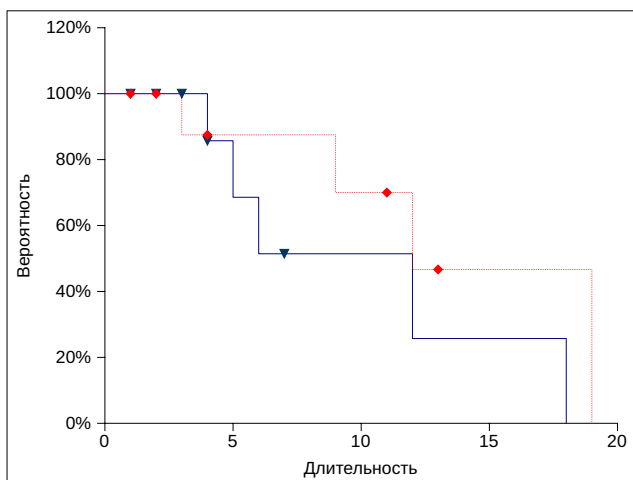


Рис. 5.52. Результаты анализа (функция выживания)

	А	В	С	Д	Е	Ф	
1	Каплан-Мейер	<i>Группа 0</i>					
2	<i>Время</i>	<i>At Risk</i>	<i>Deaths</i>	<i>Censures</i>	<i>Probability</i>	<i>Стандартное отклонение</i>	
3		0	10	0	0	100,00%	0,0000
4		1	10	0	1	100,00%	0,0000
5		2	9	0	1	100,00%	0,0000
6		3	8	0	1	100,00%	0,0000
7		4	7	1	1	85,71%	0,1224
8		5	5	1	0	68,57%	0,1719
9		6	4	1	0	51,43%	0,1792
10		7	3	0	1	51,43%	0,2069
11		9	2	0	0	51,43%	0,2534
12		11	2	0	0	51,43%	0,2534
13		12	2	1	0	25,71%	0,1567
14		13	1	0	0	25,71%	0,2216
15		18	1	1	0	0,00%	0,0000
16		19	0	0	0	#Деление на 0!	#Деление на 0!

Рис. 5.53. Численные результаты для первой группы

	A	G	H	I	J	K
1	Каплан-Мейер	<i>Группа 1</i>				
2	<i>Время</i>	<i>At Risk</i>	<i>Deaths</i>	<i>Censures</i>	<i>Probability</i>	<i>Стандартное отклонение</i>
3	0	10	0	0	100,00%	0,0000
4	1	10	0	1	100,00%	0,0000
5	2	9	0	1	100,00%	0,0000
6	3	8	1	0	87,50%	0,1094
7	4	7	0	2	87,50%	0,1169
8	5	5	0	0	87,50%	0,1383
9	6	5	0	0	87,50%	0,1383
10	7	5	0	0	87,50%	0,1383
11	9	5	1	0	70,00%	0,1715
12	11	4	0	1	70,00%	0,1917
13	12	3	1	0	46,67%	0,1968
14	13	2	0	1	46,67%	0,2410
15	18	1	0	0	46,67%	0,3408
16	19	1	1	0	0,00%	0,0000

Рис. 5.54. Хисленные результаты для второй группы

	M	N	O
		<i>Группа 0</i>	<i>Группа 1</i>
<i>Медиана:</i>		12	12
<i>Log-Rank Test:</i>			
<i>Statistics:</i>		1,00696511856098	
<i>Degrees of Freedom:</i>		1	
<i>p-Value:</i>		0,31563100206861	

Рис. 5.55. Общая статистика по группам

Глава 6

Регрессионный анализ в Gnumeric

6.1 Небольшое теоретическое введение

Всегда полезно знать, что и почему вычисляется в той или иной задаче. Поэтому сначала рассмотрим некоторые теоретические основы регрессионного анализа.

Линейный парный регрессионный анализ заключается в определении параметров эмпирической линейной зависимости (1), описывающей связь между некоторым N числом пар значений x_i и y_i , обеспечивая при этом наименьшую среднеквадратическую погрешность (метод наименьших квадратов).

$$y(x) = a \cdot x + b \quad (6.1)$$

Графически это выглядит как проведение прямой в «облаке» точек с координатами x_i, y_i так, чтобы величина всех отклонений между значениями y на этой прямой при имеющихся значениях x_i и координатами y_i имеющихся точек отвечала условию (6.2).

$$U = \sum_{i=1}^N (y_i - y(x_i))^2 \rightarrow \min \quad (6.2)$$

где $y(x_i)$ — теоретическая зависимость (6.1). Для этого нужно приравнять к нулю частные производные (6.3 и 6.4).

$$\frac{\partial U}{\partial b} = \sum_{i=1}^N (y_i - (b + a \cdot x_i)) \quad (6.3)$$

$$\frac{\partial U}{\partial a} = \sum_{i=1}^N (y_i - (b + a \cdot x_i)x_i) \quad (6.4)$$

Тогда для определения коэффициентов линейной регрессии a и b получаем систему уравнений (6.5).

$$\begin{cases} b \cdot N + a \cdot \sum_{i=1}^N x_i = \sum_{i=1}^N y_i \\ b \cdot \sum_{i=1}^N x_i + a \cdot \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i \cdot y_i \end{cases} \quad (6.5)$$

Решение этой системы даётся соотношениями 6.6 и 6.7.

$$a = \frac{\sum_{i=1}^N x_i \cdot y_i - \sum_{i=1}^N y_i \cdot \sum_{i=1}^N x_i / N}{\left(\sum_{i=1}^N x_i\right)^2 / N - \sum_{i=1}^N x_i^2} \quad (6.6)$$

$$b = \frac{1}{N} \cdot \left(\sum_{i=1}^N y_i - a \cdot \sum_{i=1}^N x_i \right) \quad (6.7)$$

Для определения отклонения связи между x_i и y_i от линейной используется коэффициент парной корреляции (6.8).

$$R = \frac{\sum_{i=1}^N x_i \cdot y_i - \left(\sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i\right) / N}{\sqrt{\frac{\sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i\right)^2 / N}{N}} \cdot \sqrt{\frac{\sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i\right)^2 / N}{N}}} \quad (6.8)$$

Если экспериментальная зависимость явно нелинейная, для её интерполяции (аппроксимации) применяются различные нелинейные зависимости (экспоненциальная, степенная с положительными или отрицательными показателями степени, полиномиальные различных порядков и пр.). При этом интерполяционная функция «линеаризуется», т.е. сводится к виду (6.1) путём замены переменных. Соответственно пересчитываются значения экспериментальных точек и коэффициент парной корреляции показывает успешность этого преобразования. Поскольку знак коэффициента парной корреляции при оценке качества линеаризации не является существенным, часто используется значение R^2 .

6.2 Реализация вычислений на модели

Вычисление параметров линейной регрессии уже рассматривалось в главе «Инструменты Gnumeric для статистиков», поэтому здесь рассмотрим подробнее процесс добавления и настройки параметров линий регрессии на график с экспериментальными данными. В качестве исходных данных используем таблицу, которая уже применялась в главе про статистику при описании инструментов предсказания и регрессии (рис. 6.1).

X	Y
1	80,54
2	54,21
3	51,01
4	25,26
5	18,43
6	13,11
7	12,75
8	9,07
9	6,4
10	4,43
11	3,39
12	2,16
13	1,7
14	1,14
15	0,65

Рис. 6.1. Исходные данные

На рис. 6.2 показан график с исходными данными (круглые точки). Поскольку линейная регрессия для таких данных, очевидно, даёт плохие результаты, будем пытаться использовать нелинейные модели. Тогда этот процесс можно будет называть "non-linear fitting" — «нелинейная подгонка».

Для добавления кривых регрессии вызовем диалог настройки графика, выберем серию исходных данных (Y) и используем кнопку «Добавить» для выбора добавляемого объекта (рис. 6.3).

Во вложенном меню «Линия тренда to Y» имеется набор классов кривых (уравнений интерполяции, рис. 6.4).

Заметим, что в списке вариантов присутствуют «Экспоненциально сглаженная кривая» и «Скользящее среднее», которые рассматривались в главе про статистику.

В качестве первой попытки описания экспериментальных данных выберем вариант интерполяции полиномом («Полиномиальная»¹) третьего порядка (рис. 6.5).

Поле «Порядок» позволяет выбрать максимальную степень аргумента (порядок) в полиноме, а на вкладке «Стиль» можно настроить внешний вид линии.

Для того чтобы узнать коэффициенты полинома ещё раз нажмём кнопку «Добавить» и увидим, что в списке объектов появился объект «Уравнение to Полиномиальная регрессия¹», как показано на рис. 6.6.

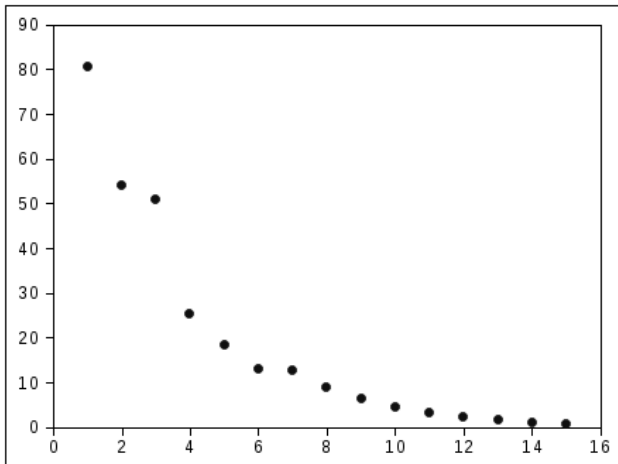


Рис. 6.2. График исходных данных

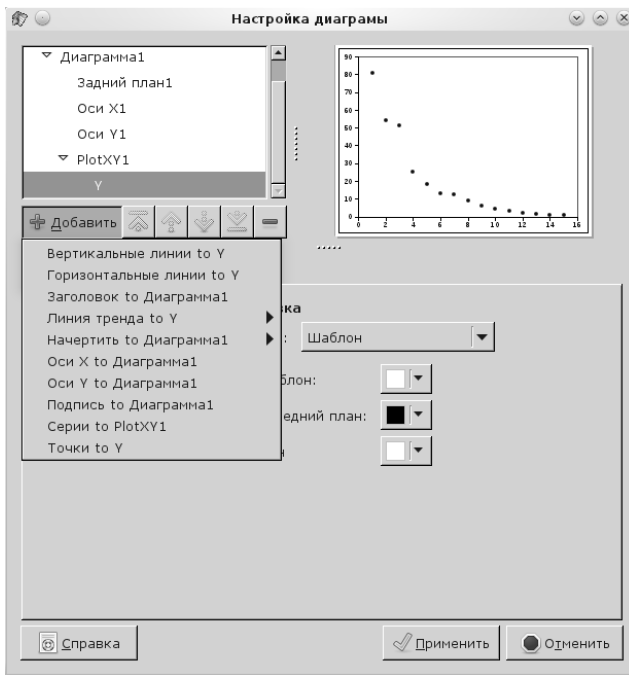


Рис. 6.3. Выбор добавляемого объекта

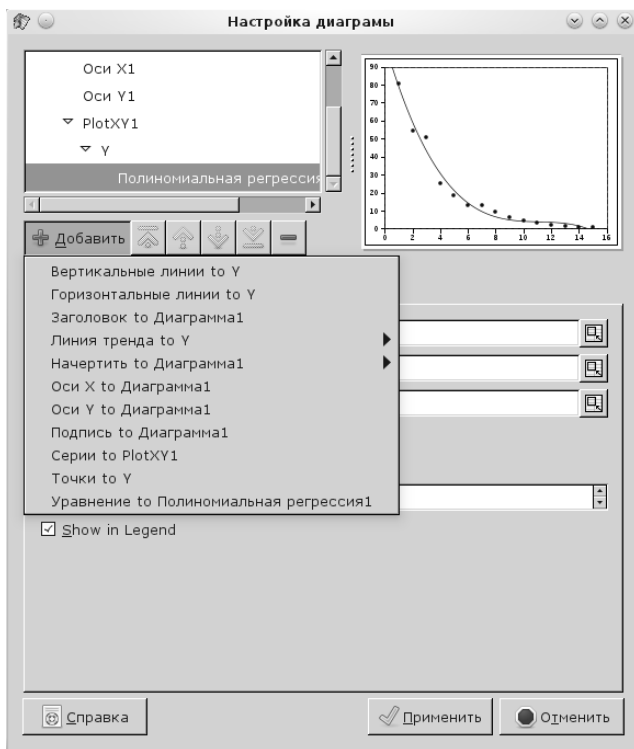


Рис. 6.6. Изменение списка добавляемых объектов в зависимости от выбранного объекта графика

Добавляемое на график уравнение кривой имеет собственный диалог настроек (рис. 6.7).

Режим «Показывать коэффициент регрессии R^2 » ("Display regression coefficient R^2 ") позволяет вывести под уравнением значение коэффициента парной корреляции, характеризующего «качество» интерполяции. Чем ближе это значение к 1, тем лучше подобрано уравнение регрессии.

На вкладке «Позиция» можно задать желаемое место уравнения на графике в относительных единицах (рис. 6.8).

На вкладках «Стиль» и «Шрифт» задаётся стиль оформления области с уравнением и шрифт для отображения уравнения.

Для сравнения добавим интерполяцию степенной функцией («Степенная»), установив стиль линии «точки» и толщину в две точки экрана (рис. 6.9). В этом случае модель имеет вид $y(x) = A \cdot x^b$, а на графике отображается линеаризованный вариант уравнения (через натуральный логарифм).

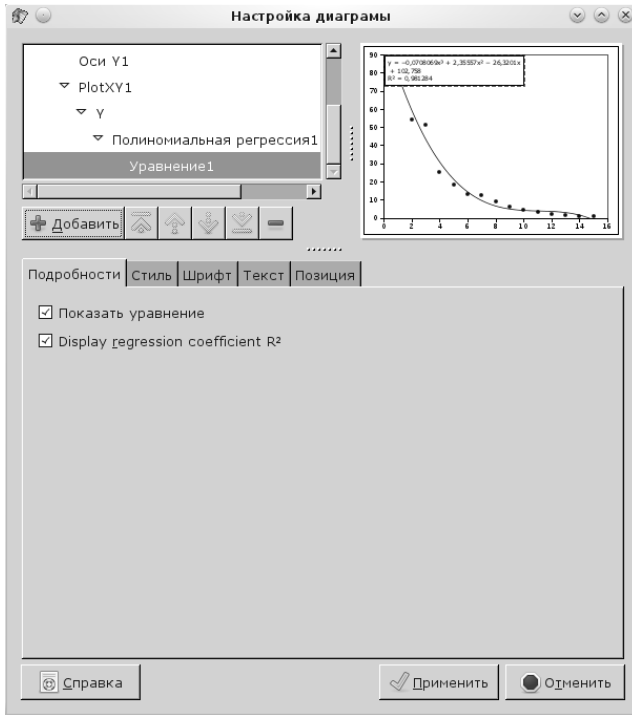


Рис. 6.7. Добавление уравнения кривой на график

Теперь можно пробовать другие варианты функций и следить за значением критерия R^2 . Наилучшим описанием будет такое, при котором это значение, как уже упоминалось, будет максимально близко к 1.

Вариант подгонки экспоненциальной зависимостью вида $y(x) = A * e^{bx}$ уже был показан в главе про статистику.

Таким образом, использование Gnumeric для подгонки экспериментальных данных даёт неплохие результаты для не очень сложных зависимостей и позволяет избежать использования громоздких и дорогостоящих математических пакетов программ.

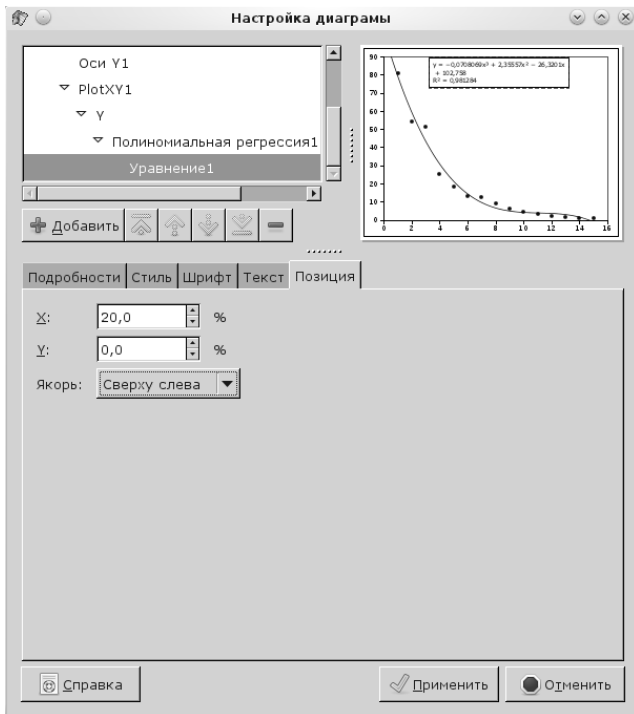


Рис. 6.8. Настройка расположения уравнения

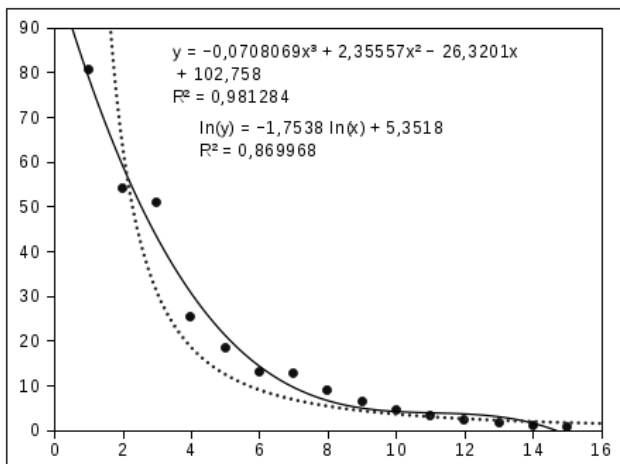


Рис. 6.9. Настройка расположения уравнения

Глава 7

Линейная оптимизация (поиск решения)

7.1 Оптимизация как задача линейного программирования

Всегда полезно знать, что и почему вычисляется в той или иной задаче. Поэтому сначала рассмотрим некоторые теоретические основы регрессионного анализа.

Пусть имеется функция, называемая целевой, линейно зависящая от некоторых переменных (факторов).

$$f(X) = c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n \quad (7.1)$$

В данном случае X — вектор неизвестных (значений факторов).

$$X = (x_1, x_2, \dots, x_n) \quad (7.2)$$

Значения неизвестных находятся из системы m линейных ограничений, которая может содержать как уравнения, так и неравенства.

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \geq b_2 \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases} \quad (7.3)$$

При этом вводится дополнительное условие неотрицательности значений переменных.

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 \quad (7.4)$$

Неотрицательное решение системы 7.3, то есть вектор X при соблюдении условий (7.3) и (7.4), называется «планом задачи».

Таблица 7.1. Таблица исходных данных для задачи

Продукция Сырьё	Продукт 1	Продукт 2	Запасы сырья
1	5	9	45
2	3	9	19
3	2	1	10
ПРИБЫЛЬ	5	6	

План X' называется оптимальным планом задачи максимизации (минимизации), если для любого плана X выполняется условие

$$f(X') \geq f(X) \quad (f(X') \leq f(X)) \quad (7.5)$$

Таким образом, поскольку решение задачи линейного программирования сводится к нахождению оптимального плана и вычислению максимального (минимального) значения целевой функции, такие задачи также называют задачами линейной оптимизации.

В современных офисных электронных таблицах для решения подобных задач имеется модуль "Solver" («Поиск решения»). Далее рассмотрим пример решения задачи линейной оптимизации в Gnumeric.

7.2 Пример задачи линейной оптимизации

Для разбора предлагается задача, взятая из учебного пособия «Математическое программирование», авторы Э.Ф.Брыжина и Э.А.Худобина, издательство СПбГИЭА, 1997 год.

Задача: Имеется три вида сырья в количествах 45 ед., 19 ед. и 10 ед. Из этого сырья нужно изготовить продукцию двух видов. Задан расход сырья каждого вида на производство единицы каждого вида продукции и прибыль от единицы продукции (см. таблицу). Требуется найти такой вариант выпуска каждого вида продукции, при котором прибыль будет наибольшей.

Обозначим через x_1 количество единиц продукции первого вида, через x_2 — второго вида. Тогда на выпуск этой продукции будет израсходовано $5x_1 + 9x_2$ ед. сырья первого вида, $3x_1 + 3x_2$ ед. сырья второго вида и $2x_1 + x_2$ — третьего. Суммарная прибыль составит $5x_1 + 6x_2$ денежных единиц. Т.к. нельзя израсходовать сырья больше, чем имеется, а суммарная прибыль зависит от количества выпущенной продукции, то получим следующую математическую модель данной задачи.

	A	B	C	I
1				
2	Продукт1	Продукт2	Прибыль	
3			0	
4				
5	Ограничения:			
6	для сырья1	0		
7	для сырья2	0		
8	для сырья3	0		

Рис. 7.1. Исходные данные для поиска решения

$$\begin{cases}
 5 \cdot x_1 + 9 \cdot x_2 \leq 45, \\
 3 \cdot x_1 + 3 \cdot x_2 \leq 19, \\
 2 \cdot x_1 + x_2 \leq 10, \\
 x_1 \geq 0, x_2 \geq 0
 \end{cases} \quad (7.6)$$

$$f(x) = 5 \cdot x_1 + 6 \cdot x_2 \rightarrow \max$$

7.3 Решение в Gnumeric

Для решения задачи в ЭТ прежде всего нужно определиться с расположением исходных данных и составить необходимые формулы. Пусть таблица будет скомпонована в соответствии с рис. 7.1).

Количества продукции 1-го и 2-го вида (которые предстоит вычислить) располагаются в ячейках A3 и B3 соответственно. В ячейку C3 записывается формула для вычисления прибыли, в ячейки B6:B8 — формулы для левых частей ограничений (см. математическую модель). Когда решение будет найдено, нули будут заменены какими-то другими значениями.

Далее вызываем модуль поиска решения (в главном меню «Сервис/Поиск решения...», рис. 7.2).

В диалоге поиска решения несколько вкладок, и в них потребуется устанавливать значения элементов ввода. На вкладке «Параметры» (рис. 7.2) требуется указать ячейку, содержащую формулу для целевой функции $f(x)$, а также ячейки, в которых должны вычисляться параметры (в нашем случае x_1 и x_2). Заполним эту вкладку в со-

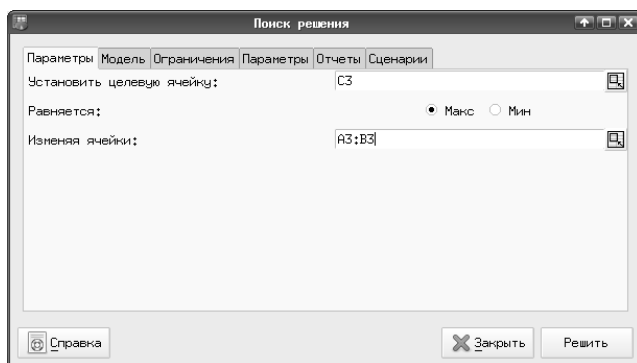


Рис. 7.2. Диалог поиска решения в Gnumeric (Solver)

ответствии с расположением исходных данных. Поскольку требуется найти максимум прибыли, установим переключатель «Равняется:» в позицию «Макс».

Адреса ячеек, содержащих целевую функцию и значения факторов, можно вписывать вручную, а можно указывать мышью, используя кнопку «свернуть/развернуть диалог» справа от полей ввода.

На следующей вкладке («Модель») устанавливаем значения в соответствии с рис. 7.3. Ключ «Предполагать неотрицательность» как раз и означает, что искать нужно только положительные решения. В самом деле, количество продукции не может быть отрицательным.

Выбор алгоритма и вида модели влияет на результат и саму возможность существования решения, так что для задач линейного программирования нужно использовать установки, показанные на рис. 7.3.

Следующая вкладка («Ограничения») позволяет задать ограничения для параметров модели. Здесь важно, что левая и правая часть ограничений задаются как адреса ячеек, содержащих формулы и значения. После того, как определены ячейка для левой части и значение для правой части ограничения и выбран знак отношения («Тип»), нажатием на кнопку «Добавить» условие добавляется (рис. 7.4).

Для правой части ограничений можно использовать как конкретные значения, так и адреса ячеек, содержащих эти значения.

Остальные вкладки не играют существенной роли в решении задачи, поэтому после ввода всех ограничений нажимаем на кнопку «Решить» и получаем сообщение о результате решения (рис. 7.5). Если чего-то не хватает или выбраны неверная модель или алгоритм, выводится соответствующее сообщение об ошибке.

Значение "Objective value" соответствует полученному в результате решения значению целевой функции, а вычисленные значения пе-

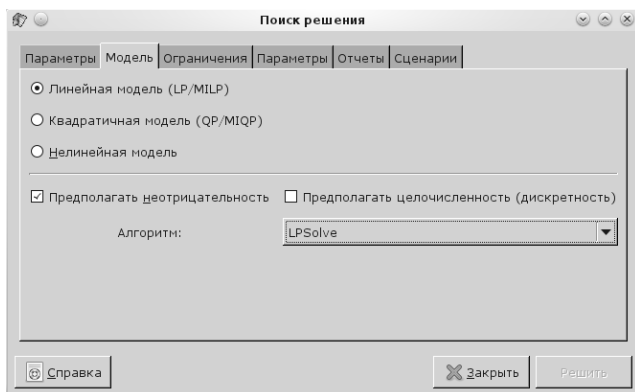


Рис. 7.3. Определение вида модели

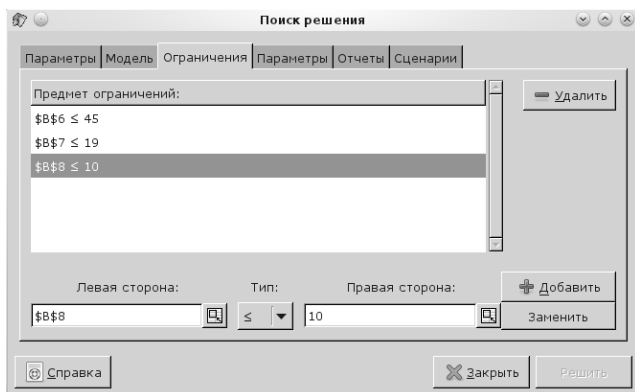


Рис. 7.4. Определение ограничений

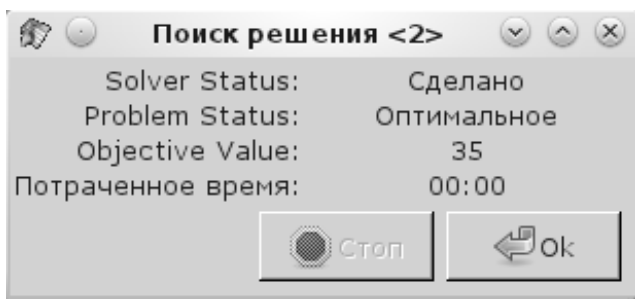


Рис. 7.5. Результат решения

	А	В	С	И
1				
2	Продукт1	Продукт2	Прибыль	
3	3	3,33333	34,99998	
4				
5	Ограничения:			
6	для сырья1	44,99997		
7	для сырья2	18,99999		
8	для сырья3	9,33333		

Рис. 7.6. Найденное решение

ременных и результирующие значения ограничений появляются вместо исходных данных на листе таблицы, как показано на рис. 7.6.

Нужно отметить, что результаты полностью соответствуют приведенным в первоисточнике значениям 3 и 10/3.

В качестве эксперимента попробуйте в диалоге настройки модели для этих же исходных данных установить режим «Предполагать целочисленность (дискретность)».

Глава 8

Подбор параметра

Существует класс задач, связанных с численным подбором решения какого-либо уравнения. Часто эти уравнения являются трансцендентными, для которых не существует аналитического решения. В этом случае решение ищется путём поиска корня уравнения по некоторому алгоритму.

В качестве простого примера рассмотрим поиск корней уравнения, описанного формулой 8.1.

$$e^x = x + 2 \quad (8.1)$$

Можно привести это уравнение к «каноническому» виду (формула 8.2).

$$x + 2 - e^x = 0 \quad (8.2)$$

Таким образом изменяя значение x по каким-то правилам, нужно найти такое значение, при котором левая часть уравнения станет равна 0 с «машинной» точностью (поскольку вещественные числа в компьютере никогда не могут быть в точности равны).

Прежде всего нужно убедиться в том, что решение существует. Для этого достаточно с помощью любой программы, позволяющей рисовать графики функций (например, средствами `Grapher`) получить на одном рисунке графики левой и правой части первоначального уравнения (рис. 8.1).

Из графиков видно, что решение существует, более того, существует два решения. Теперь рассмотрим применение `Grapher` для поиска корней нашего уравнения.

Сначала сформируем таблицу исходных данных, в которой укажем какое-то начальное значение искомой функции (для «канонического» вида уравнения), и введём формулу для функции (рис. 8.2).

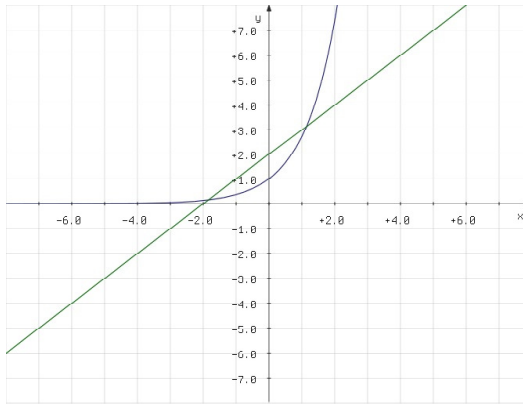


Рис. 8.1. Графики исходных функций

=C2+2-EXP(C2)			
	B	C	D
Параметр		1	
Уравнение		0,2817181715	EXP(x)=x+2

Рис. 8.2. Исходные данные для поиска решения

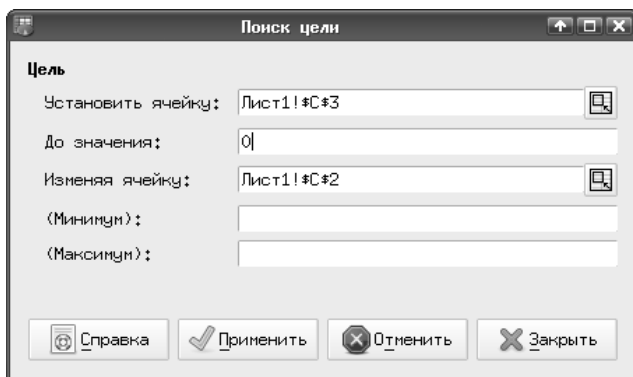


Рис. 8.3. Настройка подбора параметра

В ячейке C2 записано начальное значение параметра x , в C3 — формула для вычисления значения функции, а в D3 — исходное уравнение для справки.

Задача состоит в том, чтобы подобрать такое значение в C2, чтобы результат в C3 стал равным 0.

Для решения задачи используем инструмент «Подбор параметра...», выбрав с помощью главного меню команду «Сервис/Подбор параметра...» (рис. 8.3).

В строке «Установить ячейку:» указываем адрес ячейки, в которой записана формула функции, в строке «Изменяя ячейку:» указываем адрес ячейки, содержащей значение параметра x , а в строке «До значения:» указываем целевое значение функции.

Если не задавать значения минимум и максимум параметра при поиске решения, программа найдёт только одно из возможных решений (если их несколько). Как видно из рис. 8.1, одно решение находится в области положительных значений x , другое — в области отрицательных значений.

Нажатие на кнопку «Применить» запустит процесс поиска и в этом же окне будет показан результат (рис. 8.4).

Чтобы получить второе решение, нужно определить интервал значений x для поиска и снова запустить процесс (рис. 8.5).

Вообще говоря, достаточно указать только максимальное значение (например, искать все значения меньше 0), но для определённости зададим интервал значений от -10 до 0.

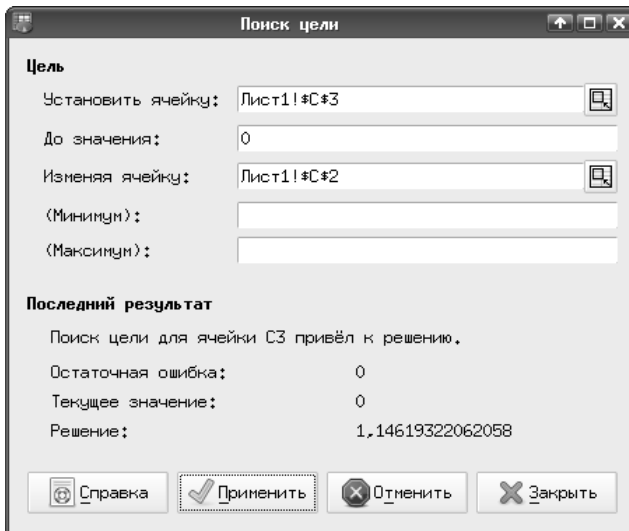


Рис. 8.4. Результат подбора параметра

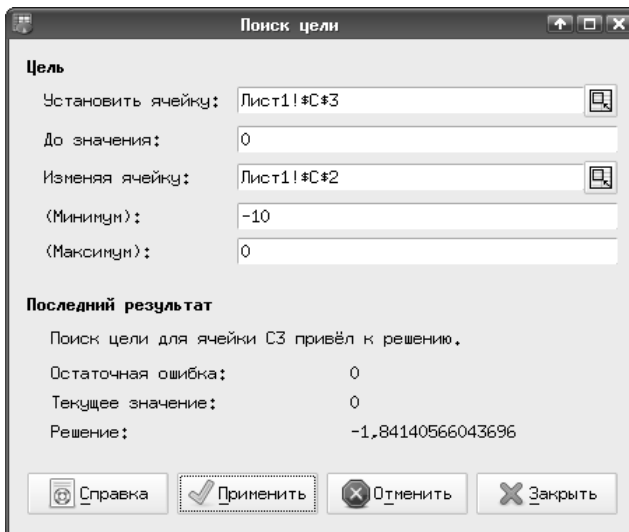


Рис. 8.5. Решение для отрицательных значений параметра

«Подбор параметра» в `Grapher` — очень простой в использовании инструмент, но для получения правильных результатов нужно соблюдать некоторые правила:

1. Предварительно убедиться в том, что решение существует;
2. Если решение не единственное, определить, какое из решений требуется по условию задачи
3. Интервалы для значений параметра выбирать так, чтобы в каждом интервале было только одно решение.

Глава 9

Моделирование рисков методом Монте-Карло

В состав Gnumeric входит средство для моделирования по методу Монте-Карло. При использовании метода Монте-Карло используются выборки случайных чисел для решения задач, в которых течение времени не играет существенной роли. Другими словами, на очередную выборку не влияют предыдущие выборки.

При моделировании по методу Монте-Карло используются функции генерации случайных чисел, а результаты представляются вместе со статистическими характеристиками, которые можно затем анализировать.

В данной главе рассмотрен пример из официального руководства по Gnumeric.

9.1 Общее описание задачи

Одна из классических расчётных задач — задача о продавцах газет. Продавцы покупают газеты за 33 цента каждую и продают по 50 центов. Непроданные газеты идут на макулатуру по 5 центов за штуку. Газеты продаются распространителям пачками по 10 штук. Спрос на газеты может быть поделён на «замечательный», «нормальный» и «плохой» с вероятностями 0.35, 0.45 и 0.20 соответственно, причём текущий спрос не зависит от предыдущего дня. Задача продавца — определить оптимальное количество газет в ситуации, когда спрос не вполне известен, то есть добиться устойчивого дохода.

Уравнение для определения дневного дохода для продавца выглядит следующим образом:

$$\text{Доход} = [(\text{Выручка}) - (\text{Себестоимость}) + (\text{Макулатура})]$$

	А	В	С	Д	Е
9	Вычисление дохода				
10					
11	Переменные	выручка	себестоимость	макулатура	доход
12	Значение	25	16,5	0	
13	Коэффициент	0,5	0,33	0,05	
14					
15	Данные	Количество			
16	Закупка	50			

Рис. 9.1. Исходная таблица вычисления дохода

Остаётся добавить, что количество закупленных от поставщика газет может изменяться от 40 до 100 включительно, а количество проданных газет также кратно 10.

9.2 Построение модели

Для построения модели в Gnumeric будем использовать два листа — лист «Доход» для вычисления дохода и лист «Таблицы спроса» для таблиц, требуемых для модельных наборов данных, задающих параметры спроса.

На листе «Доход» создадим таблицу расчёта дохода, как показано на рис 9.1.

Таблицу для вычисления дохода начнём с девятой строки. У нас есть три переменные — выручка от продаж, себестоимость газет и стоимость макулатуры, для которых на каждую единицу товара заданы коэффициенты 0.5, 0.33 и 0.05 соответственно. Запишем эти коэффициенты в ячейки от B13 до D13. В ячейках от B12 до D12 запишем формулы для дохода от продаж, себестоимости и стоимости макулатуры, как показано в таблице 9.1. В ячейке E12 запишем формулу для вычисления прибыли.

	А	В	С	Д	Е
1					
2					
3	Итог дня	Вероятность	Интегральная вероятность		
4	Занчительно	0,35	0,35		
5	Нормально	0,45	0,8		
6	Плохо	0,2	1		
7					

Рис. 9.2. Распределение уровней спроса

Таблица 9.1: Формулы для вычисления дохода

Адрес ячейки	Значение или формула
B12	=B\$13*min(B16;B20)
C12	=C13*B16
D12	=D13*max(0;B16-B20)
E12	=B12-C12+D12
B13	0,5
C13	0,33
D13	0,05
B16	50

Нужно заметить, что на этом этапе в некоторых ячейках появятся сообщения "N/A!" («нет данных»). Как только модель будет построена полностью, эти сообщения исчезнут.

В ячейке B20 будет задаваться случайное количество проданных газет («спрос»). Поскольку нельзя продать больше, чем закуплено у поставщика, выручка определяется количеством проданных газет, если закуплено больше, чем продано и ограничивается количеством закупленных газет, если спрос превышает это количество (функция `min()` при расчёте выручки).

Формула в ячейке D12 означает, что в макулатуру можно сдать только непроданные газеты, поэтому если всё продано (а также если спрос превышает количество закупленных газет), то количество макулатуры будет 0.

Начальное количество закупленных газет установим в 50.

Далее на листе «Таблицы спроса» сформируем модельные параметры спроса в соответствии с рис. 9.2, 9.3 и 9.4.

Термин «вероятность» в рассматриваемом примере означает значение функции плотности распределения, а «интегральная вероятность» — значение функции распределения.

	A	B	C	D	
9					
10	Распределение ежедневного спроса на газеты				
11	Спрос	Замечательно	Нормально	Плохо	
12	40	0,03	0,1	0,44	
13	50	0,05	0,18	0,22	
14	60	0,15	0,4	0,16	
15	70	0,2	0,2	0,12	
16	80	0,35	0,08	0,06	
17	90	0,15	0,04	0	
18	100	0,07	0	0	
19					

Рис. 9.3. Распределение спроса в зависимости от количества газет

	A	B	C	D	E	F	G	
19								
20	Функции распределения спроса							
21		Интегральная вероятность			Значения			
22	Спрос	Замечательно	Нормально	Плохо	Замечательно	Нормальнс	Плохо	
23	40	0,03	0,1	0,44	0	0	0	
24	50	0,08	0,28	0,66	0,03	0,1	0,44	
25	60	0,23	0,68	0,82	0,08	0,28	0,66	
26	70	0,43	0,88	0,94	0,23	0,68	0,82	
27	80	0,78	0,92	1	0,43	0,88	0,94	
28	90	0,93	1		0,78	0,92	1	
29	100	1			0,93	1		
30								

Рис. 9.4. Функции распределения спроса

На рис. 9.2 показаны плотность распределения и значения функции распределения уровней спроса, на рис. 9.3 — функции плотности распределения вероятности продажи заданного количества газет для каждого из вариантов спроса, а на рис. 9.4 — вспомогательные функции распределения вероятностей спроса.

Дополнительными допущениями является то, что 40 газет будут проданы при любых условиях, а вот 100 — только при наиболее благоприятных обстоятельствах.

Теперь снова перейдём на лист «Доход» и продолжим ввод формул, нужных для работы модели. Адреса ячеек и соответствующие формулы показаны в таблице 9.2.

Таблица 9.2: Адреса ячеек и формулы

Адрес ячейки	Значение или формула
B17	=rand()
C17	=if(B17<'Таблицы спроса'!C4;"Замечательно"; if(B17<'Таблицы спроса'!C5;"Нормально"; "Плохо"))
B18	=rand()
B20	=lookup(C17;\$B\$23:\$D\$23;\$B\$24:\$D\$24)
B21	=E12
B23	Замечательно
C23	Нормально
D23	Плохо
B24	=lookup(\$B\$18;'Таблицы спроса'!\$E\$23:\$E\$29; 'Таблицы спроса'!\$A\$23:\$A\$29)
C24	=lookup(\$B\$18;'Таблицы спроса'!\$F\$23:\$F\$29; 'Таблицы спроса'!\$A\$23:\$A\$29)
D24	=lookup(\$B\$18;'Таблицы спроса'!\$G\$23:\$G\$29; 'Таблицы спроса'!\$A\$23:\$A\$29)

Случайное число в ячейке B17 определяет уровень (состояние) спроса, который выводится в ячейку C17. Случайное число в ячейке B18 определяет количества проданных газет для каждого уровня спроса (ячейки B24, C24 и D24). В соответствии с ранее заданным в C17 уровнем спроса в ячейке B20 получаем текущий спрос на газеты, из которого уже рассчитывается доход.

Итоговый вариант листа «Доход» должен выглядеть примерно так, как показано на рис. 9.5.

Нажимая на клавишу F9 («Правка/Пересчитать» в главном меню), можем наблюдать за изменением чисел и, соответственно, за изменением дохода.

	А	В	С	Д	Е
9	Вычисление	дохода			
10					
11	Переменные	выручка	себестоимость	макулатура	доход
12	Значение	25	16,5	0	8,5
13	Коэффициент	0,5	0,33	0,05	
14					
15	Данные	Количество			
16	Закупка	50			
17	Итог дня	0,6873233079327	Нормально		
18		0,3227079860445			
19	Результат				
20	Спрос	60			
21	Доход	8,5			
22					
23		Замечательно	Нормально	Плохо	
24		70	60	40	
25					

Рис. 9.5. Итоговый лист для вычисления дохода

9.3 Использование модели

Понятно, что наблюдать за изменениями чисел в ячейках ЭТ при пересчёте очень увлекательно, но не очень продуктивно. Инструмент моделирования в *Gluegic* («Сервис/Моделирование...» в главном меню) позволяет автоматически проделать большое число пересчётов и проанализировать получаемые результаты.

При вызове инструмента моделирования появляется диалог «Моделирование рисков», как обычно содержащий несколько вкладок.

Вкладка «Переменные» (рис. 9.6) обеспечивает определение диапазонов входных и выходных данных. Для рассматриваемого примера эти диапазоны должны быть определены в соответствии с рис. 9.6.

На вкладке «Параметры» (рис. 9.7) настраивается режим вычислений — определяется количество вариантов модели ("Rounds"), количество пересчётов (итераций) для каждого варианта и предел времени, в течение которого проводятся вычисления (в секундах). Предел времени нужен для предотвращения слишком длительной загрузки вычислительной системы. Если модель очень сложна или составлена некорректно, вычисления прекратятся по достижении этого предела времени.

Использование вариантов модели ("Rounds") будет рассмотрено позднее.

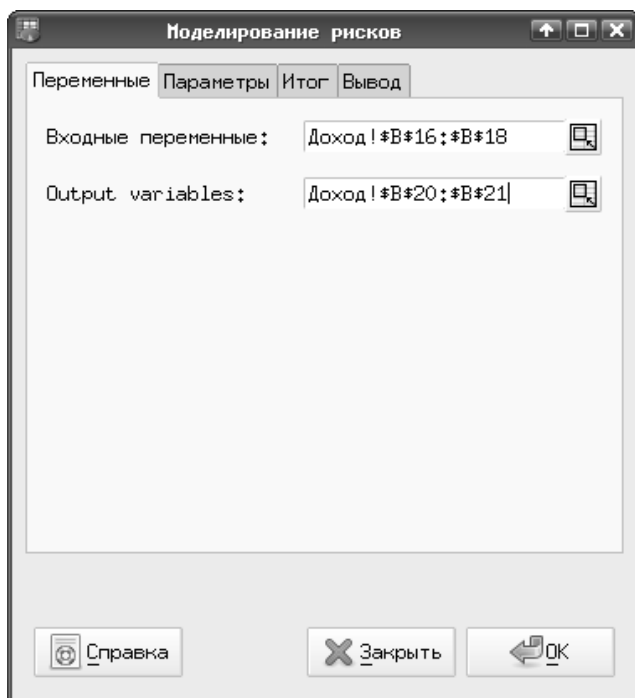


Рис. 9.6. Определение диапазонов для входных и выходных данных

Вкладка «Вывод» (рис. 9.8) является традиционной для Gnumeric и позволяет задать расположение результатов работы программы. В нашем случае целесообразно использовать вариант «Новый лист».

После нажатия на кнопку OK в диалоге «Моделирование рисков» диалоговое окно не закрывается, и на вкладке «Итог» можно увидеть, что появились какие-то результаты (рис. 9.9).

В области «Сводка результатов:» вкладки «Итог» можно увидеть значения переменных, полученные в результате выполнения указанного в области «Итог моделирования» количества итераций. Изучать эти значения в области вкладки диалогового окна не очень удобно, поэтому имеет смысл закрыть диалог и посмотреть результаты на появившемся в книге ЭТ листе «Отчёт о моделировании (1)» (рис. 9.10).

В правой части этого листа имеется ещё несколько статистических параметров, полученных в результате вычислений. Их можно получить, проделав все описанные выше действия, а назначение этих параметров (на английском языке) описано в оригинальном руководстве по Gnumeric, поэтому здесь они не приводятся.

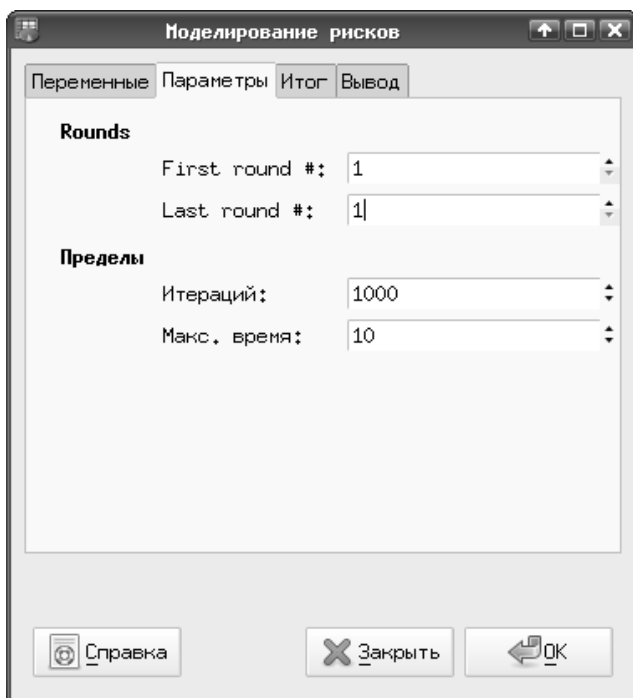


Рис. 9.7. Параметры вычислений

Полученный результат можно интерпретировать следующим образом: «Если продавец будет покупать 50 газет, то его доход будет варьироваться от 4 до 8,5 у.е. и в среднем будет составлять 7,825 у.е. Доход будет не менее 4 у.е. в самых неблагоприятных условиях, но не более 8,5 у.е. — в самых благоприятных».

Теперь можно проделать аналогичные действия, установив количество закупаемых от поставщика газет (ячейка B16 на листе «Доход») в 50, 60 и т.д. Однако процесс перебора количества закупаемых газет тоже можно автоматизировать. Для этого в Gnumeric имеется функция `simtable()`, находящаяся в категории «Случайные числа».

Для того, чтобы просчитать модель с различными количествами закупаемых газет, запишем в ячейку B16 формулу `=simtable(50;60;70;80;90)` (крайние случаи с количествами 40 и 100 брать не будем).

Поскольку в этом случае имеется 5 вариантов модели, то на вкладке настройки режима вычислений диалога «Моделирование рисков» требуется изменить значение "Last round #:" в соответствии с количеством аргументов функции `simtable()` (рис. 9.11).

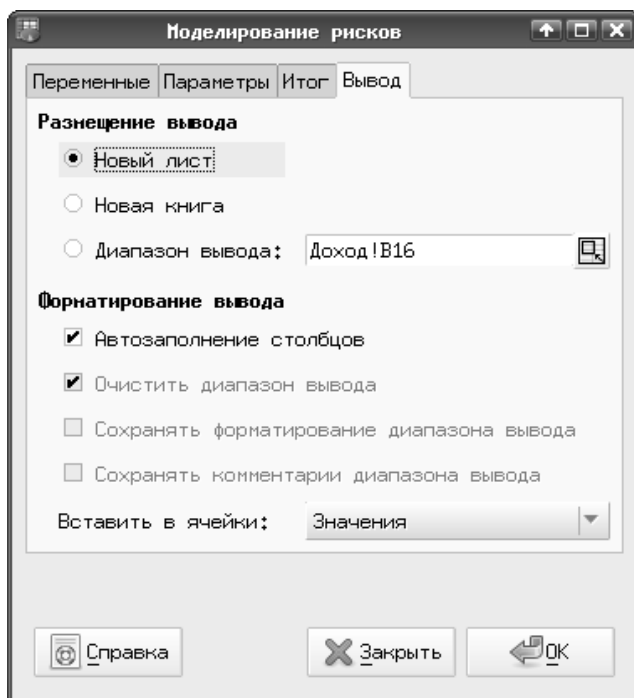


Рис. 9.8. Настройка расположения результатов моделирования

Следует заметить, что номера вариантов совпадают с номерами аргументов функции `simtable()`. Поэтому, если хочется посмотреть, что будет при количествах закупленных газет от 60 до 80, то в качестве параметра "First round #:" нужно поставить 2, а для "Last round #:" — 4.

После нажатия на кнопку OK на вкладке «Итог» окажутся активными кнопки "Next Sim." и "Prev. Sim." (рис. 9.12), которые позволяют в области «Сводка результатов» видеть результаты по каждому варианту модели (при каждом значении количества закупленных газет из набора значений, определённых в качестве аргументов функции `simtable()`).

Полная сводка результатов отображается на листе «Отчёт о моделировании (1)» (рис. 9.13), на котором поместился только фрагмент сводки результатов, поэтому основные результаты приведены в таблице 3.

Из таблицы 3 видно, что закупка 60 газет даёт наиболее надёжный доход — средний доход максимален и никогда не получается убытка.

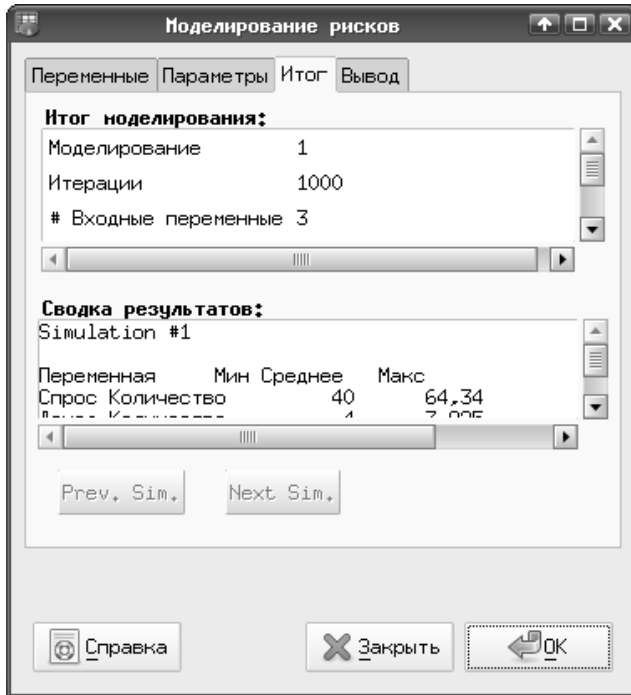


Рис. 9.9. Результаты моделирования в диалоговом окне

	A	B	C	D	E
1		gnumeric	Моделирование рисков 1.10.5	Отчет	
2		Книга: [file:///mnt/win_d/gnumeric-book/simulation.gnumeric]Доход			
3		Отчёт создан:Sun Oct 10 14:33:05 2010			
4					
5					
6		ИТОГО			
7			Мин	Среднее	Макс
8		Спрос Количество	40	64,34	100
9		Доход Количество	4	7,825	8,5
10		(Ввод) Закупка Количество	50	50	50
11		(Ввод) Итог дня Количество	5,702740021134E-05	0,50254637421886	0,9995027511702
12		(Ввод) Количество	0,00012427875973	0,50664222257099	0,9991930672988

Рис. 9.10. Фрагмент листа с отчётом о моделировании

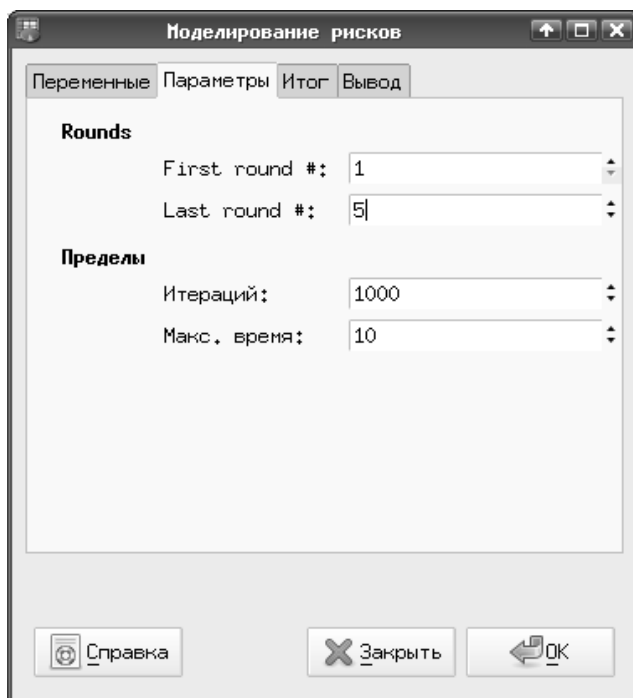


Рис. 9.11. Настройка вычислений для нескольких вариантов модели

Таблица 9.3. Результаты моделирования

Закуплено (шт.)	Доход, у.е.		
	Мин.	Средн.	Макс.
50	4	7,82	8,5
60	1,2	8,25	10,2
70	-1,6	7,36	11,9
80	-4,4	6,21	13,6
90	-7,2	3,91	15,3

При увеличении количества закупаемых газет растёт возможный максимальный доход, однако средний доход уменьшается и возрастает риск убытков.

Очевидно, что такой результат получается при параметрах модели, заданных на листе «Таблицы спроса». При других вероятностях и распределениях вероятностей результаты также будут отличаться.

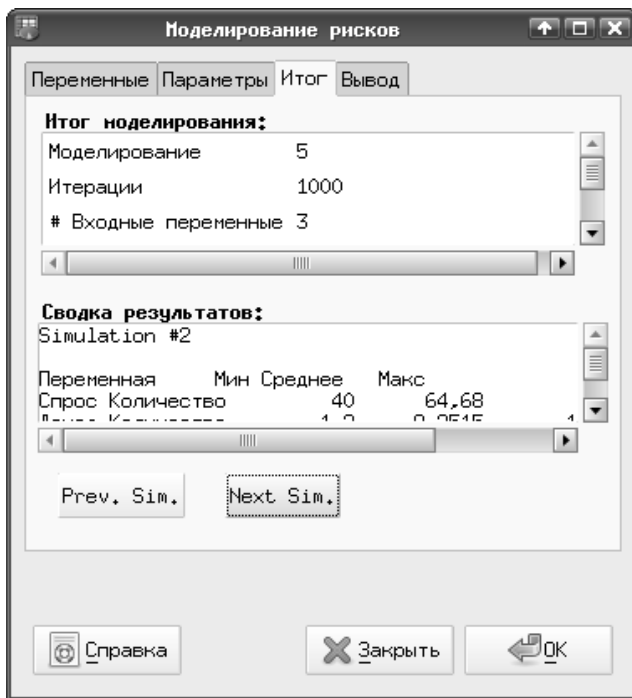


Рис. 9.12. Просмотр результатов по вариантам модели

В заключение этой главы хотелось бы заметить, что программные средства для экономического моделирования стоят тысячи долларов США даже при условии их использования в образовательных целях (а для использования в коммерческих целях — значительно дороже). Поэтому использование Gnumeric (да и вообще свободного программного обеспечения) и для такого рода задач может оказаться привлекательным.

	A	B	C	D	E
6	SUMMARY OF SIMULATION ROUND #1				
7			Мин	Среднее	Макс
8	Спрос	Количество		40	64,41
9	Доход	Количество		4	7,8205
10	(Ввод)	Закупка Количество		50	50
11	(Ввод)	Итог дня Количество	0,003352172062	0,498789753087	0,999636536864
12	(Ввод)	Количество	0,000578942701	0,499960621012	0,998962741086
13					
14					
15	SUMMARY OF SIMULATION ROUND #2				
16			Мин	Среднее	Макс
17	Спрос	Количество		40	64,68
18	Доход	Количество		1,2	8,2515
19	(Ввод)	Закупка Количество		60	60
20	(Ввод)	Итог дня Количество	0,000345595932	0,489356043302	0,998794125497
21	(Ввод)	Количество	0,001553605354	0,497441035969	0,999672653996
22					
23					
24	SUMMARY OF SIMULATION ROUND #3				
25			Мин	Среднее	Макс
26	Спрос	Количество		40	63,9
27	Доход	Количество		-1,6	7,364
28	(Ввод)	Закупка Количество		70	70
29	(Ввод)	Итог дня Количество	0,000806831029	0,500545577640	0,999709810391
30	(Ввод)	Количество	0,001793650310	0,495337517129	0,999865479445
31					
32					
33	SUMMARY OF SIMULATION ROUND #4				
34			Мин	Среднее	Макс
35	Спрос	Количество		40	65,19
36	Доход	Количество		-4,4	6,2065

Рис. 9.13. Результаты моделирования по вариантам входных данных

Список литературы

- [1] И. Хахаев, В. Машков, Г. Губкина и др. OpenOffice.org: Теория и практика. —М.: ALT Linux; БИНОМ. Лаборатория знаний. 2008.
- [2] В.Е. Гмурман. Теория вероятностей и математическая статистика. —9-е изд., стер. — М.: Высшая школа, 2003.
- [3] В.П. Бородюк, А.П. Вошинин, А.З. Иванов и др.; Под ред. Г.К. Круга. Статистические методы в инженерных исследованиях (лабораторный практикум). Учебное пособие. —М.: «Высшая Школа», 1983.
- [4] Ю.А.Горицкий, Е.Е.Перцов. Практикум по статистике с пакетами StatGraphics, Statistica, SPSS —<http://www.exponenta.ru/educat/systemat/goritskii/part2/LR7/2.asp>
- [5] А.Б.Мерков. Об анализе выживаемости —<http://www.recognition.mcsme.ru/pub/RecognitionLab.html/survival.pdf>
- [6] Портал «HR-Инструменты». Учебник по STATISTICA: ГЛАВА 14. Анализ выживаемости —<http://www.hr-portal.ru/statistica/g114/g114.php>
- [7] Г. Корн, Т. Корн. Справочник по математике для научных работников и инженеров. —М.: «Наука», 1984.
- [8] Э.Ф. Брыжина и Э.А. Худобина. Математическое программирование. Учебное пособие. —СПб.: СПбГИЭА, 1997.

Список иллюстраций

2.1	Общий вид окна ЭТ Gnumeric	9
2.2	Простой вид диалога открытия файла	10
2.3	Диалог открытия файла со строкой ввода имени и просмотром содержимого каталога	11
2.4	Диалог открытия файла с дополнительными возможностями выбора типа файла и кодировки	12
2.5	Результат импорта текстового файла	13
2.6	Диалог сохранения файла (полный вариант)	13
2.7	Контекстное меню листа ЭТ	16
2.8	Диалог управления листами в Gnumeric	16
2.9	Типовой набор цветов в Gnumeric	17
2.10	GTK-диалог выбора цвета	17
2.11	Результат управления листами	18
2.12	Вложенное меню «Формат/Лист»	19
2.13	Показ дополнительных свойств в диалоге управления листами ЭТ	19
2.14	Диалог изменения количества строк и столбцов листа ЭТ	20
2.15	Листы с различным количеством строк и столбцов в Gnumeric	20
2.16	Контекстное меню ячейки ЭТ	22
2.17	Настройка представления чисел в ячейках ЭТ	23
2.18	Настройка проверки правильности ввода данных в ячейку ЭТ	24
2.19	Сообщение об ошибке при неправильном вводе	25
2.20	Создание комментария для ячейки ЭТ	25
2.21	Просмотр комментария ячейки	26
2.22	Пример выделения отдельных символов в ячейке B4	27
2.23	Пример выделения непрерывного блока ячеек	27
2.24	Выделение нескольких произвольных ячеек	28
2.25	Результат выделения строки	28
2.26	Контекстное меню строки	28
2.27	Диалог настройки высоты строки	29
2.28	Выделение соседних строк	29

2.29	Выделение нескольких произвольных строк	30
2.30	Выделение столбца	30
2.31	Контекстное меню столбца	31
2.32	Диалог настройки ширины столбца	31
2.33	Пример недостаточной ширины столбца	31
2.34	Результат автоподбора ширины	32
2.35	Автоподбор ширины соседних столбцов	32
2.36	Выделение нескольких произвольных столбцов	32
2.37	Выделение всех ячеек листа	33
2.38	Пункт «Данные» главного меню Gnumeric	34
2.39	Вложенное меню «Заполнить»	34
2.40	Результаты использования автозаполнения	35
2.41	Настройка серий для заполнения ячеек данными	36
2.42	Результат заполнения сериями	37
2.43	Диалог генерации случайных чисел (однородное распределение)	38
2.44	Три вектора случайных чисел	39
2.45	Исходные данные для задачи о заработках	40
2.46	Результат расчёта заработка	41
2.47	Просмотр формулы в режиме редактирования	42
2.48	Таблица для вычислений с параметром	42
2.49	Неправильный результат вычислений с параметром	43
2.50	Правильные вычисления с параметром	43
2.51	Автосуммирование по столбцу	44
2.52	Диалог выбора функции (селектор функций)	45
2.53	Составление формулы)	46
2.54	Функция if () с аргументами)	47
2.55	Решение квадратного уравнения	51
2.56	Пример вычисление с датами	54
2.57	Справочник для задачи о ралли	55
2.58	Исходные данные задачи о ралли	55
2.59	Решение задача о ралли	56
2.60	Исходные данные задачи о налогах	57
2.61	Справочник к задаче о налогах	57
2.62	Решение задачи о налогах	58
2.63	Решение задачи о дискотеке	59
2.64	Решение задачи о соревнованиях	61
2.65	Задача о распределении мест	62
2.66	Пример преобразования списка	63
2.67	Создание формулы в режиме работы с матрицами	66
2.68	Пример матриц и результатов операций с ними	66
3.1	Фрагмент исходного списка	68
3.2	Меню «Данные»	68
3.3	Диалог настройки параметров сортировки	69

3.4	Определение параметров сортировки	70
3.5	Фрагмент отсортированного списка	70
3.6	Вложенное меню «Фильтр»	71
3.7	Список с включенным автофильтром	71
3.8	Варианты выбора для поля «Детей»	71
3.9	Результат поиска мужчин со средним специальным образованием	72
3.10	Поиск по началу фамилии	73
3.11	Варианты условий в автофильтре	73
3.12	Результат поиска по началу фамилии	73
3.13	Поиск слов с заданным количеством символов	74
3.14	Результат поиска по длине фамилий	74
3.15	Условие поиска в диапазоне дат	75
3.16	Результаты поиска в диапазоне дат	75
3.17	Условия выбора летних месяцев	76
3.18	Результат поиска по расчётному полю	76
3.19	Создание блока критериев	77
3.20	Определение диапазона исходных данных и блока критериев для расширенного фильтра	78
3.21	Окно для указания диапазона ячеек	78
3.22	Определение расположения результатов работы расширенного фильтра	79
3.23	Результат работы расширенного фильтра	79
3.24	Сложные условия для расширенного фильтра	80
3.25	Условия с вычисляемым критерием поиска	80
3.26	Критерии и результаты вычислений для функции dcounta()	82
4.1	Исходные данные для диаграммы	84
4.2	Выбор типа диаграммы	85
4.3	Настройка объектов диаграммы	86
4.4	Настройка оформления области графика	87
4.5	Настройка заголовка графика	88
4.6	Настройка исходных данных для диаграммы	89
4.7	Настройка подписи к диаграмме (легенды)	90
4.8	Настройка «вырезанного» сектора	91
4.9	Готовый график на листе ЭТ в режиме редактирования	91
4.10	Контекстное меню диаграммы	92
4.11	Диалог экспорта диаграммы в Gnumeric	92
4.12	Форматы для экспорта диаграммы	93
4.13	Исходные данные для множественной диаграммы	93
4.14	Исходные данные ряда данных	94
4.15	Размещение диаграммы на графике	95
4.16	Несколько диаграмм на графике	95
4.17	Исходные коммерческие данные	96

4.18	Варианты линейчатой диаграммы	97
4.19	Диаграммы областей	98
4.20	Полосковые диаграммы	99
4.21	Кольцевая диаграмма	99
4.22	Диаграмма радар	100
4.23	Отклонения значений $\cos()$ от значений $\sin()$	101
4.24	Отклонения функций в виде диаграммы пределов	101
4.25	Диаграмма XY с двумя независимыми наборами данных	102
4.26	Подтипы XY-диаграмм	103
4.27	Иллюстрация процедуры дискретизации с использованием диаграммы XY	103
4.28	Диаграмма ЦветнойXY	104
4.29	Данные для пузырьковой диаграммы	105
4.30	Варианты пузырьковой диаграммы	105
4.31	График в полярных координатах	106
4.32	Диаграмма высот	107
4.33	Настройка трёхмерного вида	108
4.34	Поверхность, построенная по сетке значений	108
4.35	Поверхность, построенная по точкам XYZ	109
4.36	Ящичковая диаграмма с выбросами для двух выборок .	110
4.37	График распределения вероятности	110
4.38	Гистограмма	111
5.1	Средства статистического анализа	112
5.2	Выбор распределения для создания исходных данных .	113
5.3	Определение параметров выборки	114
5.4	Настройка размещения результатов	114
5.5	Определение диапазона данных для обработки	115
5.6	Уточнение параметров обработки данных	116
5.7	Определение размещения результатов обработки	117
5.8	Описательные статистики для нормального распределения	117
5.9	Описательные статистики для однородного распределения	118
5.10	Исходные данные и результаты сглаживания	119
5.11	Графическое представление исходных данных и результатов сглаживания	119
5.12	Определение исходных данных для вычисления корреляции	121
5.13	Результаты вычисления корреляции выборок	121
5.14	Результат вычисления ковариации выборок	122
5.15	Результаты регрессионного анализа экспериментальных данных	123
5.16	Исходные данные и кривые регрессии на диаграмме XY	123
5.17	Гауссова кривая и её Фурье-образ	124

5.18	«Импульсная» функция и её Фурье-образ	125
5.19	Определение диапазонов значений для создания гистограммы	126
5.20	Определение правил учёта границ отрезков	127
5.21	Определение вида получаемой диаграммы и формата вывода результатов	127
5.22	Указание диапазона для вывода результатов	128
5.23	Результаты вычислений и график гистограммы	129
5.24	Определение параметров периодической выборки	129
5.25	Результат трёх испытаний для периодической выборки	130
5.26	Определение параметров случайной выборки	130
5.27	Результат трёх испытаний при случайной выборке	131
5.28	Исходные данные и расчёты рангов и процентилей	132
5.29	Определения способа вычисления рангов	133
5.30	Исходные данные для однофакторного ДА	134
5.31	Определение исходных данных для однофакторного ДА	134
5.32	Результаты однофакторного ДА	135
5.33	Исходные данные для двухфакторного ДА	136
5.34	Определение исходных данных для двухфакторного ДА	136
5.35	Результаты двухфакторного ДА	137
5.36	Настройка исходных данных для проверки гипотезы	138
5.37	Настройка режимов анализа	139
5.38	Результаты проверки гипотезы	139
5.39	Настройка режимов вычислений	140
5.40	Результаты проверки гипотезы	141
5.41	Настройка режимов вычислений	141
5.42	Результаты проверки гипотезы	142
5.43	Настройка режимов вычислений	142
5.44	Результаты проверки гипотезы	143
5.45	F-тест. Равные дисперсии	143
5.46	F-тест. Неравные дисперсии	144
5.47	Пример исходных данных для анализа выживаемости	145
5.48	Настройка исходных данных для анализа	146
5.49	Настройка групп исследуемых объектов	147
5.50	Настройка результатов анализа	147
5.51	Определение расположения результатов анализа	148
5.52	Результаты анализа (функция выживания)	149
5.53	Численные результаты для первой группы	149
5.54	Численные результаты для второй группы	150
5.55	Общая статистика по группам	150
6.1	Исходные данные	153
6.2	График исходных данных	154
6.3	Выбор добавляемого объекта	154
6.4	Определение класса интерполяционных функций	155

6.5	Настройка полиномиальной регрессии и предварительный вид графика	155
6.6	Изменение списка добавляемых объектов в зависимости от выбранного объекта графика	156
6.7	Добавление уравнения кривой на график	157
6.8	Настройка расположения уравнения	158
6.9	Настройка расположения уравнения	158
7.1	Исходные данные для поиска решения	161
7.2	Диалог поиска решения в Gnumeric (Solver)	162
7.3	Определение вида модели	163
7.4	Определение ограничений	163
7.5	Результат решения	163
7.6	Найденное решение	164
8.1	Графики исходных функций	166
8.2	Исходные данные для поиска решения	166
8.3	Настройка подбора параметра	167
8.4	Результат подбора параметра	168
8.5	Решение для отрицательных значений параметра	168
9.1	Исходная таблица вычисления дохода	171
9.2	Распределение уровней спроса	172
9.3	Распределение спроса в зависимости от количества газет	173
9.4	Функции распределения спроса	173
9.5	Итоговый лист для вычисления дохода	175
9.6	Определение диапазонов для входных и выходных данных	176
9.7	Параметры вычислений	177
9.8	Настройка расположения результатов моделирования	178
9.9	Результаты моделирования в диалоговом окне	179
9.10	Фрагмент листа с отчётом о моделировании	179
9.11	Настройка вычислений для нескольких вариантов модели	180
9.12	Просмотр результатов по вариантам модели	181
9.13	Результаты моделирования по вариантам входных данных	182

Список таблиц

2.1	Варианты курсора в Gnumeric	21
2.2	Некоторые неочевидные математические функции	47
2.3	Логические функции в Gnumeric	48
2.4	Некоторые неочевидные математические функции	50
2.5	Формулы для решения квадратного уравнения	52
2.6	Входные данные	59
2.7	Некоторые строковые функции	64
2.8	Формулы для решения квадратного уравнения	64
2.9	Функции для обработки матриц	65
3.1	Некоторые функции категории «База данных»	81
7.1	Таблица исходных данных для задачи	160
9.1	Формулы для вычисления дохода	172
9.2	Адреса ячеек и формулы	174
9.3	Результаты моделирования	180

Учебное издание

Серия «Библиотека ALT Linux»

Хахаев Иван Анатольевич

Gnumeric: Электронная таблица для всех

Оформление обложки: А. Осмоловская

Вёрстка: А. Прокудин

Подписано в печать 18.07.11. Формат 70x100/16.
Гарнитура Computer Modern. Печать офсетная. Бумага офсетная.



Редакционная подписка на журнал «Системный администратор»

Друзья!

Редакция журнала «Системный администратор» предлагает вам оформить редакционную подписку на наше издание.

С 1 июня 2011 года цена редакционной подписки на 2012 год составит:

в июне 2011года	3100 руб.
в июле 2011года	3160 руб.
в августе 2011года	3220 руб.
в сентябре 2011года	3280 руб.
в октябре 2011года	3340 руб.
в ноябре 2011года	3400 руб.
в декабре 2011года	3460 руб.
в январе 2012 года	3540 руб.

Как оформить подписку?

- 1) по факсу или по e-mail в редакцию передается заказ на подписку с кодовым словом LINUX, в котором указывается общий срок подписки и месяц, с которого должен поступать журнал;
- 2) реквизиты организации;
- 3) подробный адрес доставки с указанием контактного лица и его телефона.

Редакция работает по упрощенной системе налогообложения!

Журнал «Системный администратор» – ежемесячное издание, цель которого – предоставление максимально полной и объективной информации о решениях, продуктах и технологиях современной ИТ-отрасли.

Главная задача издания – популяризация лучших разработок ИТ-специалистов разных стран. Девяносто процентов статей в журнале носят прикладной характер, снабжены примерами, таблицами, графическим материалом. Информация, опубликованная в издании, остается актуальной в течение нескольких лет.

Именно поэтому журнал «Системный администратор» является настольным пособием для ИТ-отдела, отвечающего за развитие и функционирование ИТ-системы в компании.

тел.: (495) 687-93-57, факс: (495) 640-43-13
www.samag.ru, e-mail: subscribe@samag.ru

