

СРЕДНЕЕ
ПРОФЕССИОНАЛЬНОЕ
ОБРАЗОВАНИЕ

ОБРАБОТКА И ПРЕДСТАВЛЕНИЕ ДАННЫХ В MS EXCEL



Э. Г. Бурнаева
С. Н. Леора



ЛАНЬ

E.LANBOOK.COM



Э. Г. БУРНАЕВА, С. Н. ЛЕОРА



ОБРАБОТКА И ПРЕДСТАВЛЕНИЕ ДАННЫХ В MS EXCEL

Учебное пособие



• САНКТ-ПЕТЕРБУРГ • МОСКВА • КРАСНОДАР •
• 2021 •

УДК 004
ББК 32.973.26-018.2я723

Б 90 **Бурнаева Э. Г.** Обработка и представление данных в MS Excel : учебное пособие для СПО / Э. Г. Бурнаева, С. Н. Леора. — Санкт-Петербург : Лань, 2021. — 156 с. : ил. — Текст : непосредственный.

ISBN 978-5-8114-6919-2



Книга посвящена эффективному использованию инструментов приложения Excel для обработки и анализа многомерных данных с целью их систематизации, выявления характера и структуры взаимосвязей. Пройдя путь от структурирования данных до создания простейшей базы данных, на основе которой можно строить различные сводные таблицы для нахождения ответов на поставленные вопросы, читатель научится извлекать необходимую информацию и знание из множества данных. Дополнительно к этому, в книге рассматривается одна из наиболее востребованных задач — статистическая обработка данных, полученных в результате случайного эксперимента. Особое внимание уделено визуализации данных с помощью диаграмм и дополнительных возможностей, предоставляемых в последних версиях Excel.

Книга адресована студентам, обучающимся в колледжах по образовательным программам среднего профессионального образования. Она может быть полезна всем, кто заинтересован в получении мощного и эффективного инструмента для выполнения всевозможных математических операций при помощи компьютера.

УДК 004
ББК 32.973.26-018.2я723

Рецензенты:

Н. К. КРИВУЛИН — доктор физико-математических наук, профессор кафедры статистического моделирования Санкт-Петербургского государственного университета;

С. М. ЕРМАКОВ — профессор, зав. кафедрой статистического моделирования Санкт-Петербургского государственного университета.

Обложка

Ю. В. ГРИГОРЬЕВА

© Издательство «Лань», 2021
© Э. Г. Бурнаева, С. Н. Леора, 2021
© Издательство «Лань»,
художественное оформление, 2021

ВВЕДЕНИЕ



С развитием Интернета пользователи получили доступ к огромному объему данных. Сами по себе данные не представляют большой ценности. Необходимо научиться извлекать из них нужную, адекватную информацию, а значит, уметь эти данные структурировать, обрабатывать и, наконец, представлять.

Самым доступным инструментом для реализации всех этих задач являются электронные таблицы. Excel — программный продукт два в одном, в котором удобно сочетаются вычислительные и графические возможности.

Excel включает в себя элементы баз данных, что позволяет хранить структурированные данные, осуществлять их сортировку, фильтрацию, выборку данных, удовлетворяющих необходимым критериям.

Реализованы здесь и возможности управления многомерными данными на основе сводных таблиц, являющихся в какой-то мере аналогами мобильных таблиц в статистических пакетах (например, в SPSS). Надстройка Пакет анализа позволяет проводить статистический анализ данных.

И наконец, Excel как инструмент визуализации может использоваться для создания иллюстраций в стиле инфографики, современного направления информационного дизайна.

Именно эти продвинутые возможности являются предметом обсуждения в данном пособии. Теоретический и практический материал предназначен для работы в среде

приложений Excel 2013, которые открывает новые возможности обработки и представления данных.

Материал расположен в следующем порядке. В первой главе рассматриваются таблицы как средство структурирования и обработки данных. Во второй главе речь идет о базах данных и сводных таблицах как способах работы с многомерными данными в Excel. В третьей главе рассматриваются диаграммы — мощное средство визуализации данных, информации и знаний. И, наконец, в четвертой главе представлены возможности Excel для статистической обработки данных.

Данное пособие предназначено для проведения занятий по курсам «Информатика» и «Информационные технологии в науке и образовании» со студентами и аспирантами гуманитарных специальностей, оно может быть также полезно всем пользователям, работающим в среде Microsoft Office.

Авторы выражают большую благодарность рецензентам этого издания за их квалифицированные комментарии и замечания.





ГЛАВА 1

ТАБЛИЦЫ В EXCEL

ТЕМА 1.1.

ОСНОВНЫЕ ПОНЯТИЯ

Обработка данных в Excel обычно подразумевает работу с взаимосвязанной группой данных. Практически каждая последующая версия электронных таблиц расширяет возможности управления данными, вводя новые объекты. Поэтому необходимо обратить особое внимание на сходство и различие понятий *диапазон*, *список* и *таблица*, которые можно рассматривать как развитие различных способов структурирования данных от простого к сложному.

1.1.1.

ДИАПАЗОН

Под *диапазоном* во всех версиях Excel понимается группа ячеек (две и более), объединенных обычно некоторым смысловым значением. Для задания диапазона используется двоеточие, например, A1:D2.

Диапазон может быть смежным или несмежным. Для его выделения необходимо щелкнуть мышью по верхней левой ячейке диапазона и, не отпуская кнопку, переместить указатель в нижнюю правую ячейку диапазона. Выделение несмежных диапазонов осуществляется при нажатой клавише Ctrl.

Диапазону ячеек также можно присвоить *имя*, которое затем используется как ссылка на этот диапазон. Таким



образом, на диапазон можно ссылаться по адресу или по имени.

Смежный диапазон данных можно считать первым шагом структурирования данных. Такие инструменты, как сортировка, фильтрация, подведение итогов и другие, применимы только к смежным диапазонам, например к А3:G17 (рис. 1.1).

	A	B	C	D	E	F	G
1	Новогодние туры						
2							
3	№	Название фирмы	Страна	Кол-во дней	Стоимость	ОТЕЛЬ	Транспорт
4	1	Нева	Италия	11	447 €	***	Поезд + автобус
5	2	Нева	Германия	7	374 €	****	Самолет
6	3	Нева	Польша	7	199 €	***	Поезд + автобус
7	4	Нева	Франция	10	581 €	****	Автобус + паром
8	5	Нева	Швейцария	8	1 240 €	***	Самолет
9	6	Одиссея	Италия + Скандинавия	12	645 €	***	Автобус+паром
10	7	Одиссея	Испания	8	796 €	****	Самолет
11	8	Одиссея	Италия	12	430 €	****	Поезд
12	9	Одиссея	Польша	9	265 €	****	Поезд + автобус
13	10	Олимпиада	Финляндия	3	175 €	***	Автобус
14	11	Олимпиада	Швеция	5	800 €	***	Поезд
15	12	Олимпиада	Мальдивы	14	2 890 €	*****	Самолет
16	13	Олимпиада	Германия	12	700 €	*****	Поезд
17	14	Олимпиада	Португалия	8	1 460 €	*****	Самолет

Рис. 1.1
Список в Excel

1.1.2. СПИСОК

Во всех версиях Excel вводится понятие *списка*, основой которого являются смежный диапазон данных, удовлетворяющий следующим требованиям.

1. Список состоит из строк, называемых *записями*. По строкам описывается только одно событие (данные о конкретном человеке, фирме, сделке и т. д.).

2. Столбцы обычно называют *полями*. Каждый столбец должен содержать данные одного типа.

3. Верхняя строка списка, называемая *заголовком*, содержит имена соответствующих полей, обычно ее формат (шрифт, цвет фона и т. п.) отличается от формата записей.

4. На пересечении строк и столбцов находятся *данные*.
5. Внутри списка не должно быть полностью пустых строк и столбцов, а также объединенных ячеек.
6. От остальной части рабочего листа список отделяется хотя бы одним пустым столбцом слева и справа и хотя бы одной пустой строкой сверху и снизу.

Список можно создать вручную, введя информацию в прямоугольный диапазон с учетом вышеперечисленных требований. Такой список по виду ничем не отличается от обычной таблицы Excel. Над списками можно производить такие действия, как сортировка, фильтрация, подведение итогов и т.п.

На рисунке 1.1 приведен пример списка, который занимает диапазон A3:G17. Список имеет заголовок с именами полей. Каждая запись содержит информацию о конкретном туре. Список отсортирован по полю «Название фирмы».

1.1.3. ТАБЛИЦА

В Excel 2007/2010/2013 под *таблицей* понимается список, т. е. взаимосвязанная группа данных, обладающая дополнительными свойствами.

По структуре таблица ничем не отличается от списка. Расширились только возможности управления данными, расположенными в таблице.

Таблицы в Excel 2007/2010/2013 пользователи называют «умными» таблицами, на их основе можно организовывать и поддерживать личное хранилище данных.

Таблица, как объект, создается на основе диапазона пустых ячеек или прямоугольного диапазона данных, с помощью последовательности команд «Вставка» → «Таблицы» → «Таблица». Для создания таблицы сначала нужно выделить диапазон (пункт 1), а затем выполнить последовательность команд (пункт 2).

1. Если исходный диапазон данных удовлетворяет требованиям, предъявляемым к спискам, т. е. она не содержит пустых строк и т. д., тогда для выделения таблицы достаточно поместить курсор в любую ячейку внутри

диапазона. Если же в исходном диапазоне имеются полностью пустые строки или столбцы или вообще нет данных, то этот диапазон будет рассматриваться как таблица, и ввод данных в нее будет осуществляться по мере надобности, необходимо выделить весь этот диапазон.

2. Выполнить последовательность команд «Вставка» → «Таблицы» → «Таблица». Если первая строка выбранного диапазона содержит данные, которые являются заголовками столбцов, установите флажок «Таблица с заголовками». Если такой строки нет, строка с заголовками будет создана автоматически.

Заметим, что в таблицах не должно быть объединенных ячеек, так как это нарушает их структуру и идеологию: одна строка — одна запись, один столбец — одно поле, если использовать терминологию баз данных. Такие таблицы можно использовать как основу для создания простейших баз данных в Excel.

1.1.4. ОСОБЕННОСТИ РАБОТЫ С ТАБЛИЦАМИ

Рассмотрим новые возможности работы с таблицами, начиная с версии Excel 2007. Создадим таблицу на основе диапазона на рисунке 1.1.

Способ 1. Выделите любую ячейку диапазона A3:G17, выполните последовательность команд «Вставка» → «Таблицы» → «Таблица».

Способ 2. Для быстрого форматирования диапазона ячеек и преобразования его в таблицу выделите любую ячейку диапазона A3:G17 и выполните последовательность команд «Главная» → «Стили» → «Форматировать как таблицу» → «Выбрать стиль».

В результате произойдет следующее (рис. 1.2):

1) будет создана таблица, к которой применен стиль форматирования по умолчанию;

2) в первую строку с заголовками помещены элементы «Фильтра» (стрелочки);

3) появится новая вкладка «Конструктор» («Работа с таблицами»).

Рис. 1.2. Таблица в Excel

Рис. 1.2
Таблица в Excel

Стиль форматирования таблицы можно изменить, используя «Экспресс-стили» на вкладке «Конструктор».

«Фильтр» можно принудительно отключить, если он не нужен, на вкладке «Данные» → «Фильтр». Работа «Фильтра» будет рассмотрена далее в пункте 1.2.2.

Таким образом, таблица примет вид, как на рисунке 1.2, и она уже будет обладать новыми свойствами.

1. При выделении любой ячейки таблицы автоматически будет появляться дополнительная вкладка «Конструктор» для работы с таблицей.

2. Таблице автоматически будет присвоено имя, которое можно посмотреть на вкладке «Конструктор» → «Свойства» → «Имя таблицы». Это имя можно использовать в любых формулах.

3. Таблица автоматически подстраивается в размерах при добавлении или удалении в нее данных. Достаточно ввести данные в любую ячейку, расположенную непосредственно под таблицей или справа от таблицы.

4. Если вы введете в ячейку нового столбца формулу, она автоматически распространится на весь столбец.

5. В таблицу можно добавить строку итогов «Конструктор → Параметры стилей таблиц → установить флажок Строка итогов».

6. Чтобы вернуть таблице прежний вид, можно использовать «Конструктор» → «Сервис» → «Преобразовать в диапазон».

При работе с данными таблицы с именем, например Таблица1, допустимо использовать структурированные ссылки с указателями специальных элементов таблицы:

- =Таблица1[#Все] — ссылка на всю таблицу, включая заголовки, данные и строку итогов (если она включена);
- =Таблица1[#Данные] — ссылка только на данные (без строки заголовка);
- =Таблица1[#Заголовки] — ссылка на первую строку таблицы с заголовками столбцов;
- =Таблица1[#Итоги] — ссылка на строку итогов (если она включена);
- =Таблица1[#Эта строка] — ссылка на текущую строку, например формула =Таблица1[#Эта строка];[Стоимость]], будет ссылаться на значение «Стоимость» из текущей строки таблицы.

Пример 1.1. Имеется диапазон данных (см. рис. 1.1), требуется добавить новый столбец, чтобы вычислить среднюю стоимость одного дня каждого тура.

Для решения этой задачи сначала преобразуем диапазон данных в таблицу, как описано выше. Таблица примет вид новый вид (см. рис. 1.2). Уберем «Фильтр» и изменим форматирование.

Вставим в ячейку H4 формулу =E4/D4. При этом автоматически в таблицу добавиться новый столбец с именем Столбец1 и в ячейку H4 будет вставлена формула

=Таблица1[#Эта строка];[Стоимость]]/
Таблица1[#Эта строка];[Количество дней]].

Эта же формула автоматически будет помещена и во все остальные строки данного столбца.

Заметим, что если таблицу преобразовать в диапазон «Конструктор» → «Сервис» → «Преобразовать в диапазон», формулы изменят свой вид. В ячейке Н4 будет формула

=таблицы!\$E4/таблицы!\$D4.

Здесь таблицы — ссылка на имя листа с таблицей.

На рисунке 1.3 приведена таблица с вычисляемым столбцом и результаты расчетов. Обратите внимание, что в правом нижнем углу таблицы располагается автоматически перемещающийся маркер границы таблицы.

	A	B	C	D	E	F	G	H
1			Новогодние туры					
2								
3	№	Название фирмы	Страна	Кол-во дней	Стоимость	Отель	Транспорт	Столбец1
4	1	Нева	Италия	11	447 €	***	Поезд + автобус	40,64
5	2	Нева	Германия	7	374 €	****	Самолет	53,43
6	3	Нева	Польша	7	199 €	***	Поезд + автобус	28,43
7	4	Нева	Франция	10	581 €	****	Автобус + паром	58,10
8	5	Нева	Швейцария	8	1 240 €	***	Самолет	155,00
9	6	Одиссея	Италия + Скандинавия	12	645 €	****	Автобус + паром	53,75
10	7	Одиссея	Испания	8	796 €	****	Самолет	99,50
11	8	Одиссея	Италия	12	430 €	****	Поезд	35,83
12	9	Одиссея	Польша	9	265 €	****	Поезд + автобус	29,44
13	10	Олимпика	Финляндия	3	175 €	***	Автобус	58,33
14	11	Олимпика	Швеция	5	800 €	****	Поезд	160,00
15	12	Олимпика	Мальдивы	14	2 890 €	****	Самолет	206,43
16	13	Олимпика	Германия	12	700 €	****	Поезд	58,33
17	14	Олимпика	Португалия	8	1 460 €	****	Самолет	182,58

Рис. 1.3
Таблица с вычисляемым столбцом

	J	K	L	M
3	Количество туров	Минимальная продолжительность	Максимальная стоимость	Самолет
4	14	3	2 890 €	5

Рис. 1.4
Расчеты на основе данных из таблицы

Таблица 1.1

Расчетные формулы

Ячейка	Содержимое	Ячейка	Содержимое
J4	=СЧЕТ(Таблица1[№])	K4	=МАКС(Таблица1[Стоимость])
L4	=МИН(Таблица1[Количество дней])	M4	=СЧЕТЕСЛИ(Таблица1[Транспорт];"Самолет")

Пример 1.2. Для таблицы, приведенной на рисунке 1.3 требуется вычислить некоторые данные, используя ссылки на элементы таблицы. Например, количество туров. Результаты расчетов приведены выше (см. рис. 1.4).

В ячейки новой таблицы введены формулы (табл. 1.1).

ТЕМА 1.2. ОБРАБОТКА ДАННЫХ

В зависимости от поставленной задачи можно рассматривать три уровня обработки информации — информационно-поисковый (OLTP), оперативно-аналитический (OLAP) и интеллектуальный (Data Mining).

Основной инструментарий первого уровня — фильтрация и отбор данных, сортировки и подведение итогов. Эти операции осуществляются над списками и таблицами.

1.2.1. СОРТИРОВКА

При выполнении самых различных задач возникает необходимость упорядочить данные. Так, например, числовые данные могут быть отсортированы в порядке возрастания или убывания, а текстовые данные можно расположить в алфавитном порядке. Для этого используется средство «Сортировка», расположенное на вкладке «Данные» и на вкладке «Главная».

Результатом сортировки будет новый список (таблица), содержащий ту же информацию, но переупорядоченную. Порядок данных в новом списке (таблице) будет зависеть от условий сортировки, которые задаются в диалоговом окне «Сортировка» (рис. 1.5).

Пример 1.3. Имеется список (рис. 1.6), требуется сделать сортировку списка, распределить данные по регионам. В данном примере можно использовать список, так как нужно выполнить только сортировку и не требуется использовать дополнительные свойства, связанные с таблицами.

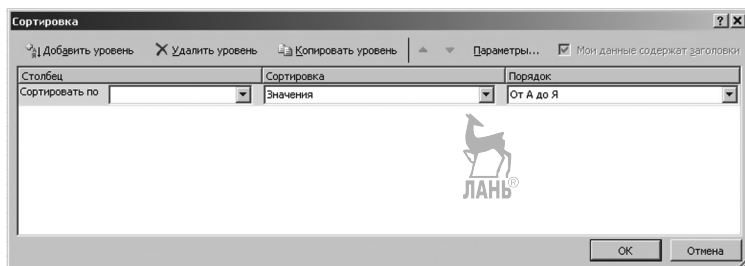


Рис. 1.5
Диалоговое окно «Сортировка»

	A	B	C	D
1	№	Город	Регион	Продажи
2	1	Санкт-Петербург	Запад	380
3	2	Мурманск	Север	250
4	3	Нижний Новгород	Центр	834
5	4	Владимир	Центр	818
6	5	Краснодар	Юг	621
7	6	Орел	Центр	595
8	7	Хабаровск	Восток	355
9	8	Калининград	Запад	767

Рис. 1.6
Исходные данные для сортировки

Последовательность действий для решения поставленной задачи следующая.

1. Поместить курсор в любое место списка (автоматически будет выделен весь список).
2. Выполнить «Данные» → «Сортировка», при этом осуществится переход к диалоговому окну «Сортировка».
3. Установить флажок «Мои данные содержат заголовки».
4. По умолчанию сортировка будет производиться по строкам. При необходимости можно установить некоторые дополнительные параметры сортировки, перейдя в диалоговое окно «Параметры» → «Параметры сортировки».
5. Выбрать способ идентификации диапазона:
 - «Сортировать по» — столбец «Регион»;
 - «Сортировка» — Значения (в выпадающем меню можно выбрать при необходимости «Цвет ячейки»,

«Цвет шифра» и «Значок ячейки», в нашем случае игнорируем);

- «Порядок» — «От А до Я».

6. Нажать **Ок**.

Пример 1.4. Имеется список (см. рис. 1.6), требуется сделать многоуровневую сортировку всего списка, распределить сначала по регионам, в пределах региона города расположить в алфавитном порядке.

Последовательность действий для решения поставленной задачи. Пункты 1–5 совпадают с предыдущим примером. Далее необходимо сделать следующее.

1. Щелкнуть «Добавить уровень».

2. Указать второе условие сортировки — «Сортировать по» (выбрать столбец «Регион»), «Сортировка» (выбрать «Значение»), «Порядок» (выбрать «От А до Я»).

3. Нажать **Ок**.

Результат сортировки показан на рисунке 1.7.

Замечание. За одно обращение к средству «Сортировка» можно произвести сортировку по любому количеству столбцов (строк).

	А	В	С	Д
1	№	Город	Регион	Продажи
2	7	Хабаровск	Восток	355
3	8	Калининград	Запад	767
4	1	Санкт-Петербург	Запад	380
5	2	Мурманск	Север	250
6	4	Владимир	Центр	816
7	3	Нижний Новгород	Центр	834
8	6	Орел	Центр	595
9	5	Краснодар	Юг	621

Рис. 1.7
Результат сортировки

Пример 1.5. В результате сортировки, выполненной в примере 1.4, нарушилась нумерация (рис. 1.7, столбец А). Требуется восстановить нумерацию, не меняя при этом содержимое всей остальной таблицы.

Для этого нужно сделать сортировку одного столбца в списке:

1) выделить полностью первый столбец, включая первую строку (заголовок);

2) выполнить последовательность команд «Данные» → «Сортировка». В диалоговом окне «Обнаружены данные вне указанного диапазона» выбрать «Сортировать в пределах указанного выделения», нажать «Сортировка»;

3) в диалоговом окне «Сортировка» выбрать направление сортировки — «По возрастанию».

1.2.2. ФИЛЬТР

Средство «Фильтр» позволяет из совокупности данных *отобрать* (отфильтровать) нужные данные, удовлетворяющие некоторым критериям отбора. Для простого отбора используется «Фильтр», для более сложного отбора используется «Расширенный фильтр».

Отличия «Фильтра» от «Расширенного фильтра»:

1) при «Фильтре» для каждого столбца можно задать не более двух условий отбора, а для «Расширенного фильтра» можно задать более сложные условия, в том числе включить в условия вычисляемые критерии с использованием формул;

2) «Фильтр» производится на месте, а результаты «Расширенного фильтра» можно поместить в другое место.

Для успешного применения средства «Фильтр» необходимо правильно задать исходную структуру данных — это должен быть список или таблица (см. Тема 1.1).

Верхняя строка таблицы должна содержать заголовки столбцов, если заголовков нет, нужно вставить пустую первую строку вместо заголовков, в нее будут помещены стрелочки «Фильтра».

Сначала нужно установить «Фильтр», потом условия отбора. При создании таблицы «Фильтр» устанавливается автоматически.

Для установления «Фильтра» потребуется:

1) задать область исходного диапазона данных, выделив всю таблицу включая заголовки;

2) выполнить последовательность команд «Данные» → «Фильтр».

При этом на именах полей появятся кнопки с изображением стрелочек вниз (рис. 1.8). Средство «Фильтр» установлено. Кнопки «Фильтра» предназначены для установки условий отбора данных.

Для выбора условий отбора данных выполните последовательно действия.

1. Нажать на стрелочку столбца, по которому будет производиться отбор данных.

	A	B	C	D
1	№	Город	Регион	Продажи
2	1	Владимир	Центр	816
3	2	Краснодар	Юг	621
4	3	Мурманск	Север	250
5	4	Нижний Новгород	Центр	834
6	5	Орел	Центр	595
7	6	Санкт-Петербург	Запад	380
8	7	Хабаровск	Восток	355
9	8	Калининград	Запад	767

Рис. 1.8
Таблица с «Фильтром»

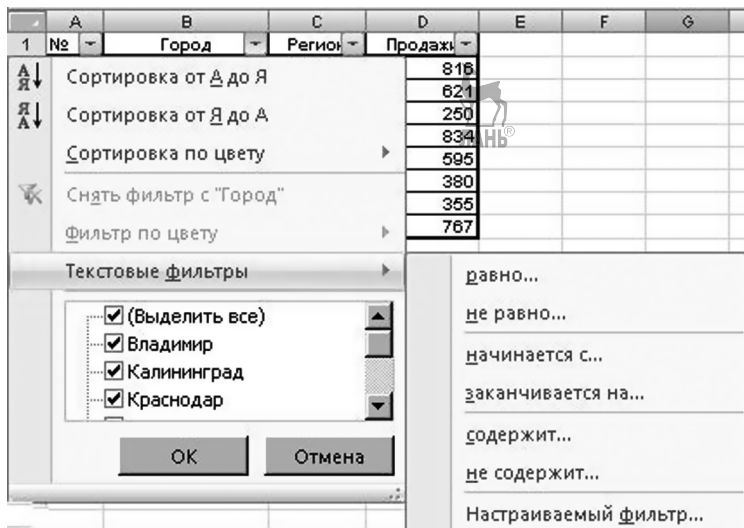


Рис. 1.9
Условия отбора для «Фильтра»

2. Выбрать условие отбора данных из выпадающего меню (рис. 1.9). Эти условия будут автоматически формироваться исходя из типа данных. В данном случае будут предложены «Текстовые фильтры».

3. Нажать **Ок**. Отобразится результат отбора. На столбцах, участвующих в отборе, появится значок «Фильтра».

4. Результат фильтрации можно скопировать в любое другое место. Это может быть любое место текущей рабочей книги или любой другой документ, например, Excel или Word.

5. Если требуется вернуться к исходной таблице, но остаться в режиме «Фильтра», выберите «Фильтр» → «Очистить». Для выхода из режима «Фильтра» необходимо щелкнуть на кнопку «Фильтр» на вкладке «Данные».

Пример 1.6. Из всех данных о продажах, расположенных в таблице (рис. 1.8), требуется отобрать продажи, произведенные только в Центральном регионе.

Так как задано только одно условие отбора, используется простой «Фильтр». Последовательность действий для решения данной задачи следующая.

1. Установить «Фильтр».

2. Нажать на кнопку столбца «Регион».

3. Из выпадающего списка выбрать критерий фильтрации — конкретное значение («Центр»). Результат изображен на рисунке 1.10.

Аналогично можно выбрать условия отбора и для других столбцов.

	A	B	C	D
1	№	Город	Регион	Продажи
2	1	Владимир	Центр	816
3	4	Нижний Новгород	Центр	834
6	5	Орел	Центр	595

Рис. 1.10
Результат отбора данных по значению

Пример 1.7. Из всех данных о продажах, расположенных в таблице на рисунке 1.8, требуется отобрать продажи, произведенные в Центральном регионе, объем которых находится в диапазоне от 800 до 900.

Заданы простые критерии отбора, можно использовать простой «Фильтр». На объем продаж установлено два ограничения. Условие отбора в этом случае является логическим выражением вида (Условие 1 И Условие 2).

Последовательность действий для решения данной задачи следующая.

1. Установить «Фильтр».

2. Нажать на кнопку столбца «Регион». Из выпадающего списка выбрать критерий фильтрации — конкретное значение (Центр).

3. Нажать на кнопку столбца «Продажи». Из выпадающего списка выбрать пункт «Числовые фильтры».

4. Из выпадающего списка выбрать «Между», появится окно «Пользовательский автофильтр».

5. В окне «Пользовательский автофильтр» установить «Больше или равно 800 И меньше или равно 900».

6. Нажать Ок.

Результат представлен на рисунке 1.11.

Рассмотрим работу с *расширенным фильтром*. Отдельной кнопки «Расширенный фильтр» нет. Диалоговое окно «Расширенный фильтр» появится, если щелкнуть по кнопке «Дополнительно» на вкладке «Данные» в группе «Сортировка и фильтр».

При использовании «Расширенного фильтра» необходимо сделать следующее.

1. Сначала подготовить области:

а) *исходный диапазон* — область таблицы данных;

б) *диапазон условий* — область, где задаются критерии фильтрации;

в) если результат необходимо поместить в новое место, тогда определить место и для третьего диапазона — области, в которую будут помещены *результаты фильтрации*.

	A	B	C	D
1	№	Город	Регион	Продажи
2	1	Владимир	Центр	816
5	4	Нижний Новгород	Центр	834

Рис. 1.11

Результат отбора данных по заданным ограничениям

2. Выполнить «Данные» → «Фильтр» → «Дополнительно». В открывшемся диалогом окне «Расширенный фильтр» задать исходный диапазон, диапазон условий, при необходимости диапазон для размещения результата (задается в окне «Поместить результат в диапазон»).

3. Нажать Ок.

4. Если требуется вернуться к исходной таблице, но остаться в режиме «Фильтра», выбрать «Фильтр» → «Очистить».

Диапазон условий должен включать:

а) заголовки, имена которых полностью совпадают с именами столбцов исходного диапазона, которые участвуют в отборе*.

б) заголовки, имена которых не совпадают с именами столбцов исходного диапазона; они предназначены для вычисляемого критерия (могут отсутствовать);

в) строки, задающие критерии отбора; на основе этих строк строится общее логическое выражение для отбора нужных значений.

При формировании диапазона условий необходимо также соблюдать следующие соглашения.

1. Диапазон условий должен состоять не менее чем из двух строк. Первая строка содержит заголовки. Все последующие строки содержат соответствующие критерии отбора строк (записей).

2. Если условия располагаются в одной строке, то это означает одновременность их выполнения, т. е. считается, что между ними поставлена логическая операция И.

3. Для истинности критерия, состоящего из условий, располагающихся в разных строках, требуется выполнение хотя бы одного из них, т. е. считается, что они соединены логической операцией ИЛИ.

4. Диапазон условий (интервал критериев) желательно располагать ниже исходного списка либо на другом рабочем листе.

5. В диапазоне условий (интервале критериев) не должно быть пустых строк.

* Их рекомендуется просто копировать из заголовков столбцов исходной таблицы. Можно полностью скопировать заголовки.

Например, диапазону условий на рисунке 1.12 будет соответствовать логическое условие отбора вида «(Условие 1 И Условие 2 И Условие 3) ИЛИ (Условие 4 И Условие 5) ИЛИ (Условие 6 И Условие 7)».

Условия на числовые данные формируются в виде равенств и неравенств. Условия на текстовые данные в виде маски. Примеры условий приведены в таблице 1.2.

	A	B	C	D	E	F
1	Заголовок 1	Заголовок 2	Заголовок 3	Заголовок 4	Заголовок 5	Заголовок 6
2	Условие 1		Условие 2		Условие 3	
3		Условие 4			Условие 5	
4			Условие 6			Условие 7

Рис. 1.12
Диапазон условий

Таблица 1.2

Примеры условий

Условие	Результат отбора
Москва	Строки, содержащие слово Москва в выбранном столбце
М*	Строки, содержащие слова, начинающиеся на букву М в выбранном столбце
М?О	Строки, содержащие слова, начинающиеся на букву М, второй символ любой, а третий символ О
<>М*	Строки, содержащие слова, не начинающиеся на букву М
>=100	Строки, содержащие числа большие или равные 100
<>20	Строки, содержащие числа, не равные 20

В «Расширенном фильтре» можно формировать *вычисляемые критерии*, которые содержат функции. В этом случае в диапазоне критериев должны быть указаны имена столбцов, которые участвуют в отборе (они должны совпадать с именами исходных заголовков), и имена столбцов, которые содержат вычисляемые критерии (они должны по названию отличаться от заголовков исходной таблицы).

Пример 1.8. Требуется отобрать туры в страны, начинающиеся на букву И, стоимость одного дня для которых меньше 60 евро. Исходные данные приведены на рисунке 1.1.

Диапазон условий может иметь, например, такой вид (рис. 1.13). Он состоит из столбца, имя которого совпадает с исходным заголовком, и столбца с новым именем, в который помещен вычисляемый критерий в ячейку J4, в данном случае формула $=E4/D4<60$, где $E4/D4$ — средняя стоимость тура.

Результат отбора данных приведен на рисунке 1.14.

	I	J
3	Страна	Средняя стоимость < 60 €
4	И*	ИСТИНА

Рис. 1.13
Условия для отбора

	A	B	C	D	E	F	G
19							
20	№	Название фирмы	Страна	Кол-во дней	Стоимость	Отель	Транспорт
21	1	Нева	Италия	11	447 €	***	Поезд + автобус
22	6	Одиссея	Италия + Скандинавия	12	645 €	***	Автобус+паром
23	8	Одиссея	Италия	12	430 €	****	Поезд

Рис. 1.14
Результат отбора данных

	A	B	C	D	E
1	ФИО	Дата Рождения	Исходный диапазон		
2	Иванов Иван Иванович	1 июл 88			
3	Петров Олег Петрович	14 дек 89			
4	Сидоров Семен Федорович	9 янв 87			
5	Белов Николай Викторович	13 авг 89			
6	Чернов Петр Михайлович	15 мар 89			
7	Петров Степан Георгиевич	13 июн 88			
8	Васина Ольга Ивановна	27 май 88			
9					
10	ФИО	День недели	← =ДЕНЬНЕД(B2;2)=5 ← =ДЕНЬ(B2)=13	Диапазон условий	
11		ИСТИНА			
12		ЛОЖЬ			
13					
14	ФИО	Дата Рождения	Результат		
15	Иванов Иван Иванович	1 июл 88			
16	Сидоров Семен Федорович	9 янв 87			
17	Белов Николай Викторович	13 авг 89			
18	Петров Степан Георгиевич	13 июн 88			
19	Васина Ольга Ивановна	27 май 88			

Рис. 1.15
Расширенный фильтр

Пример 1.9. Даны список студентов и дата их рождения. Требуется отобрать студентов, родившихся либо в пятницу, либо тринадцатого числа.

Для решения поставленной задачи воспользуемся «Расширенным фильтром» с вычисляемыми критериями.

Подготовьте «Исходный диапазон» и «Диапазон критериев». Заметьте, что в условиях заголовка столбца с вычисляемым критерием обязательно должен отличаться от заголовка соответствующего столбца исходного диапазона.

Условие =ДЕНЬНЕД(A2;2)=5 отберет тех, кто родился в пятницу, условие =ДЕНЬ(B2)=13 отберет родившихся 13-го числа.

Примените «Расширенный фильтр» → «Данные» → «Фильтр» → «Расширенный фильтр», задайте диапазоны → «Ok».

Результаты отбора приведены на рисунке 1.15.

1.2.3. ПРОМЕЖУТОЧНЫЕ ИТОГИ

Средство «Промежуточные Итоги» в меню «Данные» помогает организовать записи в группы, выводя под каждой из них итоговую информацию, другими словами, позволяет вывести на экран данные, сгруппированные по тем или иным критериям, и автоматически вычислять промежуточные и общие итоги в таблице.

Перед вставкой промежуточных итогов, список необходимо отсортировать, чтобы сгруппировать строки, по которым нужно подвести итоги. После этого можно подсчитать промежуточные итоги любого столбца, содержащего числа, используя итоговые операции (табл. 1.3).

Рассмотрим примеры использования данного средства.

Пример 1.10. Имеются данные о продажах различных категорий товаров различными продавцами. Требуется найти выручку от продажи по каждому продукту. Исходные данные приведены на рисунке 1.16.

Выполним следующие шаги.

1. Выделить любую ячейку из диапазона исходных данных.

Таблица 1.3

Итоговые операции

Сумма	Сумма чисел. Эта операция используется по умолчанию
Количество	Количество данных. Операция используется по умолчанию для подведения итогов по нечисловым полям
Среднее	Среднее значение чисел
Максимум	Максимальное число
Минимум	Минимальное число
Количество чисел	Количество данных, являющихся числами. Операция «Количество чисел» работает так же, как и функция СЧЕТ

	A	B	C	D
1	Клиент	Категория	Количество	Стоимость
2	Петров	Молочные продукты	60	9240
3	Петров	Молочные продукты	60	9240
4	Петров	Кондитерские изделия	36	13122
5	Петров	Приправы	216	20753,28
6	Майоров	Напитки	204	17442
7	Майоров	Молочные продукты	60	9240
8	Майоров	Напитки	168	13608
9	Майоров	Кондитерские изделия	12	942,36
10	Томин	Напитки	156	184977
11	Томин	Напитки	132	8316
12	Томин	Кондитерские изделия	12	942,36
13	Томин	Приправы	228	25650
14	Кротов	Кондитерские изделия	20	7290
15	Кротов	Мясо/птица	96	3218,88
16	Кротов	Напитки	156	184977
17	Кротов	Молочные продукты	48	11880
18	Кротов	Кондитерские изделия	36	13122
19	Кротов	Рыбпродукты	264	7128

Рис. 1.16
Исходные данные



2. Отсортировать таблицу исходных данных по категориям, используя средство «Данные» → «Сортировка».

3. Выполнить «Данные» → «Промежуточные Итоги».

4. Заполнить диалоговое окно «Промежуточные Итоги».

5. В результате при каждом изменении в столбце «Категории» будет выполнена операция «Сумма» при подведении итогов для данных, содержащихся в столбцах «Количество» и «Стоимость».

6. Ок.

Результат подведения итогов приведен на рисунке 1.17.

Щелкая поочередно по знакам структуры 1–2–3, можно создать итоговый отчет, скрыв подробности и отображив только итоги. Например, щелкнув мышью по цифре 2, получим только общие итоги по категориям (рис. 1.18).

1	2	3	A	B	C	D
			Клиент	Категория	Количество	Стоимость
1						
2	•		Петров	Кондитерские изделия	36	13122
3	•		Майоров	Кондитерские изделия	12	942,36
4	•		Томин	Кондитерские изделия	12	942,36
5	•		Кротов	Кондитерские изделия	20	7290
6	•		Кротов	Кондитерские изделия	36	13122
7	–			Кондитерские изделия Итог	116	35418,72
8	•		Петров	Молочные продукты	60	9240
9	•		Петров	Молочные продукты	60	9240
10	•		Майоров	Молочные продукты	60	9240
11	•		Кротов	Молочные продукты	48	11880
12	–			Молочные продукты Итог	228	39600
13	•		Кротов	Мясо/птица	96	3218,88
14	–			Мясо/птица Итог	96	3218,88
15	•		Майоров	Напитки	204	17442
16	•		Майоров	Напитки	168	13608
17	•		Томин	Напитки	156	184977
18	•		Томин	Напитки	132	8316
19	•		Кротов	Напитки	156	184977
20	–			Напитки Итог	816	409320
21	•		Петров	Приправы	216	20753,28
22	•		Томин	Приправы	228	25650
23	–			Приправы Итог	444	46403,28
24	•		Кротов	Рыбопродукты	264	7128
25	–			Рыбопродукты Итог	264	7128
26	–			Общий итог	1964	541088,88
27						

Рис. 1.17
Результаты подведения итогов

1	2	3	A	B	C	D	
			1	Клиент	Категория	Количество	Стоимость
+			7		Кондитерские изделия Итог	116	35418,72
			12		Молочные продукты Итог	228	39600
			14		Мясо/птица Итог	96	3218,88
			20		Напитки Итог	816	409320
			23		Приправы Итог	444	46403,28
			25		Рыбпродукты Итог	264	7128
			26		Общий итог	1964	541088,88

Рис. 1.18
Итоги второго уровня

ТЕМА 1.3. КОНТРОЛЬ ДАННЫХ

При наборе данных возможно появление ошибок, особенно при подготовке больших таблиц. Для контроля вводимой информации в Excel предусмотрены различные инструменты. В частности, для этих целей можно использовать условное форматирование.


1.3.1. КОНТРОЛЬ ВВОДА ДАННЫХ

При работе с данными во многих случаях критически важным является правильность их ввода. Например, вы можете неправильно ввести данные о заработной плате, неверно указать фамилию клиента или дату его рождения.

В Excel имеются средства контроля ввода данных, которые можно закрепить за определенными ячейками таблицы. Они доступны через меню «Данные» → «Работа с данными» → «Проверка данных».

Пример 1.11. Число должно попасть в заданный интервал. Проверка вводимых значений.

Требуется заполнить столбец «Возраст» в таблице (рис. 1.19). При заполнении необходимо осуществить контроль вводимого значения, так как поликлинаника детская, возраст пациентов ограничен, он может быть меньше или равен 17. Если введено число, большее 17, должно последовать сообщение о переводе пациента во взрослую поликлинику.



	А	В	С
1			
2	№	Ф.И.О.	Возраст (лет)
3	1	Петров И.А.	
4	2	Сидоров П.Р.	
5	3	Панин О.О.	
6	4	Колосова П.Р.	
7			

Рис. 1.19
Исходные данные

Сначала установить средство проверки вводимых значений.

1. Выделить ячейки С3:С6.
2. Выполнить команды «Данные» → «Работа с данными» → «Проверка данных». Появится диалоговое окно «Проверка вводимых значений» с тремя вкладками.

3. На вкладке «Параметры» установите:

- «Тип данных» — Целое число;
- «Значение» — между;
- «Минимум» — 0;
- «Максимум» — 17.

4. На вкладке «Сообщение для ввода» установить флажок «Отображать подсказку», если ячейка является текстовой и задайте:

- «Заголовок» — Ввод числа лет;
- «Сообщение» — Введите число в диапазоне 1–17.

5. На вкладке «Сообщение об ошибке» установить флажок «Выводить сообщение об ошибке» и указать тип действий при ошибке:

«Вид» — Останов; «Заголовок» — Ввод числа; «Сообщение» — Данный пациент вышел из детского возраста.

6. Попробовать ввести допустимое и недопустимое число.

При попытке ввести возраст, выходящий за установленные рамки, будет выдано предупреждающее сообщение (рис. 1.20). У пользователя есть возможность исправить свою ошибку.

Проверка данных необходима, если есть ограничения на элементы списка (список отдела или класса) или

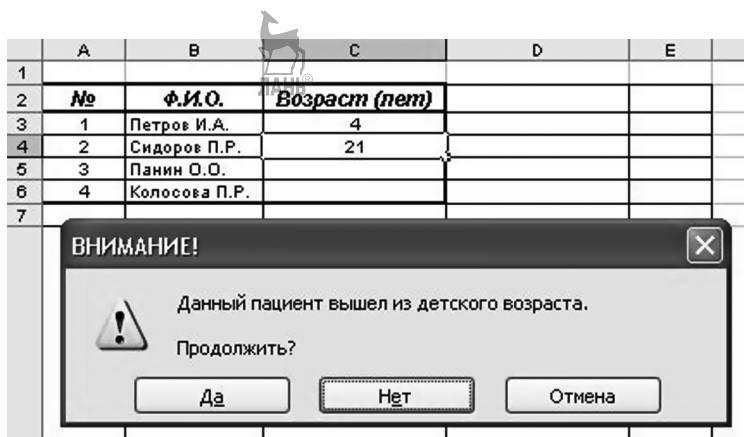


Рис. 1.20
Проверка вводимого значения

на количество символов, которые можно вводить в ячейку (номер телефона, табельный номер).

Пример 1.12. Вводимые данные могут быть только из установленного списка. Предположим, в некоторую ячейку может быть введена фамилия только сотрудника данного отдела в строго определенном формате. Нужно выполнить «Данные» → «Работа с данными» → «Проверка данных», установить «Тип данных» — Список, в поле «Источник» указать адрес блока, где находятся фамилии, или сами фамилии в нужном формате, отделяя их точкой с запятой: Иванов И. И.; Петров А. А.; Сидоров К. К. Теперь при активизации ячейки рядом справа от нее возникнет кнопка вызова сформированного списка, из которого можно выбрать нужную фамилию (рис. 1.21).

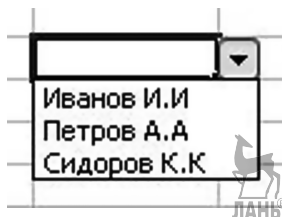


Рис. 1.21
Пример использования списка

Такой механизм не только гарантирует правильность и унификацию отображения фамилий, но и облегчит пользователю ввод данных. Достаточно выбрать нужную фамилию из предлагаемого списка.

1.3.2. УСЛОВНОЕ ФОРМАТИРОВАНИЕ

При анализе данных большое значение имеет их визуализация. Часто возникает необходимость выделить нетипичные значения, например, повторяющиеся или, напротив, уникальные. Для этого удобно использовать различные форматы представления данных в ячейке в зависимости от значений, содержащихся в них данных. В Excel для этого можно использовать *условное форматирование*,



Рис. 1.22
Условное форматирование

которое помогает дать визуальные ответы на определенные вопросы о данных.

Условное форматирование можно применить к диапазону ячеек, ко всей таблице или отчету сводной таблицы. Исходя из названия, можно сказать, что условное форматирование — это своего рода выделение ячеек, удовлетворяющих некоторым условиям.

Команда «Условное форматирование» находится во вкладке «Главная», в группе «Стили».

Условное форматирование можно применить, используя новые средства визуализации, имеющиеся в последних версиях Excel: гистограммы, цветовые шкалы и наборы значков, а также правила отбора и выделения ячеек (рис. 1.22).

Средства визуализации данных. Рассмотрим применение средств визуализации условного форматирования. Прежде всего, необходимо выделить ячейки, к которым будет применено условное форматирование.

1. Гистограммы используются для сравнения значений в выделенном диапазоне. С помощью гистограмм можно соотнести величине значения каждой ячейки цветной столбец, длина которого будет зависеть от этого значения. Большему значению будет соответствовать более длинный столбец. Можно управлять цветом столбца, использовать градиентную или сплошную заливку.

2. Цветовые шкалы помогают сравнить разброс значений анализируемых данных. Шкала может быть одно-, двух- и трехцветной. Яркость цвета возрастает с ростом значений. Например, если выбрана шкала «красный — желтый — зеленый», большим значениям будет соответствовать красный цвет, средним значениям — желтый, а меньшим значениям — зеленый.

3. Наборы значков используются для классификации данных по трем-пяти градациям. Различные значки позволяют проследить динамику изменения данных (стрелки), провести относительное сравнение, которое отражает долю рассматриваемых компонентов, как части целого (кружок с секторами), позиционное сравнение (гистограмма в ячейке).

Пример 1.13. Дана таблица основных параметров планет Солнечной системы (табл. 1.4).

Проанализируем распределение параметров, используя условное форматирование.

Для столбца «Расстояние» используем «Гистограмму». Для этого необходимо выполнить последовательность действий.

1. Выделить диапазон ячеек столбца «Расстояние».
2. Выбрать «Главная» → «Стили» → «Условное форматирование».
3. В появившемся меню выбрать пункт «Гистограммы».
4. В появившемся подменю выбрать синюю палитру.
5. Ok.

Аналогично применим средства визуализации для других столбцов. Для столбца «Масса» — Цветовую шкалу «зеленый — желтый — красный», для столбцов «Период», «Диаметр» и «Спутники» — Наборы значков. Результат приведен на рисунке 1.23. Теперь можно проанализировать распределение параметров. Видно, что планеты расположены в порядке увеличения расстояния от Солнца, самые большие диаметры у планет Юпитер и Сатурн и т. д. Условное форматирование обеспечивает наглядность при анализе данных. К сожалению, черно-белый вариант издания не позволят увидеть работу «цвета».

Планеты Солнечной системы

Таблица 1.4

№	Планета	Период	Расстояние	Диаметр	Масса	Спутники
1	Меркурий	0,241	58	4,9	0,32	0
2	Венера	0,615	106	12,1	4,86	0
3	Земля	1	150	12,8	6	1
4	Марс	1,881	288	6,8	0,61	2
5	Юпитер	11,86	778	142,6	1906,98	16
6	Сатурн	29,46	1426	120,2	570,9	17
7	Уран	84,01	2869	49	87,24	14
8	Нептун	164,8	4496	50,2	103,38	2

№	Планета	Период	Расстояние	Диаметр	Масса	Спутники
1	Меркурий	↓ 0,241	58	○ 4,9	0,32	0
2	Венера	↓ 0,615	106	○ 12,1	4,86	0
3	Земля	↓ 1,000	150	○ 12,8	6	1
4	Марс	↓ 1,881	288	○ 6,8	0,61	2
5	Юпитер	↓ 11,860	778	● 142,6	1906,98	16
6	Сатурн	↓ 29,460	1426	● 120,2	570,9	17
7	Уран	→ 84,010	2869	☾ 49	87,24	14
8	Нептун	↑ 164,800	4496	☾ 50,2	103,38	2

Рис. 1.23

Средства визуализации условного форматирования

Использование правил отбора и выделения ячеек. Правила задают ряд условий, которым должны удовлетворять данные в ячейке.

Правила выделения ячеек (рис. 1.24) позволяют применять форматирование к ячейке, выполнив сравнение содержимого этой ячейки с некоторым эталоном с помощью операторов сравнения (больше, меньше, между, равно).

Форматирование позволяет задавать шрифт, заливку, границы ячейки по усмотрению пользователя. Для одного и того же диапазона можно применять несколько правил условного форматирования.

Пример 1.14. Дан журнал успеваемости. Требуется выделить все пятёрки красным цветом с заливкой (см. табл. 1.5).

Для этого нужно сделать следующее.

1. Выделить диапазон ячеек, содержащий оценки.
2. Выполнить «Главная» → «Стили» → «Условное форматирование».

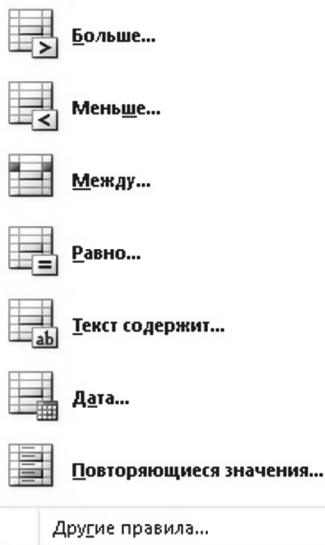


Рис. 1.24

Правила выделения ячеек

Таблица 1.5

Журнал успеваемости

Воробьев	3				5		2	
Голубев	3			3	3			
Журавлев		5			3			
Петухов		3		3	2			3
Синицын	4		2		4			

3. В появившемся меню выбрать «Правила выделения ячеек» → «Равно».

4. В появившемся подменю (рис. 1.25) установить значения.

5. Ок.

Результат отформатированного журнала приведен в таблице 1.6.

Выделение повторяющихся или уникальных значений. При работе с большими таблицами возникает необходимость выделения ячеек с повторяющимися значениями

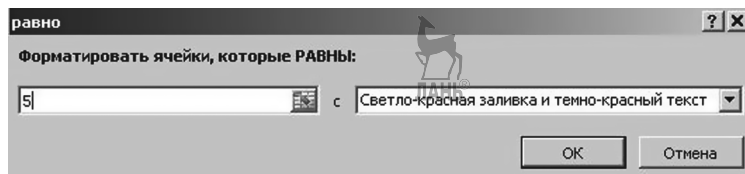


Рис. 1.25
Установка параметров

Таблица 1.6

Журнал успеваемости

Воробьев	3				5		2	
Голубев	3			3	3			
Журавлев		5			3			
Петухов		3		3	2			3
Синицын	4		2		4			
Сорокин	4		5		4	4		3

или выделить уникальные значения. Для этого нужно выполнить следующие шаги.

1. Выделить таблицу или диапазон ячеек.

2. Выполнить «Главная» → «Стили» → «Условное форматирование».

3. В появившемся меню выбрать «Правила выделения ячеек» → «Повторяющиеся значения» (рис. 1.24).

4. В открывшемся окне выбрать значение «Повторяющиеся» и указать вариант форматирования.

5. Ok.

Правила отбора (рис. 1.26) позволяют форматировать ячейку на основе правил отбора. Например, указанное количество первых или последних ячеек; содержащих значения выше или ниже среднего и т. д.

Пример 1.15. Дан журнал успеваемости. Требуется выделить все оценки, которые выше среднего значения (табл. 1.7).

Выполняется аналогично предыдущей задаче.

1. Выделить диапазон ячеек, содержащий оценки.

2. Выбрать «Главная» → «Стили» → «Условное форматирование».

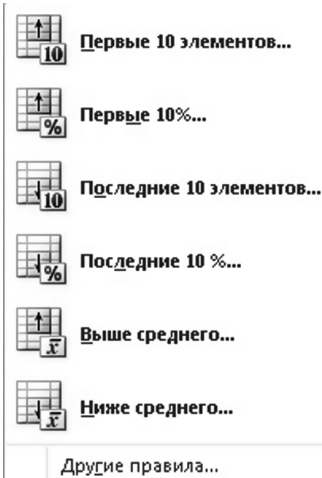


Рис. 1.26
Правила отбора первых
и последних значений

Таблица 1.7

Журнал успеваемости

Воробьев	3				5		2	
Голубев	3		3		3			
Журавлев	5				3			
Петухов	3			3	2			3
Синицын	4		2		4			
Сорокин	4	5		4	4			3

4. В появившемся подменю выбрать нужное форматирование.

Заметим, что форматирование можно проводить не только для ячеек с числовыми значениями, но и с текстовыми.

Создание собственных правил форматирования. Можно создать собственные правила форматирования. Для этого необходимо воспользоваться командой «Главная» → «Стили» → «Условное форматирование» → «Создать правило» (см. рис. 1.22). При этом откроется диалоговое окно «Создание правила форматирования».

В этом диалогом окне нужно установить «Тип» правила и соответствующее ему «Форматирование».

Если к ячейке уже было применено правило, то окно будет называться «Изменение правила форматирования».

Рассмотрим, например, полученную ранее таблицу, к которой было применено условное форматирование (см. рис. 1.23).

Выделим диапазон ячеек, содержащий данные.

Выполним «Главная» → «Стили» → «Условное форматирование» → «Управление правилами». Откроется диалоговое окно «Диспетчер правил условного форматирования» (рис. 1.27).

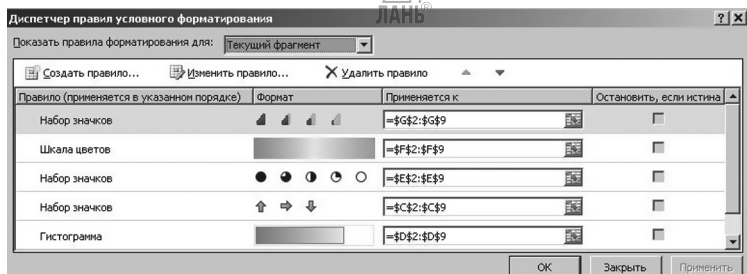


Рис. 1.27
Список правил установленного форматирования



В списке отображены все правила, примененные ранее к диапазону. Эти правила можно изменить, удалить или добавить новые.

Для отмены всего условного форматирования следует воспользоваться командой «Главная» → «Стили» → «Условное форматирование» → «Удалить правила» и выбрать в подменю нужный пункт.

Можно осуществить поиск ячеек с условным форматированием. Для этого выполните последовательность действий: «Главная» → «Найти и выделить» → «Условное форматирование».





ГЛАВА 2

ОБРАБОТКА МНОГОМЕРНЫХ ДАННЫХ



Представление данных в виде таблиц или списков значительно облегчает работу с многомерными данными, позволяя проводить над ними однотипные преобразования — сортировку, фильтрацию, подведение итогов, которые можно отнести к информационно-поисковому уровню обработки данных.

Необходимость проведения оперативно-аналитической обработки данных ставит новые задачи и требует новых инструментов. В этой главе будут рассмотрены технологии оперативной аналитической обработки данных, использующие методы и средства для сбора, хранения и анализа многомерных данных.

Разберем вопросы построения баз данных в Excel и работу со сводными таблицами.

ТЕМА 2.1. БАЗЫ ДАННЫХ В EXCEL

2.1.1. СТРУКТУРА БАЗЫ ДАННЫХ

Основу базы данных обычно составляет совокупность взаимосвязанных таблиц, на основе которых путем запроса можно создавать новые таблицы, осуществлять операции поиска, отбора, сортировки данных. Создание различных отчетов и форм на основе базы данных помогает решать задачи управления данными.

Excel, в котором данные изначально представлены в таблицах, обладающий к тому же встроенными средствами обработки данных — фильтрацией, сортировкой, подведением итогов, удобно и просто использовать для организации и поддержки небольших баз данных пользователей.

Список (или таблицу) в Excel можно трактовать как большую одноуровневую базу данных. *Запись* (строка в таблице) рассматривается как одно наблюдение, действие или результат. *Поля* (столбцы в таблице) соответствуют различным показателям или признакам, характеризующим соответствующую запись.

При обработке экономических, социологических, психологических, исторических данных запись имеет большое количество полей (столбцов в таблице), которые соответствуют различным показателям или признакам. Поэтому исходные данные необходимо представить в виде нескольких взаимосвязанных таблиц, которые и послужат основой для базы данных.

Создадим для примера небольшую базу данных оптовой торговой базы «Игрушки». Ее основу будут составлять три таблицы — «Покупатели», «Товары», «Продавцы». Каждая таблица размещается на отдельном рабочем листе с таким же названием.

Исходные таблицы представлены на рисунках 2.1–2.3.

Таблицы в базе данных должны быть взаимосвязаны, для этого используется так называемое *ключевое* поле или ключ, который является уникальным идентификатором записи.

	А	В	С	Д	Е	Ф	Г
1	КодП	Фамилия	Имя	Отчество	Адрес	День рождения	Пол
2	1	Воробьев	Иван	Петрович	СПб	10.01.1988	м
3	2	Голубева	Мария	Ивановна	Самара	11.11.1989	ж
4	3	Журавлев	Сергей	Николаевич	Рязань	12.02.1986	м
5	4	Петухов	Илья	Сергеевич	СПб	14.01.1985	м
6	5	Синицына	Марина	Петровна	Новгород	16.03.1988	ж
7	6	Сорокина	Зоя	Федоровна	Мурманск	15.05.1983	ж

Рис. 2.1
Таблица «Продавцы»

	А	В	С	Д
1	КодК	Наименование	Адрес	Контактный телефон
2	1	Лента	Выборгское ш., д.216	111-11-11
3	2	Карусель	Карбышева ул., д.9	222-11-12
4	3	Метро	Богатырский пр., 25	333-11-13
5	4	Окей	Испытателей ул., д.27	444-11-14

Рис. 2.2
Таблица «Покупатели»

	А	В	С	Д
1	Артикул	Наименование	Цена	Производитель
2	1001	Джип	80	Китай
3	1015	Трейлер	90	Китай
4	1020	Самосвал	140	Россия
5	1025	Бульдозер	100	Россия
6	2010	Катер	65	Китай

Рис. 2.3
Таблица «Товары»

	А	В	С	Д	Е	Ф
1	№П	Дата	КодП	КодК	Артикул	Единицы
2	1	01.01.2010	5	4	1025	18
3	2	01.01.2010	6	1	1025	30
4	3	01.01.2010	6	2	1001	7
5	4	01.01.2010	1	2	2010	91
6	5	01.01.2010	1	3	1020	5
990	989	30.12.2010	6	3	1015	2
991	990	30.12.2010	6	4	1015	29

Рис. 2.4
Таблица «Продажи»

В рассмотренных таблицах ключевыми полями являются: код продавца «КодП», код клиента «КодК» и артикул товара «Артикул».

Кроме основных трех таблиц, в базу данных входит таблица «Продажи», которая представляет ежедневный отчет о продажах. Всего в таблице содержится 990 записей о продажах за 2010 год. На рисунке 2.4 представлена

только часть записей этой таблицы, остальные строки скрыты для просмотра. Ключевое поле таблицы — номер сделки (продажи) № П. Количество проданных единиц товара содержится в столбце (поле) «Единицы». Для удобства в таблице указаны только значения ключей, которые обеспечивают связь с другими таблицами. Такое представление уменьшает количество ошибок при вводе исходных данных. Дополнительные средства контроля ввода данных рассмотрены в Теме 1.3.



2.1.2. ФОРМЫ В EXCEL

Для заполнения таблиц удобно использовать «Формы» — средство отображения данных на экране и управление ими. В Excel 2003 этот инструмент располагался в меню «Данные». Многие пользователи по привычке и в более поздних версиях Excel ищут их на вкладке «Данные». Однако в последних версиях «Форма» относится к группе «Команды не на ленте», которые можно при необходимости расположить на панели быстрого доступа (рис. 2.5).

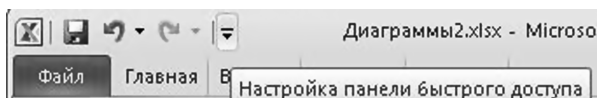


Рис. 2.5
Панель быстрого доступа



Рис. 2.6
Панель быстрого доступа
с «Формой»

Рис. 2.7
Форма «Продажи»

Для активизации «Формы» и размещения ее на панели быстрого доступа нажмите на стрелку «Настройка панели быстрого доступа» → «Другие команды...» → выберите «Команды не на ленте» → выделите «Формы» → нажмите «Добавить» → Ок. Кнопка «Форма» появится на панели быстрого доступа (рис. 2.6).

Для использования «Формы» выделите любую ячейку таблицы «Продажи» и нажмите значок «Формы». Откроется форма, в которой поля расположены в ленточном виде, удобно добавлять, удалять, корректировать записи (рис. 2.7).

2.1.3. ПОДДЕРЖКА БАЗЫ ДАННЫХ

На основе первичных таблиц «Продавцы», «Клиенты», «Товары», «Продажи» можно создавать различные производные таблицы, формировать отчеты.

Создадим, например, таблицу «Продажи2010». Вставим новый лист «Продажи2010», скопируем на этот лист содержимое листа «Продажи», вставим новые столбцы «Продавец», «Клиент», «Продукт», «Сумма» (рис. 2.8).

В поле «Продавец» необходимо вставить «Фамилию» продавца из таблицы «Продавцы», в поле «Клиент» — Наименование клиента из таблицы «Клиент», в поле «Товар» — Наименование товара из таблицы «Товары».

В столбце (поле) «Сумма» вычисляется общая сумма стоимости продаж, соответствующая номеру продаж текущей строки (записи). При вычислении суммы используется поле «Цена» таблицы «Товары», в которой указана цена единицы товара.

	A	B	C	D	E	F	G	H	I	J
1	№п	Дата	КодП	Продавец	КодК	Клиент	Артикул	Товар	Единицы	Сумма
2	1	01.01.2010	5		4		1025		18	
3	2	01.01.2010	6		1		1025		30	
4	3	01.01.2010	6		2		1001		7	
5	4	01.01.2010	1		2		2010		91	
6	5	01.01.2010	1		3		1020		5	
990	989	30.12.2010	6		3		1015		2	
991	990	30.12.2010	6		4		1015		29	

Рис. 2.8
Таблица «Продажи2010»

Таким образом, таблица «Продажи2010» взаимосвязана с остальными таблицами. Взаимосвязь между таблицами в Excel осуществляется на основе ключевого поля (ключа) с помощью встроенной функции ВПР, которая относится к группе команд «Ссылки и массивы».

Функция ВПР ищет указанное значение («Искомое_значение») из первой таблицы в первом столбце второй таблицы («Таблица») и возвращает значение ячейки, находящейся в указанном столбце («Номер_столбца») той же строки. По умолчанию вторая таблица должна быть отсортирована по возрастанию.

Синтаксис функции ВПР следующий:

ВПР(Искомое_значение;Таблица;Номер_столбца;Интервальный_просмотр).

Искомое_значение — это значение, которое должно быть найдено в первом столбце «Таблицы», оно может быть числом, ссылкой или текстовой строкой.

Таблица — таблица, в которой в первом столбце ищется *Искомое_значение*. Можно использовать ссылку на таблицу или имя таблицы, например «База Данных». В этой же таблице в указанном столбце находится значение, соответствующее искомому значению.

Столбец — это номер столбца в «Таблице», в котором должно быть найдено значение, соответствующее искомому.

Если указан номер столбца меньше 1, то функция ВПР возвращает значение ошибки #ЗНАЧ!; если номер «столбец» больше, чем количество столбцов в аргументе «таблица», тогда функция ВПР возвращает значение ошибки #ССЫЛ!.

Интервальный_просмотр — принимает значение ИСТИНА или ЛОЖЬ. Если этот аргумент имеет значение ИСТИНА (можно использовать числовое значение 1) или опущен, то возвращается приблизительно соответствующее значение; другими словами, если точное соответствие не найдено, то возвращается наибольшее значение, которое меньше, чем Искомое_значение. Если этот аргумент имеет значение ЛОЖЬ (можно использовать числовое значение 0), то функция ВПР ищет точное соответствие. Если

такое не найдено, то возвращается значение ошибки #Н/Д. Заметим, что этот аргумент необязателен, по умолчанию принимается значение ИСТИНА. Для работы с базами данных в Excel следует указывать значение этого аргумента ЛОЖЬ (или 0), так как искомое значение чаще всего является уникальным ключом и требуется поиск точного соответствия.

Заполнить таблицу «Продажи2010» можно двумя способами.

СПОСОБ 1. КОПИРОВАНИЕ «ПРОТЯГИВАНИЕМ»

Заполните поля «Продавец», «Клиент», «Товар» и «Сумма» в первой строке таблицы «Продажи2010», используя функцию ВПР.

В поле «Продавец» требуется вставить «Фамилию» продавца, имеющего указанный код «КодП». Для этого в ячейку D2 необходимо поместить нужную формулу. Выполните последовательность шагов для выполнения данной задачи.

1. Выделить ячейку D2.
2. Вызвать «Мастера функций» (кнопка fx).
3. Указать категорию «Ссылки и массивы».
4. Выбрать функцию ВПР.
5. Ок, появится диалоговое окно для заполнения аргументов функции.
6. Для удобства заполнения аргументов переместить оторванное окно «Аргументы» функции вправо.
7. В открывшемся окне задать аргументы функции ВПР:

а) Искомое_значение — переместить курсор в поле первого аргумента, щелкните по ячейке C2, которая содержит код Продавца «КодП»;

б) Таблица — переместить курсор в поле второго аргумента, щелкнуть по названию листа «Продавцы», в открывшемся листе выделить два первых столбца — «КодП» и «Фамилия», щелкнуть F4, относительные ссылки превратятся в абсолютные, это необходимо для последующего копирования формулы;

в) Номер_столбца — ввести 2, так как во втором столбце указанной таблицы содержится фамилия;

г) Интервальный_просмотр — указать 0, так как требуется найти точное соответствие;

д) Ok.

В ячейке D2 отобразится фамилия Синицына.

Аналогично в ячейку F2 поместите формулу, которая вернет «Наименование» клиента, а в ячейку H2 — «Наименование» товара. В ячейку J2 поместить формулу Цена*Единицы, где «Цена», соответствующая текущему «Артикулу», «вытаскивается» из таблицы «Товары» с помощью функции ВПР. Сначала введите вычисление «Цены» с помощью функции ВПР, а затем в строке команд введите знак умножения и щелкните по ячейке I2. В результате в ячейки первой строки таблицы будут введены формулы (табл. 2.1).

Остальные строки можно заполнить «протягиванием». Для удобства можно работать с таблицей, в которой отображаются только несколько первых и последних строк, остальные строки скрыты. Результат показан на рисунке 2.9. Столбцы с кодами можно скрыть.

Таблица 2.1

Расчетные формулы

Адрес ячейки	Формула	Значение
D2	=ВПР(C2;Продавцы!\$A\$1:\$B\$7;2;0)	Синицына
F2	=ВПР(E2;Клиенты!\$A\$1:\$B\$5;2;0)	Окей
H2	=ВПР(G2;Товары!\$A\$1:\$B\$6;2;0)	Бульдозер
J2	=ВПР(G2;Товары!\$A\$1:\$C\$6;3;0)*I2	1800

	A	B	C	D	E	F	G	H	I	J
1	№п	Дата	Кодп	Продавец	КодК	Клиент	Артикул	Товар	Единицы	Сумма
2	1	01.01.2010	5	Синицына	4	Окей	1025	Бульдозер	18	1800
3	2	01.01.2010	6	Сорокина	1	Лента	1025	Бульдозер	30	3000
4	3	01.01.2010	6	Сорокина	2	Карусель	1001	Джип	7	560
5	4	01.01.2010	1	Воробьев	2	Карусель	2010	Катер	91	5915
6	5	01.01.2010	1	Воробьев	3	Метро	1020	Самосвал	5	700
990	989	30.12.2010	6	Сорокина	3	Метро	1015	Трейлер	2	180
991	990	30.12.2010	6	Сорокина	4	Окей	1015	Трейлер	29	2610

Рис. 2.9
Таблица Продажи2010

Для «протягивания» достаточно выделить первую ячейку в столбце, например D2, поместить курсор в правый нижний угол ячейки пока не появится значок «+» и дважды щелкнуть левой кнопкой, нижние ячейки заполнятся автоматически.

СПОСОБ 2. ИСПОЛЬЗОВАНИЕ ТАБЛИЦ

Заполнение таблиц можно автоматизировать, используя новые свойства, которые получили таблицы в последних версиях Excel. Подготовленную таблицу (рис. 2.8) превратим в таблицу (см. п. 1.1.3). Для этого выделите любую ячейку в таблице и выполните последовательность команд «Вставка» → «Таблицы» → «Таблица». Автоматически выделится вся область данных A1:J991, появится окно «Создание таблицы». Нажмите Ok (рис. 2.10).

В строке заголовков появятся указатели «Фильтра», при желании их можно снять: «Данные» → «Фильтр» (рис. 2.11). В правом нижнем углу отобразится автоматически перемещающийся маркер границы таблицы.

Введите формулу в ячейку D2, как было описано в Способе 1. Теперь после ввода формула этой ячейки автоматически распространится на всю таблицу, и не понадобится лишнее действие — «протягивание». Аналогично введите формулы в ячейки F2, H2, J2.

В результате в ячейки первой строки таблицы будут введены формулы (табл. 2.2).

К полученной таблице можно применять команды, расположенные на вкладке «Данные» — «Фильтр», «Сортировка», а также осуществлять проверку данных.

Фильтр выполняет работу аналогичную запросам в базах данных, позволяя отобрать данные, удовлетворяющие определенным условиям отбора. Для создания сложных запросов используется «Расширенный фильтр».

К сожалению, для таблиц не работает средство «Подведение итогов». Чтобы им воспользоваться, приходится преобразовывать таблицу в диапазон «Работа с таблицами» → «Конструктор» → «Сервис» → «Преобразовать в диапазон». Если необходимо узнать объемы продаж и выручку

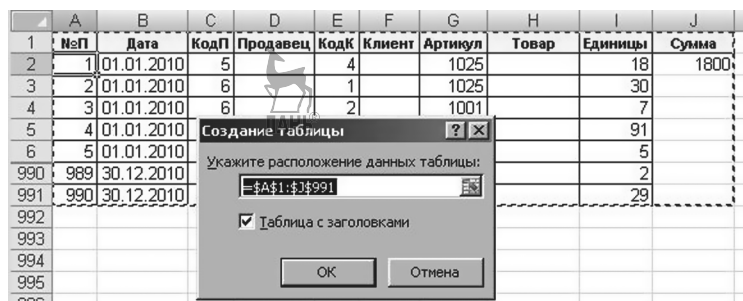


Рис. 2.10
Окно «Создание таблицы»

	A	B	C	D	E	F	G	H	I	J
1	№ *	Дата *	Код *	Продавец *	Код *	Клиент *	Артикул *	Товар *	Единиц *	Сумма *
2	1	01.01.2010	5		4		1025		18	
3	2	01.01.2010	6		1		1025		30	
4	3	01.01.2010	6		2		1001		7	
5	4	01.01.2010	1		2		2010		91	
6	5	01.01.2010	1		3		1020		5	
990	989	30.12.2010	6		3		1015		2	
991	990	30.12.2010	6		4		1015		29	

Рис. 2.11
«Фильтр»

Таблица 2.2

Расчетные формулы

Адрес ячейки	Формула
D2	=ВПР(Таблица1[[#Эта строка]; [КодП]]; Продавцы!\$A\$1:\$B\$7;2;0)
F2	=ВПР(Таблица1[[#Эта строка]; [КодК]];Клиенты!\$A\$1:\$B\$5;2;0)
H2	=ВПР(Таблица1[[#Эта строка]; [Артикул]];Товары!\$A\$1:\$B\$6;2;0)
J2	=ВПР(Таблица1[[#Эта строка]; [Артикул]];Товары!\$A\$1:\$C\$6;3;0)* Таблица1[[#Эта строка];[Единицы]]

каждого продавца, отсортируйте таблицу: «Данные» → «Сортировка» → «Продавец» → «Ok». Затем выполните «Данные» → «Структура» → «Промежуточные итоги». Выберите уровень структуры — «2» (рис. 2.12).

Работа с «Фильтром», «Сортировкой» и «Подведением итогов» описана более подробно в Теме 1.2.

1	2	3	А	В	Д	Е	Н	І	Ј
		1	№П	Дата	Продавец	Клиент	Товар	Единицы	Сумма
+		224			Воробьев Итог			10390	949000
+		450			Голубева Итог			10205	894965
+		547			Журавлев Итог			4047	368675
+		680			Петухов Итог			5633	507485
+		796			Синицына Итог			5963	532015
+		997			Сорокина Итог			9022	824420
-		998			Общий итог			45260	4076560

Рис. 2.12
Таблица Продажи2010

ТЕМА 2.2. СВОДНЫЕ ТАБЛИЦЫ

2.2.1. НАЗНАЧЕНИЕ СВОДНЫХ ТАБЛИЦ

Сводные таблицы (PivotTables) в Excel, являются в какой-то мере аналогами кубов данных, реализованных на основе OLAP-технологии (Online Analytical Processing) в базах данных, и мобильных таблиц в статистических пакетах (например, в SPSS), которые предназначены для работы с многомерными данными.

Суть заключается в том, что данные представляются в виде многомерного куба с возможностью дальнейшего произвольного манипулирования ими. Они дают возможность проводить анализ структуры взаимосвязей различных показателей, обрабатывать большие объемы информации, проводить анализ в различных разрезах и плоскостях.

Анализ данных на основе сводных таблиц позволяет аналитику получить ответы на поставленные вопросы. Например, «Какие товары были наиболее продаваемыми

в четвертом квартале текущего года?», «Кто из продавцов лидирует по объему продаж в декабре?».

Сводные таблицы дают также возможность заниматься проверкой возникающих у аналитика гипотез. Распространенным видом аналитических запросов здесь является анализ по принципу «что, если?». Например, «Насколько надо увеличить расходы на рекламу, чтобы прибыль выросла на 10%?».

Каждая последующая версия Excel добавляет новые возможности для этого замечательного инструмента. При этом меняются некоторые технологии, к которым пользователь уже привык. В данной теме рассмотрим работу со сводными таблицами в формате Excel 2010. Сделаем первый шаг на пути освоения этого средства для работы с многомерными данными. Тем, кто захочет более подробно ознакомиться со сводными таблицами, порекомендуем книгу [2].

2.2.2. СОЗДАНИЕ СВОДНЫХ ТАБЛИЦ

Сводная таблица создается на основе таблицы (списка) данных. Инструмент находится на вкладке «Вставка» в группе «Таблицы».

В качестве источника данных для формирования сводных таблиц можно использовать базу данных Excel.

Шаг 1. Переход к работе со сводными таблицами. Для создания сводной таблицы: установите курсор в любую ячейку таблицы исходных данных; выполните «Вставка» → «Таблицы» → «Сводная таблица» (в открывшемся окне диапазон для таблицы будет указан автоматически) → укажите, что отчет сводной таблицы следует поместить «На новый лист» → Ok.

В результате откроется новый лист для формирования сводной таблицы, на ленте появится дополнительная контекстная вкладка «Работа со сводными таблицами», которая включает две группы: «Параметры» и «Конструктор» (рис. 2.13). Контекстная вкладка будет отображаться при выделении любой ячейки сформированной сводной таблицы.

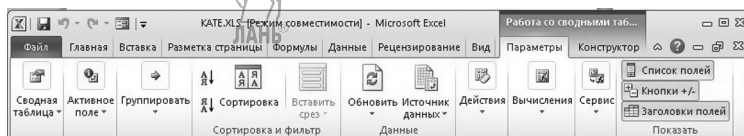


Рис. 2.13
Вкладка «Работа со сводной таблицей»

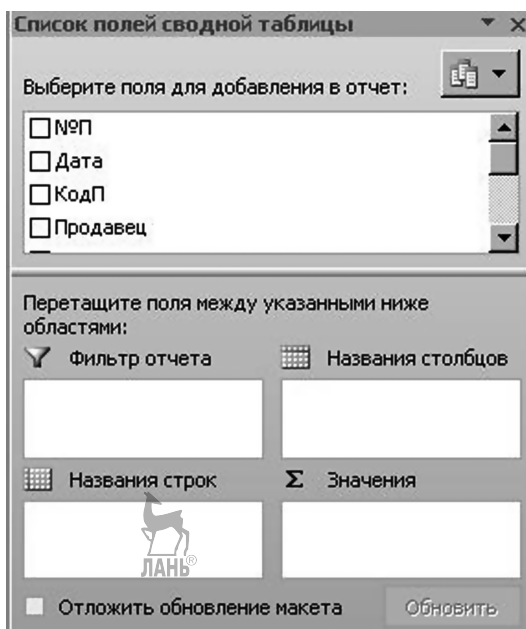


Рис. 2.14
Список полей и областей сводной таблицы

Шаг 2. Формирование макета сводной таблицы. На созданном новом листе справа размещено диалоговое окно, с помощью которого формируется логическая структура (макет) сводной таблицы (рис. 2.14).

Диалоговое окно разделено на две части — раздел полей и раздел областей. Раздел полей содержит список заголовков полей исходной таблицы (списка) данных.

Раздел областей в свою очередь разделен на четыре части: «Фильтр отчета», «Названия строк», «Названия столб-

цов», «Значения». Первые три области определяют внешние поля, последняя область «Значения» определяет внутреннее поле.

Внешние области по сути дела определяют логическую структуру сводной таблицы, внутренняя область предназначена для вычислений.

Названия строк и названия полей задают плоскую (двумерную) структуру сводной таблицы, а «Фильтр» придает таблице трехмерность, обеспечивая постраничный просмотр сводной таблицы.

Для задания макета необходимо перетянуть поля при помощи мышки в нижнюю часть диалогового окна, определяя таким образом внутреннее и внешние поля. Для внутреннего поля сразу же определяется операция обработки, по умолчанию выбирается «Сумма по полю».

Структура сводной таблицы определяется поставленной задачей и вопросами по анализу исходных данных, которые ставит перед собой пользователь.

На наш взгляд, список областей для создания сводной таблицы в Excel 2003 был представлен более наглядно (рис. 2.15). Можно сказать, что область «Значение» («Данные») задает нам четвертое измерение, определяя значение в точке с координатами (строка, столбец, страница).

Вместо «Страницы» появился «Фильтр отчета». Каждое из названий по-своему отражает назначение области. «Страница» делает акцент на то, что сводный отчет можно листать по страницам, а «Фильтр» указывает, что эта область формируется с помощью фильтра, т. е. отбора данных, удовлетворяющих некоторым условиям.



Рис. 2.15
Список областей сводной
таблицы в Excel 2003

Приведем для сравнения, как выглядят сводные таблицы в статистическом пакете SPSS, здесь они называются «Мобильными таблицами» (рис. 2.16).

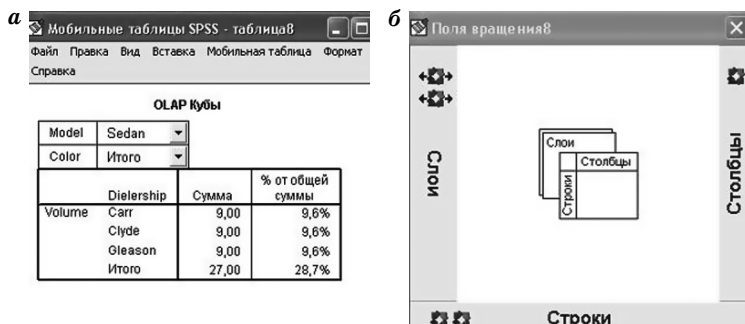


Рис. 2.16
«Мобильные таблицы» (а) и «Поля вращения» (б)

Шаг 3. Работа с макетом сводной таблицы. На этом шаге можно преобразовывать исходную сводную таблицу, меняя структуру путем различного расположения полей в областях, группировки данных, выбора формул в поле «Значение».

В сводной таблице можно проводить «Сортировку», «Промежуточные итоги», строить «Сводные диаграммы».





2.2.3. РАБОТА СО СВОДНЫМИ ТАБЛИЦАМИ

Сводные таблицы будем создавать на основе базы данных, рассмотренной в Теме 2.1 «Исходная таблица» — «Отчет2010» (рис. 2.9).

Пример 2.1. Требуется проанализировать спрос на товары и ответить на вопрос: «Какие товары и в каком объеме покупали Клиенты в 2010 году?».

Выполните Шаг 1, как указано выше. На Шаге 2 сформируйте макет сводной таблицы, удовлетворяющий поставленной задаче. Перетащите поля Клиент, Товар, Единицы в соответствующие области (рис. 2.17). По умолчанию в области Значения будет вычисляться Сумма по полю Единицы.

Перетащите поля между указанными ниже областями:

 Фильтр отчета	 Названия столбцов
Клиент	
 Названия строк	 Значения
Товар	Сумма по полю Единицы

☐ Отложить обновление макета Обновить

Рис. 2.17
Настройка областей

	А	В
1	Клиент	(Все)
2		
3		Сумма по полю Единицы
4	Самосвал	5229
5	Джип	8694
6	Бульдозер	10030
7	Трейлер	10441
8	Катер	10886
9	Общий итог	45260

Рис. 2.18
Сводная таблица

В результате будет получена сводная таблица (рис. 2.18). «Фильтр отчета» позволяет листать «страницы» нашей сводной таблицы. На каждой странице появится отчет по конкретному клиенту или по группе выбранных клиентов. Для удобства просмотра были убраны «Заголовки полей»: «Работа со сводными таблицами» → «Параметры» → «Показать и скрыть» → «Заголовки полей». Данные отсортированы: «Работа со сводными таблицами» → «Параметры» → «Сортировка».

В данном случае выведена общая сумма продаж единиц товаров по всем Клиентам.

Основа сводной таблицы создана. Теперь можно перейти к Шагу 3 и применить к таблице различные действия. Например, получить другой «Срез» — посмотреть продажи для конкретного клиента. Нажмите на стрелку в поле «Все» и выберите клиента «Карусель» (см. рис. 2.19).

	А	В
1	Клиент	Карусель
2		
3	Сумма по полю Единицы	
4	Самосвал	1410
5	Джип	2604
6	Трейлер	2610
7	Бульдозер	2763
8	Катер	2933
9	Общий итог	12320

Рис. 2.19
Выбор «Клиента»

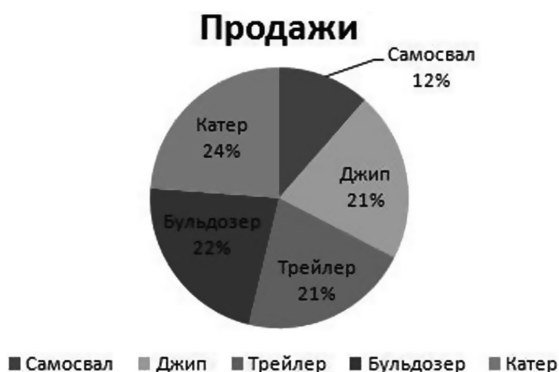


Рис. 2.20
Сводная диаграмма

Для наглядности можно на основе сводной таблицы можно сразу создать сводную диаграмму: «Работа со сводными таблицами» → «Параметры» → «Сервис» → «Сводная диаграмма» (рис. 2.20).

Полученную диаграмму можно отредактировать. При выделении сводной диаграммы появляется контекстная вкладка «Работа со сводными диаграммами», включающая вкладки «Конструктор», «Макет», «Формат» и «Анализировать». Первые три вкладки возникают при работе с любой диаграммой, а вкладка «Анализировать» — только для сводных диаграмм. На этой вкладке есть два инструмента: «Список полей» и «Фильтр сводной таблицы», которые используют для интерактивной работы с диаграммой.

Фильтр отчета для сводной диаграммы можно настраивать, как и для сводной таблицы, листая «страницы». Результаты выбора автоматически отобразятся на диаграмме (рис. 2.21).

Пример 2.2. Как были распределены продажи по продавцам?

Необходимо добавить к полученной сводной таблице продавцов. Для этого ее нужно трансформировать. Выполнить «Вращение» — перенести «Товар» в область столбцов, а в область строк поместить поле «Продавец» (рис. 2.22). Сводная таблица примет вид (рис. 2.23).

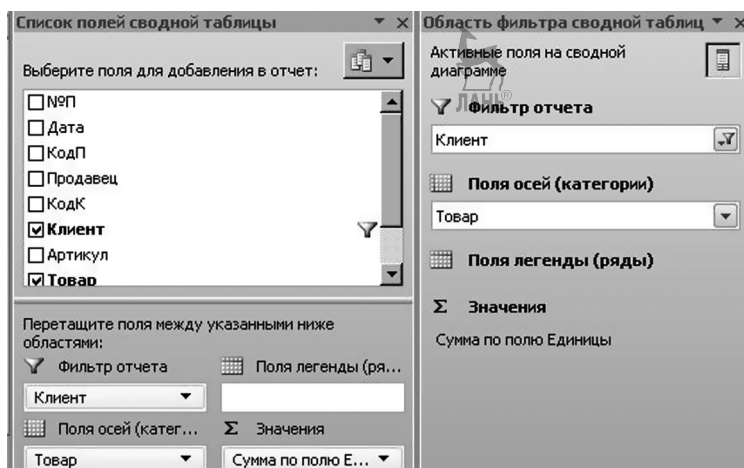


Рис. 2.21
Настройка фильтра отчета

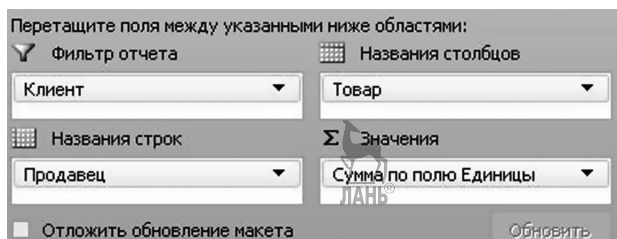


Рис. 2.22
Настройка областей

	A	B	C	D	E	F	G
1	Клиент	Карусель	✓				
2							
3	Сумма по полю Единицы						
4		Бульдозер	Джип	Катер	Самосвал	Трейлер	Общий итог
5	Воробьев	631	714	625	475	467	2912
6	Голубева	868	294	606	308	510	2586
7	Журавлев	272	140	301	92	457	1262
8	Петухов	90	391	386	266	382	1515
9	Синицына	433	415	622	176	87	1733
10	Сорокина	469	650	393	93	707	2312
11	Общий итог	2763	2604	2933	1410	2610	12320

Рис. 2.23
Сводная таблица

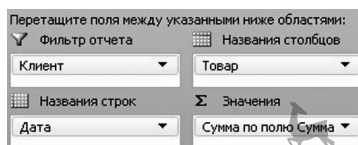


Рис. 2.24
Настройка областей

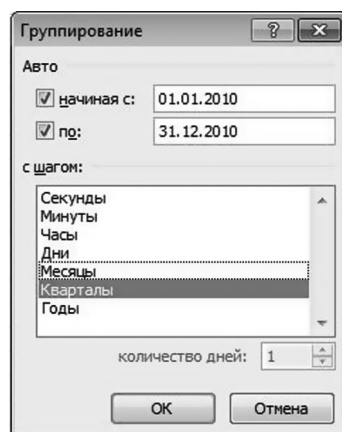


Рис. 2.25
Настройка группировки

	A	B	C	D	E	F	G
1	Клиент	Карусель	✓				
2							
3	Сумма по полю Единицы						
4		Самосвал	Джип	Трейлер	Бульдозер	Катер	Общий итог
5	Кв-л1	473	386	688	775	845	3167
6	Кв-л2	523	1206	427	985	843	3984
7	Кв-л3	161	823	765	729	479	2957
8	Кв-л4	253	189	730	274	766	2212
9	Общий итог	1410	2604	2610	2763	2933	12320

Рис. 2.26
Сводная таблица

Обратите внимание, что наименования Товаров отсортированы по алфавиту. Для этого нужно выделить любой товар и применить «Сортировку».

Пример 2.3. Теперь требуется проанализировать распределение продаж по кварталам. Для этого понадобится средство «Группировка», так в исходных данных даты продаж заданы по дням.

Необходимо сформировать новую структуру сводной таблицы (рис. 2.24), что делается очень быстро мышью.

Щелкнуть по любой дате в полученной сводной таблице. Выделится весь столбец. Выполнить «Работа со сводными таблицами» → «Параметры» → «Группировать» → «Группировка по выделенному» → откроется диалоговое окно «Группирование» (рис. 2.25) → выбрать «Кварталы» → Ок.

Результат выполнения группировки по кварталам приведен на рисунке 2.26.

Пример 2.4. Требуется сравнить прибыль от продажи товаров в 2010 году.

Для решения этой задачи создайте новую сводную таблицу, в которой отразите расчеты сумм продаж товаров по кварталам, и постройте по ней сводную диаграмму «Продажа по кварталам 2010 г.».

Структура сводной таблицы будет иметь вид (рис. 2.27), сводная таблица после группировки по кварталам изображена на рисунке 2.28.

Перетащите поля между указанными ниже областями:

<input checked="" type="checkbox"/> Фильтр отчета	<input type="checkbox"/> Названия столбцов
Клиент	Товар
<input type="checkbox"/> Названия строк	<input type="checkbox"/> Σ Значения
Дата	Сумма по полю Сумма

Рис. 2.27
Настройка областей

	A	B	C	D	E	F	G
1	Клиент	(Все)					
2							
3	Сумма по полю Сумма	Названия столбцов					
4	Названия строк	Бульдозер	Джип	Катер	Самосвал	Трейлер	Общий итог
5	Кв-л1	264500	152160	216580	212800	256950	1102990
6	Кв-л2	325400	274240	200915	241080	212130	1253765
7	Кв-л3	244700	167600	145210	152600	232470	942580
8	Кв-л4	168400	101520	143585	125580	238140	777225
9	Общий итог	1003000	695520	706290	732060	939690	4076560

Рис. 2.28
Сводная таблица

Из таблицы видно, что наибольшая прибыль получена от продажи Бульдозеров, а наименьшая от продажи Джипов. Постройте сводную диаграмму «Продажа по кварталам 2010 г.» только для двух товаров Бульдозер и Джип,

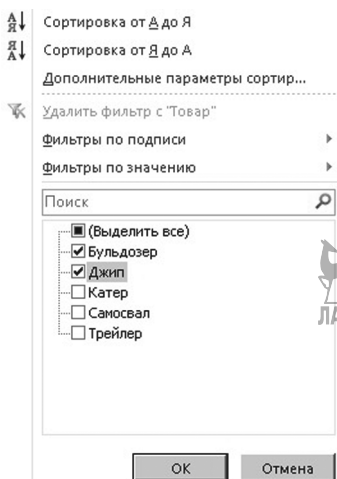


Рис. 2.29
Настройка фильтра

для сравнения разницы между объемами продаж.

Это можно сделать по-разному. Лучшее решение — использовать «Фильтр». На наличие «Фильтра» указывают стрелочки. Как видно из рисунка, доступны фильтры по «Названиям строк» и «Названиям столбцов». Нажмите стрелку фильтра по столбцам → в открывшемся окне установите нужные фильтры (рис. 2.29).

Создайте сводную диаграмму. Выделите любую ячейку в поле сводной таблицы → «Параметры» → «Свод-

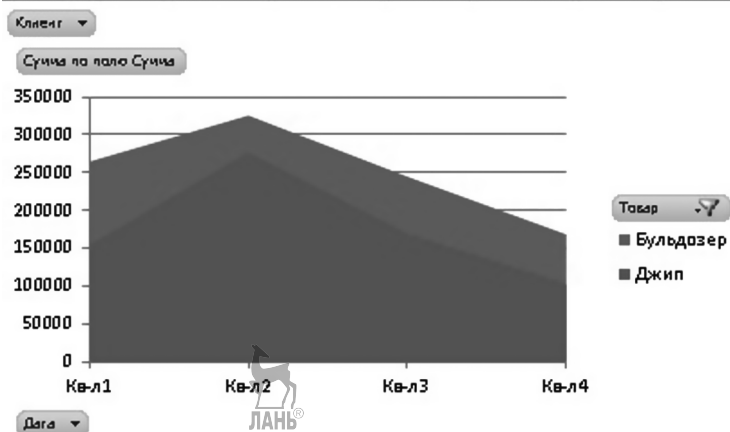


Рис. 2.30
Сводная диаграмма

ная диаграмма» → «Диаграмма с областями». Построенная сводная диаграмма изображена на рисунке 2.30. На ней, в отличие от обыкновенной диаграммы, имеются кнопки полей, которыми можно управлять, выбирая интересующий нас срез данных.

Сводная диаграмма является объектом. При ее выделении появляется дополнительная вкладка «Работа со сводными таблицами» (рис. 2.31). В дополнительной вкладке «Анализировать» можно убрать «Кнопки полей», они нужны только для настройки диаграммы, а для размещения диаграммы, например в отчет, кнопки лучше скрыть.

Вкладки «Макет» и «Формат» предназначены для форматирования диаграммы. Диаграмму легко можно привести к виду (рис. 2.32).

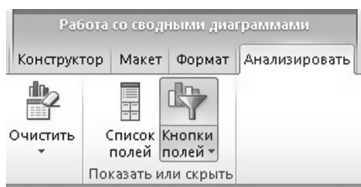


Рис. 2.31
Вкладка
«Работа со сводными таблицами»



Рис. 2.32
Диаграмма

2.2.4.

ПРИМЕНЕНИЕ «МАСТЕРА СВОДНЫХ ТАБЛИЦ»

Сводные таблицы, созданные в Excel 2007/2010, имеют свои особенности. Одна из этих особенностей — *связанность* различных сводных таблиц, созданных на основе одной и той же таблицы исходных данных.

Разобранные выше примеры 2.1–2.4 создавались на основе одной таблицы, расположенной на листе «Продажи2010». При таком построении все построенные сводные таблицы будут связаны между собой. Если, например, на листе со сводной таблицей примера 2.3 указать группировку не по кварталам, а по месяцам, тогда сводные таблицы, расположенные на других листах Excel, для примера 2.4 автоматически будут сгруппированы тоже по месяцам. Изменяться соответственно и построенные диаграммы.

Если же необходимо создать *независимую* сводную таблицу, нужно предварительно скопировать исходные данные на другой лист. Согласитесь, что это неудобно. В этом случае лучше воспользоваться уже известным инструментом Excel 2003 — «Мастером сводных таблиц». В новых версиях эта команда не размещается на ленте, ее, как и рассмотренную ранее «Форму», следует поместить на панель быстрого доступа.

Для активизации «Мастера сводных таблиц» нажмите на стрелку «Настройка панели быстрого доступа» → «Другие команды...» → выберите «Все команды» → выделите «Мастер сводных таблиц» → нажмите «Добавить» → Ок. На панели быстрого доступа появится кнопка для вызова мастера (рис. 2.33).

Для создания сводной таблицы с помощью мастера выделите любую ячейку в таблице исходных данных, напри-



Рис. 2.33
Панель быстрого доступа
с «Мастером сводных таблиц»

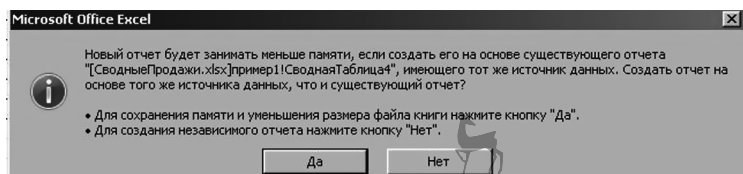


Рис. 2.34
Настройка независимого сводного отчета

мер на листе «Продажи2010» → щелкните по кнопке «Мастер сводных таблиц» на панели быстрого доступа → шаг 1 из 3 «Далее» → на шаге 2 из 3 автоматически будет указан диапазон исходных данных «Далее» → появится информационное окно (рис. 2.34).

Прочтите его внимательно, особенно последнюю строку, которая предлагает возможность создания независимого сводного отчета и нажмите «Нет» → шаг 3 из 3 «Готово». Теперь созданная сводная таблица будет независимой.

Пример 2.5. А теперь ответьте на вопрос «Кто лучше справился с продажами, мужчины или женщины?».

Воспользуйтесь «Мастером сводных таблиц» и создайте независимую сводную таблицу, затем постройте сводную диаграмму «Распределение заказов по группам продавцов м/ж».

Для начала создайте структуру сводной таблицы вида (рис. 2.35).

В поле «Значение» поместите поле Товар, по умолчанию будет установлена операция «Количество», которая

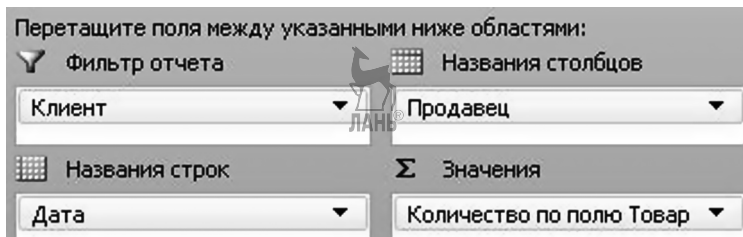


Рис. 2.35
Настройка областей

будет рассчитываться количество сделок, совершенных продавцом, а не количество проданных им единиц товара. Количество сделок равно количеству строк (записей) с фамилией продавца в таблице исходных данных.

Сгруппируйте «Дату» по годам.

Создайте две группы — Мужчины и Женщины. Здесь автоматическая группировка не подходит, придется часть работы выполнить вручную.

При группировке данных в столбцах сначала выделите в названии столбцов мужские фамилии — Воробьев, Журавлев, Петухов при нажатой клавише Ctrl. Затем выполните в группе «Группировать» → «Группировка по выделенному» появится «Группа1».

	A	B	C	D	E	F	G	H
1	Клиент	(Все)						
2								
3	Количество по полю Товар							
4		Группа1		Группа2			Общий итог	
5		Воробьев	Журавлев	Петухов	Голубева	Синицына	Сорокина	
6	Кв-л1	77	26	33	56	30	53	275
7	Кв-л2	73	28	38	75	33	49	296
8	Кв-л3	44	16	31	60	30	57	238
9	Кв-л4	28	26	30	34	22	41	181
10	Общий итог	222	96	132	225	115	200	990

Рис. 2.36
Группировка в сводных таблицах



Рис. 2.37
Сводная диаграмма

Аналогично создайте «Группу2» с женскими фамилиями. Выделите ячейку «Группа1» и наберите Мужчины. Выделите ячейку «Группа2» и наберите Женщины.

Заметьте, что глобальной группировки автоматически не произойдет. В основной таблице останутся строки соответствующие женским и мужским фамилиям. Для того чтобы убрать эти строки, щелкните на «Минус» перед названием групп, он изменится на плюс, а фамилии продавцов будут скрыты. Результат представлен на рисунке 2.36.

По сгруппированным данным можно построить сводную диаграмму (рис. 2.37).

При выборе «Клиента», который находится в области «Фильтр», сводная таблица будет автоматически перестраиваться.

Таким образом, сводные таблицы можно активно использовать при анализе многомерных данных. Можно строить трехмерные сводные диаграммы.

На первый взгляд работа со сводными таблицами кажется несколько запутанной. Поупражняйтесь и вы получите удобное, гибкое средство для анализа данных.



ГЛАВА 3

ДИАГРАММЫ



ТЕМА 3.1.

СОЗДАНИЕ ДИАГРАММ

3.1.1.

ОСНОВНЫЕ ПОНЯТИЯ

Диаграммы — удобное средство графического представления числовых данных. Добраться до диаграмм просто: вкладка «Вставка» → «Диаграммы». Здесь вы сможете ознакомиться со всеми типами диаграмм и выбрать наиболее соответствующую вашим целям. Как только диаграмма выбрана, появляется дополнительная вкладка «Работа с диаграммами», которая включает три группы команд «Конструктор», «Макет» и «Формат». Они отражают различные уровни детализации при работе с диаграммой, позволяют отформатировать и настроить диаграмму.

«Конструктор» — самый высокий уровень, на котором пользователь решает общие вопросы — как будет выглядеть диаграмма, ее тип и стиль, место расположения. Здесь же можно выбрать опорный, или если хотите «черновой» макет диаграммы из десяти предложенных, добавить новые ряды данных, а при необходимости произвести взаимную замену данных на осях.

Группы команд на ленте «Конструктор», а именно, «Тип», «Данные», «Макеты диаграмм», «Стили диаграмм», «Расположение», сгруппированы удобно (рис. 3.1), как раз в порядке вопросов, которые мы решаем при построении

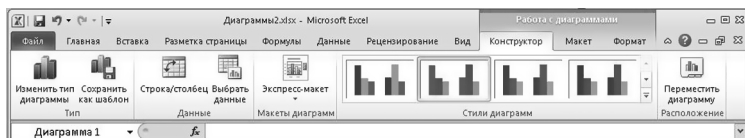


Рис. 3.1
Вкладка «Конструктор»

диаграммы. Основные операции, выполняемые этими инструментами: изменение типа и расположения диаграммы, ее данных и стиля.

«Тип» — управляет выбором типа диаграммы, команда «Сохранить как шаблон» позволяет сохранить для дальнейшего использования измененную пользователем диаграмму.

Группа «Данные» позволяет поменять строки на столбцы (можно себе это представить как транспонирование данных в таблице), а также управлять выбором данных — добавлять/удалять ряды данных, задавать их расположение на листе. После того как данные определены, у нас уже есть диаграмма, для которой макет выбран по умолчанию.

Теперь самое время оформить макет быстро, с помощью группы «Макеты диаграмм», расположенной на этой же ленте. Потом макет можно будет подправить, используя ленточную вкладку «Макет», а затем довести до совершенства, используя «Формат».

Вкладка «Макет» позволит настроить диаграмму более детально. Здесь собраны команды для работы с основными фрагментами (или элементами) диаграммы (рис. 3.2).

«Текущий фрагмент» используется для быстрого выделения фрагмента и его форматирования. Нужный элемент

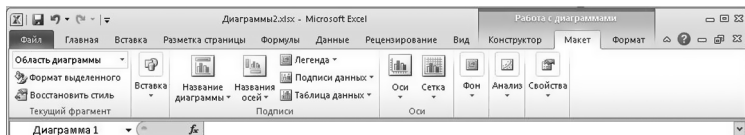


Рис. 3.2
Вкладка «Макет»

диаграмма выбирается в списке «Элементы диаграммы». Для форматирования используется команда «Формат выделенного фрагмента». При необходимости можно вернуть форматирование к первоначальному виду командой «Восстановить форматирование стиля».

Для более тонкой настройки параметров различных элементов диаграммы используются соответствующие кнопки на ленте «Вставить», «Подписи», «Оси», «Фон». Команда «Вставить» используется для вставки в диаграмму рисунков, фигур и надписей.

Кроме того, есть дополнительные команды управления диаграммой «Анализ», «Свойства».

«Формат» работает уже со стилями и включает группы «Текущий фрагмент», «Стили фигур», «Стили WordArt», «Упорядочить» и «Размер». Позволяет отформатировать отдельные элементы диаграммы, применить различные эффекты (рис. 3.3).

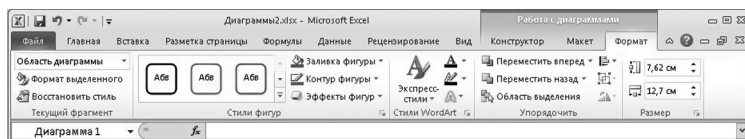


Рис. 3.3
Вкладка «Формат»

3.1.2. ПОСТРОЕНИЕ ДИАГРАММЫ

Основные шаги для построения любой диаграммы следующие.

1. Выделить область данных на листе Excel.

Диаграммы строятся на основе диапазонов данных. Структура исходных данных обычно является таблицей (списком). Поэтому достаточно выделить любую ячейку диапазона и приступить к построению диаграммы.

В некоторых случаях диапазон исходных данных необходимо выделить мышью полностью. Такая ситуация возникает, например, когда данные расположены в различных частях рабочего листа (на различных листах) или

нужно использовать для построения диаграммы только часть списка.

2. Перейти к вкладке «Вставка» и активизировать «Диаграммы».

3. Выбрать тип диаграммы. Появятся три дополнительных вкладки: «Конструктор», «Макет» и «Формат» под общим названием «Работа с диаграммами».

Тип диаграммы — это конкретный способ графического представления данных. Excel поддерживает 11 типов стандартных диаграмм. Каждый тип, в свою очередь, имеет различные модификации, называемые *видами*.

4. На вкладке «Конструктор» в группе «Макеты диаграммы» подобрать удобный макет и выбрать стиль в группе «Стили диаграмм».

5. Отформатировать полученную диаграмму, настроив параметры фрагментов, используя команды на лентах «Макет» и «Формат».

Диаграмма состоит из фрагментов. Некоторые фрагменты (в Excel 2003 они назывались элементами), например легенда, могут отсутствовать на диаграмме. Некоторые типы диаграмм имеют свои дополнительные фрагменты.

Основные фрагменты диаграммы (рис. 3.4):

- область диаграммы;
- область построения;
- заголовок;
- ряды данных;
- таблица данных;
- основная вертикальная ось;
- горизонтальная ось;
- дополнительная вертикальная ось;
- подписи диаграммы и осей;
- легенда;
- точка.

Легенда — набор маркеров, линий и текста, характеризующий каждый ряд данных. *Ряды данных* — исходные данные, по которым строится диаграмма. *Точка* — одно значение ряда данных.

Создание диаграммы обычно начинается с построения «эскиза» по шагам, указанным в пункте 3.1.2.

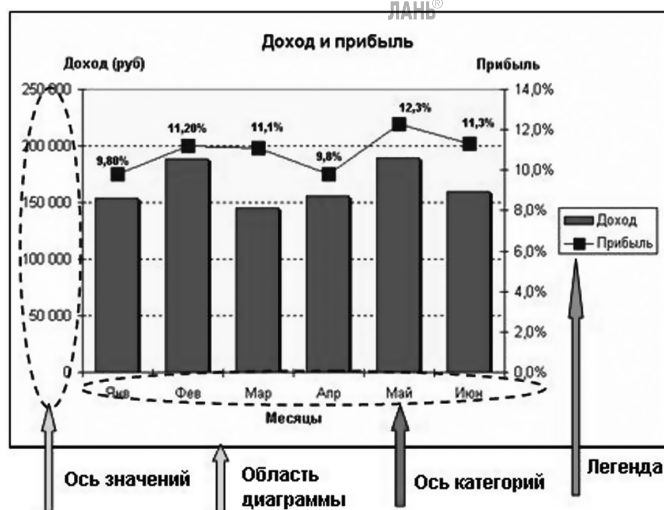


Рис. 3.4
Основные элементы диаграммы

Округ	Гостиницы %
ЦАО	38,4
СВАО	16,3
ЗАО	11,3
САО	8,4
ЮАО	6,9
ЮЗАО	7,4
Прочие	11,3

Рис. 3.5
Исходные данные

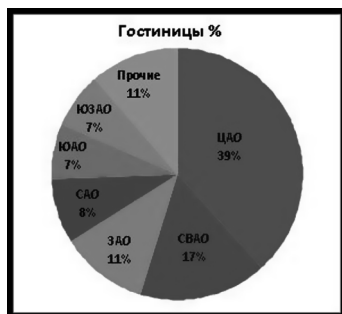


Рис. 3.6
Круговая диаграмма

Рассмотрим пример.

Пример 3.1. Требуется распределить гостиницы Москвы по округам. Диаграмма строится на основе данных (рис. 3.5).

Необходимо сделать следующее.

1. Выделить любую ячейку исходного диапазона данных.



2. Выполнить последовательность действий «Вставка» → «Диаграммы» → «Круговая».

3. На открывшейся панели выбрать первый тип «Круговая». В результате выбранная диаграмма сразу же будет построена на листе и появятся три дополнительные вкладки — «Конструктор», «Макет» и «Формат» под общим названием «Работа с диаграммами». Автоматически вы будете находиться в «Конструкторе».

4. Используя группу «Макеты диаграммы», изменить общий стиль оформления диаграммы. Выберите «Макет 1», который будет содержать заголовок и подписи данных, включая имена категорий (название) и доли. Результат приведен на рисунке 3.6.

Название диаграммы сформировано автоматически.

Основные элементы данной диаграммы: название диаграммы, область диаграммы, область построения, ряд «„Гостиницы %“», подписи данных. Они отображаются в группе «Текущий фрагмент», который есть на двух вкладках — «Макет» и «Формат».

До каждого фрагмента диаграммы можно добраться и настроить его, другими словами, отформатировать. Для этого необходимо выделить нужный элемент диаграммы левой кнопкой мыши и выполнить «Формат» → «Текущий фрагмент» → «Формат выделенного».

3.1.3. ФОРМАТИРОВАНИЕ ДИАГРАММ

После создания базового варианта диаграммы можно отформатировать ее отдельные фрагменты, подобрать подходящий стиль.

Для этого используют вкладки «Конструктор», «Макет» и «Формат» или контекстное меню, которое появляется при нажатии правой кнопки мыши после выделения фрагмента.

Для настройки диаграммы ее необходимо выделить, щелкнув мышью в области диаграммы. Автоматически активизируется и вкладка «Работа с диаграммами», диаграмма будет помещена в рамку.

1. При помощи перетаскивания курсором рамки можно изменить размеры диаграммы.

2. На вкладке «Формат» → «Текущий фрагмент» можно выбрать отдельный элемент и отформатировать его.

3. На вкладке «Конструктор» → «Тип» можно изменить тип диаграммы.

Часто возникает необходимость отформатировать ряд данных или его отдельные точки.

Для выделения ряда данных щелкните мышью по «Области построения» — выделится ряд данных. Установите курсор на нужный вам элемент (точку данных) и щелкните еще раз мышкой — выделится отдельная точка ряда. Для форматирования отдельной точки ряда теперь можно воспользоваться командами «Макет» → «Текущий элемент» → «Формат» выделенного или контекстным меню, которое вызывается нажатием правой кнопки мыши.

Приведем диаграмму, построенную в примере 2.6, к виду, как на рисунке 3.7. Выполним для этого ряд действий.

Щелкнем по диаграмме, перейдя, таким образом, в режим «Работа с диаграммами». Изменим сначала название диаграммы, а затем тип — диаграммы: «Конструктор» → «Тип» → «Изменить тип» → «Кольцевая».

Изменим заливку и границы: «Макет» → «Текущий фрагмент» → «Ряд „Гостиницы %“» → «Формат выделенного» → в окне «Формат ряда данных» установим: «Заливка» — Нет заливки, «Цвет границы» — Сплошная линия, «Цвет» — Черный → «Заккрыть».

**Более 30% гостиниц Москвы расположены
в Центральном административном округе**

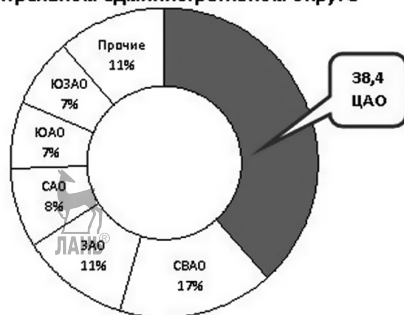


Рис. 3.7
Кольцевая диаграмма

Щелчком по точке ЦАО, выделится сектор диаграммы, соответствующий этой точке, в окне «Текущего фрагмента» появится название выделенного фрагмента. Установим для этого элемента другую заливку, для чего, выберем «Формат выделенного», в окне «Формат ряда данных» установим: «Заливка» — Сплошная заливки, «Цвет» — Синий → «Закрыть».

Удалим надпись в секторе ЦАО. В окне «Текущего фрагмента» выберем фрагмент «Ряд „Гостиницы %“» → «Подписи данных» → «Формат выделенного фрагмента», щелкните по данным ЦАО, в окне «Формат ряда данных» → «Параметры подписи» уберем все галочки → «Закрыть».

Чтобы выделить определенные аспекты диаграммы, используют пояснительные надписи и указатели.

Выполним «Макет» → «Вставить» → «Фигуры» → «Выноски». Поместим выноску рядом с выделенным фрагментом. Выделим ее правой кнопкой мыши, в появившемся контекстном окне вставим нужный текст и отформатируем рисунок — заливку и границы. В результате получим рисунок 3.7.

Для привлечения внимания к конкретному элементу на диаграмме можно использовать также надписи, различные средства рисования. Например, добавлять стрелки, пояснительные тексты.

Для этого выделите диаграмму, выберите вкладку «Вставка», в группе «Иллюстрации» выберите нужный объект: рисунок, картинку или фигуру.

ТЕМА 3.2. ПОСТРОЕНИЕ ПРОФЕССИОНАЛЬНЫХ ДИАГРАММ

3.2.1. ОСНОВНЫЕ ИДЕИ

Построение профессионально исполненных диаграмм, дело, конечно, непростое. Но мы попробуем сделать несколько шагов в этом направлении. А помогут нам в этом профессионалы своего дела — Джин Желязны и «Мистер Электронная таблица» Джон Уокенбах [5].

Для отображения одних и тех же данных можно использовать различные диаграммы. Конкретный тип диаграммы выбирается в зависимости от аспекта данных, на который вы хотите обратить внимание. Как считает Джин Желязны [3], директор по визуальным коммуникациям McKinsey & Company:

- 1) диаграмма должна отражать основную идею;
- 2) идея может быть выражена посредством одного из пяти основных типов сравнения: покомпонентного, позиционного, временного, частотного, корреляционного;
- 3) каждому типу сравнения соответствует один из пяти основных типов диаграмм: круговая, линейчатая, точечная диаграммы, гистограмма или график.

Попробуем следовать этим правилам. Для этого нужно сделать три важных шага.

Шаг 1. От данных к идее. На этом шаге необходимо сформулировать идею, которую вы хотите донести до аудитории. Диаграмма будет изображать вашу главную идею. Подумайте, о чем вы хотите сказать, на какой аспект данных обратить особое внимание? Начните с заголовка вашей диаграммы. Заголовок должен помочь вам выразить основную мысль диаграммы, причем кратко и четко.

Шаг 2. От идеи к сравнению. Обычно диаграмма предназначена для иллюстрации какой-либо закономерности и базируется на сравнении данных. Поэтому важно четко определить, на каком типе сравнения данных будет основана ваша идея: покомпонентном, позиционном, временном, частотном или корреляционном.

Шаг 3. От сравнения к диаграмме. Здесь, наконец, нужно выбрать тип диаграммы, который наилучшим образом отобразит тип сравнения данных, соответствующий вашей главной идее.

3.2.2. ОСНОВНЫЕ ТИПЫ СРАВНЕНИЯ

Диаграммы, как уже было отмечено выше, основываются обычно на одном из пяти типов сравнения данных. Рассмотрим их подробнее.

Покомпонентное сравнение называют также относительным сравнением, так как оно отражает долю рассматриваемых компонентов, как части целого, или показывает размер каждого компонента в процентах от целого. Характерным признаком такого сравнения является использование таких слов как «доля», «процент от целого», «структура».

Доли целого более наглядно отображаются с помощью «Круговой диаграммы». Однако для сравнения компонентов нескольких целых, лучше использовать «Нормированную линейчатую диаграмму» или «Нормированную гистограмму».

Позиционное сравнение отражает сравнение одних показателей с другими. Оно выявляет, как объекты соотносятся друг с другом — больше, меньше или одинаковы. Характерным признаком такого сравнения является использование таких слов, как «больше чем», «меньше чем», «равно». Для иллюстрации позиционного сравнения лучше всего подходят различные «Линейчатые диаграммы».

Временное сравнение выявляет характер изменения данных во времени. Если покомпонентное и позиционное сравнения показывают взаимосвязи данных в определенный момент времени, то временное сравнение отражает динамику изменения данных. Здесь данные могут сравниваться в различные моменты или периоды времени. Ключевые слова, характеризующие динамику, показывают как «изменяется», «убывает», «возрастает» или «снижается» некоторый показатель.

Для иллюстрации такого сравнения лучше всего использовать «График», особенно когда рассматривается изменение непрерывных данных, например изменение курса валют. Можно использовать также «Гистограммы».

Частотное сравнение показывает, сколько значений данного параметра (частотность) попадает в определенные последовательные области числовых значений. Характерным признаком такого сравнения является использование таких слов, как «частотность», «распределение», «диапазон». Здесь используются «Гистограммы».

Корреляционное сравнение характеризует взаимоотношение данных. Его используют, например, когда хотят выяснить существует ли взаимосвязь между некоторыми признаками или отсутствует. Термины, характерные для данного сравнения — «возрастает/снижается при...», «меняется при...» и т. д. Чаще всего здесь используется «Точечная диаграмма». Можно для этих целей приспособить и «Линейчатую».

3.2.3. ПОКОМПОНЕНТНОЕ СРАВНЕНИЕ

Рассмотрим на примерах соответствие основных типов диаграмм основным типам сравнения. Параллельная в предыдущем пункте между типами сравнения и типами диаграмм, отражает основной подход. На практике окончательный выбор определяется основной целью вашей иллюстрации данных.

Пример 3.2. На первый курс геологического факультета в 2010 году зачислено 111 абитуриентов, из них 71 — иногородние и иностранные студенты (табл. 3.1).

Таблица 3.1

Исходные данные

Результаты приема	2010 год, чел.
Иногородние и иностранные студенты	71
Санкт-Петербург	40

Петербургские школьники составили менее 40% приема в 2010 году



Рис. 3.8
Круговая диаграмма

Постройте диаграмму по этим данным (рис. 3.8). Отображение структуры, состава единого целого — это покомпонентное сравнение. В этом случае используются различные типы «Круговой диаграммы».

В окне «Формат подписей данных» установите параметр — «Доли». Порядок отображения данных на круговой и кольцевой диаграмме определяется порядком данных на листе. Обычно в правом верхнем углу располагают данные, на которые хотелось бы обратить внимание. Для этого возможно потребуется повернуть сектора на некоторый угол. Выполните последовательность команд: «Макет» → «Текущий фрагмент» → «Формат выделенного фрагмента» → в диалоговом окне «Формат ряда данных» → «Параметры ряда» → «Угол поворота первого сектора» → установите 130. Заголовок диаграммы тоже подчеркивает основную мысль.

Заметим, что «Круговая диаграмма» часто используется для представления данных в статистике, где ее называют секторной, так как в основе ее лежит круг, разделенный на несколько секторов, размеры которых пропорциональны вкладу значений в общую сумму.

Обычно стандартная «Круговая диаграмма» используется, если секторов не более 7. Если разброс данных велик по величине или имеется большое количество секторов, рекомендуется использовать «Круговую диаграмму» типа «Вторичная круговая» или «Вторичная гистограмма».

Пример 3.3. Требуется построить диаграмму, отражающую расходы фирмы (табл. 3.2).

Таблица 3.2

Исходные данные

Зарплата	69,80%
Налоги	19,80%
Страхование	5,90%
Обслуживание	1,50%
Накладные	1,10%
Почтовые расходы	0,80%
Междугородные звонки	0,70%
Канцтовары	0,40%

Для облегчения работы с маленькими долями диаграммы в основной диаграмме их можно объединить в один элемент, а затем разбить их в отдельную диаграмму рядом с основной.

Создайте «Круговую диаграмму» «Вторичная гистограмма». Выберите в «Конструкторе» «Макет 2» и «Стиль 1». Легенду поместите слева.

Выберите вкладку «Макет» → «Текущий фрагмент» → «Ряд 1» → «Формат выделенного фрагмента» → появится диалоговое окно «Формат ряда данных», в котором:

1) измените количество данных, которое отображается в присоединенной гистограмме «Параметры ряда» → «Вторая область построения содержит первую», установите 5;

2) установите границу для секторов «Цвет границы» → «Сплошная линия» → «Цвет» — черный.

Форматирование подписей данных удобно осуществлять через контекстное меню. Выделите левой кнопкой мыши ряд данных, нажмите правую кнопку мыши для перехода в контекстное меню. Установить параметры шрифта, перейти к «Формату подписей данных», в диалоговом окне «Формат подписей данных» установите «Число» → формат „Процентный“ → «Число десятичных знаков 2» → «Закрыть».

Диаграмма, построенная по исходным данным, приведена на рисунке 3.9. Большую роль в диаграммах играет

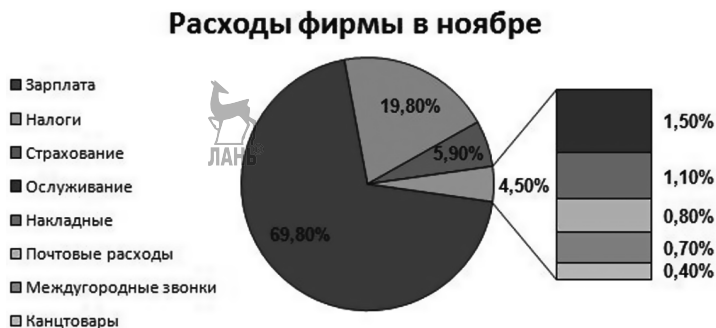


Рис. 3.9
Круговая диаграмма со вторичной гистограммой

цветовое решение, которое усиливает наглядность. В черно-белом варианте это преимущество, к сожалению, теряется.

В Excel 2007 имеется семь подтипов «Круговой диаграммы». Кроме того в эту группу можно включить и «Кольцевую диаграмму», которая в отличие от «Круговой» позволяет включать несколько рядов данных.

Пример 3.4. В таблице 3.3 приведены данные о поступлении студентов в СПбГУ. Требуется отобразить структуру данных.

В данном случае нужно использовать «Кольцевую диаграмму», которая дает возможность отображать несколько рядов данных (рис. 3.10), в данном случае двух.

Для наглядности при построении диаграммы использованы дополнительные средства — надпись и стрелка. Выполните «Макет» → «Вставить» → «Надпись (Фигура)» и «Средства рисования» к вашим услугам.

Если рядов данных более трех, то и «Кольцевая диаграмма» теряет свою наглядность. В этом случае для компонентного сравнения используют «Нормированную

Таблица 3.3

Исходные данные

Результаты приема	2010 год, %	2011 год, %
Иногородные	50	60
Санкт-Петербург	50	40

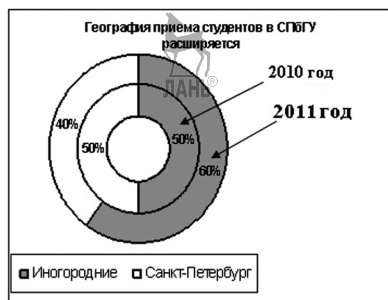


Рис. 3.10
Кольцевая диаграмма

линейчатую диаграмму с накоплением». Компоненты каждого ряда данных нанесены в соответствии с относительной (процентной) величиной компонентов независимо от абсолютных значений.

Пример 3.5. Имеются данные о населении и трудовых ресурсах в некоторых регионах (табл. 3.4).

Постройте диаграмму «Нормированная линейчатая с накоплением». На вкладке «Конструктор» выберите стиль диаграмм (Стиль 25). На вкладке «Макет» отформатируйте линии сетки, легенду поместите внизу, добавьте заголовок (рис. 3.11).

Используемая здесь «Линейчатая диаграмма» несет две смысловые нагрузки — отражает относительное соот-

Таблица 3.4

Исходные данные

	Трудовые ресурсы (млн чел.)	Численность населения (млн чел.)
Регион А	44	89
Регион В	31	63
Регион С	30	50



Рис. 3.11
Линейчатая диаграмма с накоплением



Рис. 3.12
Линейчатая диаграмма

ношение внутри каждого региона и дает возможность сравнить регионы между собой.

«Линейчатую диаграмму» удобно использовать для покомпонентного сравнения результатов опросов, когда названия категорий слишком длинные, количество вопросов более семи (рис. 3.12).

3.2.4. ПОЗИЦИОННОЕ СРАВНЕНИЕ

Для позиционного сравнения чаще всего используют различные виды «Линейчатой диаграммы». Три основных вида линейчатой диаграммы — с группировкой, с накоплением, нормированная с накоплением. Все эти виды допускают, кроме того, объемное, цилиндрическое, коническое и пирамидальное представление, а всего их 15.

«Диаграмму с группировкой» в различных вариациях используют для отображения результатов различных опросов и рейтингов. Обычно данные сортируются предварительно по убыванию или возрастанию. Чтобы показать позицию интересующего объекта, его выделяют, используя средства форматирования.

«Линейчатые диаграммы с накоплением» используют для позиционного сравнения данных, которые содержат несколько категорий. Она отражает соотношение отдельных компонентов, показывая вклад каждого значения в общую сумму по каждой категории. Категории расположены



по горизонтали, а значения по вертикали, таким образом, уделяется большое внимание сопоставлению значений.

«Нормированная линейчатая с накоплением диаграмма» чаще всего применяется, как уже отмечалось выше, для покомпонентного сравнения.

Пример 3.6. Данные о посещаемости самых популярных сайтов Рунета по состоянию на весну 2010 по версии TNS Web Index приведены в таблице 3.5.

Таблица 3.5

Исходные данные

Адрес сайта	Посетители, тыс. чел.	Адрес сайта	Посетители, тыс. чел.
Vesti.ru	528	Gismeteo.ru	1267
Kp.ru	590	LiveJournal.com	1642
LiveInternet.ru	708	Rambler.ru	3682
Rian.ru	747	Odnoklassniki.ru	5830
Kinopoisk.ru	752	Vkontakte.ru	10 812
Qip.ru	883	Mail.ru	12 045
Rbc.ru	945	Yandex.ru	13 392
Letitbit.net	979	—	—



Рис. 3.13

Линейчатая диаграмма

Создайте на основе данных «Линейчатую с группировкой» диаграмму и отформатируйте ее в соответствии с рисунком 3.13.

В данном случае речь идет о позиционном сравнении. Здесь нет общего целого, так как одни и те же пользователи посещают различные ресурсы Интернета. Диаграмма наглядно отражает позиции, которые занимают относительно друг друга популярные сайты.

Наиболее посещаемы сайты, связанные с общением. Выделите на диаграмме эти сайты, используя форматирование.

Линейные диаграммы используются также для позиционного сравнения нескольких рядов данных.

Пример 3.7. В таблице 3.6 приведены данные об успеваемости студентов. Сравните успеваемость в группах.

Таблица 3.6

Средний балл

№ группы	Математика	Информатика
Группа 7100	3,5	3,3
Группа 7110	4,3	3,9
Группа 2100	3,9	4,2
Группа 2110	4,2	3,4



Рис. 3.14

Линейчатая диаграмма с группировкой



Создайте диаграмму «Линейчатую с группировкой». В линейчатых диаграммах категории размещаются по вертикальной оси, а их значения по горизонтальной.

Заметим, что категории отобразились в обратном порядке по отношению к таблице данных. Можно расположить их в другом порядке. Выделите вертикальную ось, перейдите в режим форматирования («Формат выделенного фрагмента»), установите параметры оси «Обратный порядок категорий». При этом категории поменяют порядок, но горизонтальная ось сместится вверх, если ее нужно оставить внизу установите параметр оси «Горизонтальная ось пересекает в максимальной категории» (рис. 3.14).

Пример 3.8. Приведены данные объема продаж фруктов в городах России приведены в таблице 3.7.

Постройте «Линейчатую диаграмму с накоплением», которая отображают вклад каждого значения в общую сумму по каждой категории.

Для наглядности можно установить линии ряда данных. Выделите диаграмму, выполните «Макет» → «Анализ» → «Линии ряда данных». Линии рядов данных со-

Таблица 3.7

Объем продаж фруктов, млн руб.

Город	Груши	Яблоки	Мандарины
Псков	22	23	30
Краснодар	25	24	35
Уфа	20	20	30
Томск	15	15	40

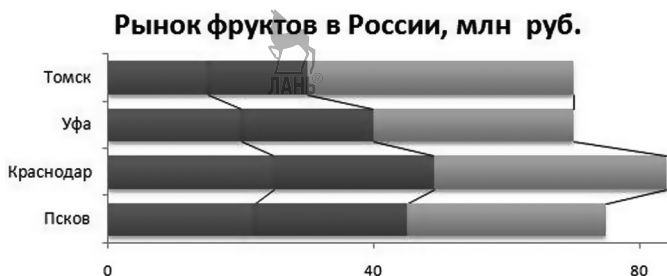


Рис. 3.15

Диаграмма с линиями рядов данных



единяют вершины каждого столбика со следующим столбиком ряда (рис. 3.15).

Однако когда сравнение идет по годам или кварталам, т. е. появляется временной аспект, в позиционном сравнении лучше использовать гистограмму, повернутую на 90 градусов. В «Линейчатой диаграмме» уделяется большее внимание сопоставлению значений и меньшее — изменениям во времени. Кроме того, более привычно, что время изменяется не сверху вниз, а справа налево. Таким образом, если данные распределены по годам или кварталам, то горизонтальное направление для категорий здесь более удобно (рис. 3.16).

Линейчатые диаграммы используют практически для всех типов сравнения — позиционного, покомпонентного, корреляционного.



Рис. 3.16
Гистограмма

3.2.5. ВРЕМЕННОЕ СРАВНЕНИЕ

Графики используются для иллюстрации непрерывных данных. Они полезны для выявления тенденций данных. Пример использования графиков приведен на рисунке 3.17.

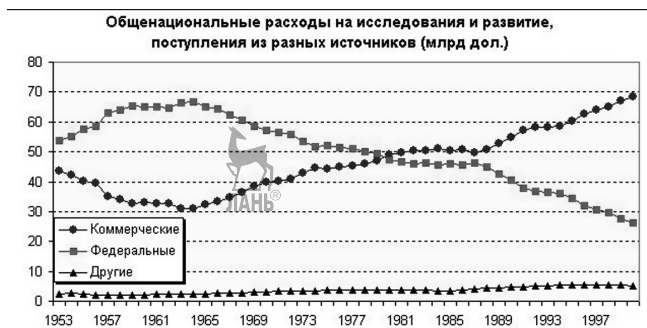


Рис. 3.17
Графики

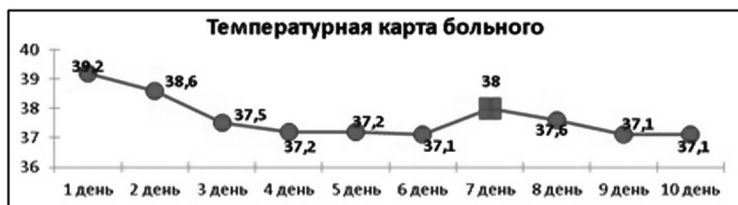


Рис. 3.18
График с маркерами

Имеется семь видов графиков, один из которых — объемный. В графиках можно применять маркеры, которые используют для привлечения внимания к некоторым точкам. Например, выделить повышение температуры (рис. 3.18).

3.2.6. КОРРЕЛЯЦИОННОЕ СРАВНЕНИЕ

Точечная диаграмма часто используется для корреляционного сравнения, которое показывает, соответствует ли соотношение двух переменных ожидаемой зависимости, т. е. для установления факта, находится ли изменчивость одного признака в некотором соответствии с изменчивостью другого. Например, диаграмма на рисунке 3.19 демонстрирует, что количество продаж напрямую зависит от количества заявок.



Рис. 3.19
Точечная диаграмма

Точечную диаграмму называют также диаграммой рассеяния, которая используется для отображения статистических данных. Обычно на одну диаграмму наносятся данные нескольких измерений для последующего сравнения. Используется для сравнения пар значений.

ТЕМА 3.3. НЕСТАНДАРТНЫЕ ДИАГРАММЫ

3.3.1. КОМБИНИРОВАННЫЕ ДИАГРАММЫ

Комбинированные диаграммы (рис. 3.20) используются для различных целей:

а) для иллюстрации взаимосвязи между разными типами информации;

б) если разница между значениями данных мала;

в) если разброс данных велик по величине.

Пример 3.9. В нем показана иллюстрация взаимосвязи между разными типами информации. Требуется построить диаграмму плана по звонкам для опроса общественного

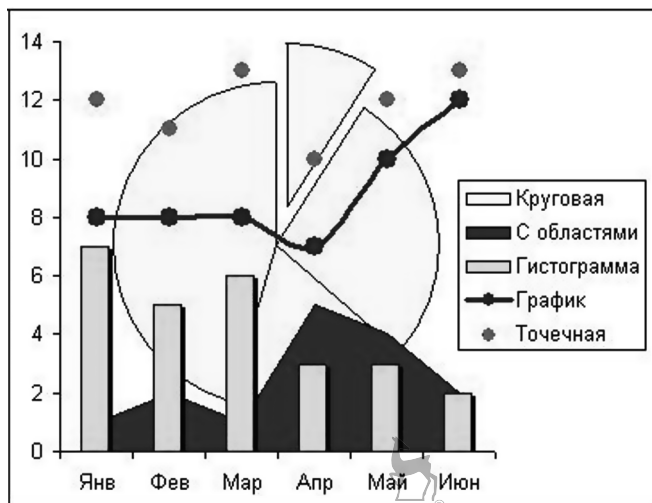


Рис. 3.20
Комбинированная диаграмма

Таблица 3.8

План — Факт

Дни недели	Запланировано	Фактически
пн	1250	1300
вт	1230	1400
ср	1500	1500
чт	2000	2000
пт	2500	1600
сб	3200	3400
вс	4500	4300

мнения на неделю и фактическое число звонков. Данные содержатся в таблице 3.8. Для создания комбинированной диаграммы необходимо выполнить ряд шагов. При построении комбинированных диаграмм обычно возникают проблемы, поэтому покажем построение пошагово.

1. Выделить ячейки, содержащие данные, включая заголовки строк и столбцов, т. е. всю таблицу.

2. Выполнить «Вставка» → «Диаграммы» → «График».

В результате получим диаграмму с графиками разного цвета для двух рядов данных. Здесь важно, чтобы была отображена легенда (рис. 3.21).

3. Щелкнуть мышью по ряду «Запланировано», ряд выделится:

а) щелкнуть правой кнопкой мыши, появится контекстное меню;

б) выбрать «Формат ряда данных», откроется меню (рис. 3.22);

в) установить в «Параметрах ряда» значение «По вспомогательной оси» → «Заккрыть», появится вспомогательная ось справа, ряд данных останется выделенным (рис. 3.23);

г) щелкнуть еще раз правой кнопкой мыши, появится контекстное меню;

д) выбрать «Изменить тип диаграммы для ряда», «Формат ряда данных», откроется меню «Изменение типа диаграммы»;

е) выбрать «Гистограмма» → «Ок» (рис. 3.24).

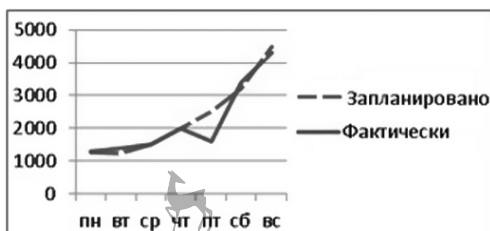


Рис. 3.21

Построение комбинированной диаграммы, шаг 1

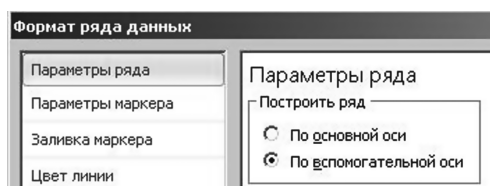


Рис. 3.22

Меню «Формат ряда данных»

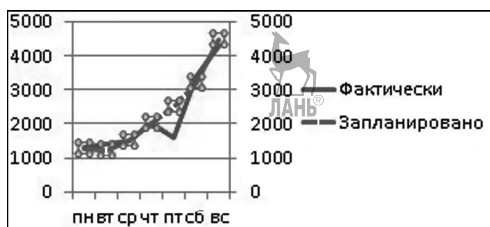


Рис. 3.23

Построение комбинированной диаграммы, шаг 2

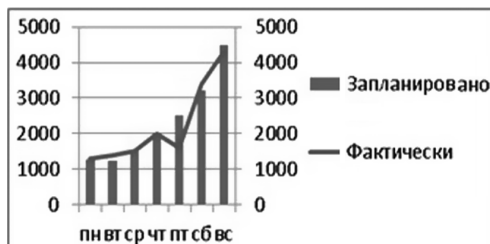


Рис. 3.24

Построение комбинированной диаграммы, шаг 3

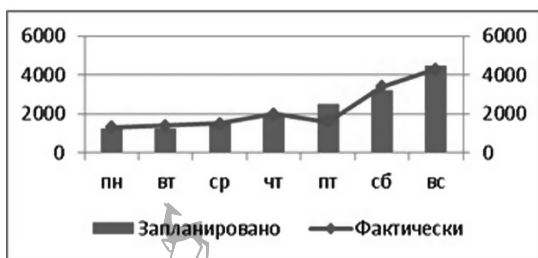


Рис. 3.25
Комбинированная диаграмма

Полученную диаграмму нужно, как говорится, «довести до ума», в данном случае — отформатировать некоторые элементы диаграммы. Например, как на рисунке 3.25.

Пример 3.10. Разница между величинами данных мала. Построить комбинированную диаграмму «График/Гистограмма» с двумя осями», отражающую количество кредитов на покупку автомобилей, выделенных в каждом месяце, и среднемесячную прибыль по займу. Исходные данные приведены в таблице 3.9, а диаграмма на рисунке 3.26.

Таблица 3.9

План — Факт

Месяц	Выдано кредитов	Норма прибыли, %
Январь	11	6,8
Февраль	39	6,78
Март	28	6,68
Апрель	25	6,50
Май	70	6,49
Июнь	52	6,47
Июль	59	6,45
Август	63	6,46
Сентябрь	62	6,47
Октябрь	60	6,48
Ноябрь	53	6,52
Декабрь	60	6,50

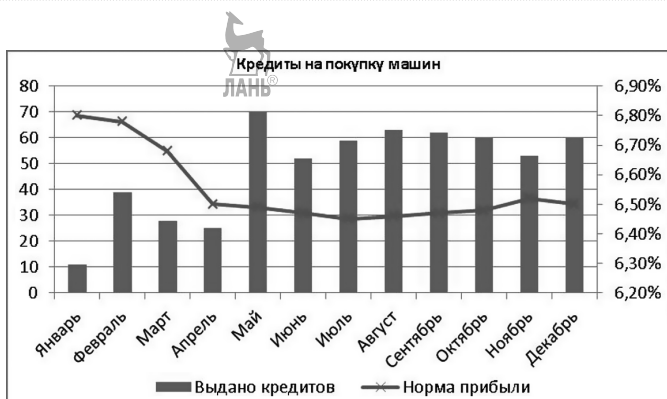


Рис. 3.26
Комбинированная диаграмма

3.3.2. ДОПОЛНИТЕЛЬНЫЕ ЭЛЕМЕНТЫ ФОРМАТИРОВАНИЯ

Можно построить и более наглядные диаграммы. Например, по исходным данным на рисунке 3.27 можно построить различные диаграммы, изображенные на рисунках 3.28, 3.29.

	А	В	С
1	ПРОДАЖИ		
2		Планируемые	Фактические
3	Янв	300	254
4	Фев	400	456
5	Мар	500	434

Рис. 3.27
Таблица исходных данных

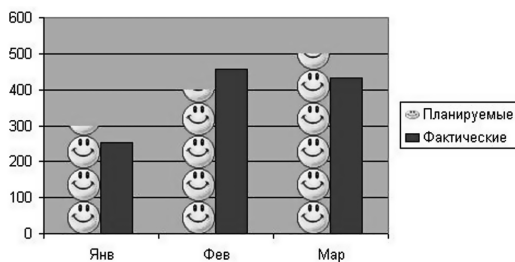


Рис. 3.28
Гистограмма

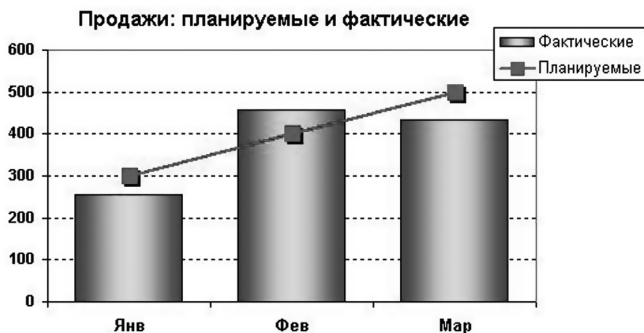


Рис. 3.29
Комбинированная диаграмма

3.3.3. ДИАГРАММЫ В ЯЧЕЙКАХ

В Excel 2010/2013 для создания диаграммы реализованы новые возможности, так называемые *спарклайны*, их называют также «диаграммы в ячейке» или «минидиаграммы».

Спарклайны расположены на вкладке «Вставка» (рис. 3.30). Они предназначены для визуализации числовых данных прямо в ячейках рабочего листа. «График» используют для наглядного отображения динамики числовых данных, а «Столбец» (аналог «Гистограммы») — для позиционного сравнения данных. Для отображения положительных и отрицательных чисел используют «Выигрыш/Проигрыш».

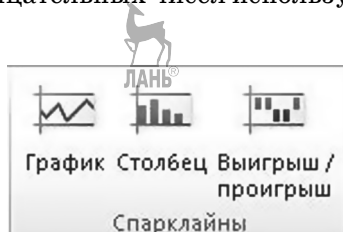


Рис. 3.30
Типы спарклайнов

Пример 3.11. Даны размеры премий, полученные сотрудниками отдела в первом полугодии (табл. 3.10), числовые данные расположены в диапазоне B1:G11.



Таблица 3.10

Список сотрудников, получивших премию

ФИО	Январь	Февраль	Март	Апрель	Май	Июнь
Иванов	5,1	8,6	4,1	3,9	1,6	6,6
Петров	4,6	6,5	4,5	1	3,4	7,5
Сидоров	4,1	7	8	1,5	0	4,2
Галкин	2,2	2,3	4,1	9,3	0,35	3,3
Синицын	0	2	6,3	2,8	6,5	5,4
Соколов	4,2	2,8	6,9	6,5	6,7	6,3
Васильев	9,8	2,2	6,2	9,5	6	1,5
Сонин	7,5	9,9	9,1	9,1	6,1	5,6
Назаров	3	3,2	9,9	2,7	1,9	5
Титов	7,8	7	7,5	3,8	6,2	1,4

Сравним размеры премий, полученные каждым сотрудником. Применим для этого спарклайн типа «Столбец», выделим максимальное значение.

Порядок действий следующий.

1. Выделить диапазон данных расположения спарклайна H2:H11.

2. Выполнить последовательность действий «Вставка» → «Спарклайны» → «Столбец».

3. В открывшемся меню «Создание спарклайнов» укажите «Диапазон данных» — B1:G11, «Диапазон расположения» H2:H11 выделяется автоматически → Ок.

4. Выделить ячейки с полученными спарклайнами, появится дополнительная вкладка на ленте — «Работа со спарклайнами». Щелкнув по ней перейти в режим работы «Конструктора».

5. Поставить галочку «Максимальная точка» в группе «Показать». Столбики, соответствующие максимальным значениям, выделятся другим цветом. Цвет столбиков можно выбирать на свое усмотрение, используя «Стили».

6. Выполнить «Стили» → «Темный тип спарклайна 1», Столбик, соответствующий максимальному значению, будет более темным по сравнению с остальными.

В столбце Н будут помещены спарклайны (рис. 3.31).

Аналогично можно построить спарклайны по столбцам. Рассмотрим, например, распределение премий в январе. Для этого нужно выполнить ту же последовательность действий 1–6. Изменятся только ссылки на диапазоны данных: диапазон расположения В12; диапазон данных В2:В11.

Аналогично можно использовать и другой тип спарклайна — «График». Результат приведен на рисунке 3.32.

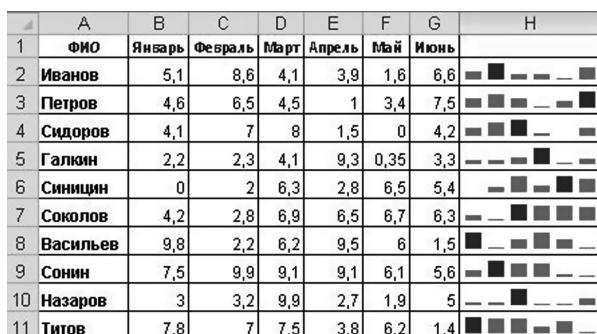


Рис. 3.31
Спарклайн-гистограмма



Рис. 3.32
Спарклайны в виде
графика и гистограммы

3.3.4. ПРИМЕРЫ ДИАГРАММ

Пример 3.12. Требуется построить диаграмму для психологов. В исследовании, посвященном проблемам ценностной реориентации, выявлялись иерархии терминальных ценностей по методике М. Рокича у родителей и их взрослых детей. Ранги терминальных ценностей, полученные при обследовании пары «мать — дочь» (матери — 66 лет, дочери — 42 года), представлены в таблице 3.11. Требуется определить, как эти ценностные иерархии коррелируют

Таблица 3.11

Терминальные ценности	Ряд 1: Ранг ценностей в иерархии матери	Ряд 2: Ранг ценностей в иерархии дочери
Активная деятельная жизнь	15	15
Жизненная мудрость	1	3
Здоровье	7	14
Интересная работа	8	12
Красота природы и искусство	16	17
Любовь	11	10
Материально обеспеченная жизнь	12	13
Наличие хороших и верных друзей	9	11
Общественное признание	17	5
Познание	5	1
Продуктивная жизнь	2	2
Развитие	6	8
Развлечения	18	18
Свобода	4	6
Счастливая семейная жизнь	13	4
Счастье других	14	16
Творчество	10	9
Уверенность в себе	3	7

друг с другом, другими словами, насколько ценности дочери соответствуют ценностям матери. Использованы данные, приведенные в книге Е. В. Сидоренко [7].

Наша задача построить диаграмму, которая поможет нам дать ответ на поставленный вопрос. Сразу приведем конечный вид диаграммы, к которому будем стремиться (рис. 3.33).

Диаграмма наглядно показывает, что ранги ценностей у матери и дочери достаточно близки. Основные расхождения приходятся только на ценности «Общественное признание» и «Счастливая семейная жизнь». Наука тоже подтверждает, что «корреляция между иерархиями терминальных ценностей матери и дочери статистически значима».

Приведенная диаграмма является диаграммой специального вида, ее называют пирамидой. Действительно, если мы отсортируем по возрастанию данные, например



Рис. 3.33
Диаграмма «Пирамида»

для матери, мы в результате получим фигуру, похожую на пирамиду.

Для построения диаграммы данные необходимо предварительно немного изменить. Данные ряда 1, который окажется слева, должны быть отрицательными числами. Приведем начало новой таблицы (табл. 3.12).

Таким образом, Ряд1 должен содержать отрицательные числа. В качестве основы будем использовать «Линейную диаграмму с накоплением». Рассмотрим построение пошагово.

Выделите данные, не включая заголовки (первую строку). Создайте «Линейчатую диаграмму с накоплением». В результате получите некоторый эскиз (рис. 3.34), который сильно отличается от требуемой диаграммы.

1. Эскиз необходимо отформатировать, т. е. изменить некоторые элементы диаграммы. Напомним, что когда диаграмма выделена, появляется дополнительная вкладка на

Таблица 3.12

Изменения исходных данных

Активная деятельная жизнь	–15	15
Жизненная мудрость	–1	3

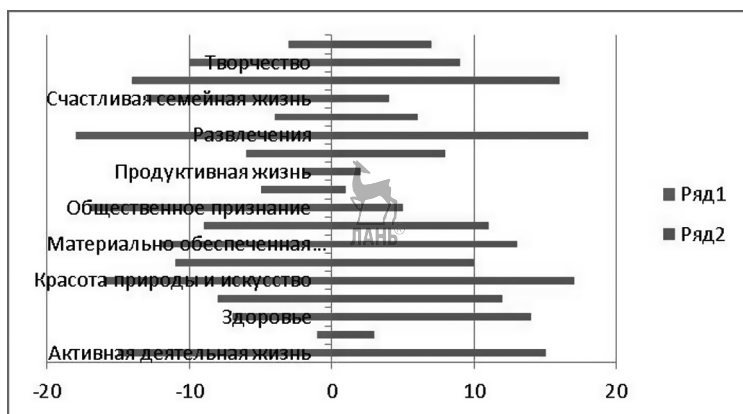


Рис. 3.34
Построение комбинированной диаграммы, шаг 1

ленте — «Работа с диаграммами», которая включает три группы команд «Конструктор», «Макет» и «Формат». Для форматирования отдельных элементов удобно использовать «Формат» → «Текущий фрагмент».

При нажатии стрелки «Область диаграммы», откроется список всех элементов данной диаграммы. Следует сначала выбрать нужный элемент, затем «Формат выделенного» (рис. 3.35), откроется диалоговое окно для форматирования выбранного элемента.

Для форматирования элементов можно использовать также контекстное меню, которое открывается при выделении элемента щелчком правой кнопки мыши. Но это не всегда бывает удобно использовать.

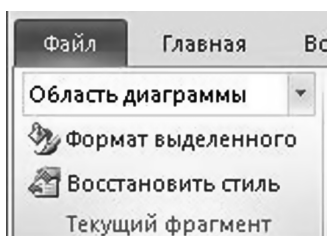


Рис. 3.35
Настройка области диаграммы

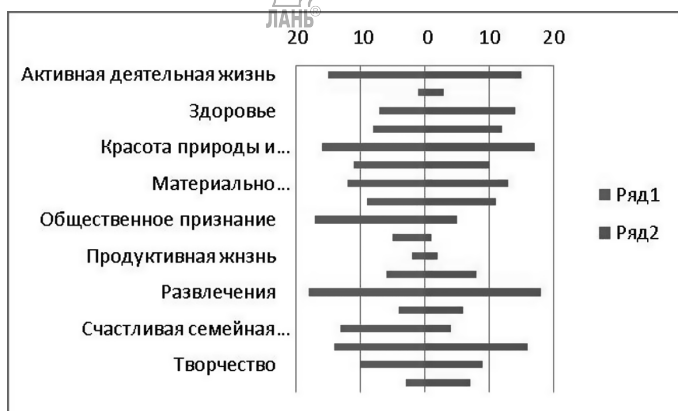


Рис. 3.36
Построение комбинированной диаграммы, шаг 2

2. Формат горизонтальной оси значений: «Макет» → «Текущий фрагмент» → «Горизонтальная ось (значений)» → «Формат выделенного». Появится диалоговое окно «Формат оси». Выберите «Число» → установите «Числовой формат» — все форматы, «Код формата» — 0; 0; 0; → «Заккрыть» диалоговое окно. Это устранил минус (рис. 3.36).

3. Формат вертикальной оси (категорий): «Макет» → «Текущий фрагмент» → «Вертикальная ось (категорий)» → «Формат выделенного». Появится диалоговое окно «Формат оси». Установите «Параметры оси»: «Обратный порядок категорий» — галочка, «Основные;» — нет, «Промежуточные:» — нет, «Подписи оси:» — внизу → «Заккрыть».

4. Формат рядов данных. «Макет» → «Текущий фрагмент» → «Ряд1» → «Формат выделенного». Появится диалоговое окно «Формат ряда данных». Установите:

а) «Параметры ряда» → «Боковой зазор» — без зазора (0%),

б) «Цвет границы» → «Сплошная линия», «Цвет» — черный.

5. Аналогично для данных Ряд2. Только цвет границы.

6. Поместите легенду сверху. «Макет» → «Легенда» → «Добавить легенду сверху». Щелкните по легенде левой кнопкой мыши, она выделится, перетащите ее немного левее.

7. Изменение названия рядов. «Конструктор» → «Выбрать данные». Откроется диалоговое окно «Выбор источника данных». Щелкните левой кнопкой мыши «Ряд1» → «Изменить» → в открывшемся окне «Имя ряда» введите «Мать» → Ок. Аналогично переименуйте Ряд2.

8. Вставьте название диаграммы. «Макет» → «Название диаграммы». Подберите шрифт. Чтобы название диаграммы не «наезжало» на легенду, можно уменьшить «Область построения».

9. Наконец, растяните диаграмму вниз, чтобы отобразились все категории. Должен получиться рисунок 3.37.

Диаграммы такого типа широко используются также в социологии (рис. 3.38).



Рис. 3.37
Диаграмма «Водопад»

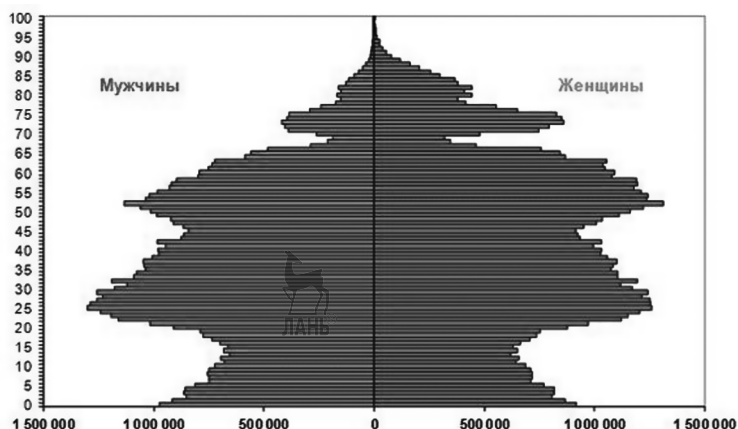


Рис. 3.38
Половозрастная пирамида населения России на начало 2013 г.
(Источник: Демографический ежегодник России, 2013 г.)

Для построения такой диаграммы можно использовать и другие средства. Например, «Условное форматирование» → «Гистограммы» (рис. 3.39).

Для создания диаграммы можно использовать также формулы (рис. 3.40).

Терминальные ценности	Мать	Дочь
Активная деятельная жизнь	15	15
Жизненная мудрость	1	3
Здоровье	7	14
Интересная работа	8	12
Красота природы и искусство	16	17
Любовь	11	10
Материально обеспеченная жизнь	12	13
Наличие хороших и верных друзей	9	11
Общественное признание	17	5
Познание	5	1
Продуктивная жизнь	2	2
Развитие	6	8
Развлечения	18	18
Свобода	4	6
Счастливая семейная жизнь	13	4
Счастье других	14	16
Творчество	10	9
Уверенность в себе	3	7

Рис. 3.39
Гистограмма в ячейке

	А	В	С	Д	Е	Ф
1	Терминальные ценности	Мать	Дочь	Мать		Дочь
2	Активная деятельная жизнь	15	15			
3	Жизненная мудрость	1	3			
4	Здоровье	7	14			
5	Интересная работа	8	12			
6	Красота природы и искусство	16	17			
7	Любовь	11	10			
8	Материально обеспеченная жизнь	12	13			
9	Наличие хороших и верных	9	11			
10	Общественное признание	17	5			
11	Познание	5	1			
12	Продуктивная жизнь	2	2			
13	Развитие	6	8			
14	Развлечения	18	18			
15	Свобода	4	6			
16	Счастливая семейная жизнь	13	4			
17	Счастье других	14	16			
18	Творчество	10	9			
19	Уверенность в себе	3	7			

Рис. 3.40
Создание гистограммы в ячейке с помощью формул

Здесь в ячейке D2 записана формула =ПОВТОР("";B2), затем эта ячейка протянута вниз. Аналогично заполняется столбец для дочери, в F2 записана формула =ПОВТОР("";C2), затем эта ячейка протянута вниз.

Пример 3.13. Требуется построить диаграмму с ве-хами. «Исторические» данные, по которым будем стро-ить диаграмму, приведены на рисунке 3.41.

Создадим две однотипные диаграммы, разделив даты на два периода: 1994–2001, 2011–2012. Приведем для удобства конечный вид диаграммы периода 1994–2001 (рис. 3.42).

1. Выделить диапазон данных A2:B6 (см. рис. 3.40). Создать диаграмму «Точечная с маркерами». Получим вот такой эскиз (рис. 3.43).

2. Формат горизонтальной оси значений (рис. 3.44): «Макет» → «Текущий фрагмент» → «Горизонтальная ось (значений)» → «Формат выделенного». Появится диалого-вое окно «Формат оси». Выбрать «Параметры оси» и уста-новите минимальное значение — 1994,0; максимальное значение — 2002,0; цена основных делений: — 1. Закрыть.

	А	В	С
1	Дата	Данные	Подписи данных
2	1994	6	Мама, сказка
3	1996	8	Каша, кошка
4	1998	10	Книжка, яркая обложка
5	1999	12	Буратино, Карабас
6	2001	14	Ранец, школа, первый класс!
7	Июнь 2011	4	Институт
8	Июль 2011	5	Экзамен, нервы
9	Август 2011	6	Конкурс
10	Сентябрь 2011	7	Лекция, курс первый!
11	Октябрь 2011	8	Тренировки, Семинары
12	Ноябрь 2011	10	Песни, танцы, тары-бары
13	Декабрь 2011	11	Прочность знаний, чет-не-чет
14	Январь 2012	15	Радость, сессия, зачет!

Рис. 3.41

Исходные данные для построения диаграммы



Рис. 3.42
Диаграмма с вехами 1994–2001

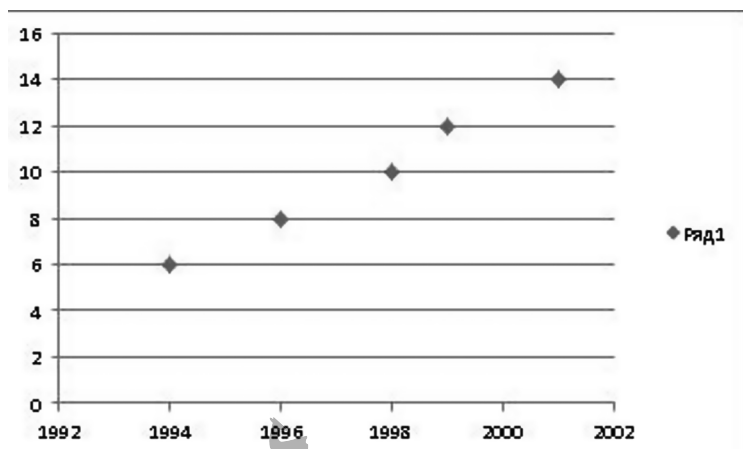


Рис. 3.43
Создание диаграммы с вехами, шаг 1

Формат оси

Параметры оси

Число

Заливка

Цвет линии

Тип линии

Параметры оси

минимальное значение: ☐ авто ☒ фиксированное 1994,0

максимальное значение: ☐ авто ☒ фиксированное 2002,0

цена основных делений: ☐ авто ☒ фиксированное 1

цена промежуточных делений: ☒ авто ☐ фиксированное 0,2

Рис. 3.44
Настройка параметров оси

3. Удалить «Легенду», «Вертикальную ось» и «Линии сетки». Для этого выделите по очереди каждый элемент левой кнопкой мыши и нажмите на клавиатуре кнопку Delete.

4. Вставить название диаграммы: «Макет» → «Название диаграммы». Подобрать шрифт. Чтобы изменить шрифт, щелкнуть правой кнопкой мыши по названию, в открывшемся контекстном меню выбрать «Шрифт». Указать «Шрифт», «Начертание», «Размер». Результат на рисунке 3.45.

5. Формат ряда данных: изменить «Параметры маркера» — установить «темно-синий» цвет и размер маркера — 10 (рис. 3.45). Как форматировать отдельный элемент, уже подробно обсуждалось в предыдущем примере.

6. Теперь — самое интересное. Нужно построить вертикальные пунктирные линии. Для этого используем «Планки погрешностей»: «Макет» → «Анализ» → «Планки погрешностей» → «Дополнительные параметры планок погрешностей». Откроется диалоговое окно «Формат планки погрешностей» (рис. 3.46).

Установить «Вертикальные планки погрешностей», как указано на рисунке 3.46, т. е. «Минус» и «Точка».



Рис. 3.45
Создание диаграммы с вехами, шаг 2

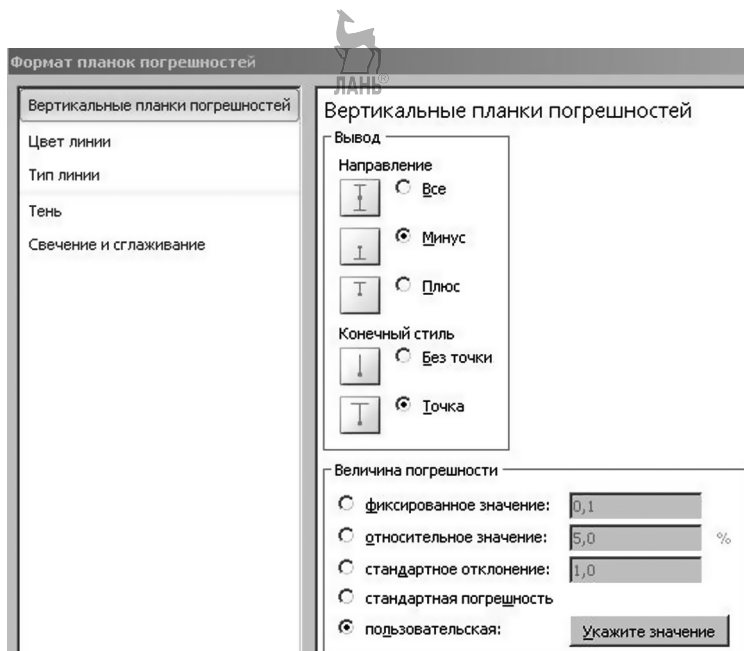


Рис. 3.46
Формат планок погрешностей

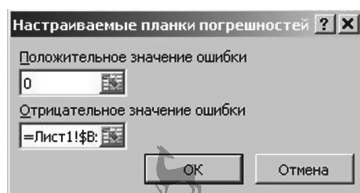


Рис. 3.47
Установка параметров
планки погрешностей

Установить «Величину погрешности». Для этого нужно выбрать «Пользовательская» и нажать кнопку «Укажите значение», откроется новое диалоговое окно (рис. 3.47).

Установить «Положительное значение ошибки» — 0, «Отрицательное значение ошибки» — ссылка на диапазон

B2:B6. Нажав **Ok**, вернуться в диалоговое окно «Формат планки погрешностей» (рис. 3.46).

7. Выполнить: «Тип линии» → установить: «Ширина» — 2 пт, «Тип штриха» — квадратные точки. «Закрывать». Результат на рисунке 3.48.

8. По умолчанию установились также горизонтальные линии погрешностей. Удалить их.

9. Добавление «Подписей данных». Здесь придется потрудиться. «Макет» → «Подписи данных» → «Сверху». Подписи, конечно, добавятся. Но это не то, что нам нужно. Придется добавлять их вручную.

Подпишем для примера первую точку, 1994 год. Для этого нужно выделить подпись только этой точки: сначала выделите все подписи данных, щелкнув по любой из них левой кнопкой мыши, и сразу щелкните еще раз по первой надписи, там пока 6. Выделится эта надпись. Мышью щелкните внутри надписи, удалите 6 и введите надпись «Мама, кошка» вручную. Чтобы надпись была в две строки, после запятой нажмите **Enter** для перехода к новой строке (рис. 3.49). Аналогично подпишите остальные данные.



Рис. 3.48
Создание диаграммы с вехами, шаг 3



Рис. 3.49
Создание диаграммы с вехами, шаг 4

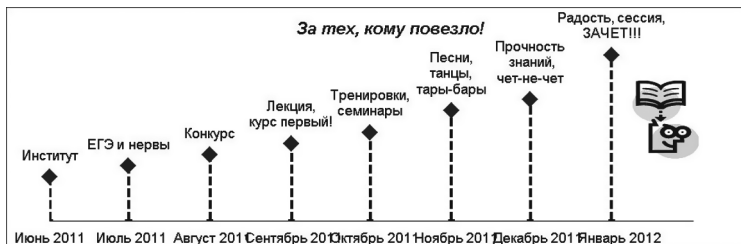


Рис. 3.50
Диаграмма с вехами 2011–2012

Обрадуем, что в Excel 2013 добавлена возможность под-
писи данных из ячеек.

Диаграмма готова (см. рис. 3.42). Диаграмму для пе-
риода 2001–2012 приводим на рисунке 3.50.

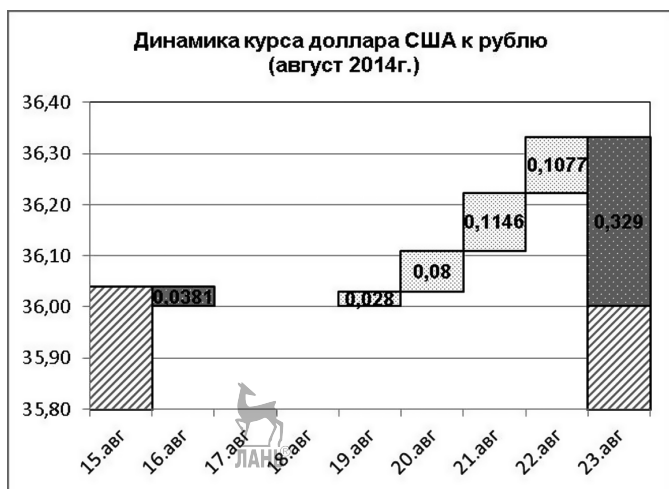
Пример 3.14. Требуется построить диаграммы для эконо-
мистов. Что волнует экономистов? Конечно, курс дол-
лара или евро. Рассмотрим колебание курса доллара по
отношению к рублю за небольшой период (рис. 3.13).

Можно проследить за курсом, используя диаграмму
«Водопад» (рис. 3.51). Идея создания такой диаграммы
приведена на сайте <http://www.planetaexcel.ru/>.

Таблица 3.13

Динамика курса доллара США к рублю (август 2014 г.)

Дата	Курс	Изменение
15 авг	36,0395	-0,1827
16 авг	36,0014	-0,0381
19 авг	36,0294	0,028
20 авг	36,1094	0,08
21 авг	36,224	0,1146
22 авг	36,3317	0,1077
23 авг	36,0027	-0,3290

Рис. 3.51
Диаграмма «Водопад»

Построим диаграмму пошагово. Для построения диаграммы нужно немного изменить таблицу данных. Необходимо создать таблицу для построения диаграммы (рис. 3.52).

1. В ячейку B2 ввести формулу $=B3-C3+D2$ и протянуть ее вниз. Получим исходные данные для построения диаграммы (рис. 3.53).

2. Выделить всю таблицу данных, включая заголовки. Построить «Гистограмму с накоплением» (рис. 3.54).

	A	B	C	D
1	Дата	Курс	Минус	Плюс
2	15.авг	36,0395		
3	16.авг		0,0381	
4	19.авг			0,028
5	20.авг			0,08
6	21.авг			0,1146
7	22.авг			0,1077
8	23.авг		0,329	

Рис. 3.52
Создание таблицы, шаг 1

	A	B	C	D
1	Дата	Данные	Минус	Плюс
2	15.авг	36,0395		
3	16.авг	36,0014	0,0381	
4	19.авг	36,0014		0,028
5	20.авг	36,0294		0,08
6	21.авг	36,1094		0,1146
7	22.авг	36,224		0,1077
8	23.авг	36,0027	0,329	

Рис. 3.53
Создание таблицы, шаг 2

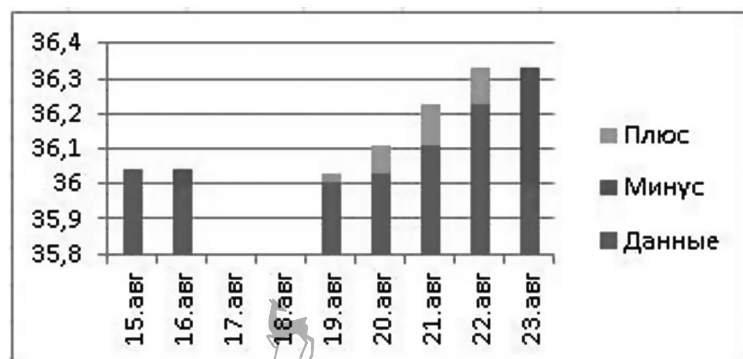


Рис. 3.54
Создание диаграммы «Водопад», шаг 1

Эта диаграмма по умолчанию будет трехцветной, ряд данных «Плюс» — зеленый, «Минус» — красный и «Данные» — синий. Нужно перейти к черно-белому варианту для наглядности.

3. Формат ряда данных «Минус». В диалоговом окне «Формат ряда данных» установите: «Заливка» → «Узорная заливка» → «90%» (темный фон с белыми точками), «Цвет границы» → «Сплошная линия», «Параметры ряда» → «Боковой зазор» — Без зазора.

4. Формат ряда данных «Плюс». В диалоговом окне «Формат ряда данных» установите: «Заливка» → «Узорная заливка» → «20%» (белый фон с черными точками), «Цвет границы» → «Сплошная линия».

5. Формат ряда данных «Данные». В диалоговом окне «Формат ряда данных» установить: «Заливка» → «Нет заливки». Должно получиться как на рисунке 3.55.

6. Выделить первый невидимый столбец «Данных». Сделаем узорную заливку типа «Широкий диагональный 2», «Цвет границы» → «Сплошная линия». Аналогично отформатируем последний невидимый столбец «Данные». Удалим легенду (рис. 3.56).

7. Добавить заголовок диаграммы. Выделить вертикальную ось, «Формат оси»: число, числовой, число десятичных знаков — 2, чтобы были видны копейки.

8. Добавить подписи данных для рядов «Плюс» и «Минус».

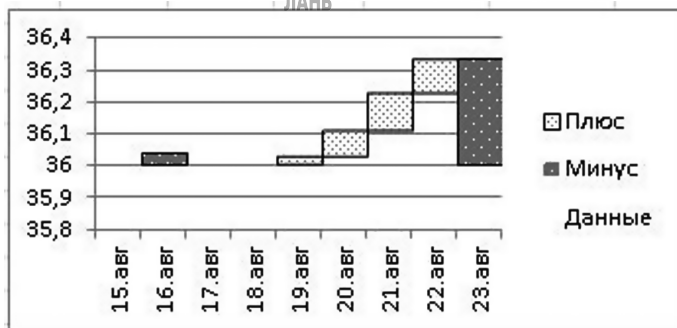


Рис. 3.55
Создание диаграммы «Водопад», шаг 2

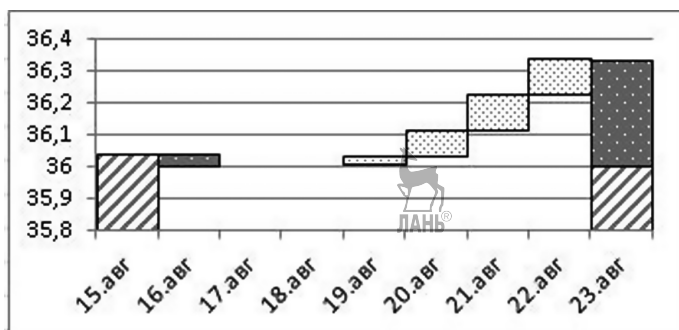


Рис. 3.56
Создание диаграммы «Водопад», шаг 3

Дата	Курс	Изменение	Гистограмма
15.авг	36,0395	-0,1827	
16.авг	36,0014	-0,0381	
19.авг	36,0294	0,028	
20.авг	36,1094	0,08	
21.авг	36,224	0,1146	
22.авг	36,3317	0,1077	
23.авг	36,0027	-0,329	

Рис. 3.57
Использование условного форматирования

9. Изменить размеры диаграммы. Сравнить с рисунком 3.51.

Чтобы проследить падение/возрастание курса, можно использовать условное форматирование. Например, построить «Гистограмму в ячейке» (рис. 3.57).

ГЛАВА 4

СТАТИСТИЧЕСКИЙ ПАКЕТ АНАЛИЗА ДАННЫХ



Современная производственная и научная деятельность людей связана с необходимостью анализа непрерывно нарастающего потока данных. Анализ данных необходим для создания моделей изучаемых явлений и правильного принятия решения. Анализ и обработка данных, как правило, ведется на разных уровнях, которые различаются как объемом, так и характером данных. Анализировать приходится небольшие массивы (< 100 данных) и громадные потоки данных, имеющих в своем составе до 10^{10} и более данных. Данные могут представлять собой числа (скаляры) или векторы (наборы чисел). В ряде случаев объемы данных настолько велики, что их полный анализ не под силу даже современным мощным компьютерам и в этом случае из общего числа данных производится выборка меньшего объема для производства анализа. Часто выборка осуществляется случайным образом.

Наука об обработке данных — это быстро развивающаяся область. В последние годы появились новые дисциплины «data mining» («раскапывание» данных), «big data analysis» (анализ больших данных) и др. Эти новые дисциплины используют, в частности, средства математической статистики и специализированные пакеты обработки данных.

Пакет Excel, рассматриваемый в данном руководстве, также может использоваться в задачах обработки данных в простейших случаях и на начальных этапах исследований. Обычно он используется для анализа скалярных данных небольшого объема.

Пакет Excel оснащен средствами статистической обработки данных, которые собраны в надстройке «Анализ данных». И хотя Excel существенно уступает специализированным статистическим пакетам, тем не менее, этот раздел математики представлен в Excel достаточно полно. В него включены основные, наиболее часто используемые статистические инструменты: средства описательной статистики, критерии различия, корреляционные и другие методы, позволяющие проводить первичный статистический анализ экономических, психологических, педагогических и медико-биологических данных.

Для проведения полноценного статистического анализа инструментов Excel может быть недостаточно. Но для проведения грамотного базового анализа, с которого обычно и начинается исследование, возможностей Excel вполне хватает. К тому же многие специализированные статистические пакеты, в частности SPSS, SAS, STATISTICA, поддерживают импорт данных из Excel.

Отметим, что данное пособие не является учебником по статистическим методам. В этой главе будут рассмотрены основные инструменты надстройки «Анализ данных», их назначение, описание входных и выходных параметров. На конкретных примерах будет рассмотрена работа с этими инструментами. Предполагается, что читатель знаком с элементарными понятиями математической статистики или познакомится с ними с помощью литературы, приведенной в конце книги.

ТЕМА 4.1. ИНСТРУМЕНТЫ НАДСТРОЙКИ. АНАЛИЗ ДАННЫХ

Средство Excel «Анализ данных» является надстройкой, которую перед первым использованием необходимо установить. Выполним для этого последовательность действий: «Файл» → «Параметры» → «Надстройки» → «Имя» → «Пакет анализа» → «Перейти» → откроется меню «Надстройки» → «Доступные надстройки» → выберите «Пакет анализа» → Ok. В результате на вкладке «Данные» в группе

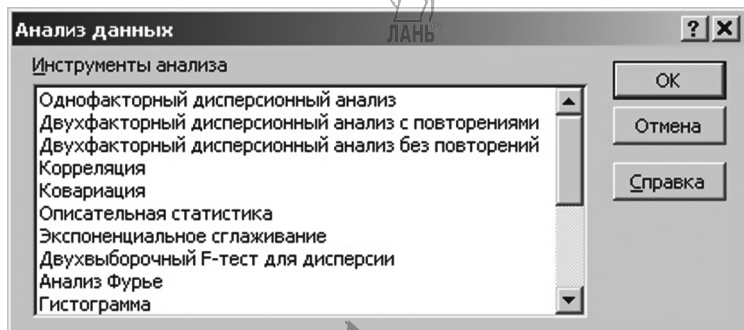
«Анализ» появится пиктограмма, которая обеспечит доступ к «Инструментам анализа».

Данная надстройка предназначена для выполнения базовых операций статистического анализа данных. Используется она и при проведении инженерных расчетов. При запуске надстройки открывается диалоговое окно, в котором можно выбрать необходимый инструмент анализа (рис. 4.1).

Всего инструментов анализа в настоящее время 19. По назначению их можно разбить на несколько групп:

- инструменты описательной статистики и построения гистограмм;

а



б

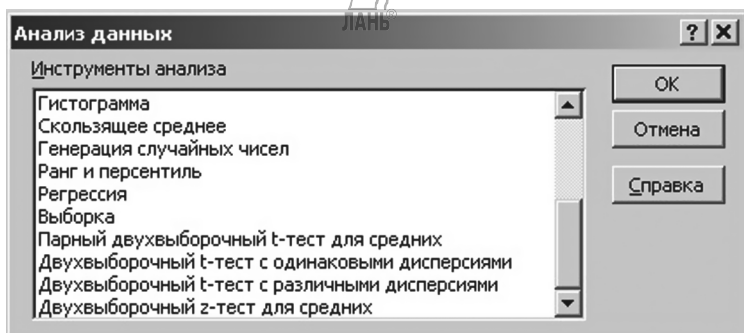


Рис. 4.1
Диалоговое окно «Анализ данных»:

а — начало списка инструментов; *б* — окончание списка.

- генератор случайных чисел для различных распределений и инструменты для создания случайной выборки;
- инструменты для подсчета рангов и персентилей;
- инструменты для вычисления корреляций и матриц ковариации;
- скользящие средние и инструменты однопараметрического экспоненциального сглаживания;
- инструменты множественной линейной регрессии;
- инструменты дисперсионного анализа, включая однофакторный, двухфакторный без повторений и двухфакторный сбалансированный с повторениями;
- парные двухвыборочные t -тесты с одинаковой и разной дисперсией;
- z -тест для средних и F -тест для дисперсий;
- быстрое преобразование Фурье.

ТЕМА 4.2.

ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Первичный анализ скалярных экспериментальных данных начинается с вычисления описательных статистик. Добавив к этому графические характеристики, получим некоторые основания для выводов о характере распределения данных исследуемой совокупности. К тому же базовый анализ дает основу для дальнейшего проведения более сложного анализа данных.

Из множества инструментов надстройки «Анализ данных» будем использовать «Описательную статистику» для получения числовых характеристик и «Гистограмму» — для графических. Заметим, что наряду с этим можно использовать также встроенные «Статистические функции», которые дублируют возможности надстройки.

Рассмотрим работу с описательной статистикой на примере.

Пример 4.1. Имеются некоторые данные о стоимости новогодних туров (рис. 4.2). Каждый из столбцов можно рассматривать как отдельный признак или переменную. Требуется провести анализ данных о продолжительности туров.

	A	B	C	D	E	F	G
1	Новогодние туры						
2	№	Название фирмы	Страна	Кол-во дней	Стоимость	Отель	Транспорт
3	1	Нева	Италия	11	447 €	***	Поезд
4	2	Нева	Германия	7	374 €	****	Самолет
5	3	Нева	Польша	7	199 €	***	Поезд
6	4	Нева	Франция	10	581 €	****	Автобус
7	5	Нева	Швейцария	8	1 240 €	***	Самолет
8	6	Одиссея	Норвегия	12	645 €	***	Паром
9	7	Одиссея	Испания	8	796 €	****	Самолет
10	8	Одиссея	Италия	12	430 €	****	Поезд
11	9	Одиссея	Польша	9	265 €	****	Поезд
12	10	Олимпиада	Финляндия	3	175 €	***	Автобус
13	11	Олимпиада	Швеция	5	800 €	***	Поезд
14	12	Олимпиада	Мальдивы	14	2 890 €	*****	Самолет
15	13	Олимпиада	Германия	12	700 €	*****	Поезд
16	14	Олимпиада	Португалия	8	1 460 €	*****	Самолет

Рис. 4.2
Таблица исходных данных

Исходные данные содержат несколько переменных, характеризующих тур. «Название фирмы», «Страна», «Транспорт» — качественные переменные, которые относятся к номинальной шкале. «Отель» — качественная переменная, которую можно отнести к порядковой шкале, так как количество звездочек отражает уровень обслуживания в отеле. «Количество дней» и «Стоимость» — количественные данные, которые относятся к метрической шкале.

Вычислим основные описательные статистики для переменной «Количество дней», которая является числовой переменной, принимающей дискретные значения. Для этого используем инструмент «Описательная статистика», входящий в «Пакет анализа».

Для перехода к описательной статистике выполните: «Данные» → «Анализ» → «Анализ данных» → «Описательная статистика» → «Ок». В открывшемся диалоговом окне «Описательная статистика» (рис. 4.3) укажите «Входной интервал», диапазон D2:D16, выберите «Групп-

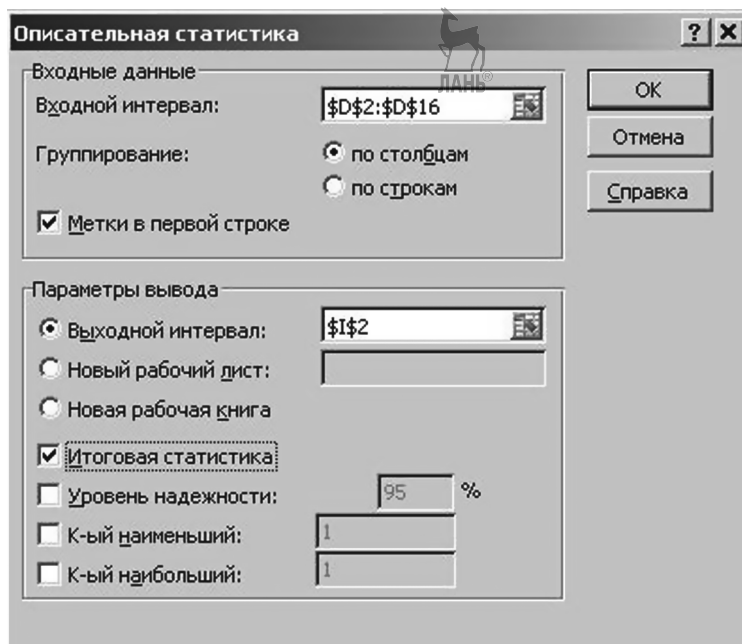


Рис. 4.3
Диалоговое окно «Описательной статистики»

пирование по столбцам», установите «Метки в первой строке», так как входной интервал содержит наименование столбца. Для «Выходного интервала» достаточно указать одну, первую, ячейку на текущем листе, как альтернативу можно выбрать «Новый рабочий лист» или «Новую рабочую книгу». И наконец, укажите хотя бы одну из выводимых статистик: «Итоговая статистика», «Уровень надежности», «К-й наименьший», «К-й наибольший».

В большинстве случаев достаточно выбрать «Итоговую статистику», которая рассчитывает основные числовые характеристики исследуемой совокупности. Три последних значения рассчитывают, только когда они действительно нужны.

«Описательная статистика» вычисляет 16 значений, из них 13 относятся к «Итоговой статистике», еще три опре-

деляют доверительный интервал и два выборочных значения.

Отметим главное — «Описательная статистика» надстройки «Анализ данных» предназначена для вычислений статистических характеристик, или статистик, одномерной *выборки* или нескольких выборок.

В литературе по статистике часто используют термин «генеральная совокупность». Обычно имеется в виду, что это множество всех доступных для наблюдения данных в противоположность «выборки» — которая подразумевает, что исследуется лишь часть данных выбранных из генеральной совокупности (может быть с помощью случайного отбора).

Обычно числовые характеристики генеральной совокупности называют *параметрами*, а числовые характеристики выборки — *статистиками*, или *выборочными характеристиками*, которые являются оценками параметров генеральной совокупности. Для более полного понимания *выборочного метода* следует обратиться к специальной литературе.

Результаты расчетов «Итоговой статистики» для переменной «Количество дней» приведены на рисунке 4.4. На этом же рисунке приведены альтернативные расчеты этих числовых характеристик с использованием встроенных функций категории «Статистические». Аргументом статистических функций является диапазон исходных данных, в данном случае D3:D16.

Таким образом, практически все расчеты «Описательной статистики» дублируются «Статистическими» функциями. Остальные характеристики можно посчитать, используя формулы. Для того чтобы на рабочем листе Excel отображались не результаты, а формулы, следует выполнить: «Формулы» → «Зависимости формул» → «Показать формулы».

Отметим некоторое отличие в применении инструментов «Анализа данных» и использовании статистических функций. При изменении значений исходных данных формулы пересчитываются, в то время как результаты, полученные с помощью инструментов «Анализа данных»,

I	J	K
Кол-во дней		Статистические функции
Среднее	9	=СРЗНАЧ(D3:D16)
Стандартная ошибка	0,81199794294115	
Медиана	8,5	=МЕДИАНА(D3:D16)
Мода	8	=МОДА(D3:D16)
Стандартное отклонение	3,038218101251	=СТАНДОТКЛОН(D3:D16)
Дисперсия выборки	9,23076923076923	=ДИСП(D3:D16)
Эксцесс	-0,290530303030302	=ЭКСЦЕСС(D3:D16)
Асимметричность	-0,268797907013457	=СКОС(D3:D16)
Интервал	11	
Минимум	3	=МИН(D3:D16)
Максимум	14	=МАКС(D3:D16)
Сумма	126	=СУММ(D3:D16)
Счет	14	=СЧЁТ(D3:D16)

Рис. 4.4
«Итоговая статистика» и «Статистические функции»

не изменяются. Чтобы обновить результаты, потребуется вызывать «Анализ данных» снова.

Числовые характеристики «Итоговой статистики» описывают средние, вариацию и форму распределения, всего 13 параметров:

- *среднее*, или *выборочное среднее*, вычисляется как среднее арифметическое наблюдаемых значений выборки;
- *медиана* определяется как значение, находящееся в середине распределения, полученного из исходного путем упорядочивания по возрастанию;
- *мода* равна наиболее часто встречающемуся значению.

Кроме того, выделяют две величины, характеризующие изменчивость, или разброс, значений распределения относительно среднего:

1) *дисперсию выборки*, или *выборочную дисперсию*, равную сумме квадратов отклонений каждого значения от среднего, деленной на $(N - 1)$, где N — число значений в распределении, или объем выборки;



2) *стандартное отклонение*, или *выборочное средне-квадратическое отклонение*, равное квадратному корню из *выборочной дисперсии*.

Дополнительными мерами изменчивости являются три простые характеристики, отражающие границы распределения данных и его размах:

- *минимум* равен наименьшему из выборочных значений;
- *максимум* равен наибольшему из выборочных значений;
- *интервал* составляет разность между максимумом и минимумом, этот параметр называют также *размахом*.

Если набор данных рассматривается как множество независимых реализаций случайной величины, то возникает вопрос, что можно сказать о функции распределения этой величины на основании выборки. Очень часто распределение оказывается нормальным или близким к нему.

Для отражения близости формы распределения к нормальному виду существует две основные характеристики:

- 1) *эксцесс*, или *выборочный коэффициент эксцесса*, который является мерой «сглаженности» распределения;
- 2) *асимметричность*, или *выборочный коэффициент асимметрии*, показывает, в какую сторону относительно среднего сдвинуто большинство значений выборки.

И наконец, *сумма* равна сумме всех выборочных значений, *счет* вычисляет объем выборки, *стандартная ошибка* равна выборочному стандартному отклонению, деленному на квадратный корень из объема выборки.

При необходимости можно вычислить три дополнительные характеристики (рис. 4.5). Результаты расчетов этих характеристик приведены на рисунке 4.6.

«К-й наибольший» выдает К-е выборочное значение, если бы выборка была отсортирована по убыванию. В рассматриваемом примере сортировка по убыванию имеет вид 14, 12, 12, 12, 11, 10 и т. д., третье значение равно 12. «К-й наименьший» выдает К-е выборочное значение, если бы выборка была отсортирована по возрастанию, это значение равно 5.

Задав «Уровень надежности», например 95%, получим значение для построения доверительного интервала для

Описательная статистика

Входные данные

Входной интервал:

Группирование: ☒ по столбцам ☐ по строкам

☒ Метки в первой строке

Параметры вывода

☒ Выходной интервал:

☐ Новый рабочий лист:

☐ Новая рабочая книга

☐ Итоговая статистика

☒ Уровень надежности: %

☒ К-ый наименьший:

☒ К-ый наибольший:

OK Отмена Справка

Рис. 4.5
Описательная статистика, дополнительные параметры

P	Q
Кол-во дней	
Наибольший(3)	12
Наименьший(2)	5
Уровень надежности(95,0%)	1,7542149

Рис. 4.6
Результаты расчетов дополнительных параметров

неизвестного математического ожидания генеральной средней с доверительным уровнем 95%. Доверительный интервал строится как выборочное среднее плюс-минус полученное значение. Обратите внимание, что граница здесь вычисляется с помощью распределения Стьюдента, что

требует достаточного количества наблюдений на каждую степень свободы.

Таким образом, к вычислению доверительных интервалов нужно относиться с осторожностью, особенно при малых выборках. Использование функции расчета доверительного интервала без понимания статистического смысла может привести к ошибкам. Начинаящим исследователям посоветуем обратиться к специальной литературе.

Например, для рассматриваемого примера полученный доверительный интервал не несет смыслового содержания.

Итак, на этапе проведения описательной статистики исследуемый ряд данных может быть как генеральной совокупностью, так и выборкой. Если для генеральной совокупности вычисляются значения параметров распределения, то для выборки находят оценки этих параметров. Рассмотрим ниже подробнее вычисление некоторых числовых характеристик в пакете Excel.

4.2.1. ДИСПЕРСИЯ

Описательная статистика, реализованная в «Пакете анализа», рассчитывает *выборочные* характеристики, т. е. *среднее* здесь — это выборочная оценка математического ожидания генеральной совокупности, из которой извлечена выборка, *дисперсия выборки* — выборочная оценка дисперсии генеральной совокупности, т. е. несмещенная оценка, *стандартное отклонение* — оценка среднеквадратического отклонения на основе несмещенной оценки дисперсии. Описательная статистика при вычислении среднего и выборочной дисперсии использует, соответственно, следующие формулы:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}; \quad S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1},$$

где N — объем выборки; x_i — i -й элемент выборки.

Описательная статистика в Excel предназначена для работы именно с выборками. Одна из статистик так и называется — дисперсия выборки. Поэтому, если расчеты

проводятся непосредственно для генеральной совокупности, для вычисления дисперсии нужно использовать специальную встроенную статистическую функцию. В последних версиях Excel 2010/2013 на это различие обратили особое внимание, добавив функции с соответствующими названиями:

- ДИСПР.Г вычисляет дисперсию для генеральной совокупности по формуле (μ — математическое ожидание):

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N};$$

- ДИСПР.В оценивает дисперсию генеральной совокупности по выборке, используя формулу:

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}.$$

При этом оставлены еще две дублирующие функции ДИСПР и ДИСП для обеспечения совместимости, чтобы не пришлось переделывать многочисленные расчеты, выполненные в более ранних версиях.

Соответственно при вычислении *среднеквадратического отклонения*, или *среднего квадратического отклонения* (СКВО), для генеральной совокупности следует использовать статистическую функцию СТАНДОТКЛОН.Г, а для его оценки по выборке нужно использовать СТАНДОТКЛОН.В.

Иногда дисперсию выборки считают по формуле:

$$\tilde{S}^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}.$$

При этом называют S^2 *исправленной выборочной дисперсией*, которая является несмещенной оценкой дисперсии генеральной совокупности, а \tilde{S}^2 — *выборочной дисперсией*, которая является смещенной оценкой дисперсии генеральной совокупности. Со свойствами, которыми должны обладать оценки, можно ознакомиться в специальной литературе.

Если объем выборки велик, то разница между значениями S^2 и \tilde{S}^2 невелика. При малых N это различие существенно. Вычислим оба значения для рассмотренного выше примера: $\text{ДИСП.В}(\text{D3:D16})=\text{ДИСП}(\text{D3:D16})=9,23077$, а $\text{ДИСП.Г}(\text{D3:D16})=\text{ДИСПР}(\text{D3:D16})=8,57143$.

В «Описательной статистике» надстройки «Анализ данных» дисперсия вычисляется с использованием функции ДИСП.В , т. е. вычисляется исправленная выборочная дисперсия S^2 . Соответственно, для вычисления среднеквадратического отклонения используется функция СТАНДОТКЛОН.В .

4.2.2. МОДА И МЕДИАНА

Мода — наиболее часто встречающееся значение во множестве наблюдений. Если такое значение только одно, распределение называется *унимодальным*, а если несколько — *полимодальным*. Изучаемая случайная величина может не иметь моды, в этом случае Excel выдает сообщение об ошибке #Н/Д.

Для вычисления моды в Excel есть несколько встроенных функций:

а) МОДА.ОДН и МОД вычисляют моду для унимодального распределения и выдают только одно значение моды, даже если распределение полимодально;

б) МОДА.НСК вычисляет моду для полимодального распределения и возвращает вертикальный массив наиболее часто встречающихся значений в указанном диапазоне, т. е. несколько значений моды.

Заметим, что при вычислении моды с помощью «Описательной статистики» используется функция МОДА.ОДН , т. е. выдается только одно значение моды, меньшее по значению. Так в рассмотренном выше примере (см. рис. 4.2) расчетное значение моды равно 8 (см. рис. 4.4), хотя числа 12 и 8 встречаются по три раза.

При вычислении моды рекомендуется сначала использовать функцию МОДА.НСК . Применение этой функции имеет свои особенности. Формулу $=\text{МОДА.НСК}(\text{диапазон})$ необходимо ввести как формулу массива.

Замечание. Ряд функций в Excel необходимо вводить как *формулу массива*, так как они возвращают несколько значений, которые выводятся в диапазон ячеек, или массив. Для ввода формулы массива выделите диапазон, в который будет сохранен результат, введите формулу и нажмите комбинацию клавиш Ctrl + Shift + Enter.

Так как заранее не известно, имеет ли исследуемая совокупность моду, а если имеет, то одну или несколько, то диапазон для вывода может содержать несколько ячеек. Найдем моду для вышеуказанного примера. Выделите диапазон F3:F6 и введите формулу =МОДА.НСК(D3:D16), где диапазон D3:D16 задает исходные данные. Эта формула отобразится также в строке формул (рис. 4.7).

Теперь нажмите одновременно комбинацию клавиш Ctrl + Shift + Enter, формула введется во все выделенные ячейки как формула массива. Отобразятся два значения моды, в остальных ячейках появится сообщение «Нет данных» (рис. 4.8).

Если известны все значения признака, то для нахождения моды не требуется проводить дополнительные расчеты, значением моды является конкретное значение признака. Расчет моды для несгруппированных данных состоит в определении наиболее часто встречающегося значения. Для дискретного ряда распределения мода соответствует

fx =МОДА.НСК(D3:D16)			
D	E	F	
Кол-во дней			
11	=МОДА.НСК(D3:D16)		
7			
7			
10			
8			
12			
8			
12			
9			
3			
5			
14			
12			
8			

Рис. 4.7

Вычисление моды

fx {=МОДА.НСК(D3:D16)}	
F	
	8
	12
#N/D	
#N/D	

Рис. 4.8

Результаты вычисления моды

значению признака, имеющего наибольшую частоту. Моду для интервального ряда распределения определяют по специальной формуле, в этом случае ее значение вычисляется приближенно.

Медиана — такое значение признака, которое делит ранжированный ряд на две равные части со значениями признака меньше медианы и со значениями признака больше медианы. Для нахождения медианы исходный ряд предварительно упорядочивают по возрастанию (ранжируют).

Для вычисления медианы в Excel есть встроенная функция МЕДИАНА (диапазон), причем исходный ряд не требуется предварительно упорядочивать.

Если известны все значения признака, ряд не сгруппирован, то для нахождения медианы не требуется проводить дополнительные расчеты. Все сводится к нахождению порядкового номера медианы. Если данные содержат нечетное число значений, то медиана есть центральное значение. Если же данные содержат четное число значений, то медиана находится как среднее арифметическое двух центральных значений. Значением моды является конкретное значение признака.

Для интервальных рядов медиана рассчитывается по специальной формуле.

Мода и медиана называются структурными средними. Кроме того, часто используют понятие «пяти базовых показателей», в которые входят минимальное значение, 1 квартиль Q_1 , 2 квартиль Q_2 , 3 квартиль Q_3 и максимальное значение. Квартили — это значения признака, делящие ранжированную совокупность на четыре равновеликие части, 2 квартиль совпадает с медианой.

4.2.3.

АСИММЕТРИЯ И ЭКСЦЕСС

Асимметрия и эксцесс являются мерами формы распределения данных. *Асимметрия* является мерой несимметричности данных. *Эксцесс* характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением.

При положительной асимметрии, $As > 0$, значения распределения скучены в области малых значений, при этом мода будет меньше медианы, а медиана меньше среднего, то есть более половины значений будут меньше среднего. При отрицательной асимметрии, $As < 0$, значения распределения скучены в области высоких значений. В этом случае среднее будет меньше медианы, а медиана меньше моды, т. е. более половины значений будут больше среднего.

Положительный эксцесс характерен для остроконечного распределения, а отрицательный — для относительно сглаженного распределения.

Для нормального распределения асимметрия и эксцесс равны нулю. Это свойство нормального распределения часто используют для приближенной проверки нормальности исследуемой совокупности данных. Дополнительно к этому строят для наглядности гистограмму. Визуальное представление гистограммы называют формой распределения.

Средство «Генерация случайных чисел» можно рассматривать как для организации случайных чисел, так и для иллюстративного поведения случайных величин с заданным законом распределения.

Воспользуемся средством «Генерация случайных чисел», которое позволяет сформировать совокупность данных, имеющих заданный закон распределения. Сформируем нормальное и равномерное распределение, каждое объемом 200, вычислим для них описательную статистику и воспользуемся инструментом «Гистограмма».

Для этого выполним последовательность действий: «Пакет Анализа» → «Генерация случайных чисел» → в открывшемся диалоговом окне установим параметры для генерации выборки из нормального распределения (рис. 4.9). Параметр «Случайное рассеивание» устанавливать не нужно, он предназначен для возможности повторить генерацию тех же самых значений.

Для полученных данных выполним «Описательную статистику». Эта процедура уже была описана выше.

Для построения гистограммы выполним «Пакет Анализа» → «Гистограмма» → в открывшемся диалоговом

Рис. 4.9

Параметры для генерации нормального распределения

окне установим параметры, задав «Входной интервал» и установив «Вывод графика». Результаты разместим на «Новый рабочий лист» (рис. 4.10). Интервал карманов можно не устанавливать, в этом случае он сформируется автоматически.

В результате на новый рабочий лист будут помещены таблица и гистограмма (рис. 4.11) для сгенерированного нормального распределения. Первый столбец таблицы задает границы карманов, а второй частоту, т. е. количество элементов выборки, попавших в указанный карман.

Гистограмма — это столбиковая диаграмма для отображения распределения частот по диапазонам значений переменной. Горизонтальная ось соответствует значениям переменной, а вертикальная частотам. Построенную гистограмму называют также гистограммой частот, в отличие от гистограммы относительных частот.

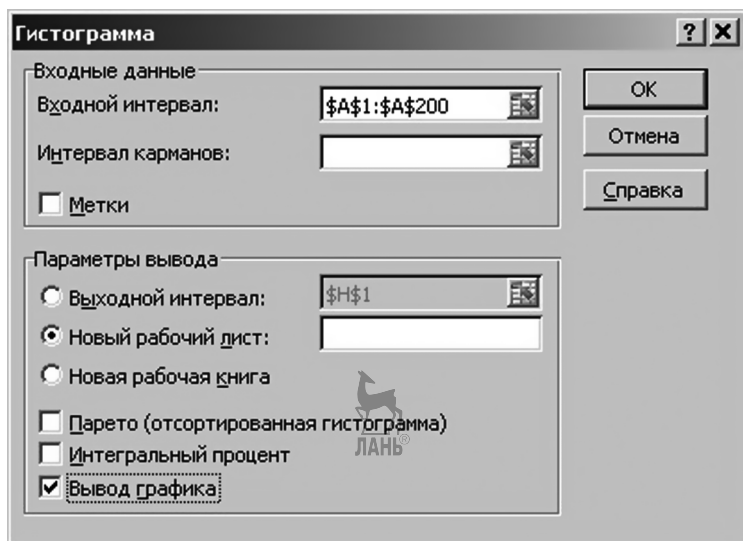


Рис. 4.10
Параметры для построения гистограммы

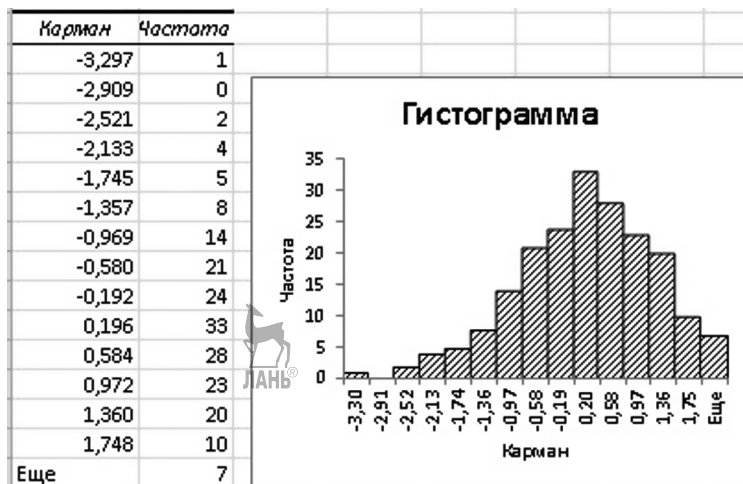


Рис. 4.11
Таблица частот и гистограмма для сгенерированного нормального распределения

Аналогичные действия выполним для генерации равномерного распределения. На рисунке 4.12 изображены таблица частот и гистограмма для сгенерированного равномерного распределения.

Возьмем еще один произвольный набор данных, тип распределения которых неизвестен, например, данные о выручке магазина за три месяца. Для него тоже посчитаем описательные статистики и построим гистограмму. На рисунке 4.13 изображены таблица частот и гистограмма, полученные для данных о выручке магазина.

Сравним описательные статистики для рассмотренных распределений (рис. 4.14). В столбце 1 расположены данные для выборки из нормального распределения, в столбце 2 — для выборки из равномерного, и в столбце 3 — для выборки с неизвестным распределением.

Предположим, нам ничего не известно о типе этих распределений, есть только описательные статистики и гистограммы. Сможем ли мы сделать некоторые выводы на основе этих данных. Какое из распределений отнести к нормальному? По виду гистограммы можно сразу исключить

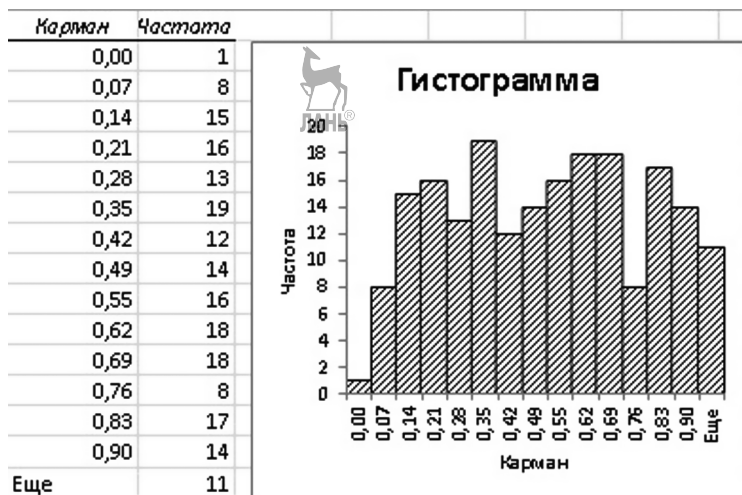


Рис. 4.12

Таблица частот и гистограмма для сгенерированного равномерного распределения

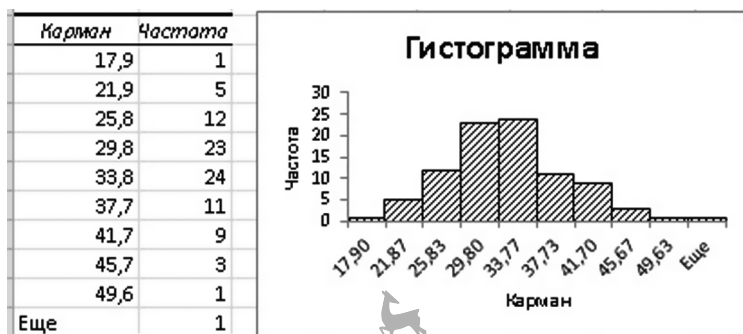


Рис. 4.13

Таблица частот и гистограмма для данных о выручке магазина

	A	B	D	E	G	H
1	Столбец 1		Столбец 2		Столбец 3	
2						
3	Среднее	0,002	Среднее	0,487	Среднее	30,68
4	Стандартная ошибка	0,073	Стандартная ошибка	0,019	Стандартная ошибка	0,71
5	Медиана	0,107	Медиана	0,501	Медиана	30,25
6	Мода	#N/Д	Мода	#N/Д	Мода	26,10
7	Стандартное отклон	1,036	Стандартное отклон	0,263	Стандартное отклон	6,73
8	Дисперсия выборки	1,072	Дисперсия выборки	0,069	Дисперсия выборки	45,30
9	Экссесс	-0,001	Экссесс	-1,107	Экссесс	0,82
10	Асимметричность	-0,412	Асимметричность	0,010	Асимметричность	0,65
11	Интервал	5,433	Интервал	0,964	Интервал	35,70
12	Минимум	-3,297	Минимум	0,004	Минимум	17,90
13	Максимум	2,136	Максимум	0,968	Максимум	53,60
14	Сумма	0,320	Сумма	97,357	Сумма	2761,10
15	Очет	200	Очет	200	Очет	90

Рис. 4.14

Описательная статистика для различных выборок

второй случай. Оставшиеся два имеют небольшую асимметрию, причем для первого набора она отрицательна, а для второго положительна.

Экссесс первого распределения довольно близок к нулю, что дает основания предположить, что данное распределение близко к нормальному. Однако это предположение следует проверить, используя критерии согласия.

Что касается третьего распределения, оно является островершинным, при этом среднее и медиана близки по значению. По виду гистограммы можно сказать, что оно

унимодальное, рассчитанное описательное значение моды не имеет содержательного смысла, так как исходные данные являются непрерывными. Данное распределение может быть близким к нормальному, однако это обязательно нужно проверить, используя различные статистические методы.



4.2.4.

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

«Описательная статистика» надстройки «Анализ данных» вычисляет доверительные интервалы. Для расчета доверительных интервалов в Excel есть еще три встроенные функции, относящиеся к разделу «Статистические». Функция **ДОВЕРИТ.СТЮДЕНТ** возвращает доверительный интервал для среднего генеральной совокупности, используя распределение Стьюдента, функция **ДОВЕРИТ.НОРМ** возвращает доверительный интервал для среднего генеральной совокупности, используя нормальное распределение, функция **ДОВЕРИТ** является аналогом функции **ДОВЕРИТ.НОРМ** и сохранена в пакетах последних версий для совместимости с более ранними версиями.

Все эти функции имеют три обязательных аргумента:

1) альфа задает уровень значимости, используемый при вычислении доверительного интервала. Он связан с «Уровнем надежности» в «Описательной статистике» формулой $\text{Уровень надежности} = (1 - \text{альфа}) * 100\%$. Например, значение аргумента альфа, равное 0,05, означает 95-процентный «Уровень надежности», который связан с доверительной вероятностью формулой $\text{Уровень надежности} = \text{Доверительная вероятность} * 100\%$;

2) **станд_откл** — выборочное стандартное отклонение для диапазона данных;

3) **размер** — объем выборки.

«Описательная статистика» надстройки «Анализ данных» при вычислении доверительных интервалов использует встроенную функцию **ДОВЕРИТ.СТЮДЕНТ**.

Для рассмотренного выше примера «Описательная статистика» выдает значение 1,7542149. Сравним: **ДОВЕРИТ.СТЮДЕНТ(0,05;N8;N16)= 1,7542149**.

Еще раз напомним, что расчет доверительного интервала и тем более интерпретация полученных результатов без понимания статистического смысла может привести к ошибкам. Начинающему пользователю необходимо предварительно обратиться к специальной литературе.

4.2.5. ПРИМЕРЫ ПРИМЕНЕНИЯ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ

Пример 4.2. В нашем распоряжении имеются выборочные данные о температуре воздуха, собранные на некоторой опытной станции за три года на небольшом острове. Нас интересует климат этого острова и насколько он комфортен для проживания. Исходные данные приведены в таблице 4.1.

Для проведения расчетов данные необходимо поместить в один столбец. Пусть она занимают диапазон A1:A100. Объем выборки равен $n = 100$.

Чтобы получить некоторые данные об изучаемой величине, воспользуемся описательной статистикой «Пакета анализа» (рис. 4.15). Исходный массив был предварительно отсортирован в порядке возрастания.

Таблица 4.1

Выборка

10,1	-4,3	-2,2	10,6	8,9	9,4	16,6	7,3	4,7	-2,7
15,2	17,1	-4,5	8,2	15,2	6,2	19,7	11,5	1,3	15,5
4,5	8,8	9,8	13,8	4,0	4,7	4,7	8,1	-3,1	14,8
13,5	20,1	12,0	7,4	8,3	6,7	-0,6	18,5	7,0	16,5
6,5	9,4	2,6	3,3	6,8	4,7	7,9	-1,0	6,6	17,2
-1,1	16,5	-9,3	2,7	10,6	9,5	11,5	-2,0	2,3	7,8
-3,0	12,3	5,7	0,5	3,8	5,0	2,3	7,9	20,1	-7,5
7,1	4,9	9,3	8,7	7,6	18,1	1,7	20,6	11,3	15,7
8,0	14,3	15,8	3,4	16,5	2,6	13,4	9,6	15,5	6,7
-2,8	-1,6	9,5	19,2	16,2	3,3	8,2	9,4	1,0	13,0



	A	B	САНЬ	D
1	10,1		Столбец1	
2	15,2			
3	4,5		Среднее	7,89
4	13,5		Стандартная ошибка	0,67
5	6,5		Медиана	7,95
6	-1,1		Мода	4,70
7	-3,0		Стандартное отклонение	6,73
8	7,1		Дисперсия выборки	45,35
9	8,0		Экссесс	-0,46
10	-2,8		Асимметричность	-0,17
11	-4,3		Интервал	29,90
12	17,1		Минимум	-9,30
13	8,8		Максимум	20,60
14	20,1		Сумма	789,10
15	9,4		Счет	100,00

Рис. 4.15
Описательная статистика

Средняя температура не радует — меньше чем в Сочи, где среднегодовая температура по поверхности России равна $+14,2^{\circ}\text{C}$, но она больше чем в Оймяконе, где всего $-15,5^{\circ}\text{C}$.

Построим гистограмму. Воспользуемся для этого средством «Гистограмма».

Интервалы карманов желательно предварительно вычислить. Для этого разбивают исходный ряд на интервалы. Обычно количество интервалов k задают, используя формулу Стерджесса $k = [1 + 3,321 \cdot \lg n]$, где $\lg n$ — десятичный логарифм, n — объем выборки. Квадратные скобки означают целую часть от числа, т. е. округление до ближайшего целого в меньшую сторону. В Excel для вычисления k можно использовать формулу $=\text{ЦЕЛОЕ}(1+3,321*\text{LOG10}(n))$. При $n = 100$ оптимальное число интервалов равно 7. Длина интервалов вычисляется по формуле

$$h = \frac{x_{\max} - x_{\min}}{k},$$

где x_{\max} — максимальное значение выборки; x_{\min} — минимальное; $h = 4,27$. Для удобства округлим значение длины интервала, примем $h = 4,3$.

Теперь воспользуемся инструментом «Гистограмма» из «Пакета анализа», который позволит посчитать частоты

и построит графические характеристики распределения — гистограмму и кумуляту. Подготовим для этого интервал карманов, диапазон I4:I9, формулы для расчетов приведены на рисунке 4.16. Интервал карманов отмечен серым фоном. Результаты вычислений приведены на рисунке 4.17. Здесь мы воспользовались некоторыми данными описательной статистики (см. рис. 4.15).

Выполним последовательность действий: «Данные» → «Анализ данных» → «Гистограмма». В открывшемся диалогом окне укажем входной интервал, интервал карманов, установим галочки «Интегральный процент» и «Вывод графика», для построения гистограммы и кумуляты. Выберем вывод результатов на новый рабочий лист. Установка параметров показана на рисунок 4.18.

В результате на новый рабочий лист выведется таблица расчета частот (рис. 4.19) и графики (рис. 4.20).

	F	G	H	I	J
1					
2	Интервал	=D11		Начало	Конец
3	Минимум	=D12		=G3	=I3+\$G\$7
4	Максимум	=D13		=I3+\$G\$7	=I4+\$G\$7
5	Счет	=D15		=I4+\$G\$7	=I5+\$G\$7
6	Кол. инт.	=ЦЕЛОЕ(1+3,321*LOG10(G5))		=I5+\$G\$7	=I6+\$G\$7
7	Длина	=ОКРУГЛ(G2/G6;1)		=I6+\$G\$7	=I7+\$G\$7
8				=I7+\$G\$7	=I8+\$G\$7
9				=I8+\$G\$7	=I9+\$G\$7

Рис. 4.16
Формулы для расчета интервала карманов

	F	G	H	I	J	K	L
1							
2	Интервал	29,90	Начало	Конец	Середина	Частота	
3	Минимум	-9,30	-9,30	-5,00	-7,15	2	
4	Максимум	20,60	-5,00	-0,70	-2,85	11	
5	Счет	100,00	-0,70	3,60	1,45	13	
6	Кол. инт.	7	3,60	7,90	5,75	22	
7	Длина интервала	4,30	7,90	12,20	10,05	25	
8			12,20	16,50	14,35	17	
9			16,50	20,80	18,65	10	

Рис. 4.17
Результаты расчетов интервала карманов

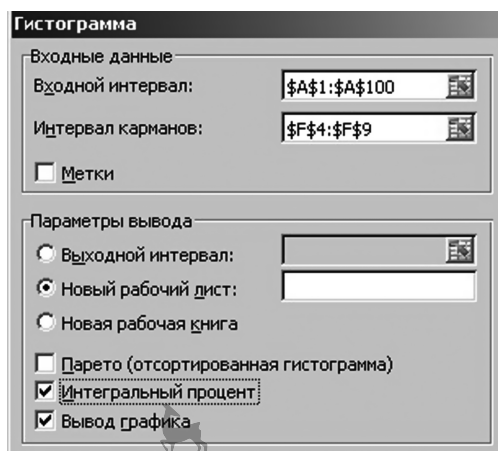


Рис. 4.18
Диалоговое окно инструмента «Гистограмма»

	А	В	С
1	Карман	Частота	Интегральный %
2	-5,00	2	2,00%
3	-0,70	11	13,00%
4	3,60	13	26,00%
5	7,90	22	48,00%
6	12,20	25	73,00%
7	16,50	17	90,00%
8	Еще	10	100,00%

Рис. 4.19
Результаты расчета частот

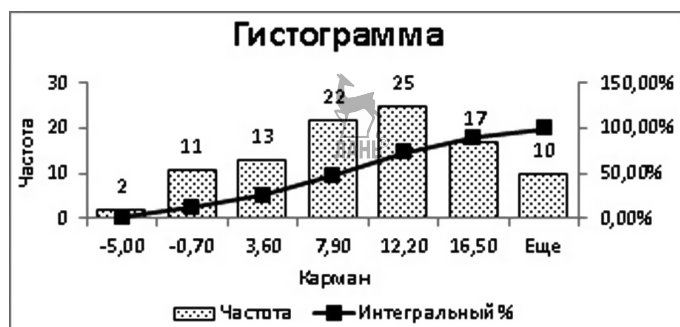


Рис. 4.20
Гистограмма и кумулята

Карман указывает верхнюю границу интервала. Интегральный процент отображает процентное содержание накопленных частот, которые используются для построения кумуляты. Графики, приведенные на рисунке 4.20 немного отредактированы, в частности, добавлены для наглядности подписи данных и значения частот.

На основании полученных расчетов можно построить интервальный ряд. *Интервальный* статистическим рядом называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами попаданий в каждый из них значений величины.

На рисунке 4.21 приведена таблица, в которой указаны границы интервалов, середины интервалов и частоты. На этом же рисунке приведен полученный интервальный ряд, где x_i — значения, n_i — частоты интервального ряда.

Таким образом, исходный диапазон выборки был разбит на интервалы, эти интервалы являются полуоткрытыми. Все выборочные значения, попавшие в интервал, заменяются серединой интервала. Полученный интервальный ряд может понадобиться для дальнейшего статистического анализа.

I	J	K	L	M	N	O	P
Начало	Конец	Середина	Частота				
-9,30	-5,00	-7,15	2				
-5,00	-0,70	-2,85	11				
-0,70	3,60	1,45	13				
3,60	7,90	5,75	22				
7,90	12,20	10,05	25				
12,20	16,50	14,35	17				
16,50	20,80	18,65	10				
x_i	-7,15	-2,85	1,45	5,75	10,05	14,35	18,65
n_i	2	11	13	22	25	17	10

Рис. 4.21
Интервальный ряд

Пример 4.3. На рисунке 4.22 представлены данные о количестве голов, забитых в матчах чемпионата России по футболу в 2013/2014. Проведем анализ полученных данных.


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		АМК	АНЖ	ВОЛ	ДИН	ЗЕН	КРА	КРЫ	КУБ	ЛОК	РОС	РУБ	СПА	ТЕР	ТОМ	УРА	ЦСК
2	АМК		1	6	3	3	4	0	4	0	1	0	3	1	2	2	4
3	АНЖ	4		0	4	3	3	1	0	4	1	1	1	3	2	1	3
4	ВОЛ	2	3		5	4	1	3	1	3	3	3	1	1	1	3	3
5	ДИН	2	3	4		2	3	2	4	4	2	0	5	1	1	3	6
6	ЗЕН	2	3	2	3		5	3	2	3	2	8	6	2	0	3	2
7	КРА	3	1	3	2	3		2	3	4	2	1	4	5	4	1	1
8	КРЫ	4	2	4	3	5	1		0	4	2	4	3	2	1	2	4
9	КУБ	3	2	4	2	5	4	4		4	4	2	4	4	2	5	4
10	ЛОК	4	0	3	1	2	4	3	1		5	0	0	3	0	3	3
11	РОС	6	2	4	5	4	4	3	0	2		0	1	3	3	2	0
12	РУБ	3	6	4	4	3	1	2	2	3	3		3	2	3	1	0
13	СПА	1	4	7	5	6	5	1	2	4	2	0		0	3	1	3
14	ТЕР	2	2	2	1	2	1	1	3	1	3	0	1		2	2	2
15	ТОМ	0	4	1	4	3	2	2	3	2	5	1	3	0		3	3
16	УРА	0	3	3	5	3	2	2	3	3	5	3	2	3	0		4
17	ЦСК	3	0	3	2	1	6	3	1	1	1	3	1	5	2	1	

Рис. 4.22
Таблица голов

Объем данных n в данном случае равен количеству проведенных матчей, $n = 240$. Изучаемая переменная X — количество голов забитых в матче. В данном примере исследуемая переменная принимает небольшое количество различных значений. Чисто визуально можно определить, что наименьшее количество голов равно нулю, а наибольшее — восьми.

Применять здесь описательную статистику пакета анализа не очень удобно из-за большого объема, так как их нужно расположить в одном столбце. Данные можно сгруппировать и построить дискретный статистический ряд.

Для построения статистического ряда необходимо указать значения, которые принимает переменная X , и посчитать частоты. Так как исходный дискретный ряд принимает конечное количество значений, для нахождения частоты повторяемости признака, воспользуемся функцией СЧЁТЕСЛИ(диапазон;критерий), которая подсчитывает количество непустых ячеек в указанном диапазоне, удовлетворяющих заданному критерию.

Подготовим таблицу для расчета частот (рис. 4.23). В строке 1 поместим количество забитых голов в порядке возрастания. В ячейку T2 введем формулу, которая отобразилась в строке формул. Напомним, что для ввода ссылок на диапазоны в формулах пользуйтесь мышью.

T2		fx = СЧЁТЕСЛИ(\$B\$2:\$Q\$17;T1)									
	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Голы	0	1	2	3	4	5	6	7	8	
2	Матчи	24	42	49	63	38	15	7	1	1	
3											
4	xi	0	1	2	3	4	5	6	7	8	
5	ni	24	42	49	63	38	15	7	1	1	

Рис. 4.23
Построение дискретного ряда

Первый аргумент функции СЧЁТЕСЛИ, диапазон \$B\$2:\$Q\$17 — ссылка на исходную таблицу данных. Вторым аргументом функции ссылается на ячейку с критерием. Для диапазона используется абсолютная адресация, тогда эту формулу можно будет поместить в соседние ячейки копированием, или «протягиванием». Таким образом, получен сгруппированный ряд, в первой строке указано количество голов, а во втором — соответствующее количество матчей. Обозначим x_i — значение переменной, или варианта, а n_i — частоту встречаемости этой переменной.

Добавим p_i — относительные частоты, получим дискретный статистический ряд (табл. 4.2).

Таблица 4.2

Дискретный статистический ряд

x_i	0	1	2	3	4	5	6	7	8
n_i	24	42	49	63	38	15	7	1	1
p_i	0,100	0,175	0,204	0,263	0,158	0,063	0,029	0,004	0,004

Относительные частоты вычислены как отношение частоты к объему выборки:

$$p_i = \frac{n_i}{n} = \frac{n_i}{240}.$$

Заметим, что $\sum n_i = 240$, $\sum p_i = 1$.

Мода найдется как значение, имеющее наибольшую частоту, в данном случае 3, т. е. чаще всего в матчах было забито три гола. Так как $n = 240$ — четное число, медиана

равна полусумме значений, стоящих в вариационном ряду на 120 и 121 местах.

Расчеты приведены на рисунке 4.24, а формулы для расчетов на рисунке 4.25.

В среднем за матч забивалось два-три гола. Так как медиана равна 3, то в среднем в половине матчей было забито не менее трех голов, а в половине не более трех.

Наглядность распределению значений данных придают графические характеристики. Для дискретных рядов

	A	B	C	D	E	F	G	H
1	xi	ni	pi	xi × ni	xi-хср	(xi-хср) ²	ni (xi-хср) ²	
2	0	24	0,100	0	0	0	0	
3	1	42	0,175	42	1	1	42	
4	2	49	0,204	98	2	4	196	
5	3	63	0,263	189	3	9	567	
6	4	38	0,158	152	4	16	608	
7	5	15	0,063	75	5	25	375	
8	6	7	0,029	42	6	36	252	
9	7	1	0,004	7	7	49	49	
10	8	1	0,004	8	8	64	64	
11	36	240	1	613	36	204	2153	Сумма
12								
13						Среднее	2,5541667	
14						Дисперсия	8,9708333	
15						СКВО	2,9951349	
16						Мода	3	
17						Медиана	3	

Рис. 4.24

Результаты расчетов основных числовых характеристик

	A	B	C	D	E	F	G	H
1	xi	ni	pi	xi × ni	xi-хср	(xi-хср) ²	ni (xi-хср) ²	
2	0	24	=B2/\$B\$11	=A2*B2	=A2-\$A\$12	=E2*E2	=B2*F2	
3	1	42	=B3/\$B\$11	=A3*B3	=A3-\$A\$12	=E3*E3	=B3*F3	
4	2	49	=B4/\$B\$11	=A4*B4	=A4-\$A\$12	=E4*E4	=B4*F4	
5	3	63	=B5/\$B\$11	=A5*B5	=A5-\$A\$12	=E5*E5	=B5*F5	
6	4	38	=B6/\$B\$11	=A6*B6	=A6-\$A\$12	=E6*E6	=B6*F6	
7	5	15	=B7/\$B\$11	=A7*B7	=A7-\$A\$12	=E7*E7	=B7*F7	
8	6	7	=B8/\$B\$11	=A8*B8	=A8-\$A\$12	=E8*E8	=B8*F8	
9	7	1	=B9/\$B\$11	=A9*B9	=A9-\$A\$12	=E9*E9	=B9*F9	
10	8	1	=B10/\$B\$11	=A10*B10	=A10-\$A\$12	=E10*E10	=B10*F10	
11	=СУММ	=СУММ	=СУММ(C2:C10)	=СУММ(D2:D10)	=СУММ(E2:E10)	=СУММ(F2:F10)	=СУММ(G2:G10)	Сумма
12								
13						Среднее	=D11/B11	
14						Дисперсия	=G11/B11	
15						СКВО	=КОРЕНЬ(G14)	
16						Мода	3	
17						Медиана	3	

Рис. 4.25

Формулы для расчетов основных числовых характеристик

строят полигон частот или относительных частот. Для сгруппированного дискретного ряда строят также гистограмму частот или относительных частот.

Полигон относительных частот, который называют также многоугольником распределения, представляет собой замкнутую ломаную линию, соединяющую точки с координатами (x_i, p_i) . Для построения полигона обычно используют «График» с маркерами. Выделите диапазон с относительными частотами C2:C10 → «Вставка» → «Диаграмма» → «График с маркерами». Построенную диаграмму нужно подкорректировать. Измените подписи данных горизонтальной оси, для этого выделите подписи оси мышью → «Конструктор» → «Выбрать данные» → «Изменить подписи горизонтальной оси» → введите диапазон A2:A10, в котором находятся значения x_i . Поменяйте параметры оси, выбрав «Положение оси:» по делениям. Добавьте название диаграммы и удалите легенду (рис. 4.26).

Многоугольник распределения является одной из форм закона распределения. Если отбросить две последние точки, которые являются выбросами, можно сказать, что значения симметрично распределяются около среднего значения.



Рис. 4.26
График полигона относительных частот



Еще одна графическая характеристика — гистограмма распределения. Построенная на частотах (рис. 4.27) или относительных частотах сгруппированного дискретного или интервального статистического ряда, она дает наглядное представление о типе распределения.

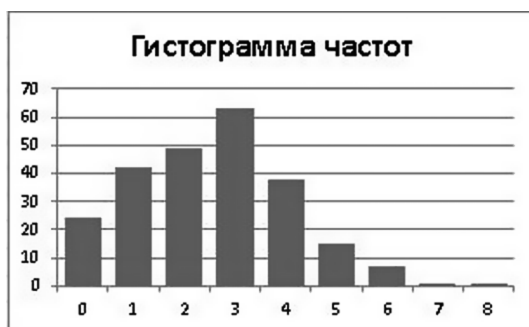


Рис. 4.27
Гистограмма частот

Пример 4.4. На рисунке 4.28 приведены выборочные данные о росте футболистов российских клубов уже предварительно отсортированные по возрастанию исследуемой переменной. На рисунке они размещены в прямоугольном диапазоне только для наглядности.

	A	B	C	D	E
1	167,8	173,0	177,1	180,3	186,0
2	168,2	173,2	177,5	181,2	186,1
3	168,5	173,7	178,1	181,3	186,9
4	168,8	173,9	178,2	181,5	187,5
5	169,1	174,2	178,8	181,7	187,9
6	170,0	174,7	179,0	182,8	189,6
7	171,8	174,9	179,3	183,2	190,0
8	172,4	175,2	179,7	184,4	190,4
9	172,8	176,3	179,9	184,9	190,5
10	172,9	176,5	179,9	185,9	191,1

Рис. 4.28
Выборка

Если исследуется непрерывный ряд данных, в этом случае значения изучаемой переменной X могут отличаться друг от друга на сколь угодно малую величину. Если к тому же объем данных велик, то сначала строится интервальный статистический, или вариационный, ряд. Затем уже его подвергают статистическому анализу. Непрерывные величины, обычно имеют единицу измерения: времени, длины, массы, объема и т. д.

Объем выборки $n = 50$. Изучаемая переменная X — рост футболистов российских клубов. Исследуемая переменная является непрерывной. По этим данным можно построить интервальный статистический ряд.

Обычно исходные данные не упорядочены. Чтобы получить ранжированный ряд в Excel, необходимо сначала, обязательно, поместить все данные в один столбец и отсортировать.

Разобьем исходный ряд на интервалы. Зададим количество интервалов k , используя формулу Стерджесса. При $n = 50$ оптимальное число интервалов равно 6.

Для расчета частот есть две альтернативы: воспользоваться встроенной функцией **ЧАСТОТА** или средством «Гистограмма» в «Пакете анализа». Рассмотрим здесь первый вариант, так как второй вариант был рассмотрен выше.

Функция **ЧАСТОТА**(массив_данных; массив_интервалов) вычисляет распределение значений по интервалам и возвращает вертикальный массив чисел, содержащий на один элемент больше, чем массив интервалов. Для использования этой функции нужно подготовить массив интервалов.

Разместим исходные данные в столбце А, это будет диапазон А2:А51. Если данные не упорядочены, выполним сортировку столбца А по возрастанию. На рисунке 4.29 отображены первые десять значений исходных данных.

Сформируем массив интервалов в столбце В следующим образом. В первую ячейку массива интервалов В2 поместим минимальное выборочное значение, в ячейку В3 введем формулу и протянем ее вниз, сформировав таким образом границы интервалов. Формулы приведены на рисунке 4.30.

	A	B	C	D	E	F	G	H
1	Массив данных	Массив интервалов	Массив частот					
2	167,80	167,80	1	≤167,80		k=	6	
3	168,20	171,68	5	(167,80 ; 171,68]		h=	3,88	
4	168,50	175,57	12	(171,68 ; 175,57]				
5	168,80	179,45	9	(175,57 ; 179,45]				
6	169,10	183,33	10	(179,45 ; 183,33]				
7	170,00	187,22	6	(183,33 ; 187,22]				
8	171,80	191,10	6	(187,22 ; 191,10]				
9	172,40		1	> 191,10				
10	172,80							

Рис. 4.29
Расчет частот

	A	B	C	D	F	G	H
1	Массив данных	Массив интервалов	Массив частот				
2	167,8	=МИН(A2:A51)	=ЧАСТОТА(A2:A51;B2:B8)		k=	=ЦЕЛОЕ(1+3,321*LOG10(50))	
3	168,2	=B2+\$H\$3	=ЧАСТОТА(A2:A51;B2:B8)		h=	=(МАКС(A2:A51)-МИН(A2:A51))/H2	
4	168,5	=B3+\$H\$3	=ЧАСТОТА(A2:A51;B2:B8)				
5	168,8	=B4+\$H\$3	=ЧАСТОТА(A2:A51;B2:B8)				
6	169,1	=B5+\$H\$3	=ЧАСТОТА(A2:A51;B2:B8)				
7	170	=B6+\$H\$3	=ЧАСТОТА(A2:A51;B2:B8)				
8	171,8	=B7+\$H\$3	=ЧАСТОТА(A2:A51;B2:B8)				
9	172,4		=ЧАСТОТА(A2:A51;B2:B8)				
10	172,8						

Рис. 4.30
Расчет частот, формулы

Теперь заполним массив частот. Функция ЧАСТОТА здесь должна быть введена как формула массива, поскольку данная функция возвращает массив значений. Для этого выполните следующие шаги.

1. Выделить мышью диапазон C2:C9, который содержит на одну ячейку больше, чем массив интервалов.

2. Ввести формулу =ЧАСТОТА(A2:A15;B2:B8), она автоматически будет вводиться в C2, первую ячейку выделенного диапазона.

3. Нажать одновременно клавиши CTRL + SHIFT + ENTER.

В ячейки C2:C9 будет введена одна и та же формула массива (рис. 4.30). Признаком формулы массива являются фигурные скобки, в режиме «Показать формулы»

фигурные скобки не отображаются. Результаты расчетов частот приведены на рисунке 4.29. Там же, в столбце Е, указаны для наглядности границы интервалов, соответствующие частотам.

Обратите внимание, что в первом и последнем интервале оказалось по единице. Добавим эти единицы к прилегающим интервалам. Единица в последнем интервале появилась из-за приближенных вычислений, так как в ячейке В8 на самом деле расположено значение чуть меньшее, чем 191,10. С этим можно, конечно, побороться. Например, поместить в ячейку В8 максимальное значение выборки или при расчете длины интервала вычислить его, используя округление вверх.

Функция **ЧАСТОТА** относится к «Статистическим» функциям, ее лучше вводить, используя мастера функций. На рисунке 4.31 приведено диалоговое окно функции.

Данные для интервального ряда вычислены. Окончательный вид интервального ряда приведен на рисунке 4.32.

Для дальнейшего статистического анализа в качестве значений будут использоваться середины полученных

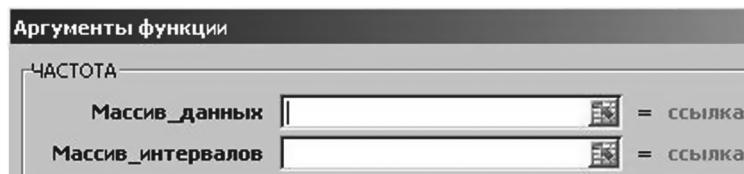


Рис. 4.31
Диалоговое окно функции **ЧАСТОТА**

	А	В	С	Д	Е	Г	Н
1	№						
2	интервала					Середина	Частота
3	1		[167,80 ;	171,68]		169,74	6
4	2		(171,68 ;	175,57]		173,63	12
5	3		(175,57 ;	179,45]		177,51	9
6	4		(179,45 ;	183,33]		181,39	10
7	5		(183,33 ;	187,22]		185,28	6
8	6		(187,22 ;	191,10]		189,16	7

Рис. 4.32
Интервальный ряд

интервалов. Напомним, что середина интервала находится как полусумма значений его концов.

Объем выборки небольшой, равен 50, поэтому данный ряд можно было не разбивать на интервалы, а применить описательную статистику к исходному ряду данных, используя все имеющиеся данные. Выделим диапазон с данными A2:A51 и проведем описательную статистику из «Пакетом анализа». Результаты приведены в таблице 4.3.

Как видим, все посчитано, даже мода, которая в данном случае не является содержательной характеристикой.

Таблица 4.3

Описательная статистика

Столбец1	Значения
Среднее	178,972
Стандартная ошибка	0,93684
Медиана	178,9
Мода	179,9
Стандартное отклонение	6,62445
Дисперсия выборки	43,8833
Эксцесс	-0,9176
Асимметричность	0,1543
Интервал	23,3
Минимум	167,8
Максимум	191,1
Сумма	8948,6
Счет	50



Интервальный ряд

Таблица 4.4

x_i	169,74	173,63	177,51	181,39	185,28	189,16
f_i	6	12	9	10	6	7
f_i^*	6	18	27	37	43	50

Значение среднего, моды и медианы близки, что является признаком симметричного распределения. В данном случае интервальный ряд (табл. 4.4) может дать более содержательные характеристики. Здесь f_i — частоты, f_i^* — накопленные частоты.

Мода и медиана для интервального ряда считаются по специальным формулам. Для расчета модального значения определим модальный интервал, это интервал с максимальной частотой. Наибольшая частота равна 12. Следовательно, второй интервал (171,68; 175,57] — модальный. Для интервального ряда вычисляется приближенное значение моды по формуле:

$$Mo = x_{Mo} + h \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}.$$

Здесь x_{Mo} — нижняя граница модального интервала; h — величина модального интервала; f_{Mo} — частота модального интервала; f_{Mo-1} — частота интервала, предшествующего модальному; f_{Mo+1} — частота интервала, следующего за модальным:

$$Mo = 171,68 + 3,88 \cdot \frac{12 - 6}{(12 - 6) + (12 - 9)} \approx 174,27.$$

Рост примерно 174 см является наиболее распространенным среди футболистов.

Для расчета медианы определим медианный интервал, это интервал в котором содержатся 25–26-й элементы ряда, т. е. срединные элементы, так как объем выборки равен 50. Следовательно, медианным является третий интервал (175,57; 179,45], так как накопленная частота этого интервала равна 27. Для интервального ряда вычисляется приближенное значение медианы по формуле:

$$Me = x_{Me} + h \cdot \frac{0,5 \sum f - S_{Me-1}}{f_{Me}}.$$

Здесь x_{Me} — нижняя граница медианного интервала; h — величина медианного интервала; $\sum f$ — сумма всех частот; S_{Me-1} — накопленная частота интервала, предше-

ствующего медианному; f_{Me} — частота медианного интервала:

$$Me = 175,57 + 3,88 \cdot \frac{25-18}{9} \approx 178,58.$$

Половина спортсменов имеют рост не более 178,58, а другая половина — не менее 178,58.

Графические характеристики интервального ряда — гистограмма относительных частот, как эмпирический аналог функции распределения и кумулята, как эмпирический аналог функции распределения, если исследуемая случайная величина является непрерывной.

В таблице 4.5 приведены результаты расчетов относительных частот p_i и накопленных относительных частот p_i^* для построения гистограммы (рис. 4.33) и кумуляты (рис. 4.34).

Таблица 4.5

Интервальный ряд

x_i	169,74	173,63	177,51	181,39	185,28	189,16
f_i	6	12	9	10	6	7
p_i	0,12	0,24	0,18	0,20	0,12	0,14
p_i^*	0,12	0,36	0,54	0,74	0,86	1



Рис. 4.33

Гистограмма относительных частот



Рис. 4.34
Кумулята относительных частот

Оба графика показывают, что распределение близко равномерному. То есть при подборе спортсменов рост играет немаловажную роль.

ТЕМА 4.3. ВЫРАВНИВАНИЕ ВРЕМЕННЫХ РЯДОВ

Для выравнивания временных рядов «Анализ данных» имеет два инструмента: «Скользящее среднее» и «Экспоненциальное сглаживание». Во временных рядах фактором, от которого зависит переменная прогнозирования, является время.

4.3.1. СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

Метод «Скользящего среднего» является одним из наиболее используемых способов выравнивания значений временного ряда. Обычно используется для сглаживания краткосрочных колебаний и выделения основных тенденций, или тренда. Метод заключается в том, что значения исходного ряда заменяются прогнозными значениями, которые считаются по формуле



$$\tilde{x}_t = \frac{1}{k} \sum_{i=0}^{k-1} x_{t-i},$$

где k — количество наблюдений, по которым будет вычисляться усреднение. Прогнозных значений будет меньше на $(k-1)$ единицу. Заметим, что эта формула используется инструментом «Скользящее среднее». На практике для усреднения применяют и другие формулы.

Пример 4.5. Пусть имеются данные за три года о продажах соков некоторой торговой компанией (табл. 4.6). Требуется выявить тенденцию временного ряда, используя скользящее среднее.

Визуальный анализ показывает, что к третьему кварталу продажи увеличиваются. Данный временной ряд содержит сезонные колебания периодичностью 4. Применим четырехточечную скользящую среднюю.

Выполним последовательность действий: «Данные» → «Анализ данных» → «Скользящее среднее». В открывшемся диалогом окне укажем «Входной интервал А3:М3», установим галочку «Метки в первой строке», так как мы захватили заголовок ряда данных.

Для параметра «Интервал» укажем 4, так как усреднение будем проводить по четырем точкам. По умолчанию для вычисления скользящего среднего используется три точки.

Укажем «Выходной интервал», здесь достаточно указать одну, начальную ячейку для вывода данных. Исходные данные могут находиться в строке или столбце. Так как исходный интервал расположен в строке, выходные данные тоже будут расположены в строке.

Таблица 4.6

Данные о продажах

Год	2010				2011				2012			
Квартал	1	2	3	4	1	2	3	4	1	2	3	4
Объем продаж (тыс. л)	96	144	192	64	104	160	200	76	108	168	212	88

Выберем дополнительно «Вывод графика» и «Стандартные погрешности». Установка параметров показана на рисунке 4.35.

В результате на рабочем листе, начиная с указанной ячейки В5, будут выведены прогнозные значения, рассчитанные по методу скользящего среднего, и значения стандартных погрешностей (рис. 4.36), а также графики фактических и прогнозных значений (рис. 4.37).

Как мы видим, прогнозных значений будет на три меньше. Эти значения легко посчитать вручную. Например, для первой прогнозной точки получим $(96 + 144 + 192 + 64) : 4 = 124$. Погрешность вычисляется корень из суммы квадратов разностей между исходными и расчетными k значениями, деленной на число k . Т. е. в ячейку Н6, например, автоматически помещается формула $=КОРЕНЬ(СУММКВРАЗН(Е3:Н3;Е5:Н5)/4)$.

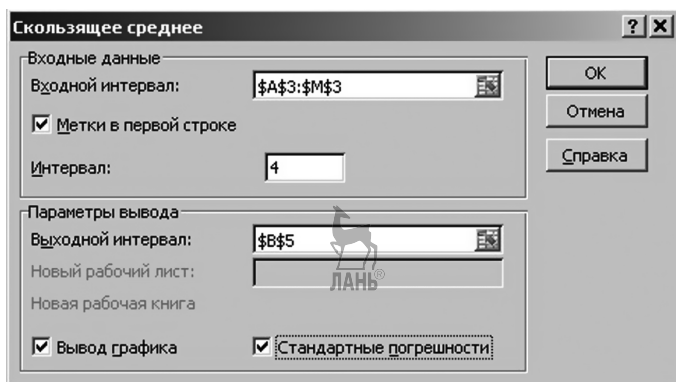


Рис. 4.35
Диалоговое окно инструмента «Скользящее среднее»

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Год	2010				2011				2012			
2	Квартал	1	2	3	4	1	2	3	4	1	2	3	4
3	Объем продаж (тыс. литров)	96	144	192	64	104	160	200	76	108	168	212	88
4													
5	Прогноз	#Н/Д	#Н/Д	#Н/Д	124	126	130	132	135	136	138	141	144
6	Погрешности	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	49,010	48,706	49,470	49,470	50,512	49,651

Рис. 4.36
Результаты расчетов с применением «Скользящего среднего»



Рис. 4.37
График «Скользящее среднее»

4.3.2. ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ

«Экспоненциальное сглаживание» также применяется к временным рядам. В этом методе сглаживания учитывается «старение» данных — в процессе сглаживания больший вес имеют последние данные. Сглаженные значения вычисляются по формуле

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t,$$

где α — коэффициент сглаживания, $0 < \alpha < 1$; \hat{y}_{t+1} — прогнозное значение за период $(t + 1)$; \hat{y}_t — прогнозное значение за период t ; y_t — фактическое значение за период t , причем $\hat{y}_1 = y_1$.

Выберем теперь «Экспоненциальное сглаживание» и установим параметры в диалоговом окне (рис. 4.38). Укажем «Входной интервал» и «Выходной интервал», установим «Фактор затухания» равным 0,5, отметим «Выбор графика» и «Стандартные погрешности».

Выполним еще раз «Экспоненциальное сглаживание», теперь с «Фактором затухания» 0.8. Результаты сглаживания и исходные данные приведены на рисунке 4.39. Соответствующие графики изображены на рисунке 4.40 и рисунке 4.41.

Экспоненциальное сглаживание

Входные данные

Входной интервал:

Фактор затухания:

☐ Метки

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

☒ Вывод графика ☒ Стандартные погрешности

OK Отмена Справка

Рис. 4.38
Диалоговое окно «Экспоненциальное сглаживание»

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Год	2010				2011				2012			
2	Квартал	1	2	3	4	1	2	3	4	1	2	3	4
3	Объем продаж (тыс. литров)	96	144	192	64	104	160	200	76	108	168	212	88
4	$\alpha=0,5$												
5		#Н/Д	96	120	156	110,00	107,00	133,50	166,75	121,38	114,69	141,34	176,67
6		#Н/Д	#Н/Д	#Н/Д	#Н/Д	72,92	67,54	61,40	49,22	71,80	65,41	61,26	51,68
7	$\alpha=0,8$												
8		#Н/Д	96	106	123	111,1	109,7	119,7	135,8	123,8	120,7	130,1	146,5
9		#Н/Д	#Н/Д	#Н/Д	#Н/Д	66,42	60,5	44,9	54,84	64,67	58,5	44,97	55,36

Рис. 4.39
Результаты расчетов с применением
«Экспоненциального сглаживания»



Рис. 4.40
График «Экспоненциальное сглаживание» при $\alpha = 0,5$

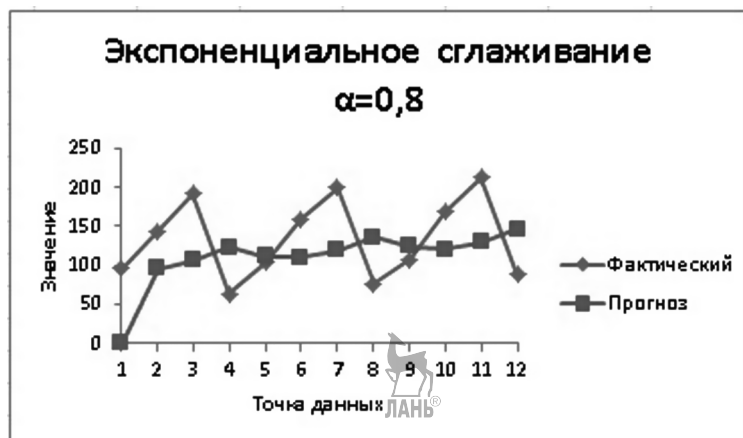


Рис. 4.41

График «Экспоненциальное сглаживание» при $\alpha = 0,8$ 

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА



1. *Васильев, А. Н.* Числовые расчеты в Excel. — СПб. : Лань, 2014. — 608 с.
2. *Вуколов, Э. А.* Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL : учеб. пособие. — М. : ФОРУМ : ИНФА-М, 2004. — 464 с.
3. *Джелен, Б.* Сводные таблицы в Microsoft Excel 2013 / Б. Джелен, М. Александер. — М. : Вильямс, 2014. — 448 с.
4. *Желязны, Д.* Говори на языке диаграмм. Пособие по визуальным коммуникациям для руководителей. Институт комплексных стратегических исследований. — М., 2004.
5. *Закс, Л.* Статистическое оценивание. — М. : Статистика, 1976. — 598 с.
6. *Уокенбах, Д.* Excel 2010. Библия пользователя. — М. : ООО «ИД „Вильямс“», 2011. — 1094 с.
7. *Корнелл, П.* Анализ данных в Excel. Просто как дважды два. — М. : Эксмо, 2007. — 224 с.
8. *Левин, Д. М.* Статистика для менеджеров с использованием Microsoft Excel / Д. М. Левин, Д. Стефан, Т. С. Кребиль. — М. : ИД «Вильямс», 2004.
9. *Наследов, А. Д.* Математические методы психологического исследования. Анализ и интерпретация данных : учеб. пособие. — СПб. : Речь, 2007. — 392 с.
10. *Сидоренко, Е. В.* Методы математической обработки в психологии. — СПб. : ООО «Речь», 2007.
11. *Хили, Дж.* Социологические и маркетинговые исследования / под общей ред. к. ф.-м. н. А. А. Руденко. — Киев : ООО «ДиаСофтЮП»; СПб. : Питер, 2005. — 638 с.

ОГЛАВЛЕНИЕ

Введение	5
ГЛАВА 1	
Таблицы в Excel	7
<i>Тема 1.1</i>	
Основные понятия	7
1.1.1. Диапазон	7
1.1.2. Список	8
1.1.3. Таблица	9
1.1.4. Особенности работы с таблицами	10
<i>Тема 1.2</i>	
Обработка данных	14
1.2.1. Сортировка	14
1.2.2. Фильтр	17
1.2.3. Промежуточные Итоги	24
<i>Тема 1.3</i>	
Контроль данных	27
1.3.1. Контроль ввода данных	27
1.3.2. Условное форматирование	30
ГЛАВА 2	
Обработка многомерных данных	38
<i>Тема 2.1</i>	
Базы данных в Excel	38
2.1.1. Структура базы данных	38
2.1.2. Формы в Excel	41
2.1.3. Поддержка базы данных	42
<i>Тема 2.2</i>	
Сводные таблицы	48
2.2.1. Назначение сводных таблиц	48
2.2.2. Создание сводных таблиц	49
2.2.3. Работа со сводными таблицами	52
2.2.4. Применение «Мастера сводных таблиц»	60

ГЛАВА 3**Диаграммы 64***Тема 3.1***Создание диаграмм 64**

3.1.1. Основные понятия 64

3.1.2. Построение диаграммы 66

3.1.3. Форматирование диаграмм 69

*Тема 3.2***Построение профессиональных диаграмм 71**

3.2.1. Основные идеи 71

3.2.2. Основные типы сравнения 72

3.2.3. Покомпонентное сравнение 74

3.2.4. Позиционное сравнение 79

3.2.5. Временное сравнение 83

3.2.6. Корреляционное сравнение 84

*Тема 3.3***Нестандартные диаграммы 85**

3.3.1. Комбинированные диаграммы 85

3.3.2. Дополнительные элементы форматирования 89

3.3.3. Диаграммы в ячейках 90

3.3.4. Примеры диаграмм 93

ГЛАВА 4**Статистический пакет анализа данных 110***Тема 4.1***Инструменты надстройки. Анализ данных 111***Тема 4.2***Описательная статистика 113**

4.2.1. Дисперсия 120

4.2.2. Мода и медиана 122

4.2.3. Асимметрия и эксцесс 124

4.2.4. Доверительные интервалы 130

4.2.5. Примеры применения описательной статистики ... 131

*Тема 4.3***Выравнивание временных рядов 147**

4.3.1. Скользящее среднее 147

4.3.2. Экспоненциальное сглаживание 150

Рекомендуемая литература 153