

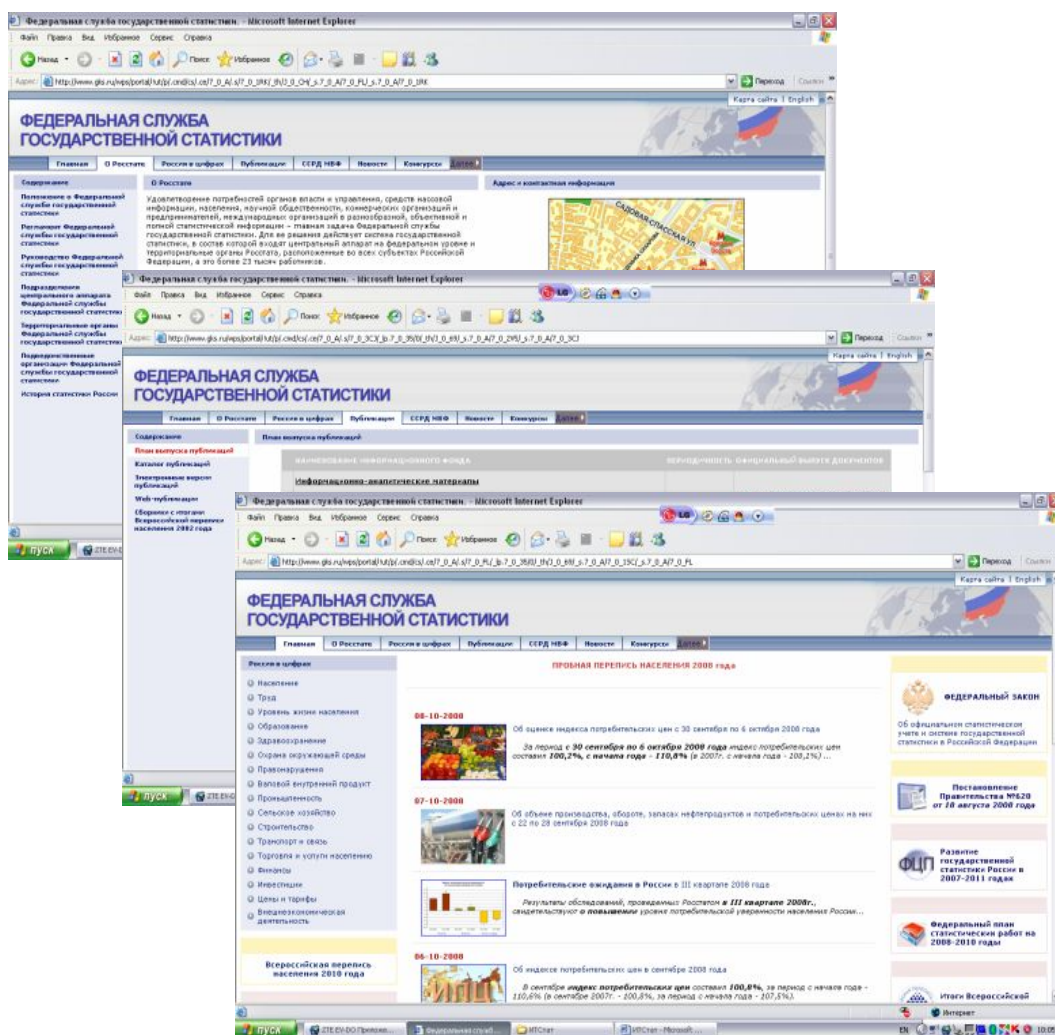


Государственное образовательное учреждение
высшего профессионального образования
«Магнитогорский государственный технический университет
им. Г.И. Носова»

Кафедра вычислительной техники и прикладной математики

О.С. Логунова
Е.А. Ильина
В.В. Королева

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В СТАТИСТИКЕ:
ПРАКТИКУМ ПО ДИСЦИПЛИНЕ
ДЛЯ СТУДЕНТОВ СПЕЦИАЛЬНОСТИ 080601 «СТАТИСТИКА»**



Магнитогорск
2010

УДК 311.2:004.9 (075)

Рецензенты:

Профессор, кандидат педагогических наук
ГОУ ВПО «Магнитогорский государственный университет»

Г.А. Лисьев

Заведующий кафедрой промышленной кибернетики и систем управления,
доцент, кандидат технических наук

ГОУ ВПО «Магнитогорский государственный технический университет
им. Г. И. Носова»

С.М. Андреев

Логунова О.С., Ильина Е.А., Королева В.В.

Информационные технологии в статистике: практикум по дисциплине для студентов специальности 080601 «Статистика» [Электронный ресурс] / О.С. Логунова, Е.А. Ильина, В.В. Королева; ГОУ ВПО «Магнитогорский государственный технический университет им. Г.И. Носова». – Электрон. текстовые дан. (1,16 Мб). – Магнитогорск : ГОУ ВПО «МГТУ», 2010. – 1 электрон. опт. диск (CD-R). – Систем. требования : IBM PC, любой, более 1 GHz ; 512 Мб RAM ; 10 Мб HDD ; MS Windows XP и выше ; Adobe Reader 8.0 и выше ; CD/DVD-ROM дисковод ; мышь. – Загл. с титул. экрана.

Практикум по дисциплине «Информационные технологии в статистике» предназначен для студентов направления подготовки 080600 – «Статистика», специальности 080601 – «Статистика». Может быть, полезен аспирантам, соискателям экономических и технологических специальностей, а также всем, кто интересуется статистической обработкой данных различной природы средствами вычислительной техники.

Практикум включает пять лабораторных работ, каждая из которых содержит задания и описание примеров их выполнения, позволяющих сформировать умения и навыки по использованию информационных технологий при решении задач статистической обработки данных. Большое внимание уделено процессу постановки задачи по исследованию социально-экономического явления, алгоритмизации процесса исследования, проектированию форм документации, определяющих набор исходных данных и отчетов, исследованию наличия общегосударственных кодов и изучению их структуры. В практикуме приведено описание структуры официального сайта Государственного комитета статистики РФ и предложены задания для поиска информации о социально-экономических процессах, нормативно-справочных данных, истории развития статистики в России и т.п.

Работа выполнена при поддержке Федерального агентства по науке и инновациям в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы, контракт № 02.740.11.0422.

УДК 311.2:004.9 (075)

© Логунова О.С., Ильина Е.А.,
Королева В.В., 2010

© ФГБОУ ВПО «Магнитогорский
государственный технический
университет им. Г.И. Носова», 2010

Содержание

Введение	4
Лабораторная работа №1. Технология обработки статистических данных и содержание типовых операций	7
Лабораторная работа №2. Технология проектирования носителей, кодов и классификаторов статистической информации	26
Лабораторная работа №3. Технология реализации этапов обработки статистической информации программными средствами.....	42
Лабораторная работа №4. Технология оценки экономической эффективности автоматизированной обработки статистической информации.....	57
Лабораторная работа №5. Технология применения официального сайта Государственного комитета статистики для добычи статистических данных.....	64

Введение

Появление новых информационных технологий потребовала введения в учебный процесс новых дисциплин, которые позволяют применить эти технологии к различным предметным областям. Не является исключением и новая дисциплина «Информационные технологии в статистике».

Предмет курса «Информационные технологии в статистике» – это совокупность научных знаний об основах проектирования и функционирования системы сбора и обработки статистической информации в органах государственной статистики с применением экономико-математических методов, современных средств вычислительной и организационной техники и средств связи.

Статистическая информация – это разновидность экономической информации, содержащей различные сведения социально-экономического характера, которые можно фиксировать, передавать, преобразовывать, хранить и использовать для решения социальных и экономических задач.

Для социально-экономической информации характерны массовость (большие объемы), повторные циклы получения и обработки данных в установленные сроки, большой удельный вес данных, используемых для дальнейшей обработки или длительного хранения.

Статистическая информация имеет следующие специфические особенности:

- единство методологии исчисления и обработки статистических показателей;
- многократность применения (показатели отчетных периодов могут быть использованы как показатели соответствующих прошедших периодов и как элементы динамического ряда);
- длительный срок хранения (текущая отчетность – 2 года, годовая – до 15 лет и т.д.).

Для информации такого вида особенно важно выполнить рациональную разработку носителей, выделить коды и составить классификаторы.

Содержание курса составляют сведения о развитии системы государственной статистики, статистической информации как объекте автоматизированной обра-

ботки, об основных научных принципах создания новых информационных технологий и организации обработки статистических данных на различных уровнях системы Госкомстата РФ.

Авторы учебного пособия не преследовали цель дать полное изложение статистических методов или полное описание возможностей программного продукта Statistica 6.0. Целью работы является показать возможность использования этого программного обеспечения для реализации стандартных статистических задач, с которыми чаще всего сталкивается студент высшего учебного заведения и инженер, занимающийся обработкой наблюдений в ходе своей профессиональной деятельности.

Выбор программного обеспечения Statistica 6.0 в качестве основного программного продукта для работы студентов обусловлен наличием и установкой лицензионной версии в центре информационных технологий ГОУ ВПО «Магнитогорский государственный технический университет им. Г.И. Носова» с января 2006 года.

Выбор пакета для анализа данных зависит от характера решаемых задач, объема обрабатываемого материала, квалификации пользователей, имеющегося оборудования и т. д.

Для пользователей, имеющих дело со сверхбольшими объемами данных или узкоспециализированными методами анализа, пока нет альтернативы профессиональным западным пакетам. Среди них самыми широкими возможностями обладает SAS. Для создания собственной системы обработки данных можно обратиться к библиотеке IMSL, содержащей сотни тщательно и квалифицированно составленных статистических подпрограмм. Несколько меньшими возможностями обладают универсальные пакеты. Вместе с тем, их стоимость значительно ниже, чем профессиональных. При приобретении такого пакета не мешает, однако, лишний раз убедиться, что он содержит требуемые методы обработки.

Авторами предлагаются варианты заданий к практическим занятиям по указанному курсу. Всего предлагается провести пять лабораторных работ, тематика и план проведения которых приведены в таблице.

*План проведения практических занятий по курсу
«Информационные технологии в статистике»*

<i>Номер</i>	<i>Название лабораторной работы</i>	<i>Оценка времени, акад. час</i>	<i>Программное обеспечение</i>
1	Технология обработки статистических данных и содержание типовых операций	6	MS Word BPWin
2	Технология проектирования носителей, кодов и классификаторов статистической информации.	4	MS Word
3	Технология реализации этапов обработки статистической информации программными средствами	4	MS Excel MS Word Statistica
4	Технология оценки экономической эффективности автоматизированной обработки статистической информации	2	MS Word MS Excel
5	Технология применения официального сайта Государственного комитета статистики для добычи статистических данных	2	MS Explorer MS Excel MS Office Statistica
Итого		18	

Желаем успехов в работе!

Лабораторная работа №1.
Технология обработки статистических данных
и содержание типовых операций

Выполнив эту лабораторную работу, Вы:

- узнаете на каких принципах основывается проектирование технологического процесса автоматизированной обработки статистической информации;
- узнаете из каких этапов и операций состоит технологический процесс автоматизированной обработки статистической информации;
- научитесь применять методы контроля достоверность вводимых данных;
- узнаете какими документами регламентируется технологический процесс;
- научитесь строить информационно-технологические и инструкционные карты процесса по статистическим задачам.

Информация¹

Технологический процесс автоматизированной обработки статистической информации представляет собой совокупность операций, выполняемых в определенной последовательности – от сбора исходной информации до получения результатной информации.

Под *операцией* в данном случае понимается комплекс действий, выполняемых над информацией на одном рабочем месте.

Разработка технологического процесса должна обеспечить максимальную автоматизацию процессов обработки информации при использовании различных технологических средств и высокую достоверность получения результатной информации при минимальных трудовых и стоимостных затратах,

Состав операций и последовательность их выполнения зависят от характера решаемых статистических задач и имеющегося комплекса технических средств на каждом уровне обработки (районном, региональном и федеральном).

¹ Источник теоретической информации: Информационные технологии в статистике: Учебник / Под ред. проф. В.П. Божко и проф. А.В. Хорошилова. – М.: Финстатинформ, 2002. – 144 с.

Характер задач в основном определяется объемами обрабатываемой информации, периодичностью решения, а также сложностью алгоритмов ее преобразования. Решение задачи на каждом уровне обработки может выполняться разными способами с применением различных вычислительных средств и методов контроля.

При определении состава операций технологического процесса создается возможность выбора наиболее рационального способа обработки информации.

В технологическом процессе автоматизированной обработки статистической информации при решении различных задач прослеживаются три основных этапа: первичный, основной и заключительный.

Операции первичного этапа связаны со сбором и подготовкой статистических отчетов к машинной обработке. От качества выполнения этих операций во многом зависит эффективность и достоверность непосредственно автоматизированной обработки.

Операции основного этапа обеспечивают ввод и обработку первичных отчетов, а также печать сводных отчетов в соответствии с заданным алгоритмом решения статистической задачи. Автоматизированная обработка статистической информации может осуществляться с применением отдельных типов ПЭВМ или ПЭВМ, входящих в состав локальной информационно-вычислительной сети (ЛВС).

На заключительном этапе выполняются операции контроля и выпуска сводных отчетов с целью проверки качества их обработки и печати, дальнейшего размножения и рассылки различным потребителям.

Так как при выполнении отдельных операций технологического процесса обработки информации могут быть использованы различные средства вычислительной техники и в различных сочетаниях, в практике существует множество вариантов технологических процессов. Эффективность проектирования машинной обработки статистической информации требует их стандартизации.

Под *стандартизацией технологии автоматизированной обработки информации* понимается разработка комплекса детализированных и максимально уни-

фицированных схем технологических процессов, в которых строго установлены состав и последовательность выполнения операций.

Ориентация на стандартные технологические процессы приводит к улучшению качества проектирования, облегчению и ускорению внедрения проектов машинной обработки данных. Поэтому подробнее рассмотрим типовую схему технологии обработки отчетности на районном и региональном уровнях, представленную на рис. 1.1.

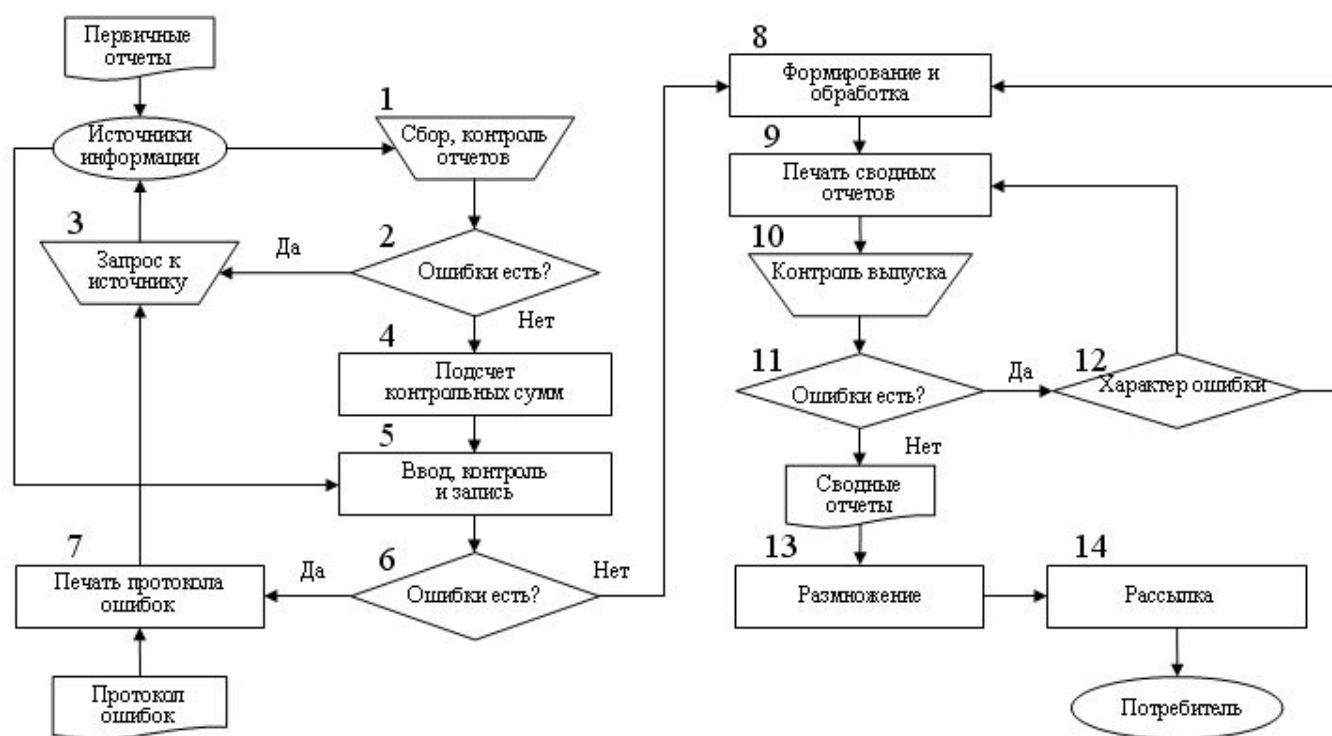


Рис. 1.1. Общая схема технологического процесса обработки статистической информации на районном и региональном уровнях

Первичный этап при обработке первичной отчетности начинается с ее сбора, контроля полноты и качества заполнения первичных отчетов (опер. 1) экономистами соответствующих отделов информационного обеспечения от источников информации, (предприятий и организаций). Если обнаруживаются ошибки, то экономисты делают запросы к источникам информации (опер. 3).

В противном случае по первичным отчетам производится подсчет и регистрация контрольных сумм (опер. 4). Контрольные суммы необходимы в дальнейшем для проверки правильности ввода данных в ПЭВМ, и в зависимости от

структуры первичного отчета они могут подсчитываться по каждой строке, графе или по отчету в целом.

Основной этап (обычно привлекают операторов) начинается с ввода первичных отчетов, их контроля (логический, арифметический и синтаксический) и независимо от его результатов записи информации на магнитные носители (опер. 5). При этом осуществляется проверка полноты записи и внесения корректировок в первичные отчеты на магнитных носителях.

Арифметический контроль заключается в проверке баланса данных в строках и графах первичного отчета с применением заранее подсчитанных контрольных сумм. При логическом контроле проверяются допустимые отклонения от наиболее вероятных данных. С помощью синтаксического контроля проверяется структура заполнения документа различными данными. Если при контроле обнаружилось нарушение, производится печать протокола ошибок (опер. 7), в котором указывается характер нарушений, конкретная ошибка и ее местонахождение в массиве информации.

На основании этого протокола делается запрос к источникам информации (опер. 3), и по его результатам осуществляется корректировка информации на магнитных носителях (опер. 5).

Только после внесения всех изменений и положительного контроля полноты записи первичных отчетов на магнитные носители приступают к их обработке (опер. 8) согласно заданному алгоритму решения задачи.

Далее осуществляется непосредственная печать сводных отчетов (опер. 9) либо в виде рабочих таблиц без шапок и заголовков, либо в полностью оформленном виде.

На заключительном этапе технологического процесса обработки информации экономистами осуществляется контроль сводных отчетов (опер. 10) на полноту вхождения исходных данных в заданный разрез обработки и качества их печати. Учитывая характер ошибки (опер. 12), сводные отчеты обрабатываются или печатаются заново.

Затем осуществляется размножение сводных отчетов с помощью различных копировально-множительных средств (опер. 13) в зависимости от заказанного различными потребителями тиража и рассылка их местным органам управления и вышестоящим уровням системы Госкомстата РФ (опер. 14).

В рассмотренной схеме в случае обработки срочной отчетности первичная информация от источника может сразу поступать на ввод и контроль (опер. 5) в зависимости оттого, каким образом она будет передаваться (по факсу или по электронной почте). При этом рассылка сводной информации (опер. 14) на вышестоящий уровень системы Госкомстата РФ будет осуществляться по электронной почте статистики (ЭПС).

Основными документами, регламентирующими технологический процесс обработки статистической информации на вычислительных установках системы Госкомстата, являются информационно-технологические и инструкционные карты.

Информационно-технологическая карта представляет собой описание последовательно выполняемых операций технологического процесса по каждой статистической задаче.

На каждую операцию технологического процесса разрабатывается *инструкционная карта*, в которой указываются сведения об исходной информации, о конечных результатах и порядке выполнения конкретной работы.

Задание 1.1

Определить предметную область для сбора статической информации при этом:

- 1) определить наименование предметной области;
- 2) определить вид документа – источника информации;
- 3) выбрать реквизиты документа для сбора статических данных;
- 4) подготовить исходные данные для последующей обработки;
- 5) определить способы контроля исходных данных;
- 6) определить перечень решаемых статических задач на основе выбранных исходных данных, в которые должны войти не менее пяти задач по разделам

группировки данных, расчета обобщенных показателей и графической визуализации результатов.

Задание 1.2

Выполнить построение общей схемы технологии обработки статистической информации с подробным уточнением всех блоков, а именно, блоки с номерами 1, 2, 3, 4, 6, 8 – 11. Для блока 12 выполнить разветвление структуры по возможным характерам ошибок (см. рис. 1).

Задание 1.3

Выполнить построение информационно-технологической карты технологического процесса для каждой статистической задачи (см. пример).

Задание 1.4

Выполнить построение инструкционной карты процесса для каждой статистической задачи (см. пример) с помощью спецификации и блок-схемы.

Задание 1.5

Оформить результаты выполнения заданий 1.1 – 1.4 в виде отчета.

Пример выполнения задания 1.1

1. Пусть в качестве предметной области определена пассажирские перевозки для Российских авиакомпаний.

2. В качестве документа – источника информации о перевозках рассматривается форма пассажирского билета для авиаперевозки, который содержит 49 полей для заполнения. Форма билета приведена на рис. 1.2.

The image shows a detailed form for an air passenger ticket. It includes sections for passenger information (name, address, age), flight details (origin, destination, date, time, class), and payment information (amount, currency, date of payment). The form is divided into several numbered sections (1-18) and contains a barcode at the bottom. The text 'Please staple into the Universal Ticket Cover' is visible on the left side.

Рис. 1.2. Форма билета для пассажирской авиаперевозки

3. Для статистического исследования выберем реквизиты: код и серийный номер билет, код авиакомпания, фамилия и имя пассажира, пол пассажира, стоимость билета, направление прямого вылета, направление обратного вылета, номер прямого рейса, номер обратного рейса, дата и время прямого рейса, дата и время обратного рейса.

На рис. 1.2 указаны поля, приведенные в табл. 1.1.

Таблица 1.1

Наименование полей авиабилета

<i>Номер поля</i>	<i>Содержание</i>	<i>Номер поля</i>	<i>Содержание</i>
1	Фамилия пассажира	12	Аэропортовые таксы
2	Маршрут следования	13	Общая стоимость авиабилета
3	Код авиакомпании-перевозчика / Номер рейса	14	Форма оплаты авиабилета
4	Дата вылета	15	Информация о первоначальной выписке авиабилета
5	Время вылета	16	Номер авиабилета
6	Статус бронирования	17	Код данного бронирования в компьютерной системе
7	Ограничения срока действия билета (если имеются)	18	Валидатор
8	Норма бесплатного провоза багажа	19	Класс бронирования
9	Указание на ограничения по данному тарифу	20	Вид тарифа
10	Величина тарифа	21	Расчет тарифа по участкам путешествия
11	Эквивалент тарифа	22	Дополнительный билет

4. При оформлении исходных данных рекомендуется использовать прикладное программное обеспечение. В первую очередь речь должна идти о применении статистических пакетов, например, Statistica² любой доступной версии. В случае отсутствия таких средств возможно использование электронных таблиц или математических пакетов класса MathCad любой доступной версии. Для выбранной задачи должна быть сформирована таблица, содержащая 14 столбцов и не менее 50 строк. При оформлении отчета приводится полный набор данных.

² В ГОУ ВПО «МГТУ им. Г.И. Носова» установлен пакет Statistica версии 6.1, рекомендованный для использования.

5. В качестве способов контроля должны быть выбраны:

- а) отсутствие пустых ячеек в каждой строке и каждом столбце таблицы;
- б) суммарная стоимость всех билетов;
- в) количество символов в записи кода и серии билет – 10 цифр;
- г) отсутствие цифр в записи фамилии и имени пассажира;
- д) диапазон значений для времени и даты определить способы контроля исходных данных;
- е) использование в записи времени, даты, стоимости билета только цифр и специальных разделителей по формату данных;
- ж) использование двух значений для указания пола пассажира.

6. Для выбранных исходных данных должны быть решены следующие статистические задачи:

а) группировки данных по признакам:

- количества пассажиров, перевезенных в каждую дату выбранного периода;
- количества человек, перевезенных в данную дату выбранного периода;
- стоимость перевозки по рейсам и датам;
- количество пассажиров каждой авиакомпании за данный период времени;
- стоимость перевозок каждой авиакомпании по датам.

б) определения обобщающих показателей:

- отношения, характеризующие структуру данных по авиакомпаниям³;
- отношения, характеризующие структуру данных по датам периода;
- средние значения перевозок по группам⁴, указанным в п. а;
- относительная величина координации по группам⁵, указанным в п. а;
- относительный показатель динамики⁶ для каждой авиакомпании по количеству и стоимости перевозок.

³ Относительная величина структуры совокупности (ОВСС) – это отношение части к целому. Выражается в процентах или долях.

⁴ Определяется среднее арифметическое, среднее гармоническое, среднее квадратическое и среднее геометрическое. Устанавливается свойство мажорантности средних.

⁵ Относительная величина координации (ОВК) – это соотношение между частями одного целого.

в) графической визуализации итогов:

- диаграмма структуры перевозок авиакомпаниями по количеству пассажиров;
- диаграмма структуры перевозок авиакомпаниями по стоимости перевозок;
- временной ряд количества, перевозимых пассажиров;
- диаграмма структуры перевозок авиакомпаниями по направлениям;
- диаграмма темпа изменения стоимости перевозок по датам и авиакомпаниям;
- столбиковые или круговые диаграммы для всех обобщающих показателей.

Пример выполнения задания 1.3

Информация

Одной из современных технологий построения информационно-технологических карт является SADT – технология структурного анализа и проектирования, построенная на основе стандартов IDEF.

SADT (Structured Analysis and Design Technique) – одна из самых известных методологий анализа и проектирования систем, введенная в 1973 г. Россом (Ross). SADT успешно использовалась в военных, промышленных и коммерческих организациях для решения широкого спектра задач, таких как программное обеспечение телефонных сетей, системная поддержка и диагностика, долгосрочное и стратегическое планирование, автоматизированное производство и проектирование, конфигурация компьютерных систем, обучение персонала, встроенное ПО для оборонных систем, управление финансами и материально-техническим снабжением и др. Данная методология широко поддерживается Министерством обороны США, которое было инициатором разработки стандарта IDEFO как подмножества SADT. С точки зрения SADT модель может основываться либо на функциях системы, либо на ее предметах (планах, данных, оборудовании, информации и т.д.).

⁶ Относительные величины динамики (ОВД) показывают во сколько раз увеличился или уменьшился уровень показателя текущего периода по сравнению с каким-либо предшествующим (базисным) периодом.

Основным рабочим элементом при моделировании является диаграмма. Модель SADT объединяет и организует диаграммы в иерархические древовидные структуры, при этом чем выше уровень диаграммы, тем она менее детализирована. В состав диаграммы входят блоки, изображающие активности моделируемой системы, и дуги, связывающие блоки вместе и изображающие взаимодействия и взаимосвязи между блоками. SADT требует, чтобы в диаграмме было 3-6 блоков: в этих пределах диаграммы и модели удобны для чтения, понимания и использования. Вместо одной громоздкой модели используются несколько небольших взаимосвязанных моделей, значения которых взаимодополняют друг друга, делая понятной структуризацию сложного объекта.

Блоки на диаграммах изображаются прямоугольниками и сопровождаются текстами на естественном языке, описывающими активности. В отличие от других методов структурного анализа в SADT каждая сторона блока имеет вполне определенное особое назначение: левая сторона блока предназначена для **Входов**, верхняя – для **Управления**, правая – для **Выходов**, нижняя – для **Исполнителей**. Такое обозначение отражает определенные принципы активности: *Входы* преобразуются в *Выходы*, *Управления* ограничивают или предписывают условия выполнения, *Исполнители* описывают, за счет чего выполняются преобразования.

Дуги в SADT представляют наборы предметов и маркируются текстами на естественном языке. Предметы могут состоять с активностями в четырех возможных отношениях: *Вход*, *Выход*, *Управление*, *Исполнитель*. Каждое из этих отношений изображается дугой, связанной с определенной стороной блока, таким образом стороны блока чисто графически сортируют предметы, изображаемые дугами. Входные дуги изображают предметы, используемые и преобразуемые активностями. Управляющие дуги обычно изображают информацию, управляющую действиями активностей. Выходные дуги изображают предметы, в которые преобразуются входы. Исполнительские дуги отражают (по крайней мере частично) реализацию активностей.

Блоки на диаграмме размещаются по «ступенчатой» схеме в соответствии с их доминированием, которое понимается как влияние, оказываемое одним блоком

на другие. Кроме того, блоки должны быть пронумерованы, например, в соответствии с их доминированием. Номера блоков служат однозначными идентификаторами для активностей и автоматически организуют эти активности в иерархию модели.

Взаимовлияние блоков может выражаться либо в пересылке *Выхода* к другой активности для дальнейшего преобразования, либо в выработке управляющей информации, предписывающей, что именно должна делать другая активность. Таким образом, диаграммы SADT являются предписывающими диаграммами, описывающими как преобразования между *Входом* и *Выходом*, так и предписывающие правила этих преобразований.

В SADT требуются только пять типов взаимосвязей между блоками для описания их отношений: *Управление*, *Вход*, *Управленческая Обратная Связь*, *Входная Обратная Связь*, *Выход – Исполнитель*. Отношения *Управления* и *Входа* являются простейшими, поскольку они отражают интуитивно очевидные прямые воздействия. Отношение *Управления* возникает тогда, когда *Выход* одного блока непосредственно влияет на блок с меньшим доминированием. Отношение *Входа* возникает, когда *Выход* одного блока становится *Входом* для блока с меньшим доминированием. Обратные связи более сложны, поскольку они отражают итерацию или рекурсию – *Выходы* из одной активности влияют на будущее выполнение других функций, что впоследствии влияет на исходную активность. *Управленческая Обратная Связь* возникает, когда *Выход* некоторого блока влияет на блок с большим доминированием, а отношение *Входной Обратной Связи* имеет место, когда *Выход* одного блока становится *Входом* другого блока с большим доминированием. Отношения *Выход – Исполнитель* встречаются нечасто и представляют особый интерес. Они отражают ситуацию, при которой *Выход* одной активности становится средством достижения цели другой активностью.

Дуги SADT, как правило, изображают наборы предметов, поэтому они могут разветвляться и соединяться вместе различным образом. Разветвления дуги означают, что часть ее содержимого (или весь набор предметов) может появиться в каждом ответвлении дуги. Дуга всегда помечается до разветвления, чтобы дать

название всему набору. Кроме того, каждая ветвь дуги может быть помечена в соответствии со следующими правилами: считается, что непомеченная ветвь содержит все предметы, указанные в метке перед разветвлением; каждая метка ветви уточняет, что именно содержит эта ветвь. Слияние дуг указывает, что содержимое каждой ветви участвует в формировании после слияния объединенной дуги. После слияния дуга всегда помечается для указания нового набора, кроме того, каждая ветвь перед слиянием может помечаться в соответствии со следующими правилами: считается, что непомеченные ветви содержат все предметы, указанные в общей метке после слияния; каждая метка ветви уточняет, что именно содержит эта ветвь.

При создании модели одна и та же диаграмма чертится несколько раз, что создает различные ее варианты. Чтобы различать различные версии одной и той же диаграммы, в SADT используется схема контроля конфигурации диаграмм, основанная на их номерах. Если диаграмма замещает более старый вариант, то предыдущий номер помещается в скобках для указания связи с предыдущей работой.

SADT, как и другие методологии проектирования, целесообразно использовать на ранних этапах ЖЦ: для понимания системы до ее воплощения. SADT позволяет сократить дорогостоящие ошибки на ранних этапах создания системы, улучшить контакт между пользователями и разработчиками, сгладить переход от анализа к проектированию. Несомненное достоинство SADT заключается в том, что она легко отражает такие характеристики как управление, обратная связь и исполнители⁷.

Технология SADT чаще и проще всего и реализуется средствами BPWin любой доступной версии.

Указания к выполнению

Рассмотрим несколько диаграмм декомпозиции для выбранной задачи обработке статистической информации по авиаперевозкам Российских авиакомпаний. На рис. 1.3 приведена контекстная диаграмма A0 верхнего уровня для проекта

⁷ Калянов Г.Н. Case – технологии. Консалтинг при автоматизации бизнес-процессов. 2-е изд. перераб. и доп. – М.: Горячая линия – Телеком, 2000. – 320 с.

статистической обработки информации по авиаперевозкам Российскими компаниями.

Диаграмма, приведенная на рис. 1.3, определяет наименование задачи, цели ее решения и точку зрения. *Вход* (справа) определяет информацию, которая используется для обработки и формируется на основе исходных документов. *Выход* (слева) указывает результаты решения основных статистических задач. Управление (сверху) определяется регламентирующие, которые определяют правила формирования исходной и выходной информации. Механизмы (снизу) определяют основных исполнителей для формирования исходной информации и ее обработки.

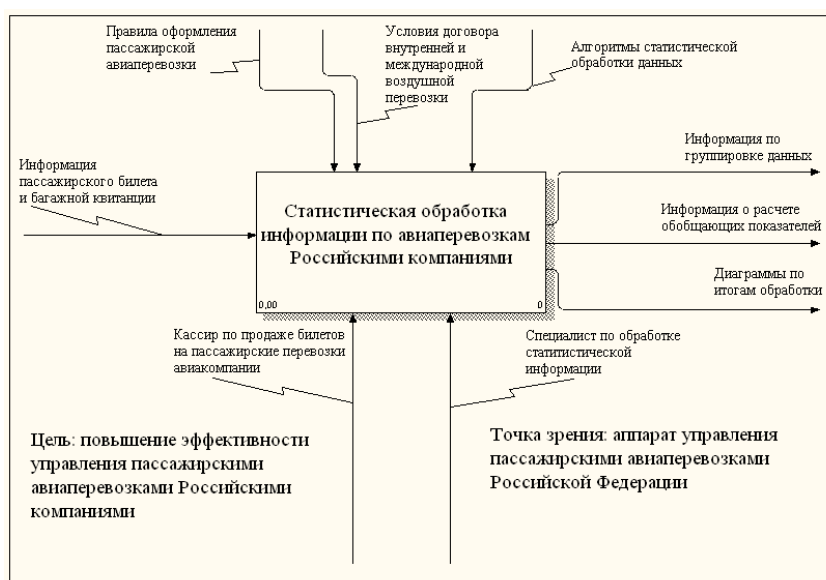


Рис. 1.3. Контекстная диаграмма А0 верхнего уровня для проекта статистической обработки информации по авиаперевозкам Российскими компаниями

На рис. 1.4 выполнена декомпозиция проекта на основные задачи обработки данных. При этом сохраняются входные, выходные и управляющие потоки информации, а также исполнители-механизмы решения этих задач.

Каждый из этапов декомпозиции также является довольно сложным и нуждается в уточнении-декомпозиции. На рис. 1.5 приведена декомпозиция этапа по выполнению алгоритмов статистической обработки данных. Следует обратить внимание на появление «туннельного» выхода, для которого выходные данные

формируются только на этом этапе и могут передаваться на следующий, а именно, формирование сводных таблиц по результатам работы.

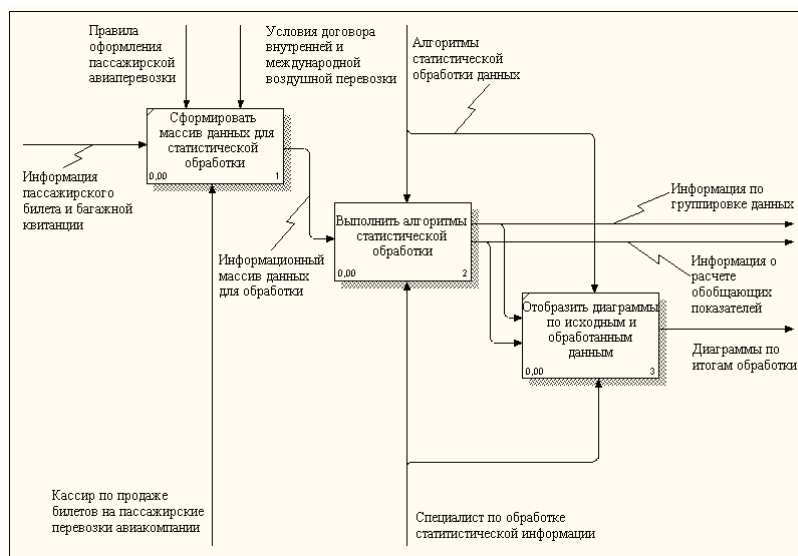


Рис. 1.4. Диаграмма декомпозиции функциональной модели проекта статистической обработки информации по авиаперевозкам Российскими компаниями (основные этапы обработки данных)

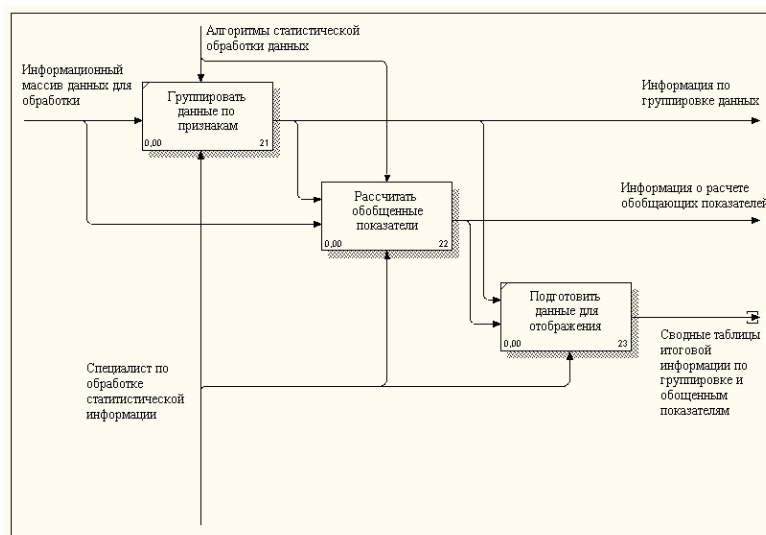


Рис. 1.5. Диаграмма декомпозиции функциональной модели проекта статистической обработки информации по авиаперевозкам Российскими компаниями (этап выполнения алгоритмов обработки)

В свою очередь этап расчета обобщенных показателей также разбивается на отдельные шаги по количеству выбранных показателей (см. рис.1.6). При боль-

шем количестве показателей они могут быть сгруппированы по смысловому значению или однотипным операциям вычислений.

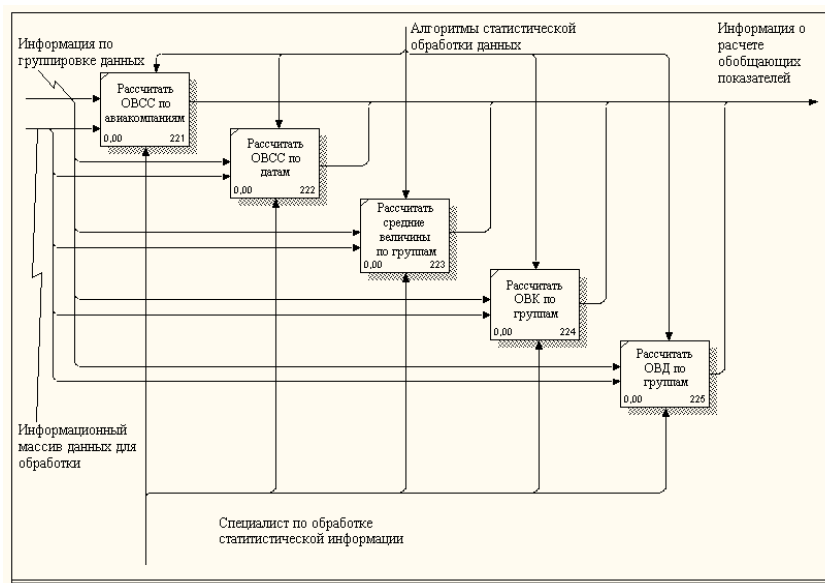


Рис. 1.5. Диаграмма декомпозиции функциональной модели проекта статистической обработки информации по авиаперевозкам Российскими компаниями (этап выполнения расчета обобщенных показателей)

Аналогичным образом выполняется декомпозиция всех остальных шагов на каждом из этапов. В результате формируется дерево проекта SADT для задачи обработки статистической информации. Фрагмент древовидной структуры, сформированной средствами BPWin 4.0 приведен на рис. 1.6.

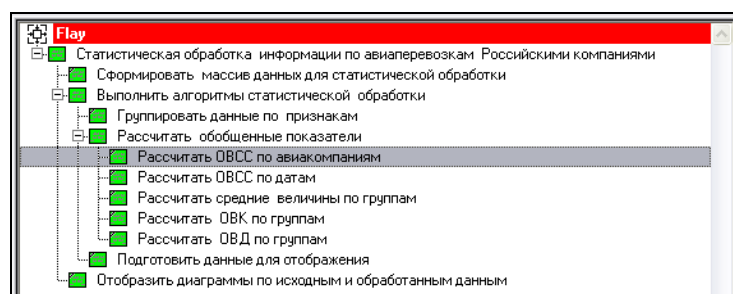


Рис. 1.6. Фрагмент древовидной структуры проекта обработки статической информации по обработке информации о пассажирских авиаперевозках

Таким образом, построение информационно-технологических карт проекта позволяет определить основные задачи и их части для последовательной реализации.

Пример выполнения задания 1.4

Информация

Информационно-технологическая карта позволяет получить общее представление об этапах решения поставленной задачи. Детализация каждого шага выполняется на основе инструкционных карт. Одной наиболее часто применяемых технологий по построению этих карт является представление алгоритмов в виде блок-схем или спецификации процессов.

Спецификация процесса (СП) используется для описания функционирования процесса в случае отсутствия необходимости детализировать его с помощью DFD (т.е. если он достаточно невелик, и его описание может занимать до одной страницы текста). Фактически СП представляют собой алгоритмы описания задач, выполняемых процессами: множество всех СП является полной спецификацией системы. СП содержат Номер и/или имя процесса, списки входных и выходных данных и тело (описание) процесса, являющееся спецификацией алгоритма или операции, трансформирующей входные потоки данных в выходные. Известно большое число разнообразных методов, позволяющих задать тело процесса, соответствующий язык может варьироваться от **структурированного естественного языка** или псевдокода до **визуальных языков проектирования** (типа **FLOW-форм и диаграмм Насси-Шнейдермана**) и формальных компьютерных языков.

Независимо от используемой нотации спецификация процесса должна начинаться с ключевого слова (например, **@СПЕЦПРОЦ**). Требуемые входные и выходные данные должны быть специфицированы следующим образом:

@ВХОД = <имя символа данных>

@ВЫХОД = <имя символа данных>

@ВХОДВЫХОД = <имя символа данных>, где <имя символа данных> – соответствующее имя из словаря данных.

Эти ключевые слова должны использоваться перед определением СП, например:

@ВХОД = СЛОВА ПАМЯТИ

@ВЫХОД = ХРАНИМЫЕ ЗНАЧЕНИЯ

@СПЕЦПРОЦ

Для всех СЛОВ ПАМЯТИ выполнить:

Распечатать ХРАНИМЫЕ ЗНАЧЕНИЯ

@ КОНЕЦ СПЕЦИФИКАЦИИ ПРОЦЕССА

Ситуация, когда символ данных является одновременно входным и выходным, может быть описана двумя способами: либо символ описывается два раза с помощью @ВХОД и @ВЫХОД, либо один раз с помощью

@ВХОДВЫХОД.

Иногда в СП задаются **пред-** и **постусловия** выполнения данного процесса. В предусловии записываются объекты, значения которых должны быть истинны перед началом выполнения процесса, что обеспечивает определенные гарантии безопасности для пользователя. Аналогично, в случае наличия постусловия гарантируется, что значения всех входящих в него объектов будут истинны при завершении процесса.

Спецификации должны удовлетворять следующим требованиям:

- для каждого процесса нижнего уровня должна существовать одна и только одна спецификация;
- спецификация должна определять способ преобразования входных потоков в выходные;
- нет необходимости (на данном этапе) определять метод реализации этого преобразования;
- спецификация должна стремиться к ограничению избыточности – не следует переопределять то, что уже было определено на диаграмме или в словаре данных;
- набор конструкций для построения спецификации должен быть простым и стандартным⁸.

⁸ Калянов Г.Н. Case – технологии. Консалтинг при автоматизации бизнес-процессов. 2-е изд. перераб. и доп. – М.: Горячая линия – Телеком, 2000. – 320 с.

Приведем пример спецификации процесса для расчета отношения, характеризующие структуру данных по авиакомпаниям.

@ВХОД=СПИСОК АВИКОМПАНИЙ

@ВХОД=КОЛИЧЕСТВО ПЕРЕВОЗОК, ВЫПОЛНЕННЫХ КАЖДОЙ АВИКОМПАНИЕЙ

@ВЫХОД=СПИСОК ОВСС

@СПЕЦПРОЦ 2.1.ПОЛУЧИТЬ СПИСОК ОВСС

ВЫПОЛНИТЬ

определить суммарное количество перевозок по всем авиакомпаниям

количество присвоить 1

ДОТЕХПОРПОКА *количество <=размер списка авиакомпаний*

ВЫПОЛНИТЬ

определить ОВСС компании равное отношению количества перевозок выбранной авиакомпании к общему количеству перевозок

поместить рассчитанное ОВСС в список ОВСС

перейти к следующей авиакомпании

КОНЕЦВЫПОЛНИТЬ

КОНЕЦВЫПОЛНИТЬ

@КОНЕЦ СПЕЦИФИКАЦИИ ПРОЦЕССА 2.1

Подобная спецификация составляется для каждого элементарного шага в схеме декомпозиции.

Для самостоятельной работы предлагается:

- 1) выбор предметной области для исследования – 2 часа;
- 2) изучение инструментальных средств среды проектирования BPWin 4.0 – 4 часа.
- 3) расширение знаний по стандартам IDEF0 – 2 часа.

Вопросы для самопроверки:

1. Определите понятие технологического процесса автоматизированной обработки статистической информации.

2. Перечислите основные этапы автоматизированной обработки статистической информации.
3. Укажите функции первичного этапа обработки информации.
4. Поясните процесс стандартизации процесса автоматизированной обработки статистической информации.
5. Укажите основные этапы типовой схемы обработки статистической информации на региональной и районном уровнях.
6. Перечислите возможные виды ошибок при сборе информации.
7. Какие методы контроля могут быть использованы для первичной информации?
8. Определите назначение информационно-технологической карты?
9. Какие известны технологии построения информационно-технологических карт?
10. Каковы основные принципы построения схем декомпозиции процесса?
11. Каково назначение инструкционных карт?
12. Какие известны методы построения инструкционных карт?
13. Каковы требования к построению спецификации?

Лабораторная работа №2.

Технология проектирования носителей, кодов и классификаторов статистической информации

Выполнив эту лабораторную работу, Вы:

- узнаете какие требования предъявляются к кодированию информации;
- узнаете зачем нужны общегосударственные классификаторы;
- научитесь применять технологии кодирования к статистической информации;
- узнаете какие требования предъявляются к статистической отчетности;
- научитесь проектировать машиночитаемые документы;
- научитесь проектировать массивы информации.

Информация⁹

Под *носителем информации* понимается физическая среда, используемая для записи и накопления информации. По способу записи и считывания информации носители подразделяются на ручные, машинные и гибридные.

К ручным носителям относятся бумажные документы, заполненные информацией и воспринимаемые непосредственно человеком без использования технических средств.

К машинным носителям относятся носители, на которые записывается и с которых считывается информация с использованием технических средств (магнитные и оптические носители).

Гибридные носители в зависимости от способа записи и считывания представляют собой машиночитаемые документы, информацию на которые заносит пользователь, а считывание производится с помощью специальных технических средств.

⁹ Информационные технологии в статистике: Учебник / Под ред. проф. В.П. Божко и проф. А.В. Хорошилова. – М.: Финстатинформ, 2002. – 144 с.

В настоящее время наиболее распространенным носителем информации является *документ*. Статистическая информация, возникающая из учетных данных предприятий и организаций, фиксируется в *документах-формах первичной статистической отчетности*, а статистическая информация со сведениями обобщающего характера, используемая для принятия управленческих решений, отражается в *документах-формах сводной статистической отчетности*. *Отчетностью* называются сведения, предоставляемые в обязательном порядке по установленным формам и в определенные сроки.

Проектирование первичных отчетов происходит в 3 этапа.

На первом этапе нужно определить необходимый состав разрабатываемых показателей, при этом избежать лишней информации, но сохранить ту, которая необходима для решения целого комплекса задач. Особое внимание здесь уделяется отбору постоянной и переменной информации.

На втором этапе приступают к размещению реквизитов на бланке отчета. Практикой проектирования с целью эффективного использования различных средств вычислительной техники рекомендуется следующий порядок размещения реквизитов: справочные реквизиты-признаки, группировочные реквизиты-признаки и другие реквизиты-признаки, а также абсолютные и относительные реквизиты-основания.

Указанные группы реквизитов располагаются в определенных частях отчета и могут иметь разные способы размещения: анкетный, линейный и табличный.

Линейный способ предусматривает размещение реквизитов в двух строках: в верхней располагаются наименования, а в нижней – их значения.

Анкетный способ применяется для размещения реквизитов в вертикальной последовательности обычно в виде двух граф: левая графа предназначена для наименований, а правая – для их значений.

Табличный способ предусматривает размещение различных реквизитов в виде таблицы с графами по вертикали и строками по горизонтали. Табличный способ используется чаще всего для размещения реквизитов-признаков в сочетании с реквизитами-основаниями.

На третьем этапе после окончательного уточнения разрабатываемого первичного отчета и согласования с размещением его реквизитов на машинном носителе приступают к расчету площади бланка. Бланки документов разрабатываются строго определенного, предусмотренного ГОСТом формата (мм): $A_1 - 841 \times 594$, $A_2 - 594 \times 420$, $A_3 - 420 \times 297$, $A_4 - 297 \times 210$, $A_5 - 210 \times 148$, $A_6 - 148 \times 105$ и $A_7 - 105 \times 74$.

Для большинства первичных статистических отчетов, представляемых по почте, принята типовая структура размещения информации, состоящая из трех частей: заголовочной, содержательной и оформительной.

Заголовочная часть для большинства статистических отчетов является стандартной. В ней приводятся постоянные справочные и группировочные реквизиты-признаки и дается полное наименование отчета с индексом формы.

Для построения заголовочной части используются анкетный и линейный способы размещения информации.

Содержательная часть отчета содержит наименования реквизитов-признаков (строк и граф) и их числовые значения (реквизиты-основания), представленные в табличном виде. Это удобно для визуального контроля и анализа, а также позволяет вводить необходимые итоги по строкам или графам и осуществлять балансовую увязку показателей.

Оформительная часть предназначена для проставления подписей должностных лиц, отвечающих за составление отчета, печати и даты составления отчета.

При проектировании первичных отчетов обязательно учитываются возможности применения современных технических средств, чтобы автоматизировать процесс их дальнейшего ввода в ПЭВМ.

Одним из перспективных способов подготовки данных является применение считывающих устройств, позволяющих осуществлять непосредственное считывание информации с первичных документов.

Широкому внедрению читающих устройств предшествует разработка специальных форм документов, которые получили название *машиночитаемых доку-*

ментов. При их проектировании предъявляются особые требования к форме, размеру, качеству и цвету бумаги, а также к порядку размещения реквизитов и их количественному составу.

В настоящее время в системе Госкомстата РФ наибольшее распространение получило читающее устройство для считывания информации, нанесенной в виде графических отметок и нормализованных цифр в определенных позициях специального бумажного документа. Устройство применяется в подготовке материалов переписи населения и других единовременных работах при вводе машиночитаемых документов в ПЭВМ. Бланки этих документов печатаются типографским способом на плотной белой бумаге обычно размером 210 × 297 мм. *Информационное поле* представляет собой матрицу, состоящую из 1000 позиций (40 строк × 25 колонок), в каждой из которых располагается заданное значение (цифры и метки), наносимое вручную.

Задание 2.1

Выполнить проектирование первичных отчетов по результатам группировки, определенных в задании 1.1.

Пример выполнения задания 2.1

Подготовим первичный отчет по количеству пассажиров, перевезенных авиакомпаниями, за каждую дату выбранного периода.

Первый этап. Отчет должен включать реквизиты: название документа и его номер, начальная дата периода, конечная дата периода, перечень дат внутри периода, показатель количества пассажиров в каждую дату периода, полученных в ходе группировки данных, дата подготовки документа и данные ответственного исполнителя.

Второй этап. В заголовочной части линейным способом размещаются: название документа, начальная и конечная даты периода. Содержательная часть представляет собой таблицу, включающей поля дата и количество пассажиров. В последней строке рассчитывается суммарное количество пассажиров. В оформительной части линейным способом располагаются дата подготовки документа и данные ответственного лица.

Третий этап. Отчет размещается на белой бумаге формата А4: 210 × 297 мм с полями по 20 мм с каждой стороны.

Макет отчета приведен на рис. 2.1.

[illegible]

Рис. 2.1. Макет первичного отчета о количестве пассажиров

Аналогичным способом описываются и макетируются все первичные отчеты для решаемой задачи.

Информация

Наиболее распространенной и удобной для пользователей (экономистов) формой вывода результатов является сводный отчет.

Проектируемые сводные статистические отчеты должны содержать необходимый состав показателей для качественного удовлетворения органов управления, иметь четкий порядок размещения реквизитов и конкретное название отчета, которое полностью отражает его содержание.

Проектирование сводных отчетов производится в три этапа.

На первом этапе определяется общий состав сводных показателей с целью установления возможностей получения максимального их числа машинным способом.

На втором этапе устанавливается конкретный состав показателей, включаемых в отдельные сводные отчеты. Возможность объединения нескольких показателей в один отчет решается после того, как установлено, что они имеют однородные признаки и должны быть получены в одинаковые сроки и для одного круга потребителей.

На третьем этапе проводится размещение реквизитов и разрабатывается бланк формы сводного отчета.

При проектировании формы сводного отчета обычно выделяются заголовочная и содержательная части.

В *заголовочной части* указывается наименование и индекс формы сводного отчета, а также ряд групп признаков, по которым сводятся одноименные показатели.

В *содержательной части* размещаются группировочные реквизиты-признаки (подлежащее таблицы) и наименования граф (сказуемое таблицы) и их числовые значения.

При проектировании форм сводной отчетности учитывается только требование, связанное с ограничением ширины листа, выдаваемого на печать.

В условиях совершенствования статистической методологии важное значение придается унификации, устанавливающей единые требования к содержанию и построению отчетно-статистических документов, используемых в экономике.

Унификация отчетности обеспечивается единым законодательным порядком разработки и централизованного утверждения органами Госкомстата РФ форм отчетов и инструкций по их заполнению, а также единой методологией приспособления статистических форм к их машинной обработке. В качестве руководящего материала при унификации используются государственные стандарты на систему отчетно-статистической документации.

При организации обработки отчетности особое внимание уделяется созданию различных массивов информации, под которыми понимается совокупность данных, хранящихся на машинных носителях. Их проектирование осуществляется в три следующих этапа.

На первом этапе определяется общий состав информационной базы, необходимой для решения требуемого комплекса задач, для чего производится анализ содержания проектируемых первичных и сводных отчетов.

В результате анализа устанавливается, в какой степени перечень реквизитов первичной отчетности обеспечивает разработку всех сводных отчетов. При необходимости вносятся соответствующие изменения за счет включения новых или устранения дублирующих реквизитов при сохранении минимальной избыточности данных и целостности информационной базы.

На втором этапе устанавливаются перечень и содержание массивов исходя из стабильности (постоянные и переменные) и сроков поступления данных. Для выполнения этого этапа необходимо знать не только состав комплекса решаемых задач, но и располагать всеми их объемно-временными характеристиками.

На третьем этапе проектируется внутримашинная структура записи массива и выбирается машинный носитель.

При проектировании внутримашинной структуры необходимо определить состав реквизитов массива и по каждому реквизиту обозначить тип записи и максимальную его значность.

При выборе машинного носителя необходимо руководствоваться такими критериями, как тип используемого компьютера, информационная емкость носителя, скорость записи и считывания данных и обеспечение определенного уровня достоверности и надежности данных.

Задание 2.2

Выполнить проектирование сводного отчета, содержащего информацию об обобщенных показателях статистической информации по заданию 1.1.

Пример выполнения задания 2.2

Первый этап. В задании 1.1 был определен перечень обобщающих показателей статистических данных:

- отношения, характеризующие структуру данных по авиакомпаниям;
- отношения, характеризующие структуру данных по датам периода;

- средние значения перевозок по группам, охватывающие даты, рейсы и авиакомпания;
- относительная величина координации по аналогичным группам;
- относительный показатель динамики для каждой авиакомпании по количеству и стоимости перевозок.

Второй этап. В сводный отчет включаются показатели, относящиеся к одному объекту и одному периоду времени: отношение, характеризующее структуру данных по авиакомпаниям; средние значения перевозок по авиакомпаниям; относительная величина координации по количеству перевозок для каждой авиакомпании. Остальные результаты группировки нарушают правило однородности данных.

Третий этап. Отчет размещается на белой бумаге формата А4: 210 × 297 мм с полями по 20 мм с каждой стороны. Ориентация страницы – альбомная. Макет документа приведен на рис. 2.2.

Аналогичным образом рассматривается структура каждого сводного отчета. Можно разрабатывать многостраничные отчеты с графическим отображением результатов.

Код компании	Наименование компании	Показатели структуры			Средняя стоимость одной перевозки, руб.	Относительная величина координации по количеству перевозок													
		По количеству пассажиров, чел.	По стоимости перевозок, руб.	555		Код компании													
555	Аэрофлот																		

Дата: _____
Исполнитель: _____

Рис. 2.2. Макет сводного отчета о работе авиакомпаний

Информация

Под *кодом* понимают систему условных обозначений объекта или элементов информационной совокупности в виде знака или группы знаков, выраженных цифрами, буквами и символами.

Процесс присвоения объектам кодовых обозначений называется *кодированием*. Основная цель кодирования состоит в однозначном обозначении объектов, а также представлении их в компактной форме.

При проектировании кодов к ним предъявляется ряд требований:

- охват всех объектов, подлежащих кодированию, и возможность их расширения без изменения правил обозначения;
- удобство восприятия и запоминания кодовых обозначений;
- максимальная информативность кода при минимальной его значности;
- возможность автоматического контроля кодовых обозначений с целью обнаружения ошибок.

Разработка кодов состоит из четырех этапов.

На первом этапе определяются объекты (номенклатура), подлежащие кодированию, которыми могут быть министерства и ведомства, номенклатура промышленной и сельскохозяйственной продукции и т.д. Затем по каждой номенклатуре устанавливается полный перечень всех позиций, подлежащих кодированию.

На втором этапе каждая номенклатура систематизируется по определенным классификационным признакам на основе выбранной системы классификации.

Упорядоченное расположение классифицируемых элементов на основе установленных взаимосвязей между признаками составляет систему классификации. В практике проектирования кодовых обозначений применяются две системы классификации: иерархическая и многоаспектная.

По *иерархической системе классификации* устанавливаются отношения соподчинения между классификационными признаками путем их деления на классы, подклассы, группы, подгруппы и т.д., т.е. осуществляется последовательный

переход от общих признаков к детализированным. Достоинство иерархической системы классификации – в простоте построения и ее большой емкости.

Однако такая жесткая структура классификации требует строгой фиксированности признаков и порядка их следования, что исключает возможность группировать объекты по не предусмотренному заранее сочетанию признаков.

В отличие от иерархической *многоаспектная (фасетная) система классификации* предполагает деление множества объектов одновременно по нескольким, независимым признакам без их соподчиненности.

При многоаспектной системе классификации множество объектов классификации может одновременно и независимо друг от друга образовывать исходя из алгоритмов решения задач любое подмножество классификационных признаков (группировок), которые внутри *фасетов* (наборов признаков) в основном располагаются в виде произвольного перечисления.

При построении многоаспектной системы классификации требуется, чтобы признаки, используемые в различных фасетах, не повторялись. В связи с группировкой признаков в независимые фасеты классификация информации приобретает гибкую блочную структуру, позволяющую вводить новые и исключать старые фасеты. Недостаток такой системы классификации — в сложности построения при чрезмерном увеличении количества признаков классификации.

На третьем этапе на основании проведенной классификации определяют правила обозначения объектов в соответствии с выбранной системой кодирования. *Выбор системы кодирования* в основном зависит от количества классификационных признаков и структуры их построения.

При проектировании машинной обработки статистической информации широко применяют следующие системы кодирования: порядковую, серийно-порядковую, позиционную (разрядную) и комбинированную.

Порядковая система кодирования заключается в последовательной регистрации объектов только с одним признаком. Для этой системы характерна простота присвоения кодовых обозначений, и ее рекомендуется применять для устойчивых и простых номенклатур.

Серийно-порядковая система применяется для кодирования двух признаков номенклатур, находящихся в определенной соподчиненности. Старшему признаку выделяется серия номеров с учетом возможного расширения позиций объекта, а младшему присваиваются порядковые номера в пределах выделенной серии. В пределах каждой серии предусматриваются резервные номера.

Позиционная (разрядная) система применяется для кодирования объектов, имеющих два и более классификационных признака, для каждого из которых выделено определенное количество позиций (разрядов). Достоинство позиционной системы – в четком выделении любого числа классификационных признаков, что придает стройность и логичность построению структуры кода.

Комбинированная система представляет собой сочетание нескольких систем кодирования, что дает возможность сократить общую значность кода.

На последнем, четвертом этапе осуществляется непосредственное присвоение объектам кодовых обозначений, которые затем оформляются в виде справочников. На этом же этапе составляются методические положения по ведению кодовых обозначений и внесению в них изменений.

Для обмена статистической информацией между различными ЭИС осуществляется разработка и внедрение классификаторов.

Классификатор представляет собой систематизированный свод наименований и кодовых классификационных группировок данной системы классификации. В зависимости от применения они делятся на три группы:

- общегосударственные, используемые во всех отраслях экономики;
- отраслевые, используемые в пределах определенной отрасли экономики;
- локальные, используемые в пределах отдельных предприятий и организаций.

Все общегосударственные классификаторы являются составной частью Единой системы классификации и кодирования технико-экономической информации (ЕСКК ТЭИ), используемой в различных ЭИС, и их применение в формах статистической отчетности и при обработке информации средствами вычислительной

техники является обязательным условием для предприятий, организаций и учреждений всех отраслей экономики.

В органах государственной статистики функционирует около 20 общегосударственных классификаторов, относящихся к следующим группам:

- классификаторы информации о трудовых и природных ресурсах;
- классификаторы информации о продуктах труда, производственной деятельности и услугах;
- классификаторы информации о структуре экономики и административно-территориальном делении страны;
- классификаторы управленческой информации и документации. Большинство общегосударственных классификаторов имеют блочную структуру, что создает возможности обработки информации на ПЭВМ при использовании кодовых обозначений отдельных блоков или их частей.

В качестве примера рассмотрим структуру общегосударственного классификатора промышленной и сельскохозяйственной продукции (ОКП) и единого государственного регистра предприятий и организаций (ЕГРПО), имеющих различную степень сложности.

Объектами ОКП являются кодовые обозначения промышленной и сельскохозяйственной продукции. Он состоит из двух блоков, представленных в табл. 2.1.

Таблица 2.1

Структура ОКП

Блок идентификации						Блок наименования продукции
Идентификационный код					КЧ	
Класс	Подкласс	Группа	Подгруппа	Вид		
XX	X	X	X	X	X	XX...X

Блок идентификации построен с применением шестиступенчатой иерархической системы классификации (класс – 2 разряда, подкласс – 1 разряд, группа – 1 разряд, подгруппа – 1 разряд, вид – 1 разряд и одноразрядное контрольное число (КЧ)), все ступени которой построены по позиционной системе кодирования. КЧ рассчитывается по специальным формулам и применяется для защиты классифи-

каторов от ошибок, возникающих в процессе кодирования и ввода данных в ПЭВМ. Блок наименования содержит полное наименование промышленной и сельскохозяйственной продукции.

ЕГРПО представляет собой единую систему государственного учета и идентификации всех субъектов хозяйственной деятельности на территории Российской Федерации, объектами которой могут быть юридические лица, филиалы, представительства и индивидуальные предприниматели.

Структура ЕГРПО включает три блока и. может быть представлена в виде табл. 2.2.

Таблица 2.2

Структура ЕГРПО

<i>Блок идентификации</i>	<i>Блок наименования и местонахождения пред- приятия, организации</i>		<i>Блок классификационных признаков</i>				
			<i>Коды признаков</i>				
<i>Идентифика- ционный код ОК- ПО</i>	<i>КЧ</i>		<i>ООГУ</i>	<i>СОАТА</i>	<i>ОКОНХ</i>	<i>КФС</i>	<i>КОПФ</i>
<i>XXXXX</i>	<i>X</i>	<i>XX...X</i>	<i>XXXXX</i>	<i>XXXX</i>	<i>XXXXX</i>	<i>XX</i>	<i>XX</i>

Каждый объект ЕГРПО идентифицируется семиразрядным кодом общегосударственного классификатора предприятий и организаций (ОКПО), построенного по серийно-порядковой системе кодирования (серии выделены по отдельным отраслям экономики), и одноразрядным КЧ.

Блок наименований и местонахождения объектов включает полное наименование предприятий (организаций), принятое в официальных документах, и отражает непосредственную связь предприятий с местом их конкретного расположения. Этот блок состоит из записей переменной длины, содержащих не более 200 символов.

Блок классификационных признаков построен по фасетной системе классификации и содержит:

- четырехразрядные коды министерств и ведомств, закодированных по СО-ОГУ (система обозначений органов государственного управления), которым подчинены предприятия;

- четырехразрядные коды территорий, закодированных по СОАТО (система обозначений административно-территориальных объектов), на которых расположены предприятия;

- пятиразрядные коды отраслей экономики, закодированных по ОКОНХ (общегосударственный классификатор отраслей народного хозяйства);

- двухразрядные коды форм собственности, закодированных по КФС (классификатор форм собственности);

- двухразрядные коды организационно-правовых форм, закодированных по КОПФ (классификатор организационно-правовых форм).

Коды министерств и ведомств по СООГУ построены по серийно-порядковой системе: старшие четыре разряда обозначают конкретное министерство, а младший – форму его подчинения (общегосударственное, местное и т.д.).

Коды территорий по СОАТО построены по позиционной системе кодирования. В данном случае из десятиразрядного кода СОАТО чаще всего используются только четыре старших разряда: коды республик, входящих в РФ, краев и областей. Остальными шестью младшими разрядами и СОАТО кодируются объекты регионального подчинения (административные районы, города, поселки городского типа и сельские органы власти), которые представляют интерес для местных и территориальных органов управления. Полный код СОАТО используется в ограниченных случаях, например при организации обработки данных переписи населения.

Коды отраслей экономики по ОКОНХ построены по позиционной системе кодирования, разряды которой размещаются по старшинству и выражают следующую классификацию: отрасль, подотрасль, вид, группа, подгруппа.

Коды форм собственности и организационно-правовой формы построены по серийно-порядковой системе, отражающей конкретную форму и ее виды.

Задание 2.3

Определить все коды, которые включены в основной информационной массив для задания 1.1. Описать вид кода, его структуру. Привести фрагмент классификатора (не менее 10 записей).

Задание 2.4

Измените вид кода для одного из показателей. Построить классификатор по новому виду кода. Укажите преимущества и недостатки нового классификатора.

Пример выполнения задания 2.3

В массив исходных данных входят поля коды: код перевозчика, код тарифа, номер авиабилета, кодирования и серийный номер, валидатор, вид тарифа.

Номер билета имеет серийно-порядковую структуру и состоит из 13 разрядов (см. табл. 2.3).

Таблица 2.3

Структура кода для номера билета

<i>Коды признаков</i>	
<i>Код авиакомпании</i>	<i>Порядковый номер билета</i>
<i>XXX</i>	<i>XXXXXXXXXX</i>

Пример фрагмента классификатора авиакомпаний:

555 – Аэрофлот,

298 – Ютейр.

Билеты нумеруются в порядке и выпуска.

Аналогичным образом описываются все поля, являющиеся кодами.

Вопросы для самопроверки

1. Какие реквизиты характеризуют качественные свойства отображаемого явления?
2. Из каких реквизитов состоит показатель?
3. Укажите классификационные признаки, имеющие принципиальное значение для машинной обработки информации.
4. На каких уровнях системы Госкомстата РФ осуществляется децентрализованная обработка?

5. В каком виде чаще всего выражается кодовое обозначение?
6. Какие преимущества дает применение комбинированной системы кодирования?
7. С использованием какой системы классификации построен общегосударственный классификатор продукции?
8. Какова степень стабильности информации, содержащейся в заголовочной части первичного отчета?

Лабораторная работа №3.

Технология реализации этапов обработки статистической информации программными средствами

Выполнив эту лабораторную работу, Вы:

- научитесь выполнять подготовку исходных данных для последующей обработки в электронных таблицах и пакете Statistica 6.1;
- узнаете какие существуют способы обработки статистической информации;
- узнаете какие существуют средства для расчета обобщающих показателей статистической информации;
- научитесь использовать средства электронных таблиц и пакета Statistica 6.1 для расчета обобщающих показателей статистической информации;
- узнаете какие существуют средства для графической визуализации результатов обработки статистической информации;
- научитесь использовать средства электронных таблиц и пакета Statistica 6.1 для графической визуализации результатов обработки статистической информации.

Информация

Рынок компьютерных программ анализа данных обширен и разнообразен. На нем представлены продукты более тысячи наименований. Такое разнообразие отражает многоплановость задач анализа в различных областях человеческой деятельности. Обзоры указанных программ приводятся в специальных справочниках, где содержатся краткие описания их назначения, требования к техническим характеристикам компьютера, сведения о дополнительных сервисных возможностях, цены и адреса фирм-поставщиков. Это весьма объемные издания, публикуемые в западной прессе¹⁰.

¹⁰ Логунова О.С., Филиппов Е.Г., Ильина Е.А. и др. Программные статистические комплексы. – Магнитогорск: ГОУ ВПО «МГТУ им. Г.И. Носова», 2007. – 190 с.

Информация о последних версиях программ регулярно помещается в популярных компьютерных журналах и газетах типа «PC Magazine», «PC World», «BYTE», «PC Week» и др. Известны аналогичные отечественные публикации. Они представлены, в основном, в журнале «Мир ПК».

Вместе с тем необходимо отметить, что значительная часть публикуемой информации быстро устаревает. Это связано со стремительными темпами развития отрасли. На рынке программного обеспечения в условиях жесткой конкуренции происходит процесс консолидации, и положение на сегодняшний день заметно отличается от ситуации, скажем, трехлетней давности. Возглавляют процесс консолидации (как недавно сказал Jack Noonan, президент корпорации SPSS) те, кто может предложить наилучший продукт и сделать это быстрее всех. Для тех же, кто испытывает трудности при переходе к новым операционным системам, процесс объединения оказывается фатальным. Пример такой консолидации – слияние SPSS с фирмой SYSTAT и приобретение в январе 1996 года корпорацией SPSS одного из крупнейших конкурентов, фирмы BMDP Statistical Software Inc.

В настоящее время наиболее популярными программными средствами для обработки статистической информации являются электронные таблицы и универсальные статистические пакеты. Сравнительный анализ возможностей программных средств приведен в таб. 3.1.

Таким образом, использование каждого класса программного обеспечения является вполне возможным и правомочным. Однако, следует предавать предпочтение специализированным программным продуктам по праву их назначения к данному типу задач.

Таблица 3.1

*Сравнительный анализ средств программного обеспечения для обработки
статистической информации*

<i>Показатель</i>	<i>Универсальные статистические программы</i>	<i>Электронные таблицы</i>
Группировка статистических данных	Сортировка данных по выбранному показателю	Средства фильтрации, группы и структуры
Подведение итогов по группе	Расчет блочных описательных статистик, гистограммы с расчетом заданных параметров	Итоги, сводная таблица
Расчет обобщенных показателей	Определение описательных статистик, возможности группировки, использование формул рабочего листа, макросы	Пакет анализа данных: описательный статистики; организация вычислений на рабочем листе
Индексный анализ данных	Использование формул рабочего листа, макросы	Организация вычислений на рабочем листе
Корреляционный анализ данных	Средства анализа: парные и частные корреляции	Стандартные встроенные функции, надстройки пакета анализа; пакет анализа данных: корреляционный анализ
Регрессионный анализ данных	Средства анализа: множественный анализ, обобщенные линейные и линейные модели, общие регрессионные модели, нелинейное оценивание, множественная нелинейная регрессия, моделирование структурными уравнениями	Стандартные встроенные функции, надстройки пакета анализа; пакет анализа данных: регрессионный анализ
Оценка временных рядов	Средства анализа: временные ряды и прогнозирование, моделирование структурными изменениями, нелинейное оценивание	Стандартные встроенные функции; графические возможности: линия тренда
Оценка динамических рядов	Средства анализа: временные ряды и прогнозирование, моделирование структурными изменениями, нелинейное оценивание	Стандартные встроенные функции; графические возможности: линия тренда
Визуализация результатов обработки	Средства графики: гистограммы, диаграммы рассеяния, матричные графики, категоризированные графики, графики блоковых данных, пиктографические графики и т.д.	Набор стандартных и нестандартных диаграмм

Задание 3.1

Организовать хранение исходных данных для выбранной задачи средствами электронных таблиц и статистического пакета Statistica.

Задание 3.2

Выполнить расчет обобщающих показателей для всей совокупности и групп статистических данных, выбранной задачи.

Задание 3.3

Выполнить графическое отображение структуры, рядов динамики и временных рядов для исходных данных средствами электронных таблиц и статистического пакета Statistica.

Задание 3.4

Оформить результаты решения заданий 3.1 – 3.3 в виде исследовательского отчета.

Указания по выполнению задания 3.1

Несмотря первоначальную схожесть в интерфейсе программных продуктов, исходные данные в электронных таблицах и статистическом пакете Statistica имеют существенные различия, приведенные в табл. 3.2.

Таблица 3.2

Сравнительный анализ отличий пользовательских средств статистического пакета Statistica и электронных таблиц MS Excel

<i>Показатель</i>	<i>Статистический пакет Statistica</i>	<i>Электронные таблицы MS Excel</i>
Размер рабочего листа	Определяется размерностью исходных данных	256 столбцов и 65536 строк
Расположение исходных данных	Заполняют весь рабочий лист	Располагаются в любом диапазоне рабочего листа
Построение заголовка таблицы	Запись в строке заголовка переменных. Названия столбцов используются как названия переменных	Первая строка таблицы, построенной на рабочем листе
Форматирование данных	Отдельно для каждого столбца	Отдельно каждой ячейки или выделенного диапазона
Запись формул	Отдельно для каждого столбца	Отдельно каждой ячейки или выделенного диапазона
Построение диаграмм	Для выделенного столбца, строки или блока	Для любого произвольного диапазона
Сохранение исходных данных и результата	Сохранение в отдельных файлах различного формата (STA, STW)	Сохранение в одной рабочей книге в файле с расширением XLS

Указанные отличия требуют их учета при организации исходных данных на рабочих листах электронных таблиц и статистического пакета. Примеры представления исходных данных приведены на рис. 3.1 и 3.2.

1	2	3	4	5	6	7	8	9	10	11	12	13
	Билет	Компания	Имя	НапрПр	НапрОб	Стоим	НомПр	НомОбр	ДатПр	ДатОбр	ВрПр	ВрОбр
555	6100619603	Кримина	Юлия	Москва	Берлин	20354	326	327	16/08	09/10/08	17:30	5:05
555	6100619604	Хрипунова	Екатерина	Москва	Берлин	20354	326	327	16/08	09/10/08	17:30	5:05
324	7200000001	Евдокимов	Роман	Киев	Москва	10257	564	456	29/08	15/09/08	18:32	22:58
324	7200000002	Гальцев	Иван	Киев	Москва	10257	564	456	29/08	15/09/08	18:32	22:58
345	1000258001	Шеркова	Анастасия	Псков	Москва	8479	689	690	7/9/08	21/09/08	8:45	10:25
345	1000258002	Хатисов	Сергей	Псков	Москва	8479	689	690	7/9/08	21/09/08	8:45	10:25
231	5231250007	Зинатулина	Эльвира	Псков	Москва	8479	689	690	7/9/08	21/09/08	8:45	10:25
231	5231250008	Муромцева	Николай	Псков	Москва	8479	689	690	7/9/08	21/09/08	8:45	10:25
555	6100619605	Корсаков	Евгений	Москва	Владивосток	19555	328	329	10/9/08	20/09/08	18:33	6:03
555	6100619606	Галкин	Денис	Москва	Владивосток	19555	328	329	10/9/08	20/09/08	18:33	6:03

Рис 3.1. Пример представления исходных данных в статистическом пакете Statistica

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Номер записи	Код	Билет	Фамилия	Имя	НапрПр	НапрОб	Стоим	НомПр	НомОбр	ДатПр	ДатОбр	ВрПр	ВрОбр
2	1	324	7200000006	Бугайчук	Ольга	Москва	Киев	9524	556	564	12.09.2008	18.09.08	21:33	18:32
3	2	324	7200000007	Пастухов	Константин	Москва	Киев	9524	556	564	12.09.2008	18.09.08	21:33	18:32
4	3	324	7200000008	Матвеева	Кристина	Москва	Киев	9524	556	564	12.09.2008	18.09.08	21:33	18:32
5	4	324	7200000009	Янгизина	Руфина	Москва	Киев	9524	556	564	12.09.2008	18.09.08	21:33	18:32
6	5	555	6100619603	Кримина	Юлия	Москва	Берлин	20354	326	327	09.01.2008	09.10.08	17:20	5:05
7	6	555	6100619604	Хрипунова	Екатерина	Москва	Берлин	20354	326	327	09.01.2008	09.10.08	17:20	5:05
8	7	324	7200000001	Евдокимов	Роман	Киев	Москва	10257	564	456	09.02.2008	15.09.08	18:32	22:58
9	8	324	7200000002	Гальцев	Иван	Киев	Москва	10257	564	456	09.02.2008	15.09.08	18:32	22:58
10	9	324	7200000010	Галимзянов	Денис	Киев	Москва	10257	564	456	09.02.2008	15.09.08	18:32	22:58
11	10	324	7200000003	Зинатулина	Эльвира	Киев	Москва	10257	564	456	09.02.2008	15.09.08	18:32	22:58
12	11	324	7200000004	Бычинский	Николай	Киев	Москва	10257	564	456	09.02.2008	15.09.08	18:32	22:58

Рис. 3.2. Пример представления исходных данных на рабочем листе электронных таблиц MS Excel

Указания по выполнению задания 3.2

Для расчета обобщающих показателей следует использовать в первую очередь встроенные возможности статистического пакета и электронных таблиц. Рассмотрим примеры использования таких средств в пакете Statistica и MS Excel для расчета среднего значения стоимости авиабилета по всем компаниям и каждой компанией отдельно.

Для расчета среднего значения по всей совокупности в пакете Statistica легче всего использовать средства описательных статистик: *Анализ* → *Основные статистики и таблицы* → *Описательные статистики*. В диалоговом окне описательных статистик следует определить ряд показателей, которые должны быть определены (см. рис. 3.3).

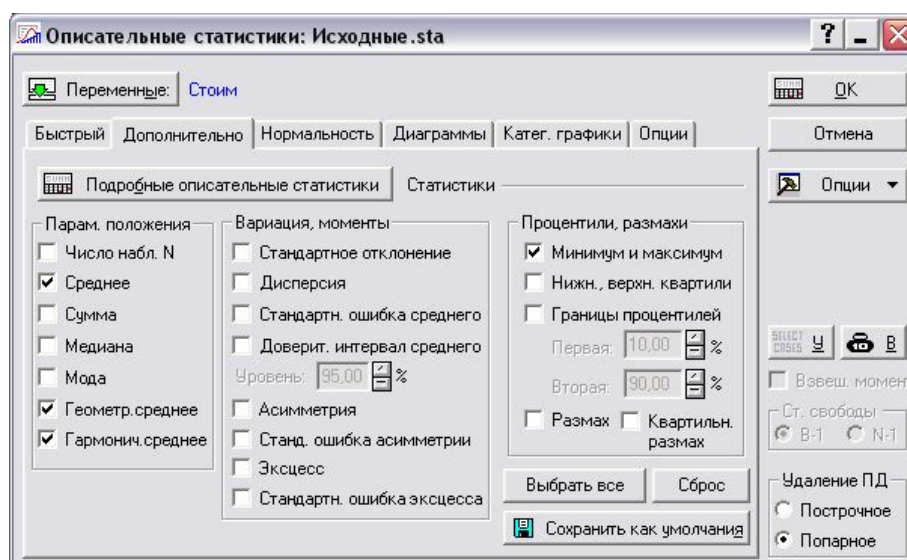


Рис. 3.3. Вид диалогового окна для определения дополнительных параметров описательных статистик

Результаты расчета отображаются в отдельном окне (см. рис. 3.4.).

Описательные статистики (Исходные.sta)					
Переменная	Среднее	Геометр. Среднее	Гармонич. Среднее	Минимум	Максимум
Стоим	14799,12	13148,33	11973,81	8479,000	35241,00

Рис. 3.4. Вид рабочего листа с результатами расчета описательных статистик

Для определения среднего значения по группам используются *Основные статистики и таблицы* → *Группировка и однофакторый ДА* (дисперсионный анализ). При использовании этого средства следует определить:

1) на закладке *Отдельные таблицы* список зависимых и группирующих переменных. Для зависимых переменных определяются обобщенные показатели, а группирующие определяют признак группировки (см. рис. 3.5);

2) на закладке *Списки таблиц* выбираются виды показателей, которые рассчитываются для каждой группы (см. рис. 3.6).

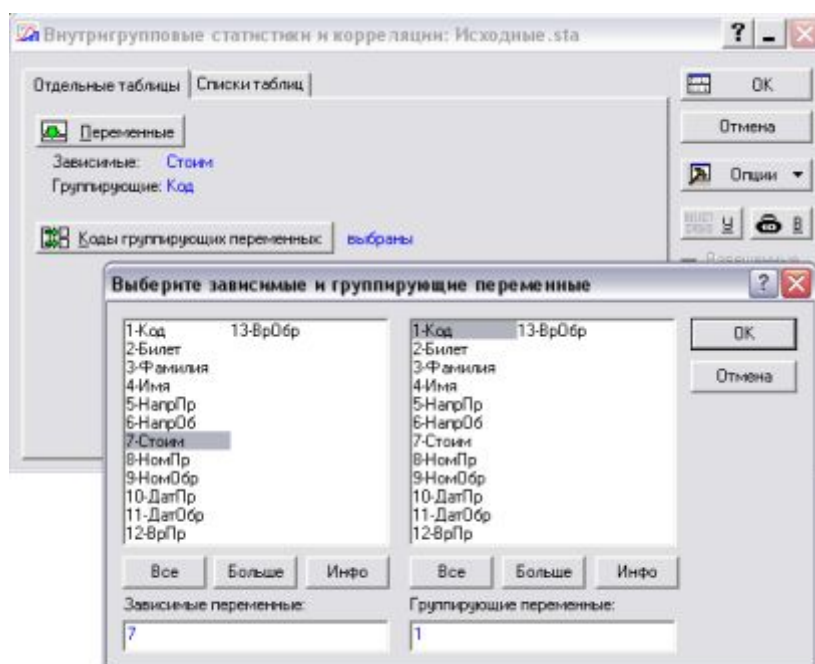


Рис. 3.5. Вид диалогового окна для определения переменных по группировке

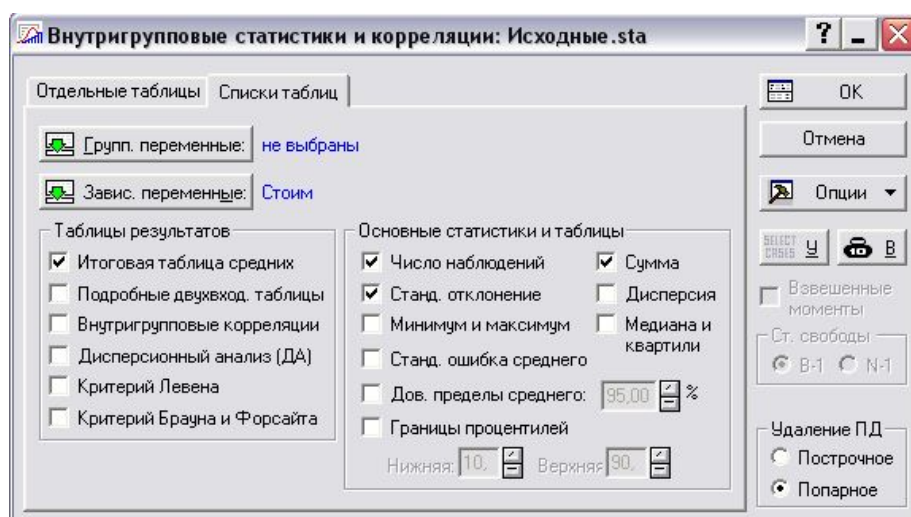


Рис. 3.6. Вид диалогового окна для выбора вида результата группировки

На рис. 3.7 приведены результаты расчета обобщенных показателей по группам авиакомпаний.

Итоговая таблица средних (Исходные.sta) N=50 (Нет пропусков в завис. перем.)				
Код	Стоим Среднее	Стоим N	Стоим Сумма	Стоим Ст.откл.
231	8525,50	8	68204,0	28,700
324	9963,80	10	99638,0	378,520
345	8499,67	9	76497,0	31,000
555	21548,57	23	495617,0	7674,685
Всего	14799,12	50	739956,0	8144,991

Рис. 3.7. Вид рабочей таблицы с результатами расчета групповых показателей

При расчете показателей структуры можно воспользоваться формулами рабочего листа с результатами. Для этого добавить переменную *Структура* после переменной V5 и для нее записать формулу

$$=V4/739956.0.$$

В результате получим дополнение таблицы, приведенной на рис. 3.7 (см. рис. 3.8).

Таким образом, по результатам итоговой таблицы (см. рис. 3.8) можно утверждать, что

1) всеми авиакомпаниями было выполнено 50 перевозок пассажиров на сумму 739956,00 руб.;

2) наибольшее количество пассажиров перевезено компанией с кодом 555 - Аэрофлот и доля перевозок в общей сумме составила 67 % при средней стоимости одной перевозки в этой авиакомпании в 21548, 57 руб.;

3) наименьшее количество пассажиров перевезено компанией с кодом 231 и доля перевозок в общей сумме составила 9 % при средней стоимости одной перевозки в этой авиакомпании в 8525,50 руб.

Итоговая таблица средних (Исходные.sta)					
N=50 (Нет пропусков в завис. перем.)					
Код	Стоим Среднее	Стоим N	Стоим Сумма	Стоим Ст.откл.	Структура =v4/739956
231	8525,50	8	68204,0	28,700	9%
324	9963,80	10	99638,0	378,520	13%
345	8499,67	9	76497,0	31,000	10%
555	21548,57	23	495617,0	7674,685	67%
Всего	14799,12	50	739956,0	8144,991	100%

Рис. 3.8. Вид рабочей таблицы с результатами расчета показателей структуры

При расчете обобщенных показателей средствами электронных таблиц MS Excel рекомендуется использовать средства встроенных функций (см. табл. 3.3).

Пример использования указанных функций приведен на рис. 3.9 и результаты вычислений на рис. 3.10.

Таблица 3.3

*Встроенные функции MS Excel для расчета обобщающих показателей
статистических данных*

<i>Наименование задачи</i>	<i>Синтаксис функции</i>	<i>Описание аргументов</i>
Расчет среднего значения	СРЗНАЧ(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется среднее.
Расчет общей дисперсии	ДИСП(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется дисперсия.
Расчет стандартного отклонения	СТАНДОТКЛОН(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется стандартное отклонение
Расчет максимального элемента	МАКС(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется максимум
Расчет минимального элемента	МИН(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется минимум
Определение моды	МОДА(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется мода
Определение медианы	МЕДИАНА(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется медиана
Расчет среднего гармонического	СРГАРМ(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется среднее гармоническое
Расчет среднего геометрического	СРГЕОМ(число1; число2; ...)	Число1, число2, ... – это от 1 до 30 аргументов, для которых вычисляется среднее геометрическое
Подсчет количества по критерию	СЧЁТЕСЛИ(диапазон; критерий)	Диапазон – диапазон, в котором нужно подсчитать ячейки. Критерий – критерий в форме числа, выражения или текста, который определяет, какие ячейки надо подсчитывать.
Подсчет суммы по критерию	СУММЕСЛИ(диапазон; критерий; диапазон_суммирования)	Диапазон – диапазон вычисляемых ячеек. Критерий – критерий в форме числа, выражения или текста, определяющего суммируемые ячейки
Подсчет частот по интервалам	ЧАСТОТА(массив_данных; массив_интервалов)	Массив_данных – массив или ссылка на множество данных, для которых вычисляются частоты.

	P	Q	R	S	T
	Коды	Количество	Суммарная стоимость	Коэффициент структуры по количеству	Коэффициент структуры по сумме
1					
2	231	=СЧЕТЕСЛИ(\$B\$2:\$B\$51;P2)	=СУММЕСЛИ(\$B\$2:\$B\$51;P2;\$H\$2=Q2/\$Q\$6		=R2/\$R\$6
3	324	=СЧЕТЕСЛИ(\$B\$2:\$B\$51;P3)	=СУММЕСЛИ(\$B\$2:\$B\$51;P3;\$H\$2=Q3/\$Q\$6		=R3/\$R\$6
4	345	=СЧЕТЕСЛИ(\$B\$2:\$B\$51;P4)	=СУММЕСЛИ(\$B\$2:\$B\$51;P4;\$H\$2=Q4/\$Q\$6		=R4/\$R\$6
5	555	=СЧЕТЕСЛИ(\$B\$2:\$B\$51;P5)	=СУММЕСЛИ(\$B\$2:\$B\$51;P5;\$H\$2=Q5/\$Q\$6		=R5/\$R\$6
6	Итого	=СУММ(Q2:Q5)	=СУММ(R2:R5)		

Рис. 3.9. Фрагмент рабочего листа с формулами для расчета показателей структуры по количеству и по стоимости перевозок

	P	Q	R	S	T
	Коды	Количество	Суммарная стоимость	Коэффициент структуры по количеству	Коэффициент структуры по сумме
1					
2	231	8	68204	16,00%	9,22%
3	324	10	99638	20,00%	13,47%
4	345	9	76497	18,00%	10,34%
5	555	23	495617	46,00%	66,98%
6	Итого	50	739956		

Рис. 3.10. Фрагмент рабочего листа с результатами расчета показателей структуры по количеству и по стоимости перевозок

Для отсортированных данных возможно применение возможностей *Итого* в меню *Данные*.

Указания по выполнению задания 3.3

Алгоритмы обработки статистической информации подразумевают построение различных графиков. Графики, наряду со статистическими таблицами, являются важным средством выражения и анализа статистических данных, поскольку наглядное представление облегчает восприятие информации. Графики позволяют мгновенно охватить и осмыслить совокупность показателей – выявить наиболее типичные соотношения и связи этих показателей, определить тенденции развития, охарактеризовать структуру, степень выполнения плана, оценить географическое размещение объектов. Этим объясняется широкое применение графиков для пропаганды статистической информации, характеризующей результаты развития различных сфер национальной экономики и социальных отношений.

В настоящее время разработаны пакеты прикладных программ компьютерной графики, которые облегчают задачу исследователя в практическом применении графиков. Наиболее распространенными пакетами прикладных программ являются: «Harvard graphics», «Statgraf», «Supercalc», «MS Excel».

Для графического изображения статистических данных используются самые разнообразные виды графиков. Классификация графиков представлена на рис.

3.11. На этой схеме графики подразделяются по двум признакам классификации: по способу построения и по цели использования.

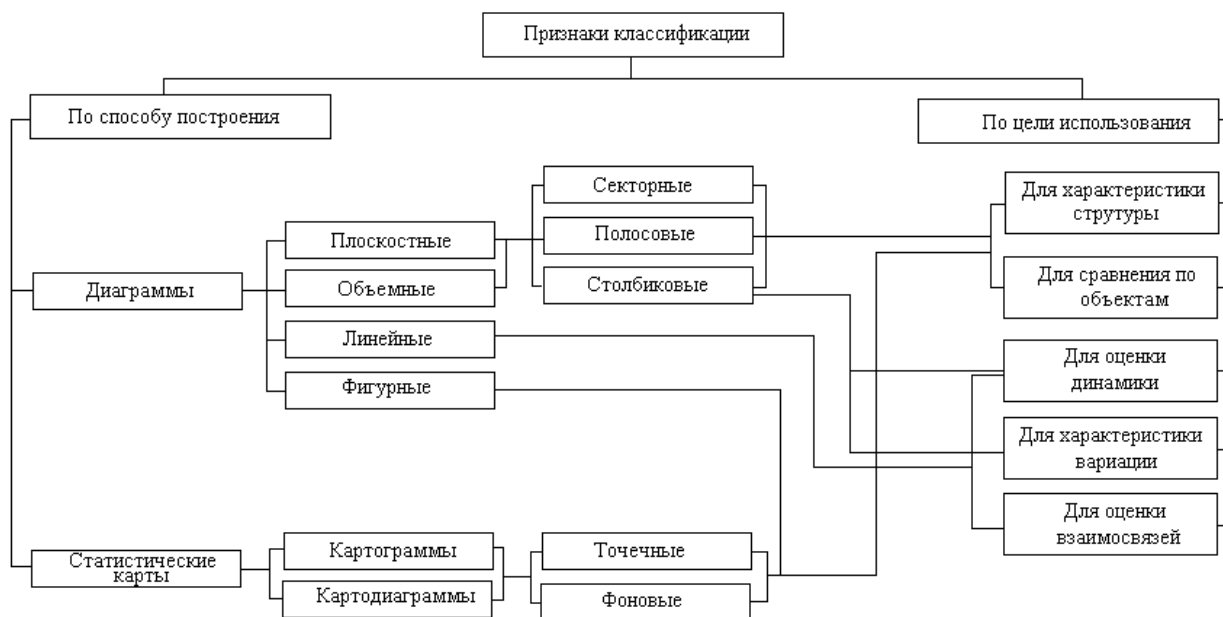


Рис. 3.11. Классификация диаграмм для графического отображения исходных статистических данных и результата¹¹

Несмотря на многообразие видов графических изображений, при их построении выполняются общие правила. Так, во-первых, в соответствии с целью использования выбирается графический образ, т. е. вид графического изображения. Во-вторых, определяется поле графика – то пространство, в котором размещаются геометрические знаки. В-третьих, задаются масштабные ориентиры с помощью масштабных шкал (равномерных или неравномерных). В-четвертых, выбирается система координат, необходимая для размещения геометрических знаков в поле графика. Наиболее распространенной системой координат при построении статистических графиков является система прямоугольных координат. При этом наилучшее соотношение масштаба по осям абсцисс и ординат, равное 1,62:1 называется «Золотым сечением».

¹¹ Ефимова М.Р. Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. – М.: Инфра-М, 1998. – 416 с.

Для отображения структуры исходных данных в пакете Statistica можно использовать гистограммы и круговые диаграммы. Например, для отображение структуры по количеству перевозок достаточно выбрать: *Графика* → *2М Графики* → *Гистограмма* или *Графика* → *2М Графики* → *Круговая диаграмма*. В качестве переменной выбирается *Код* авиакомпании и для интервалов частот все коды. Результаты построения диаграмм приведены на рис. 3.12 и 3.13.

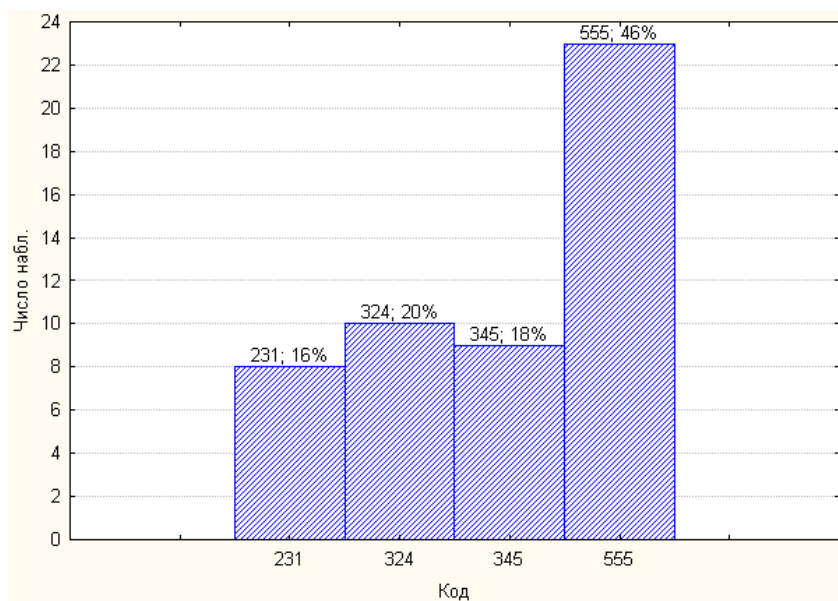


Рис. 3.12. Вид столбчатой диаграммы для структуры количества перевозок по авиакомпаниям

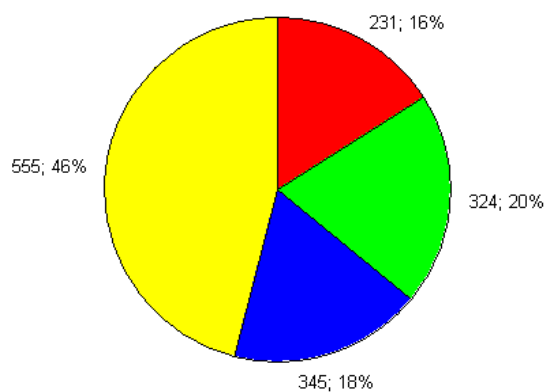


Рис. 3.13. Вид круговой диаграммы для структуры количества перевозок по авиакомпаниям

Для построения диаграмм структуры по сумме перевозок необходимо:

- 1) выполнить группировку данных и рассчитать показатели структуры (см. рис. 3.8);

- 2) создать новую таблицу, в которую скопировать результаты группировки;
- 3) выполнить построение столбчатой или круговой диаграммы.

Результаты построения таких диаграмм приведены на рис. 3.14 и 3.15.

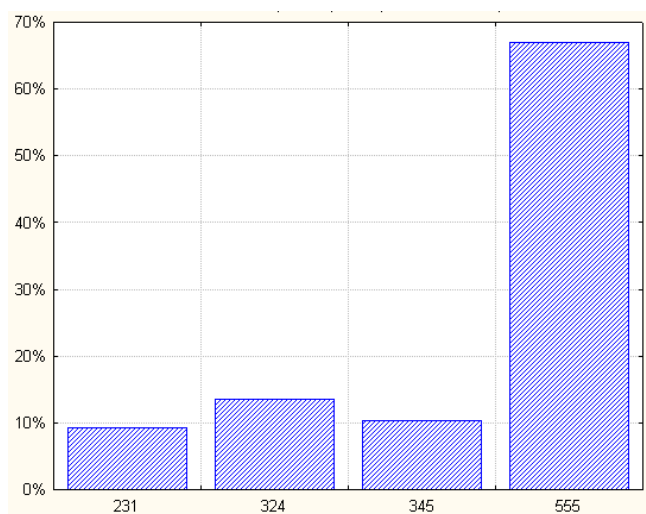


Рис. 3.14. Вид столбчатой диаграммы для структуры объема перевозок по авиакомпаниям, руб.

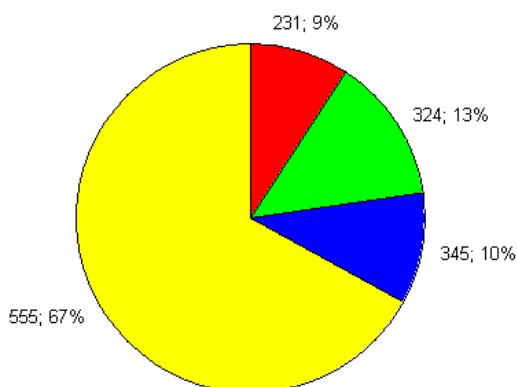


Рис. 3.15. Вид круговой диаграммы для структуры объема перевозок по авиакомпаниям, руб.

Построение временного ряда для статистических данных можно выполнить, используя средства *Анализ* → *Временные ряды и прогнозирование*, либо *Линейный график* по столбцу выделенного диапазона, либо *Диаграмма рассеяния*. Первоначально должна быть выполнена группировка данных по выбранным показателям. Пример построения временного ряда по количеству перевезенных пассажиров приведен на рис. 3.16.

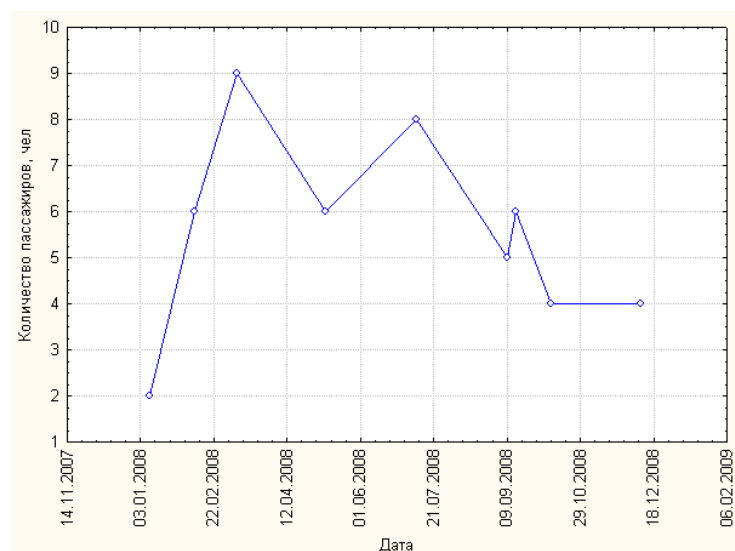


Рис. 3.16. Вид временного ряда для отображения зависимости количества перевезенных пассажиров во времени

На основании рис. 3.16 можно утверждать, что ряд имеет убывающую тенденцию, начиная с февраля месяца 2008 года.

Для построения подобных диаграмм в электронных таблицах первоначально следует подготовить исходные данные и выбрать соответствующий вид диаграммы.

Для самостоятельной работы предлагается:

- 1) рассмотрение использования средств пиктографики и категоризированных диаграмм для отображения многомерных зависимостей – 2 часа;
- 2) изучение инструментальных средств для проведения индексного анализа данных – 4 часа.
- 3) расширение знаний по стандартам проектированию отчетов с графически отображенными данными – 2 часа.

Вопросы для самопроверки

1. Укажите программное обеспечение, которое может применяться для обработки статистической информации
2. Перечислите основные возможности электронных таблиц для обработки статистических данных.
3. Укажите преимущества статистических пакетов перед электронными таблицами при обработке статистических данных.

4. Укажите основные отличия представления исходных данных на рабочем листе статистических пакетов и электронных таблиц.

5. Перечислите встроенные функции электронных таблиц для обработки статистической информации.

6. Укажите средства статистических пакетов, которые можно применять для расчета показателей структуры статистической информации.

7. Приведите классификацию диаграмм для графического отображения статистической информации.

8. Укажите отличия столбчатой диаграммы и гистограммы в статистическом пакете.

9. Опишите средства построения временных рядов с помощью программных продуктов.

10. В каких целях используются секторные и столбиковые диаграммы?

11. Приведите примеры диаграмм, которые можно использовать для отображения многомерных зависимостей.

Лабораторная работа №4.

Технология оценки экономической эффективности автоматизированной обработки статистической информации¹²

Выполнив эту лабораторную работу Вы:

- научиться выполнять расчет экономической эффективности автоматизированной обработки статистической информации;
- научиться рассчитывать основные показатели прямой эффективности автоматизированной обработки статистической информации.

Информация

Мероприятия, связанные с внедрением современных средств вычислительной техники, требуют значительных стоимостных и трудовых затрат. Обязательным условием применения этих средств является получение экономической эффективности от их внедрения. Для этого требуется проведение расчетов экономической эффективности и ее обоснования.

Экономическое обоснование эффективности автоматизированной обработки статистической информации позволяет:

- определить необходимость и целесообразность затрат на создание и внедрение автоматизированной системы обработки информации на различных уровнях системы Госкомстата РФ;
- установить основные направления и очередность автоматизации обработки информации исходя из специфики различных уровней системы Госкомстата РФ, а также методов получения, передачи и обработки информации на каждом из них;
- выбрать экономически более эффективные варианты технологических процессов обработки информации.

¹² Источник теоретической информации: Информационные технологии в статистике: Учебник / Под ред. проф. В.П. Божко и проф. А.В. Хорошилова. – М.: Финстатинформ, 2002. – 144 с.

Экономическая эффективность автоматизированной обработки информации обеспечивается за счет следующих факторов:

– высокой скорости выполнения различных операций, достигнутой за счет высокой производительности современных технических средств и максимального сокращения времени на их обработку;

– повышения качества выполнения статистических работ за счет создания единой информационной базы и многофункционального ее использования;

– улучшения информационного обслуживания различных уровней системы Госкомстата РФ за счет сокращения продолжительности отчетных периодов и сроков получения отчетности.

Исходя из вышеизложенного следует сказать, что *общая эффективность* автоматизированного решения статистических задач находится в прямой зависимости от снижения затрат на обработку информации (*прямая эффективность*) и достигаемого уровня информационного обслуживания (*косвенная эффективность*).

Прямая эффективность выражается в снижении трудовых и стоимостных затрат на обработку статистической информации и непосредственно, т.е. количественно, влияет на экономические показатели деятельности вычислительных установок, занимающихся ее обработкой.

Показатели прямой экономической эффективности определяются в результате сравнения затрат на обработку информации нескольких вариантов проектных решений. Важно, чтобы они рассматривались в равных условиях, т.е. при совпадении состава решаемых

Косвенная эффективность характеризует качественные изменения, происходящие в результате применения средств вычислительной техники. Они выражаются в повышении качества статистических работ, увеличения состава получаемой информации и повышении ее достоверности и оперативности и т.д.

Следует сказать, что в настоящее время существуют трудности при расчете косвенной эффективности, поскольку нет методов для оценки доли эффективности от автоматизированной обработки информации в общей эффективности, получаемой в результате различных мер

задач, объема исходных данных и степени достоверности.

В упрощенном походе это сравнение двух вариантов: спроектированного и базового. За *базовый* вариант принимается существующая система машинной обработки, а за *спроектированный* – результат модификации существующей системы.

по улучшению информационного обслуживания различных потребителей.

В этом случае для расчета ожидаемой косвенной эффективности рекомендуется применять метод экспертных оценок.

Для оценки прямой эффективности используются три вида затрат:

- стоимостные затраты;
- капитальные затраты;
- трудовые затраты.

Расчет показателей эффективности для стоимостных затрат СИС¹³

1. Величина снижения затрат на обработку информации:

$$\Delta C = C_0 - C_1,$$

где C_0 – стоимость обработки информации при базовом варианте;

C_1 – при проектируемом варианте.

2. Коэффициент эффективности затрат:

$$K_c = \frac{\Delta C}{C_0} \text{ или } K_c = \frac{\Delta C}{C_0} \times 100\%.$$

Коэффициент эффективности затрат показывает, какая часть затрат будет сэкономлена при проектируемом варианте (или на сколько процентов снижаются затраты).

3. Индекс изменения затрат:

¹³ Источниками эффективности компьютерных технологий являются:

- уменьшение времени разработки приложений;
- сокращение трудоемкости при хранении и обработке информации;
- сокращение затрат на перевозку и обслуживание образцов оборудования и технологий, полная замена их на электронные образы.

$$I_c = \frac{C_0}{C_1}.$$

Индекс изменения затрат показывает, во сколько раз снизятся затраты при проектируемом варианте.

Расчет показателей эффективности для капитальных затрат СИС¹⁴

1. Величина снижения капитальных затрат на обработку информации:

$$K_{\partial} = K_1 - K_0,$$

где K_0 – капитальные затраты обработки информации при базовом варианте;

K_1 – при проектируемом варианте.

2. Срок окупаемости дополнительных капитальных затрат:

$$t = \frac{K_{\partial}}{\Delta C}.$$

Капитальные затраты считаются экономически целесообразными в том случае, если они окупаются экономией текущих затрат в пределах нормативного срока окупаемости.

3. Коэффициент экономической эффективности капитальных затрат:

$$E_p = \frac{1}{t}.$$

Коэффициент экономической эффективности капитальных затрат определяет долю окупаемости этих затрат за год.

Расчет показателей эффективности трудовых затрат СИС

1. Абсолютный показатель снижения трудовых затрат:

$$\Delta T = T_0 - T_1,$$

где T_0 – годовая трудоемкость обработки информации при базовом варианте;

T_1 – при проектируемом варианте.

2. Коэффициент снижения трудовых затрат:

¹⁴ Капитальные затраты – «...инвестиции в основной капитал (основные средства), в том числе затраты на новое строительство, расширение, реконструкцию и техническое перевооружение действующих предприятий, приобретение машин, оборудования, инструмента, инвентаря, проектно-изыскательские работы и другие затраты...».

$$K_T = \frac{\Delta T}{T_0} \text{ или } K_T = \frac{\Delta T}{T_0} \times 100\%.$$

3. Индекс изменения трудовых затрат:

$$I_T = \frac{T_0}{T_1}.$$

Задание 4.1

Определить перечень статей для определения стоимостных, капитальных и трудовых затрат для разработки или модификации СИС.

Указания к выполнению задания 4.1

Одним из сложных этапов определения эффективности является определение списка статей расходов по стоимостным, капитальным и трудовым затратам.

При построении этого списка следует учесть следующие виды затрат:

1) стоимостные на:

- 1.1) эскизное проектирование документов и интерфейса;
- 1.2) модификацию структуры данных;
- 1.3) разработку программного обеспечения;
- 1.4) разработка проектной документации;
- 1.5) внедрение модифицированной системы;

2) капитальные на:

- 2.1) модификацию технических средств;
- 2.2) организацию рабочего места;
- 2.3) стоимость расходных материалов;

3) трудовые на:

- 3.1) обработку первичной информации (ввод данных);
- 3.2) подготовку промежуточных и итоговых отчетов;
- 3.3) передачу информации.

Задание 4.2

Определить расходы по каждой статье затрат, указанных в задании 4.1 для базовой существующей системы и предполагаемой модифицированной.

Указания к выполнению задания 4.1

По каждой статье затрат выполняется расчет расходов. При этом следует учесть существующие рекомендации и показатели оценки работ различного вида.

Например, возможное изменение структуры входных форм может привести к изменению времени на ввод информации: нажатие клавиши на клавиатуре или манипуляторе мышь – 0,2 с. Ввод данных возможен с клавиатуры или выбором из списка (календаря). Ввод даты по краткому формату – 8 символов или 1,6 секунды + ментальная подготовка по 1,35 с на каждый символ. Итого $1,6 + 8 \cdot 1,35 = 12,4$ с. Выбор из календаря: 4 щелчка кнопкой манипулятора «мышь» + 4 ментальные подготовки. Итого: 6,2 с. Экономия времени на подготовку первичной информации: $12,4 - 6,2 = 6,2$ на каждое измененное поле.

При модификации структуры данных следует выбирать типы с наименьшим объемом памяти (см. табл. 4.1).

Таблица 4.1

Характеристика типов переменных для оценки объема структур, используемых в СИС

<i>Тип данных</i>	<i>Размер, байт</i>	<i>Тип данных</i>	<i>Размер, байт</i>
Логический	2	Вещественный длинный	4
Целый короткий	2	Денежный	8
Целый длинный	4	Дата / время	8
Вещественный короткий	4	Текстовый	1+1 на каждый символ

Результаты расчета затрат следует оформить в виде таблицы (см. табл. 4.2).

Таблица 4.2

Сводная таблица затрат на модификацию СИС

<i>Статья затрат</i>	<i>Базовый проект</i>	<i>Модифицированный проект</i>
1. Стоимостные		
1.1.		
1.2.		
...		
Итого стоимостных		
2. Капитальные		
2.1.		
...		

Задание 4.3

Выполнить расчет абсолютных и относительных показателей экономической эффективности разработки или модификации СИС. Оформить полученные результаты в виде отчета.

Для самостоятельной работы предлагается:

- 1) поиск полного состава статей расходов при модификации СИС – 2 часа;
- 2) изучение методик расчета стоимости разработки программного обеспечения – 2 ч.
- 3) расширение знаний по внедрению СИС на автоматизированных рабочих местах – 2 часа.

Вопросы для самопроверки

1. Определите понятие прямых затрат при модификации и разработке СИС.
2. Определите понятие косвенных затрат при модификации и разработке СИС.
3. Укажите порядок расчета абсолютных и относительных показателей эффективности стоимостных затрат при модификации и разработке СИС.
4. Укажите порядок расчета абсолютных и относительных показателей эффективности капитальных затрат при модификации и разработке СИС.
5. Укажите порядок расчета абсолютных и относительных показателей эффективности трудовых затрат при модификации и разработке СИС.
6. Укажите факторы, оказывающие влияние на формирование стоимостных затрат. Поясните причину влияния.
7. Укажите факторы, оказывающие влияние на формирование капитальных затрат. Поясните причину влияния.
8. Укажите факторы, оказывающие влияние на формирование трудовых затрат. Поясните причину влияния.

Лабораторная работа №5.

Технология применения официального сайта

Государственного комитета статистики для добычи статистических данных

Выполнив эту лабораторную работу Вы:

- изучите структуру официального сайта Государственного комитета статистики Российской Федерации;
- ознакомьтесь с основными нормативными документами по организации Государственного комитета статистики и доступа к информации;
- научитесь использовать ресурсы сайта Государственного комитета статистики Российской Федерации для поиска и добычи статистической информации.

Информация

Удовлетворение потребностей органов власти и управления, средств массовой информации, населения, научной общественности, коммерческих организаций и предпринимателей, международных организаций в разнообразной, объективной и полной статистической информации – главная задача Федеральной службы государственной статистики. Для ее решения действует система государственной статистики, в состав которой входят центральный аппарат на федеральном уровне и территориальные органы Росстата, расположенные во всех субъектах Российской Федерации¹⁵.

В настоящее время мир переживает переход от «индустриального» общества к «информационному». Уровень развития информационного пространства общества решающим образом влияет на экономику, обороноспособность и политику страны. От этого уровня в значительной степени зависит поведение людей, формирование общественно-политических движений и социальная стабильность.

Целями информатизации во всем мире и в том числе в России является наиболее полное удовлетворение информационных потребностей общества во

¹⁵ Информация с официального сайта Государственного комитета статистики Российской Федерации

всех сферах деятельности, улучшение условий жизни населения, повышения эффективности производства, содействие стабилизации социально-политических отношений в государстве на основе внедрения средств вычислительной техники и телекоммуникаций.

В настоящее время созданы необходимые предпосылки формирования и развития информационного общества в России. Во всех развитых странах удовлетворение всех основных социальных и индивидуальных потребностей осуществляется за счет распространения и использования информационных ресурсов общества, получения доступа к ним посредством современных технологий и развитой информационно-коммуникационной структуры¹⁶.

Одной из технологий распространения информации и предоставления к ней доступа является технологий использования глобальных сетей с разработкой официальных сайтов государственных организаций.

Государственный комитет статистики использует указанные технологии для реализации государственной программы по информатизации общества в России. Адрес официального сайта Государственный комитет статистики Российской Федерации – www.gks.ru.

Сайт содержит ряд WEB-страниц, которые имеют тематическую группировку материалов (см. рис. 5.1). Назначение основных страниц можно определить следующим образом:

– *Главная* – ссылки на основные показатели социально-экономической статистике по Российской Федерации, ссылки на информацию об основных событиях в социально экономической сфере, ссылки на Федеральные законы, постановления и общегосударственные классификаторы (см. рис. 5.1).

– *О Росстате* – ссылки на информацию, содержащую сведения о положениях и регламенте Федеральной службы статистики России, подразделениях службы, исторические сведения и развитии статистики в мире и в России (см. рис. 5.2).

¹⁶ Волокитин А.В., Маношкин А.П., Курносов И.Н. и др. Практические аспекты информатизации. Стандартизация, сертификация и лицензирование. – М.: ФИОРД-ИНФО, 2000. – 270 с.

– *Россия в цифрах* – ссылки на информацию, содержащую сведения об основных показателях социально-экономической статистики в России по различным отраслям: статистика населения, уровня жизни и трудовых ресурсов и т.д. (см. рис. 5.3).

– *Публикации* – ссылки на информацию, содержащую сведения об издании информационно-аналитических материалов (см. рис. 5.4).

– *ССРД МВФ* – ссылки на информацию, содержащую сведения о состоянии международного валютного фонда и распространения информации данного вид (см. рис. 5.5).

– *Новости* – ссылки на информацию, содержащую сведения официальной хроники, официальных обращений к гражданам и о вакансиях (см. рис. 5.6).

– *Конкурсы* – ссылки на информацию, содержащую сведения о закупках различного вида в различных областях народного хозяйства (см. рис. 5.7).

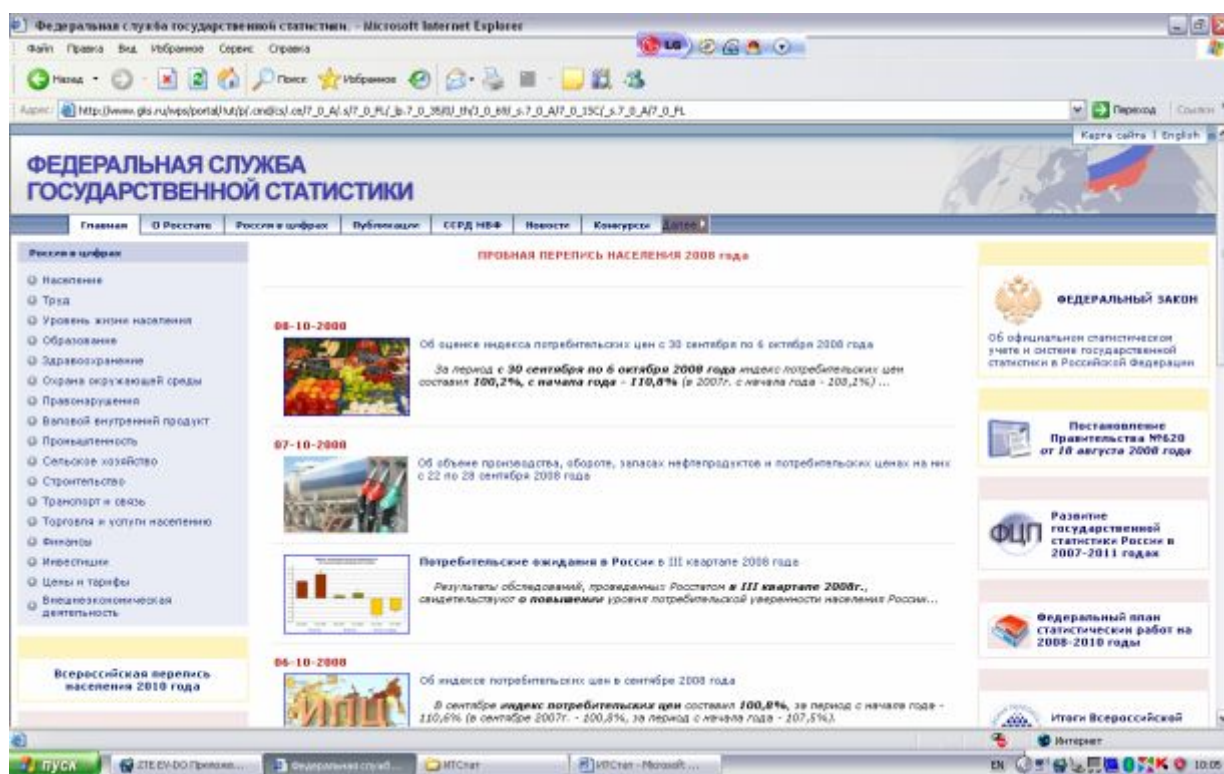


Рис. 5.1. Вид главной страницы официального сайта Государственного комитета статистики Российской Федерации

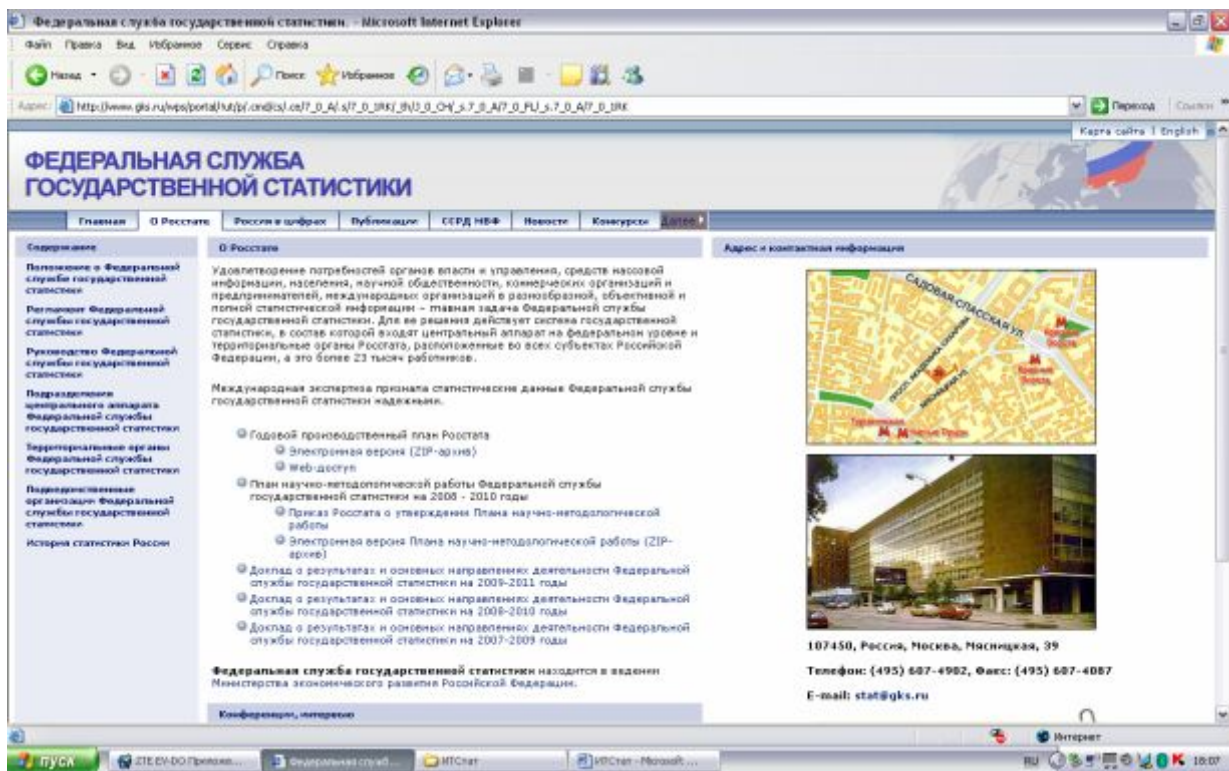


Рис. 5.2. Вид страницы О Росстате официального сайта Государственного комитета статистики Российской Федерации

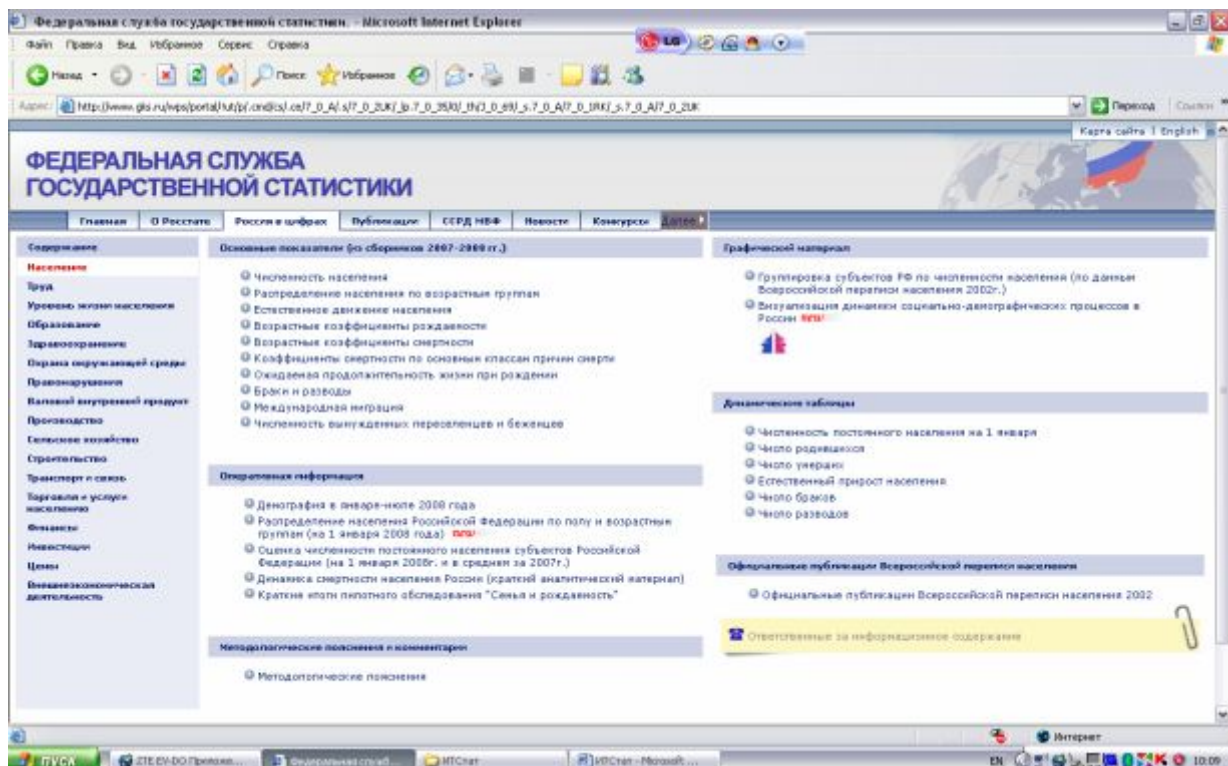


Рис. 5.3. Вид страницы Россия в цифрах официального сайта Государственного комитета статистики Российской Федерации

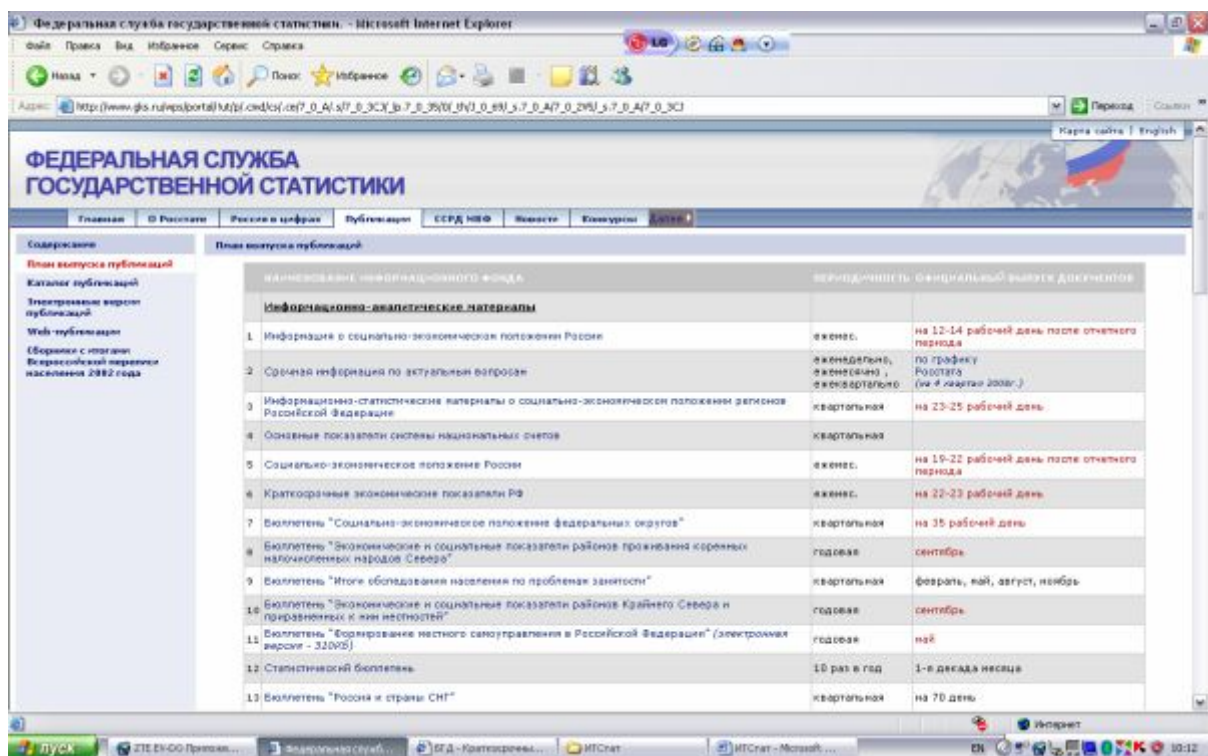


Рис. 5.4. Вид страницы Публикации официального сайта Государственного комитета статистики Российской Федерации

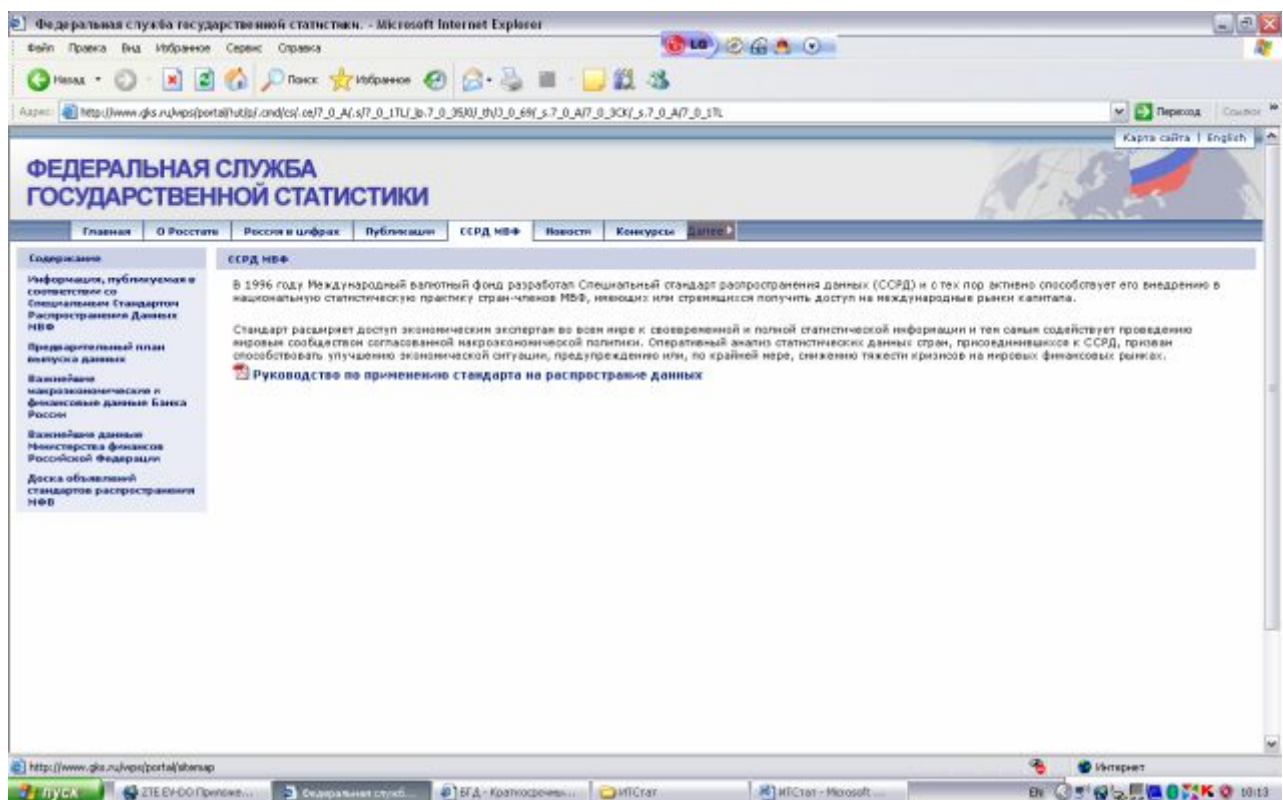


Рис. 5.5. Вид страницы ССРД МВФ официального сайта Государственного комитета статистики Российской Федерации

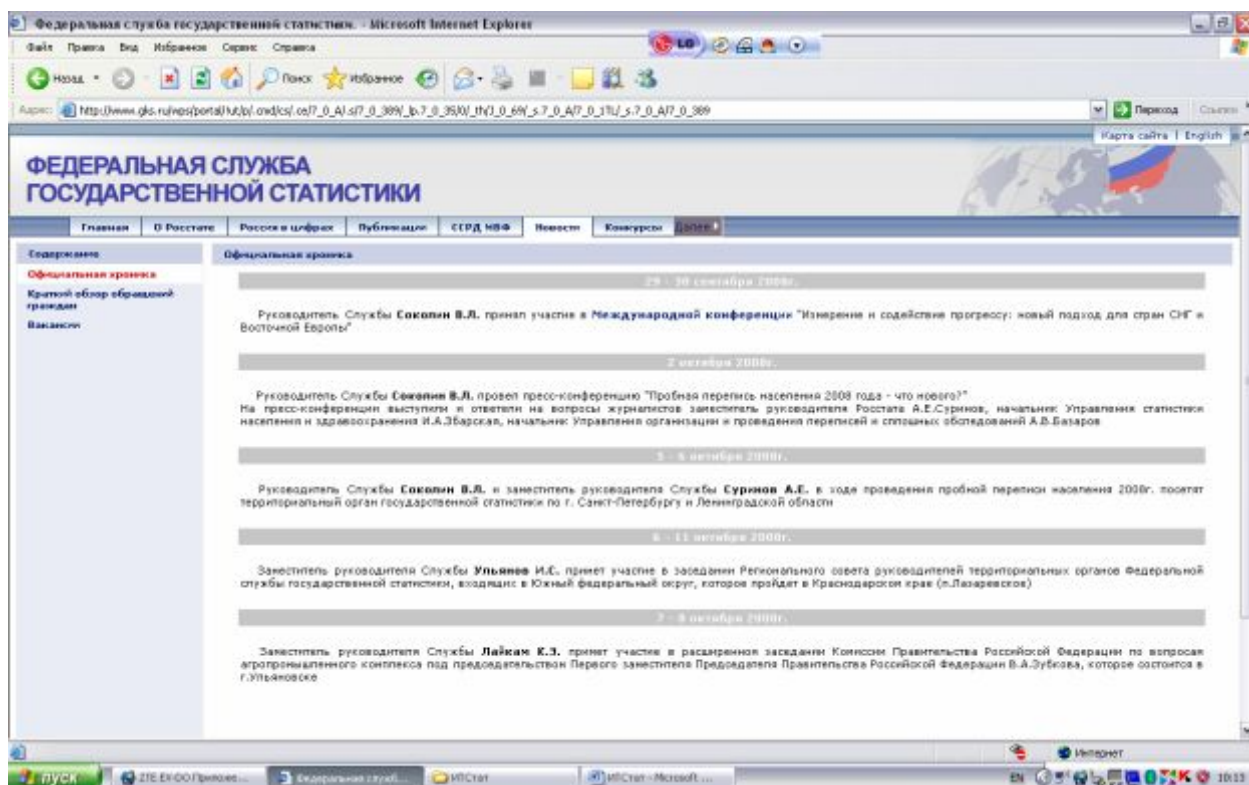


Рис. 5.6. Вид страницы новостей официального сайта Государственного комитета статистики Российской Федерации

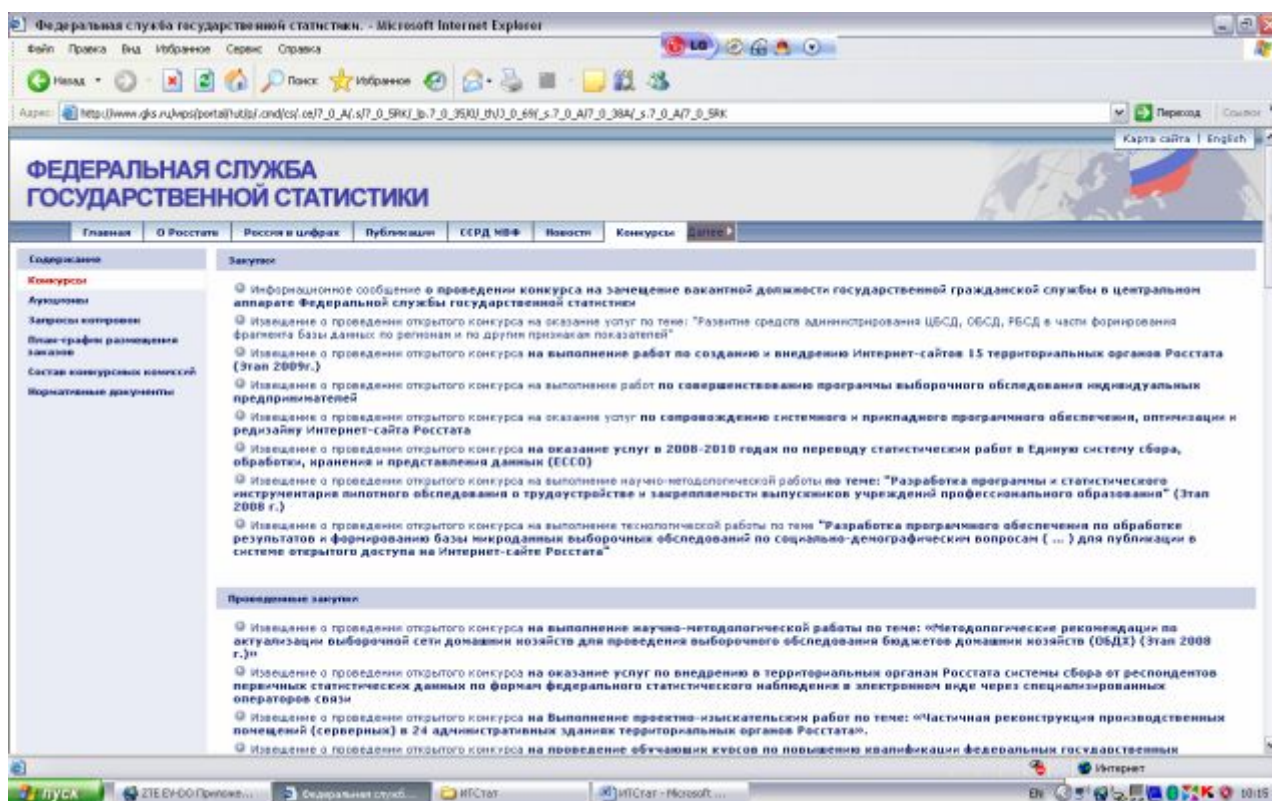


Рис. 5.7. Вид страницы Конкурсы официального сайта Государственного комитета статистики Российской Федерации

Задание 5.1.

С помощью ресурсов страницы *Главная* официального сайта Государственного комитета статистики определить:

- код ОКВЭД для обучения в образовательных учреждениях высшего профессионального образования (университетах, академиях, институтах и в др.);
- код ОКВЭД для среднего (полного) общего образования;
- дату и данные постановления введения ОКВЭД.

Задание 5.2

По материалам страницы *О Росстате* ответить на вопросы:

1. Какие подразделения имеет Государственный комитет статистики Российской Федерации?
2. Каковы реквизиты территориального органа Федеральной службы государственной статистики по Челябинской области?
3. Каковы основные правила организации деятельности территориальных органов Росстата?

Задание 5.3

На странице *Россия в цифрах* найти сведения и выполнить задания с найденными данными:

1. Построить диаграмму численности населения на 1 января различных лет.
2. Определить динамику количества террористических актов на территории Российской Федерации.
3. Построить диаграмму индексов металлургического производства и производства готовых металлических изделий.
4. Определить продукты с наибольшей долей потребления в ближайший временной период.
5. Построить диаграмму количества браков с 1997 по 2007 год по Российской Федерации и Челябинской области.

Задание 5.4

Используя страницы официального сайта Государственного комитета статистики:

1. Указать основные положения статистического обзора по Российской Федерации.
2. Определить состав срочной информации по актуальным вопросам.
3. Ознакомиться с содержанием руководства по применению стандартов на распределение данных.
4. Определить основные новости на текущую дату.
5. Определить конкурсы: на разработку средств обработки, по разработке методологии, по кадровому составу.