

С.Н. МАРТЫШЕНКО, Н.С. МАРТЫШЕНКО, Д.А. КУСТОВ

**СОВЕРШЕНСТВОВАНИЕ МАТЕМАТИЧЕСКОГО И
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ОБРАБОТКИ ПЕРВИЧНЫХ
ДАННЫХ В ЭКОНОМИЧЕСКИХ И СОЦИОЛОГИЧЕСКИХ
ИССЛЕДОВАНИЯХ**

Излагаются принципы и структура, разработанного комплекса программных средств анализа анкетных данных, встраиваемого в EXCEL . Программный комплекс включает средства анализа, не входящие в распространенные пакеты статистической обработки информации.

Любое маркетинговое или социологическое исследование неотрывно связано с исходной целью проведения исследования. В большинстве случаев, такой целью является создание информационной базы, используемой для обоснования управленческих решений.

Как маркетинговые, так и социологические исследования включают два этапа: сбор первичной информации и анализ информации. В этих исследованиях опрос является основным методом получения информации. Он используется более чем в 90% исследований [7].

В нашей работе мы ориентировались в большей степени на экономические системы и соответственно основной областью приложения своей работы считаем не социологические, а маркетинговые исследования. Поскольку с методологической точки зрения эти два вида исследований имеют много общего и часто очень сильно пересекаются, как возможную область применения, мы рассматриваем, и социологические исследования. Еще одна из причин, по которой нельзя обойти вниманием социологические исследования, состоит в том, что в социологии в настоящее время гораздо более высокий уровень проработанности данной проблемы. Об этом свидетельствуют многочисленные научные публикации отечественных ученых в этой области [3,5,8,9,10]. Для зарубежных исследований так же характерно, что наиболее

значимые результаты в области маркетинговых исследований были получены именно социологами [7, 11].

В маркетинговых исследованиях применяются несколько видов опросов: опрос экспертных групп, опрос методом фокус-групп, глубинное интервью, анкетирование [7]. Первые три могут рассматриваться, как предварительные этапы, предшествующие анкетному опросу. На предварительных этапах производятся качественные исследования, а результатом анкетного опроса являются количественные данные (табл.1).

Качественные исследования позволяют понять суть ситуации, сложившейся вокруг изучаемой проблемы и сформулировать гипотезы, которые могут быть проверены в ходе анкетного опроса.

Таблица 1

Характеристики качественного и количественного исследования

Элементы исследования	Качественные исследования	Количественные исследования
Цель	Определить качественное понимание скрытых мотивов и побуждений	Представление данных в количественной форме и обобщение результатов анализа данных на всю генеральную совокупность
Выборка	Небольшое количество объектов	Большое количество объектов
Сбор информации	Неструктурированный	Структурированный
Анализ информации	Нестатистический	Статистический
Результат	Начальное представление и формулировка гипотез	Рекомендации для принятия управленческого решения

Данная работа посвящена совершенствованию методов и технологий обработки количественной информации, поэтому далее мы будем рассматривать только системы сбора первичной информации, основанные на проведении анкетных опросов.

В России анкетные опросы пока не приобрели столь массового характера, как в странах с развитой рыночной экономикой. Большая часть анкетных опросов до последнего времени производится в научных целях. Можно

привести несколько причин, по которым использование анкетных опросов в маркетинговых исследованиях у нас не получило большего распространения.

Первая причина – нежелание большинства респондентов идти на контакт и предоставлять информацию. Если в развитых странах население, в большинстве своем считает, что предоставленная ими информация может положительно отразиться на их жизни, то у нас недоверие к любым контактам с незнакомыми людьми преодолеть очень трудно. За последнее десятилетие у большинства людей в нашей стране сложилось устойчивое мнение, что его везде и всегда обманывают. Сохраняющаяся в нашей стране высокая криминальная напряженность тоже не способствует к установлению контактов. По этим причинам наиболее трудно организовать опросы по почте или по телефону. Поэтому в ближайшие годы основным методом опроса будет оставаться личный опрос с заполнением анкетных форм на бумажном носителе. Неизбежно будут развиваться электронные опросы [7].

Вторая причина состоит в том, что маркетинговая деятельность у нас пока не очень развита. Чтобы получить действительно полезную информацию для бизнеса в ходе опросов, необходимо накопить большой опыт работы, который складывается годами. Сегодня многим российским предприятиям не по средствам организовывать собственные маркетинговые исследования, а тем более содержать подразделение, работающее на перспективу. Можно было бы заказывать исследования специализированной фирме, но таких фирм у нас пока тоже очень мало.

В странах с развитой рыночной экономикой, если предприятие не может организовать собственные маркетинговые исследования, оно всегда может обратиться за помощью в многочисленные специализированные фирмы.

Третья причина состоит в том, что большинство маркетологов не владеют в достаточной степени программным обеспечением по обработке статистических данных. Большинство программных средств по обработке статистических данных требует очень высокой математической грамотности. Отечественных программных средств, пригодных для этих целей крайне мало.

Разработка таких программ в последние годы даже уменьшилась. По всей видимости, это связано с тем, что до сих пор у нас не привыкли закупать программное обеспечение. Нелицензированные зарубежные пакеты по статистике легко доступны. Однако, большинство из них, не имеют документации на русском языке. Даже профессионалу, далеко не всегда можно разобраться в обилии параметров, определяемых при вызове программ и выводимых в результате работы программ.

Четвертая причина состоит в отсутствии доступных и понятных широкому кругу маркетологов методик и технологий компьютерной обработки анкетных данных. Методики должны быть рассчитаны не на научных работников, а на практических работников. Практическую работу отличает большое количество повторяющихся рутинных операций.

Пятая причина состоит в том, что данные анкетных опросов имеют свою специфику, которая не всегда укладывается в рамки классической статистики [2, 3]. Отличительной чертой данных является то, что они в большинстве своем носят нечисловой характер, либо являются разнотипными данными. Наиболее распространенные пакеты по статистике в основном рассчитаны на данные, измеренные в одной относительной шкале.

Шестая причина, которая отмечается известным специалистом в области анализа анкетных данных [10], состоит в том, что у нас крайне мало литературы по данному вопросу, хотя на западе этой проблеме уделяется несоизмеримо большее внимание. Тем не менее, тот же автор отмечает, что многие методы разбросаны в разных источниках и проблема не может быть решена только переводом зарубежных источников.

Теория обработки нечисловых данных находится в стадии развития. В последние годы эта проблема привлекает внимание ученых статистиков, но проводимые исследования больше носят научный, чем практический характер [3, 9, 10].

Данная работа имеет целью обеспечить широкий круг практиков дополнительными средствами анализа анкетных данных, учитывающими спе-

цифику решаемых задач. Разработанный комплекс компьютерных программ это не просто набор независимых функций обработки данных, а инструментальное средство проведения системных исследований. То есть разрабатываемые средства предназначаются не столько для тех, кто использует анкетный опрос для решения частной задачи, а для тех, кто производит или собирается проводить анкетные опросы на профессиональной основе.

Профессиональное использование анкетных опросов предполагает проведение ни одного опроса, а множества. Один опрос является только базой для дальнейших исследований. Качественные результаты можно получить только при накоплении достаточного собственного опыта. И хотя в литературе можно найти множество рекомендаций по проведению опросов [7, 1], без собственного опыта не избежать ошибок. Слишком велик спектр возможных источников ошибок. На рис. 2 представлены компоненты общей ошибки анкетного опроса. Подробный анализ ошибок можно найти в работе Нэреша К. Малхотры [7].

Предлагаемая в работе система анализа данных разрабатывалась для обслуживания крупномасштабных исследований, когда в опросах участвуют тысячи респондентов, анкеты могут включать сто, и более вопросов. Для обработки таких анкет даже простые операции могут превратиться в проблему. В этой ситуации особенно важна технологичность выполнения любой операции. Особые требования выдвигаются к эффективности работы программного обеспечения. То есть выдвигаются требования не просто получить результат, а получить его в рамках реального времени.

Характерной особенностью масштабных исследований является то, что для сбора данных необходимо привлекать большое количество интервьюеров. Это, как правило, штатные сотрудники, часто неквалифицированные, не всегда ответственно относящиеся к порученной им работе. Как у нас в стране, так и за рубежом к этой работе чаще всего привлекаются студенты. Между тем, многие исследователи отмечают необходимость учета личности интервьюера [6, 10]. Зарубежные исследователи, не смотря на удорожание

услуг по сбору данных, постоянно повышают уровень требований к лицам, производящим опросы.

В этих условиях особое внимание должно быть уделено анализу достоверности информации, предоставляемой каждым отдельным интервьюером. Для решения этой задачи необходимы специальные методы и программные средства. Такие средства ни в один известный пакет по обработке данных не входят.

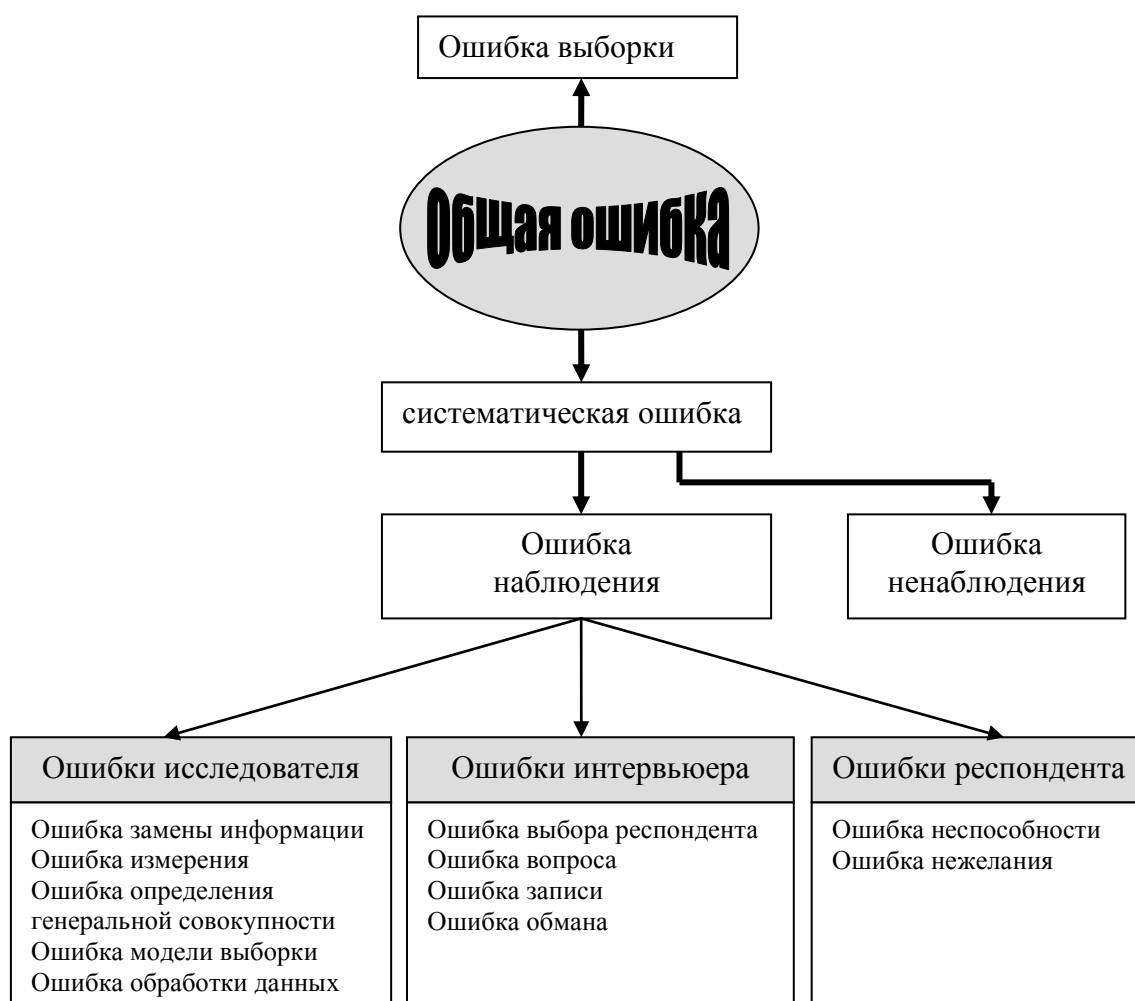


Рис. 2. Источники ошибок при проведении анкетного опроса

Обработка данных может занимать относительно длительное время (месяц и более). Огромное количество результатов обработки требует систематизированного хранения иначе в них можно просто запутаться.

Маркетинговые исследования на основе таких анкетных опросов необходимо рассматривать, как развивающуюся систему накопления знаний. По-

этому организация всей работы должна быть построена с соблюдением принципов системного подхода.

В этом случае систему маркетинговых исследований удобно рассматривать, как совокупность двух подсистем: подсистемы сбора первичного данных и подсистемы анализа данных.

Анкета – один из основных элементов системы сбора первичных данных, в которую входят кроме методов составления анкет, методы организации опросов, методы стимулирования респондентов и интервьюеров, методы кодирования и компьютерного представления данных.

Повышение эффективности обработки анкетных данных может быть достигнуто за счет использования новой компьютерной технологии обработки данных. Работоспособность технологии во многом определяется тем, насколько в ней выдержаны основные принципы системного подхода. Рассмотрим основные элементы компьютерной технологии с точки зрения удовлетворения этим принципам.

Любая компьютерная технология разрабатывается с целью повышения эффективности решения определенного класса задач. Совокупность методов и алгоритмов обработки информации еще не дают компьютерной технологии. Большое значение при создании компьютерной технологии имеет проблема согласования элементов технологии по способу представления и передачи информации при переходе от одного этапа обработки к другому.

Рассмотрение компьютерной технологии анализа анкетных данных начнем с рассмотрения укрупненных блоков (рис. 3). Анализ взаимосвязи укрупненных блоков позволяет определить логику построения или применения каждого из них.

Система задач, решаемых по данным анкетных опросов, определяет систему сбора данных. Совместно они определяют требования к средствам обработки данных. Но и средства обработки оказывают свое влияние на систему сбора.

Основным элементом новой компьютерной технологии является комплекс программных средств, реализующий новые методы обработки специфических данных, каковыми являются данные анкетных опросов. Специализированный комплекс программных средств не решает все задачи, которые могут возникнуть при анализе анкетных данных, а только обеспечивает новые дополнительные возможности при решении ряда специфических задач.

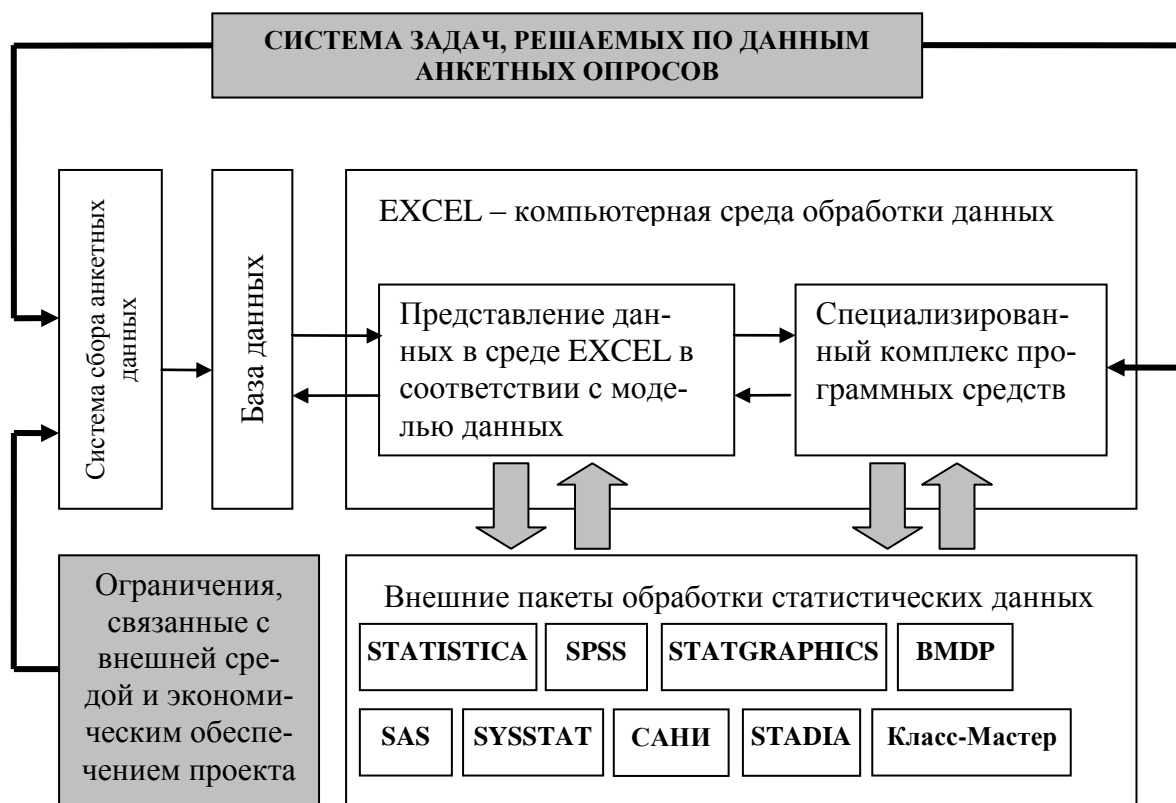


Рис. 3. Компьютерной технологии анализа анкетных данных

Новый комплекс программных средств создается на базе существующего универсального программного средства, используемого для анализа данных. В качестве такой среды, для встраиваемого комплекса, в данной работе была выбрана программная среда EXCEL. Выбор EXCEL обусловлен, в первую очередь тем, что этот пакет нашел самое широкое распространение среди пользователей. Кроме того, EXCEL постоянно пополняется новыми статистическими функциями и обеспечен доступной документацией на рус-

ском языке. Владение EXCEL позволяет пользователю быстрее освоить работу с новой надстройкой.

Однако, в EXCEL представлены далеко не все методы анализа статистических данных. Более полно средства статистического анализа представлены в специализированных пакетах статистической обработки данных, таких как SPSS, STATISTICA, STATGRAPHICS, STADIA и других. Среди этих пакетов очень трудно отдать предпочтение тому или иному пакету. Использование статистических пакетов позволяет решать многие дополнительные задачи, которые сегодня невозможно решить в EXCEL. Но и эти пакеты, в силу своей универсальности, не позволяют эффективно решать ряд специфических задач, возникающих при обработке анкетных данных.

Специализированный комплекс обеспечивает не только решение новых задач, но и повышает эффективность обработки данных с помощью программных средств, представленных в EXCEL и облегчает взаимодействие с внешними пакетами статистического анализа.

Специализированный программный комплекс нельзя рассматривать только, как набор некоторых дополнительных функций или инструментальных средств анализа данных в EXCEL. Эти средств не только увязаны со средой, но и взаимосвязаны между собой некоторой общей идеологией.

Ряд средств специализированного комплекса, предполагает соблюдение некоторых дополнительных условий представления данных, непредусмотренных в EXCEL. Эти условия связаны с такими понятиями, как “проект анкетного опроса” и “модель данных”. Понятие проекта и модели данных в явной или неявной форме присущи любому специализированному пакету программных средств.

Подчинение данных некоторым дополнительным условиям позволяет существенно упростить обращение к программным средствам специализированного комплекса. Таким образом, если соблюдаются определенные правила в представлении данных, их не надо описывать при каждом обращении к программам, входящим в комплекс. С другой стороны, количество условий

должно быть не очень велико и они должны быть достаточно простыми и понятными, иначе будет затруднено, как пользование программами, так и согласование с программами среды и внешними пакетами.

Уточним смысл, вкладываемый в понятия проект и модель данных анкетного опроса при разработке специализированного программного комплекса.

Сбор анкетных данных, а затем обработка могут производиться достаточно длительный срок – от полугода до года, а при повторении опросов и несколько лет. Обработка данных проходит множество этапов. Исходные данные и результаты, полученные на различных этапах, хранятся на нескольких страницах в одном файле EXCEL. Такой файл ассоциируется с проектом по анализу данных конкретного анкетного опроса. Одновременно один исследователь может сопровождать множество анкетных опросов и соответственно разрабатывать множество проектов.

В процессе обработки в данных могут происходить изменения (корректировка, пополнение выборки), что может привести к потере актуальности некоторых результатов (таблиц, признаков классификации). Часть результатов при изменениях в данных пересчитывается автоматически. Однако, отдельные результаты требуют пересчета (актуализации). Например, результаты, полученные с помощью внешних программ. Поэтому исследователю, часто необходимо знать в какой последовательности и когда он производил те, или иные действия. Чтобы исследователю самому не производить записи по сопровождению проекта, будет лучше, если большинство из них возьмет на себя компьютерная программа, которая будет выполнять эту функцию, представляя результаты по сопровождению проекта в систематизированном виде.

При планировании новых проектов исследователю могут быть очень полезны статистические данные о ходе выполнения предшествующих проектов. Анализируя эту информацию, он может избежать многих ошибок. Процесс сбора статистических данных о ходе выполнения проекта в большой степени может быть автоматизирован.

В процессе работы над проектом, исследователю часто приходится разрабатывать некоторые новые инструментальные средства анализа данных, учитывающие специфику данных конкретного проекта. Сопровождение таких средств также может выполнять программа.

Все функции по сопровождению проекта берет на себя специальная программа – мастер сопровождения проектов.

В случае, если исследователь работает с небольшими проектами, разрабатываемыми в сжатые сроки, ему совершенно нет необходимости использовать все функции мастера. То есть, мастер предоставляет дополнительные возможности по обслуживанию проекта, но пользователь может их и не использовать.

Автоматическое сопровождение проекта основано на определенной структуре проекта. Под структурой проекта понимается создание определенного перечня элементов проекта, оформленных в виде отдельных листов EXCEL. Элементы проекта заполняются информацией по мере развития проекта. Они содержат информацию, которая может понадобиться исследователю в ходе работы над проектом. Структура проекта представлена на рис. 4.

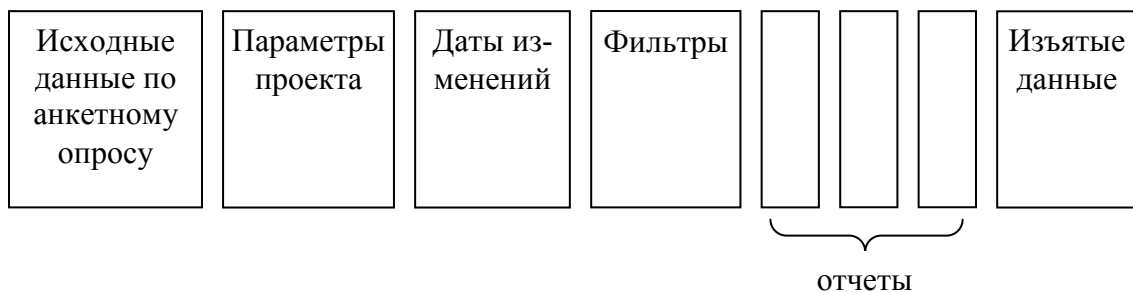


Рис. 4. Основные элементы структуры проекта

Большинство параметров проекта являются необязательными. Однако без установки обязательных параметров программный комплекс работать не будет. Некоторые параметры необходимы только для использования определенных функций программного комплекса. Рассмотрим параметры и содержание элементов проекта.

На первом листе проекта должны быть размещены исходные данные анкетного опроса, оформленные в соответствии с моделью данных (модель данных будет рассмотрена ниже). После ввода или экспорта исходных данных из внешней базы данных, пользователь должен создать проект (функция СП). Программа запросит определить имя проекта и указать лист EXCEL, на котором размещены исходные данные. Это обязательные параметры (обязательные параметры выделяются красным цветом).

После инициализации проекта программа создаст следующие три элемента проекта и внесет в них начальные сведения о проекте. Рассмотрим элемент проекта “Параметры проекта”.

Лист “Параметры проекта” имеет пять разделов. Первый раздел “Заголовки проекта” расположен в верхней части листа. В нем содержатся следующие данные: название проекта, число признаков в таблице данных анкетного опроса, количество анкет в выборке, дата создания проекта, дата последней корректировки данных.

Следующий раздел содержит характеристики данных (табл. 2.). При создании проекта, в таблице автоматически заполняются первый, второй и четвертый столбец. В пятом столбце в исходном состоянии для всех строк установлен знак пропуска “-”. Остальную информацию должен ввести пользователь.

Таблица 2

Характеристики данных

Номер	Название признака	Шкала Измерения	Представление	Признак отсутствия данных	Частота отсутствия данных	Зависимости
1	Признак 1	Номинальная	числовое	“-1”	100	
2	Признак 2	Номинальная	Текстовое	“н/д”	300	
3	Признак 3	Номинальная	Числовое	–	–	
4	Признак 4	Относительная	Числовое	“-1”	–	

Признак отсутствия данных формализован в рамках пакета вполне определенным образом. **А именно, для данных в числовом представлении**

признаком отсутствия данных служит значение “-1”. Уникальность значения “-1” основана на том, что в анкетах числовые данные принимают только положительные значения. Для текстовых данных признаком отсутствия данных служит текст “н/д”. Для тех признаков, для которых невозможно определить признак отсутствия данных в графе признак отсутствия данных проставляется символ “-”.

Если признак отсутствия данных определен, то частота встречаемости значения “отсутствие данных” рассчитывается автоматически. Столбец зависимости не формализован и заполняется произвольно. В нем полезно перечислить признаки, предопределяющие некоторые значения данного признака. Кроме указанных, данная таблица содержит даты последних корректировок признаков.

Третий раздел листа “Параметры проекта” содержит список зарегистрированных отчетов (табл. 3), содержащих результаты анализа данных. Результаты анализа данных, как правило, имеют форму таблиц или графиков отображающих данные, представленные в этих таблицах. Проект может содержать множество таблиц с результатами. Однако, регистрировать все таблицы совершенно необязательно. Рекомендуется регистрировать только те таблицы, которые предполагается включить в сводный отчет по проекту. Остальные таблицы можно считать промежуточными результатами.

Таблица 3

Список, зарегистрированных отчетов

Номер	Название отчета	Дата создания	Дата корректировки	Размещение	Актуализация	Использованные признаки
1	Отчет 1					
2	Отчет 2					
3	Отчет 3					
4	Отчет 1					

При регистрации таблицы в списке отчетов вводится его имя, используемые признаки и выделяется блок размещения данных таблицы. Даты создания и корректировки определяются системой автоматически. Поле актуа-

лизация определяется пользователем в формате “требуется”, “не требуется”. Это поле указывает на то, могут или не могут автоматически обновляться результаты таблицы при изменении данных.

Четвертый раздел листа “Параметры проекта” содержит список зарегистрированных вариантов классификации. Приемлемый вариант классификации или группировки объектов выборки может быть получен исследователем в результате длительной работы, возможно с использованием внешних программ.

Классификацию или типизацию данных удобно представлять, как дополнительный признак. Этот признак может содержать числовое значение – номер класса или текстовое имя, общее для всех объектов одного класса. Если количество объектов выборки возрастет, то процесс классификации должен быть произведен повторно. Повторять классификацию необходимо и при корректировке данных. Однако, не все корректировки данных могут потребовать актуализации классификации или расчетных таблиц. Пересчет необходим только в том случае, если в расчетах использовались признаки, в данных которых произошли изменения.

Необходимость контроля актуальности классификаций возникает в случае, когда проект содержит множество различных вариантов классификации. Регистрация признака классификации осуществляется аналогично регистрации расчетных таблиц. Совместно с названием классификации хранится дата регистрации (создания классификации), место размещения, состав признаков, используемых при классификации, средство классификации (в форме комментария).

Если проект не очень масштабный пользователь может не регистрировать ни отчеты, ни классификации.

Следующим (третьим) элементом проекта (рис. 3) является лист EXCEL, содержащий даты внесения изменений в таблицу основных данных. Размерность таблицы дат изменений совпадает с размерностью таблицы исходных данных. При завершении каждого сеанса работы с проектом проис-

ходит обновление дат корректировки. Данные по датам изменения данных могут быть очень полезны для расчета статистики по ходу выполнения проекта.

Четвертый элемент проекта – “Фильтры”. На этой странице хранятся все логические фильтры, связанные с проектом. Логические фильтры позволяют существенно повысить достоверность данных. На разработку таких фильтров может потребоваться достаточно большое время. Автоматизация разработки и обслуживания логических фильтров осуществляется с помощью специальной программы, входящей в состав специализированного комплекса.

Листы EXCEL, содержащие результаты обработки данных составляют пятый элемент проекта – “Отчеты”.

Шестой элемент проекта – “Изъятые данные”, содержит анкеты, которые были по тем или иным причинам изъяты из выборки.

После рассмотрения структуры проекта, определим понятие модели данных. Модель данных – это совокупность правил, регламентирующих представление основных данных на листе проекта “Исходные данные”.

Модель данных предполагает, что данные анкет располагаются на листе по строкам. Первая строка содержит названия признаков (или вопросы анкеты). Первый столбец таблицы данных содержит номера анкет в опросе.

Целесообразно (но не обязательно) последние два столбца таблицы данных отводить под дату опроса и фамилию интервьюера.

Таблица данных не должна содержать никаких дополнительных строк, связанных с обработкой данных, кроме основных данных (например, итогов и т.п.). На листе с данными кроме данных анкет могут быть представлены дополнительные столбцы с расчетами и классификациями. Такие столбцы могут быть созданы только после инициализации параметров проекта.

Данная модель данных полностью согласуется с моделями данных, принятыми в большинстве статистических пакетов, что облегчает экспорт и импорт данных во внешнюю программную среду.

Понятие проекта анкетного опроса и модели данных являются основой разрабатываемой компьютерной технологии обработки анкетных данных. Расширение возможностей исследователя по анализу данных достигается за счет использования компьютерных программ, реализующих конкретные алгоритмы обработки данных.

Алгоритмы нацелены на решение проблем, возникающих при анализе анкетных данных. Разрешение таких проблем, либо невозможно с использованием известных программных средств, либо крайне затруднительно.

Эффективность обработки анкетных данных повышается не только за счет использования отдельных модулей, но и за счет совместного их использования. Программный комплекс охватывает все основные этапы обработки данных – от подготовительного этапа до визуального представления результатов.

Алгоритмы и программы комплекса представлены четырьмя разделами (рис. 4.). Каждый из них связан с одной из проблем, возникающих в практической работе с анкетными данными.

Первая проблема – это отсутствие ответов на конкретные вопросы анкеты (или отсутствие данных). Эта проблема обсуждается во многих работах отечественных и зарубежных авторов [6, 10]. Для успешного разрешения этой проблемы необходимо понять причину, которая могла вызвать отсутствие ответа, затем разрабатывать методы устранения ошибок.

Методы восстановления пропущенных данных обсуждаются так же в работах [2, 7].

Следующая проблема, которая может повлиять на достоверность статистического вывода, является проблема засоренности данных. Засоренность данных проявляется в наличии явных выбросов и логически непоследовательных ответов.

Методы разрешения двух указанных проблем (отсутствие данных и засоренность) известны, как методы проверки состоятельности данных [7]. На схеме эти методы представлены блоком “Алгоритмы повышения достовер-

ности анкетных данных”. Этот блок представлен двумя типами алгоритмов: статистическими и логическими.

Хотя, обе выделенные проблемы, в литературе отмечаются, как важнейшие, их решение не нашло отражения в современных пакетах анализа статистических данных. Какие либо методики, охватывающие проблему в целом, тоже отсутствуют. Между тем, решения этих проблем не являются тривиальными. Во-первых, речь идет о многомерных выборках, во-вторых, признаки могут быть представлены в различных шкалах. Сложность их решения возрастает многократно при большом количестве вопросов анкеты и массовых опросах. В этом случае необходимо не просто разработать алгоритмы и программы, а и создать специальные технологии решения задач.



Рис. 4. Основные разделы специализированного программного комплекса

Следующая проблема, которая стоит перед исследователем анкетных данных, это создание типологий и классификации многомерных данных. Ес-

ли модули по кластерному анализу присутствуют в специализированных статистических пакетах, то алгоритмы построения типологий вообще в пакетах не представлены. Однако, и представленные в пакетах программы классификации, имеют весьма ограниченное применение для обработки анкетных данных, поскольку они работают с числовыми признаками, измеренными в шкале отношений. В реальных анкетных данных такие признаки встречаются достаточно редко. Чаще всего, пространство признаков представлено разнообразными шкалами. В этом случае может оказаться весьма полезным непараметрический алгоритм классификации, рассмотренный и реализованный в данном комплексе.

Вспомогательные средства анализа данных имеют исключительно важное значение при решении практических задач. Неслучайно большинство средств анализа, реализованных в статистических пакетах, относятся к этой категории. Различные формы представления данных весьма важны для формирования содержательных гипотез о структуре и взаимосвязи данных.

На схеме рис.4 вспомогательные средства представлены двумя блоками: “Процедуры преобразования данных” и “Сервисные и визуальные средства анализа данных”. Ряд процедур преобразования данных можно найти в пакетах по обработке данных. Наш программный комплекс содержит ряд нестандартных преобразований. Те же, которые можно найти в пакетах, у нас имеют расширенные функции и обеспечивают расчет дополнительных характеристик. Кроме того, эффективность их работы в рамках проекта выше, чем стандартных средств.

Сервисные и визуальные средства представленные в комплексе программ являются проблемно-ориентированными и предоставляют исследователю дополнительные возможности анализа специфических данных анкет.

Кроме рассмотренных выше разделов, комплекс программ включает блок моделирования многомерных выборок. Этот блок полезен при исследовании возможностей собственных программ и тестирования внешних программ. Большинство статистических пакетов содержат генераторы данных.

Однако все генераторы рассчитаны на моделирование только одного признака и поэтому для тестирования программ многомерного анализа не пригодны.

Рассмотренная структурная схема, является основой новой компьютерной технологии обработки анкетных данных. Компьютерная технология призвана расширить спектр решаемых задач или повысить эффективность решения отдельных задач по сравнению со стандартными средствами.

Литература

1. Айвазян С.А., Бухштабер В.М., Енюков И.С. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 422 с.
2. Адамов С.Ю. Система анализа нечисловой информации “САНИ” // Социология: 4М (методология, методика, математическое моделирование). 1991. № 2. С.86-104.
3. Анализ нечисловой информации в социологических исследованиях /Под ред. В.Г. Андреевкова, А.И. Орлова, Ю.Н. Толстой. – М.: Наука, 1985. – 220 с.
4. Анфилатов В.С. Системный анализ в управлении. – М.:Новый век, 2003. – 368 с.
5. Березин И.С. Проведение массовых опросов //Маркетинг и маркетинговые исследования в России. 1996. №5.
6. Ключина Н.А. Причины, вызывающие отказ от ответа // Социол. исслед. 1990. №1. с. 98-105.
7. Малхотра Нэреш К. Маркетинговые исследования. – М.: Вильямс, 2002. – 960с.
8. Мартышенко Н.С. Методическое обеспечение анализа поведения потребителей на региональном туристском рынке // Вестник Тихоокеанского государственного экономического университета. – 2005. - №4. С. 19-31.

9. Орлов А.И. Нечисловая статистика / А.И.Орлов. – М.: МЗ-Пресс, 2004. – 513 с.
10. Толстова Ю.Н. Анализ социологических данных. Методология, дескриптивная статистика, изучение связей между номинальными признаками. – М.: Научный мир, 2000. - 352с.
11. Черчилль. Маркетинговые исследования. – СПб.: Питер, 2002. – 752с.