

## **Анализ проблем, возникающих при использовании данных анкетных опросов для исследования социально-экономических процессов**

**С.Н. Мартышеко, Н.С. Мартышенко, Д.А. Кустов**

Для повышения уровня обоснованности управленческих решений на всех уровнях экономики требуется качественная и достоверная информация. Одним из ее основных источников в экономических и социологических исследованиях являются данные анкетных опросов.

Если для нас область исследования мотивации потребления на основе анкетного опроса является молодым направлением, то в странах с развитой рыночной экономикой это научное направление уже давно перешло в ранг классического знания и широко представлено работами Ф. Котлера, Р. Блэкуэлла, Д. Хокинса, Н. Малхотры, Дж. О'Шонесси и др. Среди отечественных ученых, специализирующихся в маркетинге и уделявшим внимание вопросам сбора и обработки первичных данных, можно назвать А.В. Алешина, Г.Л. Багиев, И.С. Белявский, Е.П. Голубков, Л.А. Козырев, С.Г. Светуных. Однако большинство работ в этой области носит концептуальный теоретический характер, а не методический.

В нашей стране вопросами изучения социально – экономических явлений методами анкетного опроса все-таки больше занимались ученые в области социологии (И.С. Березин, Ю.Н. Толстова, В.А. Ядов, Г.Г. Татарова, Г.И. Саганенко и др.), а не маркетологи.

В большинстве случаев анализ анкетных данных ограничивается лишь анализом частотных рядов одномерных признаков. То есть, большой объем полезной информации, содержащейся в данных, фактически не используется. Для выделения такой информации требуется использование сложных методов обработки многомерных данных.

Однако многие методы многомерного анализа ориентированы на обработку некоторых идеализированных выборок и не учитывают особенностей данных, полученных в ходе анкетного опроса. Соответственно и программы средства, реализующие известные многомерные методы обработки данных чаще всего не могут быть применимы к анкетным данным в их исходном виде, то есть без предварительной обработки.

Для того, чтобы выработать направления разработки методов анализа анкетных данных необходимо выделить особенности данных и проблемы, с которыми приходится сталкиваться в реальной ситуации.

Одной из главных особенностей анкетных данных является то, что они включают признаки различной природы. Многообразие типов данных обусловлено разнообразием шкал, используемых в анкетах для измерения характеристик объектов исследования. Наличие разнообразных шкал вызвано не прихотью исследователей, а их стремлениями получить от респондентов более достоверную информацию. Поскольку не респондент, а исследователь заинтересован

в получении информации, то ему приходится подстраиваться под респондента, предоставляя последнему вопрос в приемлемой для него форме. Многомерные статистические методы можно использовать только к признакам, представленной в одной шкале. Следовательно, необходимо, во-первых, иметь набор методов для различных типов данных, а во-вторых, средства преобразования данных, приводящие их к единому представлению.

Вторая особенность анкетных данных отличающая их от идеальной выборки состоит в том, что они включают множество источников ошибки. Достоверность данных определяется с одной стороны желанием респондента предоставлять информацию, а с другой – спецификой восприятия вопросов анкеты каждым индивидуумом. Различие восприятия анкеты респондентами проявляется, во-первых, в том, что некоторых вопросы остаются без ответов (проблема отсутствия данных), а во-вторых, могут появиться данные, резко отличающие или противоречащие всем остальным.

Сложность проблемы усугубляется еще и тем, что признаки, полученные по анкетным данным, необходимо рассматривать не по одному, а в совокупности или по группам. То есть имеет место проблема многомерности.

Таким образом, прежде чем переходить к содержательному анализу данных они должны быть очищены от грубых выбросов и ошибок. Для выявления многомерных выбросов признаков, представленных в различных шкалах, необходимо иметь свой набор инструментальных средств. Такие средства должны позволять производить сравнение различных многомерных объектов.

При выявлении некачественных данных у исследователя может быть два пути. Первый – это исключить часть данных из рассмотрения. Второй – попытаться восстановить недостоверные данные, используя информацию многомерной выборки.

При разработке алгоритмов восстановления данных принципиальным является оценка точности восстановления. Если исследователь не может восстановить данные с требуемой точностью, то лучше такие данные вообще не использовать в дальнейшем анализе. Специфика данных состоит в том, что исследователь чаще всего может восполнить отброшенные данные при дополнительном опросе, предприняв меры по устранению возможных причин возникновения ошибок.

Одной из особенностей анкетных данных является то, что эти данные могут содержать качественную неструктурированную информацию. Такая информация возникает при включении в опрос открытых вопросов, которые могут оказаться гораздо информативнее, чем обычные стандартные вопросы.

Реальный анкетный опрос отличает и то, что в нем часто используется очень большие выборки при значительном количестве признаков. Если рассмотреть эту проблему в контексте

очень большого количества задач, решаемых по анкетным данным, то становится очевидным, что отдельные средства обработки данных должны быть объединены единой технологией.

Потребность широкого круга исследователей в результатах анализа данных и наличия методов еще не решает проблемы. Необходимы средства анализа, воплощенные в конкретных компьютерных технологиях. Здесь практика сталкивается с большим дефицитом таких средств или их полным отсутствием.

Если ранее на рынке программных продуктов еще присутствовали некоторые отечественные пакеты, обрабатывающие статистические данные, то теперь они почти сошли со сцены, а новые практически не разрабатываются. Присутствующие на рынке зарубежные пакеты по обработке информации являются далеко не идеальными средством анализа анкетных данных. Они больше приспособлены для применения классических статистических методов анализа к данным числовой природы, а так же требуют некоторых идеализированных данных.

Последние годы авторы занимались разработкой алгоритмов и программных средств, направленных на преодоление указанных проблем. Основные результаты опубликованы в работах авторов [1,2,3].

Актуальность проведенного исследования подтверждается тем, что оно выполнялось в рамках научно-исследовательской работы “Исследование взаимодействия в системе "биологический объект - внешняя среда" на основе моделирования и обработки данных статистики в условиях ограниченности и неопределенности исходной информации” (грант РФФИ – ДВОРАН № 06-05-96017) и научно-исследовательской работы “Построение математических моделей этнических миграций на примере переселения корейцев из районов Центральной Азии на Дальний Восток России в 90-е годы XX века” (грант РФФИ – ДВОРАН № 06-06-96004),

### **Список используемой литературы**

1. Мартышенко С.Н. Совершенствование математического и программного обеспечения обработки первичных данных в экономических и социологических исследованиях / С.Н. Мартышенко, Н.С. Мартышенко, Д.А. Кустов // Вестник ТГЭУ. – 2006. – № 2 – С. 91–103.
2. Мартышенко С.Н. Многомерные статистические методы повышения достоверности маркетинговых данных / С.Н. Мартышенко, Н.С. Мартышенко, Д.А. Кустов // Практический маркетинг –2007.– № 1 (119) – С. 20 – 30.
3. Мартышенко С.Н. Применение алгоритмов многомерной классификации и распознавания образов в решении задач анализа анкетных данных/ С.Н. Мартышенко, Н.С. Мартышенко, Д.А. Кустов // Известия высших учебных заведений. Поволжский регион. Сер. Технические науки – 2006. – № 6 – С. 305-310

Поступила в редакцию: дата, месяц, год