

## ТЕХНОЛОГИИ ПОВЫШЕНИЯ КАЧЕСТВА ДАННЫХ В АНКЕТНОМ ОПРОСЕ

Н.С. Мартышенко, С.Н. Мартышенко,

Владивостокский государственный университет экономики и сервиса

Анкетный опрос – один из важнейших методов сбора информации в маркетинговых исследованиях. Данные, полученные в результате опросов, обладают рядом специфических особенностей, которые необходимо учитывать при разработке технологии их обработки. В работе [1] выделено более десяти особенностей анкетных данных. Здесь мы рассмотрим более подробно одну из особенностей. А именно сосредоточим внимание на механизме возникновения ошибок в процессе опроса. Анкетные данные имеют множество источников ошибок (рис.1). Определения ошибок, представленных на рис.1 можно найти в работе [2].

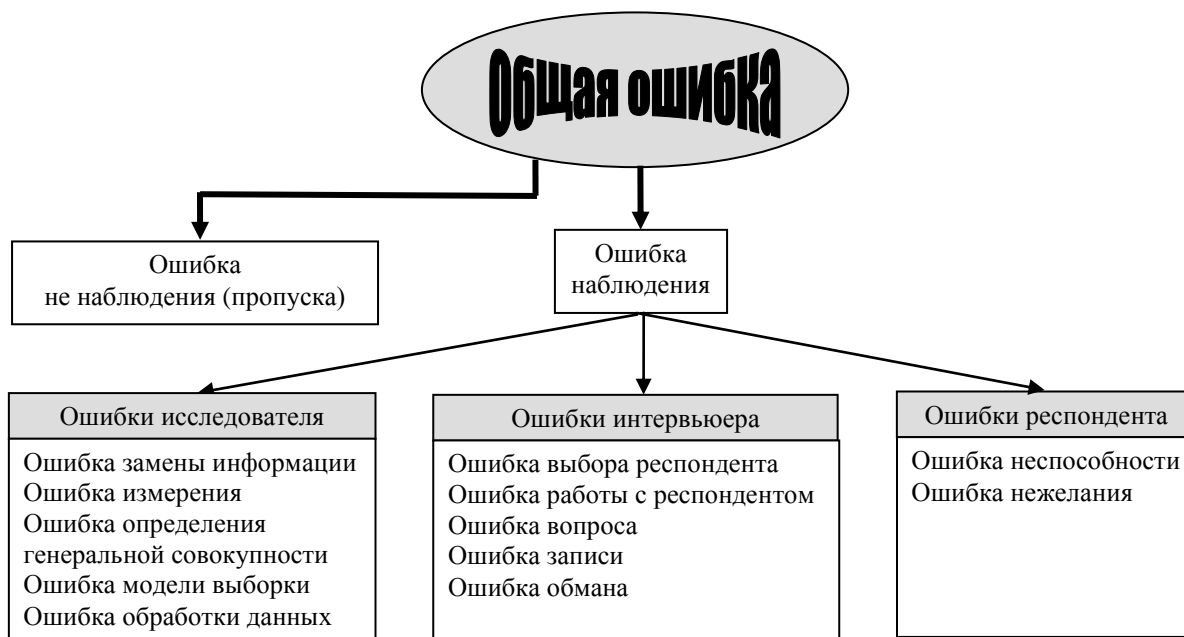


Рис. 1. Источники ошибок при проведении анкетного опроса

Информационные технологии особенно необходимы для исследований, производимых на профессиональной основе. Эту работу отличает то, что исследователь использует в своей работе множество анкет, в опросах участвует большое количество респондентов и интервьюеров, опросы по одним и тем же отработанным анкетам повторяются с некоторой периодичностью (мониторинг меняющейся ситуации). При этом отдельные отработанные модули вопросов могут включаться в несколько анкет.

Построить эффективную технологию обработки данных, обладающих сложной структурой ошибки невозможно без анализа всего процесса формирования данных. Технология должна способствовать снижению уровня ошибок на всех этапах сбора и обработки информации.

Любой анкетный опрос производится с определенной целью. Задачи, которые необходимо решить для достижения цели, в начале формулируются несколько более размыто. Со временем задачи и методы их решения постоянно пересматриваются и уточняются. Конечный результат – выводы, формулируемые в результате обработки данных, является продуктом некоторого возобновляемого процесса. По мере накопления знаний появляются новые задачи, и совершенствуется механизм сбора данных. То есть, в процессе обработки информации накапливается некоторая база знаний. Основные этапы процесса формирования базы знаний представлены на рис. 2.

На первых этапах ошибки, которые допускает респондент (неспособность и нежелание), исследователь относит на качество анкеты и уточняет формулировки и форму представления вопросов. На этом же этапе формируются средства выявления ошибок операторов, осуществляющих ввод данных в компьютер. Эффективность работы операторов повышается при совершенствовании форм ввода данных.

Ключевым звеном в системе сбора данных являются респонденты. Качество их работы зависит от качества анкеты. Но, как показывает опыт, достоверность данных, предоставляемых респондентом, еще в большей степени зависит от качества работы интервьюера.

Привлечение к работе в качестве интервьюеров студентов вузов является достаточно распространенной практикой, а для исследователей, преподающих в вузах, просто нормой. Однако результативность работы таких интервьюеров, как и успеваемость по учебным дисциплинам, может существенно различаться у разных студентов.

Работа с интервьюерами — студентами подчиняется определенному ритму, связанному с учебным процессом. Поэтому у исследователя, организующего опрос, в короткий период появляется очень большое количество анкет. В такой сжатый срок исследователь не в состоянии проанализировать качество первичных данных, содержащихся в анкетах. Чтобы обоснованно судить о качестве данных, информацию, во первых, необходимо ввести в компьютер, затем произвести анализ качества с помощью специальных программ [3,4]. На это требуется достаточно много времени. И к моменту оценки качества данных, студенты представившие информацию выпадают из поля зрения исследователя (обратная связь теряется). Если этот процесс не контролируется, то со временем качество работы интервьюеров-студентов ухудшается.

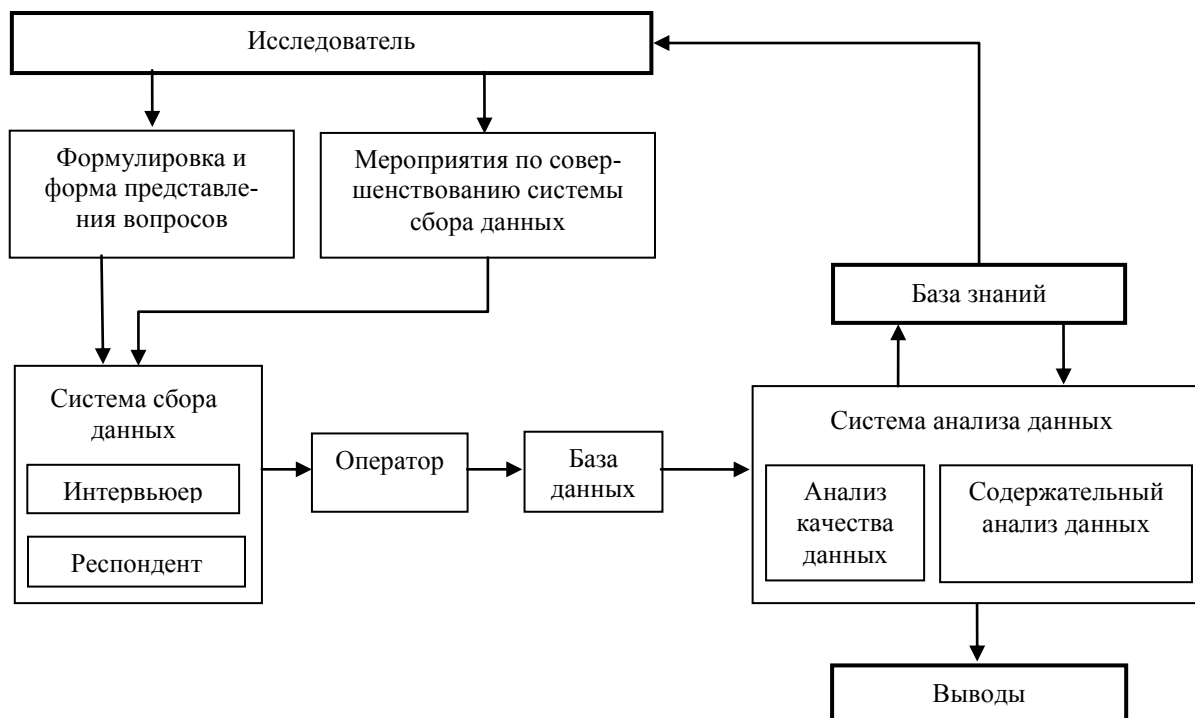


Рис.2. Технологическая схема обработки анкетных данных

Для организации более качественной работы интервьюеров необходимо разрабатывать специальные инструкции и осуществлять контроль за процессом сбора информации. Существенно повысить качество удастся, если студентов обязывают, кроме предоставления анкет на бумажном носителе, предоставлять отчет о произведенной ими работе.

Отчет в компактном виде в форме таблиц — содержит информацию о том где, как и когда собирались данные. Такой подход, во-первых, повышает ответственность работы интервьюеров, во-вторых, обращает внимание на технологию сбора, в-третьих, позволяет исследователю очень быстро оценить качество работы интервьюера. При некачественной работе она может быть не принята или должна быть произведена повторно.

Исследователь должен уделить особое внимание инструктажу интервьюеров перед проведением опроса. Инструктаж включает:

- обсуждение цели опроса с организатором;
- изучение структуры анкеты;
- заполнение пробной анкеты и уточнение у организатора неясных вопросов;
- изучение перечня ошибок, которые чаще всего допускают интервьюеры в работе с респондентами;
- изучение перечня ошибок, которые чаще всего допускают респонденты при заполнении анкет;
- изучение функции интервьюера при проведении опроса;
- изучение возможных схем опроса;
- изучение требований, которым должны отвечать респонденты;
- изучение формы отчетности интервьюера, предоставляемой по завершении опроса.

Наш опыт показал, что качество работы интервьюера повышается, если он в ходе заполнения таблицы по способам сбора данных (табл. 1). В этой таблице определены возможные варианты схем сбора данных.

Таблица 1. Форма таблицы отчета по способу сбора анкетных данных (КАК?)

№ Анкеты	СХЕМЫ СБОРА АНКЕТ					Количество н/д
1						
2						
3						
4						
5						
...	.....	.....	.....	.....	.....	.....
39						
40						
Итого						

При проведении анкетного опроса, возможно использовать несколько наиболее распространенных схем. Рассмотрим варианты схем опроса представленные в таблице 1.

1. Личное интервью с заполнением анкеты (А) респондентом (Р) в присутствии интервьюера (И). Прямой контакт в режиме реального времени (рис. 3).

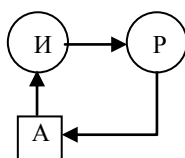


Рис. 3. Прямой контакт в режиме реального времени

2. Сбор данных с заполнением анкеты респондентом без присутствия интервьюера. Интервьюер передает анкету лично респонденту. Респондент заполняет анкету в свободное время и возвращает анкету интервьюеру через некоторый промежуток времени (Т). Прямой контакт с задержкой заполнения (рис. 4).

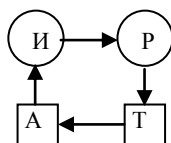


Рис. 4. Прямой контакт с задержкой заполнения

3. Личное интервью с заполнением анкеты интервьюером со слов респондента в его присутствии. Прямой контакт с заполнением анкеты интервьюером в режиме реального времени (рис.5).

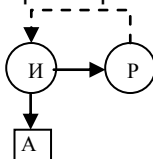


Рис. 5. Прямой контакт с заполнением анкеты интервьюером

4. Сбор данных с заполнением анкеты респондентом без присутствия интервьюера. Респондент передает анкету интервьюеру без личного контакта, через посредника (П). Респондент заполняет анкету в свободное время и возвращает анкету интервьюеру через некоторый промежуток времени (Т) через посредника. Передача анкеты с задержкой заполнения без личного контакта с интервьюером (рис. 6).

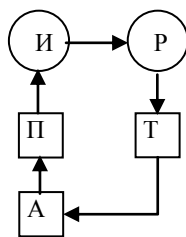


Рис. 6. Передача анкеты с задержкой заполнения без личного контакта с интервьюером

5. Личное интервью по телефону с заполнением анкеты интервьюером со слов респондента в режиме реального времени. Контакт по телефону с заполнением анкеты интервьюером в режиме реального времени (рис.7).

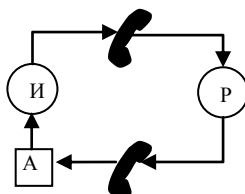


Рис. 7. Контакт по телефону с заполнением анкеты интервьюером в режиме реального времени

6. Интервьюер передает пакет анкет другому лицу – посреднику, имеющему устойчивые контакты с группой потенциальных респондентов. Посредник в своей работе по сбору анкет может использовать одну из вышеперечисленных схем. Сбор анкет без прямого контакта с интервьюером (рис. 8).

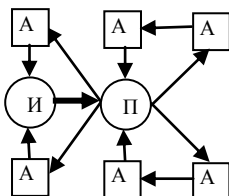


Рис. 8. Сбор анкет без прямого контакта с интервьюером

При заполнении таблицы 1 необходимо придерживаться следующих правил. Для каждой анкеты заполняется строка таблицы. В первом столбце указывается фамилия интервьюера или привлеченного посредника. В одном из следующих четырех столбцов проставляется знак “+” для схемы, использованной при опросе. В последнем столбце указывается количество незаполненных позиций анкеты. Интервьюер должен стремиться к тому, чтобы незаполненных позиций не было. По завершении опроса всех респондентов заполняется последняя строка итогов.

Отчет включает помимо таблицы 1 еще несколько форм. При подготовке отчета студенту становится достаточно сложно заимствовать или генерировать логически увязанную информацию. Для того чтобы умышленно сфальсифицировать результаты, необходимо иметь практический опыт работы по сбору данных. Поэтому студенту, чаще всего, проще собрать достоверную информацию, чем задумываться над тем, как ее исказить.

Разработанная нами технология обработки анкетных данных основывается на принципах системного подхода. Чтобы удовлетворить требованиям системного подхода анкетный опрос связывается с понятием проекта. Проект, помимо данных, включает описание параметров проекта и базу знаний. Проект вынуждает исследователя, соблюдая определенные правила, систематизировать информацию, исключает многие возможные ошибки. Основные элементы проекта представлены в работе [5]. Сама система обработки данных реализована как приложение к EXCEL.

Одним из основных блоков системы является блок анализа качества данных. Методы повышения качества данных неотрывно связаны с понятием грубой ошибки. Этому понятию невозможно дать однозначное формализованное определение. Поэтому попытаемся уточнить его через некоторые его свойства. Грубой ошибкой можно считать многомерное наблюдение, которое резко отличается на фоне всех остальных. Совокупность значений признаков можно считать грубой ошибкой, если они совместно воссоздают абсурдный, с содержательной точки зрения, объект или его поведение. При этом значения одномерных признаков могут быть вполне правдоподобными. Определить грань, за которой наступает абсурдность объекта, может только сам исследователь в процессе содержательного анализа многомерного объекта.

Существует множество причин возникновения и вариантов проявления грубых ошибок. Кроме того, данные, представленные в анкете, имеют различные шкалы измерения. Поэтому для обнаружения ошибок необходимо иметь набор инструментальных средств. В арсенал средств борьбы с ошибками входят алгоритмы, которые работают по принципу многомерных фильтров. Программы позволяют выделить анкеты, которые являются “подозрительными” на содержание грубой ошибки или выброса. Исследователь должен подвергнуть выделенные критические анкеты углубленному содержательному анализу, после чего принять решение о том, как поступить с такими анкетами. Исследователь может оценить ситуацию как допустимую или как недопустимую. В последнем случае он может либо отбросить данные анкеты как недостоверные и только искажающие конечный результат, либо попытаться восстановить отдельные значения признаков по многомерной выборке. В отдельных случаях он может интерпретировать необъяснимое значение как ситуацию отсутствия данных, или пропуск. Отбрасывание небольшой части данных низкого качества никак не сказывается на репрезентативности выборки. Тем более, что при недостатке данных мы можем произвести опрос дополнительной группы респондентов.

Для анализа качества данных большое значение имеет формализация представления пропущенных данных для различных типов данных. Для анализа качества данных мы используем статистические и логические методы [3, 6]. Большинство статистических алгоритмов позволяют производить не только анализ грубых выбросов по отдельным анкетам, но и позволяют производить сравнительный анализ пакетов анкет, представленных различными интервьюерами.

Отличительная особенность разработанной нами системы анализа анкетных данных состоит в том, что она включает большой блок обработки качественной информации. Возможности сбора информации об изучаемом явлении существенно расширены с помощью средств обработки качественных данных. В систему введен ряд новых понятий, связанных с представлением открытых вопросов. Например, введено понятие составного вопроса. На сегодняшний день основными средствами обработки качественной информации, представленными в программном комплексе, являются средства, позволяющие производить типизацию качественных данных [7]. Для типизации используется база знаний в форме специальных словарей замен. Программные средства типизации часто оказываются полезными и при выявлении ошибок в данных.

Разработанная технология обработки анкетных данных и программный комплекс прошли апробацию при обработке нескольких массовых опросов. Только в одном опросе, проводившемся в течение пяти лет, участвовало около шести тысяч респондентов. В процессе апробации была подтверждена высокая эффективность программного обеспечения.

### Литература

1. Мартышенко С.Н, Мартышенко Н.С. Информационные технологии повышения качества данных, полученных при исследовании социально-экономических систем // Надежность и качество: труды Международного симпозиума: в 2-х т. – Пенза, – 2007. – т.1. – С. 40-43.
2. Малхотра Нэреш К. Маркетинговые исследования. – М.: Вильямс, 2002. – 960с.
3. Мартышенко С.Н. Цензурирование при обработке анкетных данных / С.Н. Мартышенко, Н.С. Мартышенко, Д.А. Кустов // Известия высших учебных заведений. Поволжский регион. — 2006. — №6(27). С. 170–177.
4. Мартышенко С.Н. Многомерные статистические методы повышения достоверности маркетинговых данных / С.Н. Мартышенко, Н.С. Мартышенко, Д.А. Кустов // Практический маркетинг – 2007. – № 119(1. 2007)– С. 20-30.
5. 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.
6. Мартышенко Н.С., Власенко А.А. Устранение ошибок в данных маркетинговых исследований с помощью логических фильтров // Наука и образование – 2007. Материалы V Международной научно-практической конференции. Том 7 Экономические науки. — Днепропетровск. 2007. — С. 76-78.
7. Мартышенко С.Н. Средства разработки типологий по данным анкетных опросов в среде EXCEL / С.Н. Мартышенко, Н.С. Мартышенко, Д.А. Кустов // Академический журнал Западной Сибири. – 2007. – № 1 – С. 75–77