

МАРКЕТИНГОВЫЕ ИССЛЕДОВАНИЯ

Подготовка данных к анализу

- 1. Подготовка к анализу и редактирование данных**
- 2. Кодирование**
- 3. Преобразование и очищение данных**
- 4. Статистическая коррективировка данных**
- 5. Выбор стратегии анализа данных и классификация статистических методов**

1. Подготовка к анализу и редактирование данных

Процесс подготовки данных к анализу представлен этапами:

1. Подготовка предварительного плана проведения анализа данных
2. Проверка анкет
3. Редактирование
4. Кодирование
5. Преобразование

6. Очищение данных

7. Статистическая корректировка
данных

8. Выбор стратегии анализа
данных

Проверка анкет заключается в их проверке на полноту заполнения и качество интервьюирования.

Часто

проводится параллельно с полевыми работами. Если процедура выполняется по контракту специализированным агентством, необходимо провести независимую проверку после ее завершения.

Анкеты могут быть не приняты по следующим причинам:

1. Не заполнены отдельные части.
2. Респондент не понял вопросов, не точно следовал инструкциям по заполнению анкеты.
3. Ответы варьируются очень незначительно. (все средние оценки)

4. Анкета является неполной.
5. Получена с нарушением сроков.
6. Отвечал респондент, не входящий в репрезентативную группу.

Если были определены
конкретные квоты респондентов,
либо назначены размеры
ячеечных групп, все принятые
анкеты необходимо
соответствующим образом
классифицировать и подсчитать.

Редактирование - обработка собранных анкет для повышения точности и аккуратности данных, заключается в просмотре анкет, выявлении нечитабельных, неполных, логически непоследовательных или неоднозначных ответов.

Анкеты с ответами

неудовлетворительного качества:

1. Отправляются обратно на места сбора данных для уточнения
2. Назначаются пропущенные значения
3. Отбраковываются и не включаются в анализ

2. Кодирование

Процедура кодирования -
присваивание кода, обычно
цифрового, каждому возможному
варианту ответа по каждому
вопросу, Информация,
образующая код, распределяется
по столбцам.

Код включает указание на положение столбцов (полей) и информации, которая в них содержится. Пол респондентов может кодироваться следующим образом: 1 - для женщин и 2 - для мужчин.

Поле отображает единичный элемент данных, например пол респондента.

Запись состоит из ряда соответствующих полей: пол, семейное положение, возраст, состав семьи, занятие респондента и т.д.

Все демографические и
личностные характеристики
респондента, как правило,
содержатся в одной
регистрационной записи. Все
записи (т.е. данные) по всем
респондентам хранятся в
компьютерном **файле. Все**
данные должны соответствовать
выбранной **схеме кодирования.**

Если анкета включает структурированные вопросы и незначительное количество неструктурированных вопросов, она кодируется до полевых работ. Если неструктурированные вопросы, коды присваиваются после возвращения заполненные анкеты с места сбора данных.

Код респондента и номер записи необходимо указывать для каждой записи данных.

Дополнительные коды для каждого респондента: код проекта, код интервьюера, код даты и времени, проверочный код.

Фиксированные коды полей -
номер записей по каждому
респонденту одинаков, и в одних и
тех же колонках по всем
респондентам указываются одни
и те же данные.

По всем пропущенным данным
следует использовать
стандартные особо
отличающиеся коды. **НАПРИМЕР,**
переменную, указываемую в
одном столбце, можно
закодировать 9; для переменной,
указываемой в двух столбцах,
назначить код 99 и т.д.

Кодирование
структурированных вопросов -
простая процедура, поскольку
варианты ответов определяются
заранее. Если предусмотрен
только один ответ, то каждому
возможному варианту ответа
присваивается код и указывается
соответствующая запись и
столбец, в который эти коды будут

Если же существует большое
количество вариантов ответов,
для каждого возможного варианта
следует выделить отдельный
столбец. К таким вопросам
относятся, например, вопросы об
использовании торговых марок, о
предпочтениях респондентов и т.
д.

**Кодирование
неструктурированных
вопросов представляет сложную
задачу.**

- 1) ответы респондентов дословно
записываются в анкету,**
- 2) для ответов разрабатываются
коды, которые присваиваются
конкретным ответам.**

- 3) составляется список, включающий 50-100 наиболее частых вариантов ответов на неструктурированные вопросы,
- 4) определяются категории, подлежащие кодированию,
- 5) кодировщик присваивает записанным в словарной форме ответам коды.

Коды категорий должны быть взаимоисключающими и
взаимоисчерпывающими, когда
каждому ответу присваивается
только один код. Категории не
должны перекрывать одна другую.

В категорию "другое" или "ни один из предложенных вариантов" должна войти только незначительная часть ответов (не больше 10%). Подавляющая часть ответов должна относиться к значимым категориям.

По наиболее важным вопросам коды категорий должны присваиваться даже в том случае, если они не упоминались ни одним из респондентов. Иногда важно знать именно то, что никто из отвечающих не дал тот или иной вариант ответа.

НАПРИМЕР, руководство решило выяснить, нравятся ли потребителям упаковка. При кодировании ответов на вопрос "Что вам больше всего не нравится в этой марке мыла?" была включена отдельная категория "упаковка". Данные кодируются для того, чтобы сохранить как можно больше деталей ответов.

Кодировочная книга содержит инструкции по кодированию, а также необходимую информацию о переменных, используемых в конкретном наборе данных, помогает правильно определять и располагать переменные.

Кодировочная книга включает
следующую информацию: номер
столбца, номер записи, номер
переменной, название
переменной, номер ответа,
инструкции по кодированию.

3. Преобразование и очистление данных

Преобразование данных
заключается в переносе
закодированных данных из
анкеты непосредственно в
компьютер. Данные можно
вводить в процессе считывания
меток или маркеров с бланков,
оптического сканирования или
компьютеризированного
сенсорного анализа.

Для метода считывания меток с
бланков, ответы должны
специально записываться в
конкретные поля анкеты.
Оптическое сканирование
заключается в прямом машинном
считывании кодов с
одновременным
преобразованием данных.
НАПРИМЕР: преобразование

При методе ввода ответов с клавиатуры, полностью избежать ошибок трудно, необходимо проверять введенные массивы данных. Для проверки правильности введенных данных применяется проверочный компьютер и второй оператор.

Процедура очищения данных
заключается в проверке
состоятельности собранных
данных и работе с пропущенными
ответами. Эта проверка намного
точнее и тщательнее
предварительной проверки,
поскольку выполняется с
использованием компьютерной
техники.

**Проверка состоятельности
данных позволяет выявить
данные, выходящие за пределы
определенного диапазона, и
логически непоследовательные
ответы либо определить
экстремальные значения.**

Значения, выпадающие из
конкретного диапазона, нельзя
использовать в анализе, их
необходимо исправить.

Компьютерные программы могут
выявить по каждой переменной
значения, выходящие за пределы
определенного диапазона, а также
распечатать необходимые коды.

Различается несколько типов логической непоследовательности ответов. НАПРИМЕР: респондент указывает, что часто пользуется каким-то товаром, и в то же время никогда его не приобретал. Чтобы выявить непоследовательные ответы и провести корректировку, всю необходимую информацию нужно распечатать.

Необходимо тщательно
анализировать экстремальные
значения. Иногда они указывают
на проблемы с качеством
собранных данных. Например,
чрезмерно заниженная оценка
торговой марки может быть
результатом, что респондент
поставил 1 по всем ее
характеристикам (шкала от 1 до 7).

Пропущенные ответы - значения переменных, которые остались неизвестными исследователю либо потому, что ответы респондентов были неоднозначны, либо неправильно или неразборчиво записаны.

Методы работы с пропущенными ответами:

1) Замена пропущенного значения нейтральным (обычно среднее значение по данной переменной, которое остается неизменным, а другие статистические данные, искажаются незначительно).

2) Замена пропущенного значения условным (вычисляется подходящий ответ через структуру ответов по другим вопросам).

3) Исключение объекта целиком (если многие респонденты не отвечают на те или иные вопросы, метод может вызвать значительное сокращение выборки).

4) Попарное исключение переменных (во всех вычислениях учитываются только наблюдения или респонденты, по которым есть полные ответы).

4. Статистическая корректировка данных

Процедуры статистической корректировки данных

включают: 1) взвешивание, 2) переопределение переменной, 3) преобразования шкалы.

Корректировки необязательны,
но, применяя их, можно
значительно повысить качество
анализа.

Взвешивание - метод статистической корректировки данных, каждому наблюдению или респонденту в базе данных присваивается весовой коэффициент, отображающий степень его значимости по сравнению с другими наблюдениями или респондентами.

Взвешивание применяется: 1)
чтобы выборочные данные
максимально точно представляли
конкретные характеристики
генеральной совокупности; 2)
корректировка выборки, чтобы
повысить
значимость ответов респондентов
с определенными признаками.

НАПРИМЕР: присвоение весового коэффициента 3,0 покупателям, которые покупают продукцию чаще всех, коэффициента 2,0 - тем, кто пользуется ею в средних пределах, 1,0 - тем, кто приобретает этот товар редко либо не пользуется им никогда.

Переопределение переменной -
преобразование данных для
создания новых переменных либо
изменения существующих с
тем, чтобы они точнее
соответствовали основным
задачам исследования.

НАПРИМЕРЫ: 1) показатель использования продукции – 10 категорий ответов, можно сократить до четырех категорий: часто, средне, редко, никогда.

2) сравнивая количество покупок и количество покупок в кредит получаем новую переменную - доля покупок в кредит.

Фиктивные переменные
используются для
переопределения категорий,
часто также называются
двоичными, дихотомическими,
инструментальными,
качественными. Могут принимать
только два значения, например, 0
или 1. Долю мужчин в выборке
можно получить на основе данных
о количестве женщин в выборке.

Преобразование шкалы -
манипулирование значениями
шкалы, чтобы иметь возможность
сравнивать ее с другими шкалами
либо каким-либо другим образом
преобразовывать данные и
делать их подходящими для
анализа.

Существует общая процедура преобразования шкал **нормализация** или **нормирование** - корректировка данных для приведения их к одной и той же шкале вычитанием выборочного среднего и деления полученного значения на стандартное отклонение.

5. Выбор стратегии анализа данных и классификация статистических методов

Процесс выбора стратегии анализа данных:

1. Предыдущие этапы процесса
МИ
2. Известные характеристики
данных
3. Конкретные свойства
статистических методов
4. Опыт и методология
исследователя

5. О

Цель анализа данных - получить информацию, которая поможет решить определенную проблему
МИ

В анализе известных характеристик данных на выбор статистического метода сильно влияет, какие шкалы измерений используются в ходе исследования.

Методы анализа значительно связаны со структурой исследования. **Дисперсионный анализ** подходит для работы с экспериментальными данными, полученными от проведения причинно-следственного исследования.

Статистические методы делятся на **одно- и многомерные**.

Одномерные методы - применяются для анализа данных, если существует единый измеритель для оценки каждого элемента выборки, либо если этих измерителей несколько, но каждая переменная анализируется отдельно.

Одномерные методы
классифицируются на
метрические или неметрические.
Метрические данные по своей
природе интервальные или
относительные. **Неметрические
данные** получаются на основе
измерений по номинальной или
порядковой шкале.

Многомерные методы
применяются для анализа
данных, если для оценки каждого
элемента выборки используется
два или больше измерителя и эти
переменные анализируются
одновременно. Можно
определить взаимосвязь между
двумя или больше явлениями.

Многомерные методы отличаются от одномерных тем, что при их использовании центр внимания смещается с уровней (средних показателей) и распределений (дисперсий) явлений и сосредотачивается на степени взаимосвязи (корреляции или ковариации) между явлениями.

Одномерные методы можно классифицировать на метрические или неметрические.

Метрические данные измеряются по интервальной шкале или относительной шкале.

Неметрические данные оцениваются по номинальной или порядковой шкале.

Многомерные методы можно разделить на методы зависимости и методы взаимозависимости.

Методы зависимости применяются в случаях, когда одна или больше переменных идентифицированы как зависимые, а остальные - как независимые.

Методы взаимозависимости -
группируют данные по лежащему
в
основе сходству, что позволяет
интерпретировать разные
структуры данных. При этом
переменные не подразделяются
на зависимые и независимые. При
исследовании взаимозависимости
чаще всего применяется