

6. ЛЕКЦИЯ «ОБРАБОТКА ИНФОРМАЦИИ, ПОЛУЧЕННОЙ ПРИ ПРОВЕДЕНИИ МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЙ»

6.1. Проблемы количественного измерения

Упорядочение множества реальных объектов, изучаемых в ходе сбора первичной информации, производится с помощью шкалирования, т.е. использования различных шкал. Шкалы – это алгоритмы, по которым производится отображение изучаемых социальных объектов в ту или иную числовую математическую систему: различаются по высоте – степени своей сложности – и по объему тех математических действий, которые можно производить с полученными в результате наблюдений значениями переменных. Шкалы подразделяются на метрические (интервальная и пропорциональная) и неметрические (номинальная и ранговая).

Типы шкал (уровней измерения)

Номинальная шкала. С помощью номинальной шкалы измеряются такие переменные, которые не могут количественно отличаться друг от друга. Другое название этого уровня измерений – шкала наименований, что довольно точно отражает его сущность: каждое значение здесь представляет собой отдельную категорию, и значение является просто своего рода ярлыком или именем. Для данных номинального уровня измерение центральной тенденции производится с помощью определения моды. Модой или модальной категорией называется то значение переменной, которое встречается среди данных наиболее часто.

Ранговая шкала. Данные рангового уровня измерений включают в себя категории наблюдения, которые размещены по порядку (от большего значения измеряемого признака к меньшему его значению или наоборот – от меньшего к большему). В дополнение к определению моды для выявления центральной тенденции в распределении значений переменной, измеренной по ранговой шкале, может быть выявлена медиана – категория или значение в распределении значений, лежащих выше и ниже того уровня, на который приходится половина всех частот.

Интервальная шкала. В отличие от номинальных или ранговых измерений, значения переменных, измеряемых с помощью интервальных шкал, изменяются непрерывно, они представляют собой численные величины, а не сами по себе категории, поэтому может реально существовать такое большое число различных наблюдаемых значений, что частоты и процентные отношения не в состоянии эффективно просуммировать данные.

Пропорциональная шкала (шкала отношений) – это та же интервальная шкала, но с зафиксированной точкой начала отсчета.

6.2. Этапы и процедуры первичной обработки данных

Результаты проведения маркетингового исследования представляют собой массив «сырых», необработанных данных: стопы заполненных анкет, пачки карточек наблюдений, кодировочных таблиц контент-анализа или иных способов фиксации анализа документов. В них, разумеется, содержится важная для

исследователя информация, однако прочесть ее с первого взгляда, сделать по ней какие-то выводы невозможно. Именно поэтому ее называют первичной информацией.

Казалось бы, первое, что необходимо проделать с этой информацией – ввести ее в компьютер в соответствии с кодировкой. Однако перед вводом следует произвести проверку данной информации на точность и полноту, а также на качество заполнения каждого документа. Упорядочив таким образом первичную информацию, можно приступить к ее вводу.

Кодирование – это присвоение числовых кодов вопросам тем или иным их значениям (т.е. вопросам и вариантам ответов). Его задачей является упорядочивание первичной информации и обеспечение удобств ее ввода в компьютерную базу данных, которая и подлежит затем статистической обработке.

При кодировании альтернативных вопросов кодируют номер самого вопроса и номера каждого из вариантов ответа, причем каждому варианту соответствует своя цифра (код).

При кодировании неальтернативных вопросов необходимо кодировать каждый вариант ответа как отдельную переменную. Вариантов ответов на каждый из таких вопросов всего два, и кодируются они при компьютерном вводе очень просто: 1 – есть ответ и 0 – нет ответа.

По завершении кодирования необходимо проверить правильность порядка следования номеров кодов переменных один за другим.

Кодированные переменные вводятся в матрицу программы SPSS¹. В приводимой ниже блок-схеме (табл. 13) справа приведены команды, используемые в программе SPSS, а слева – пояснения и комментарии. Мы не будем разъяснять значения каждой из команд, учитывая, что смысл их фактически разъясняется напротив – в соответствующей клетке левого столбца.

Таблица 13

Блок-схема обработки данных

| № | Команды SPSS | Содержание операции |
|---|--|--|
| 1 | Анализ→Описательные статистики→Частоты | Расчеты линейного частотного распределения всех переменных |
| 2 | Анализ→Описательные статистики→Таблицы сопряженности | Расчет кросстабов |
| 3 | Данные→Отобрать наблюдения | Отбор необходимых единиц наблюдения |

6.3. Методы математической и статистической обработки баз данных

Одним из первых шагов анализа является выявление центральной тенденции – уточнения того, какие из значений переменных встречаются в линейных распределениях наиболее часто, а значит, определяют общие (или цен-

тральные) закономерности. Для того, чтобы сделать выводы из результатов произведенных измерений одной переменной, используется так называемая описательная статистика. Соответствующие такому анализу таблицы называют линейными или одномерными распределениями.

При обработке эмпирических данных и анализе полученных результатов необходимо принимать во внимание шкалу, с помощью которой производилось измерение той или иной переменной. Данные, полученные по номинальной шкале, обычно резюмируются с помощью простого частотного распределения. В частотном распределении необходимо прибегать к указанию не только на частоту каждого значения переменной в ее абсолютном выражении, но и в процентном выражении, поскольку проценты раскрывают центральную тенденцию более наглядно и убедительно.

Критериями центральной тенденции для интервального и пропорционального уровней измерений выступают и мода, и медиана, и среднее арифметическое – сумма значений переменной, поделенная на число значений:

$$\bar{x} = \frac{\sum x_i}{N} = \frac{x_1 + x_2 + \dots + x_i}{N}, \quad (2)$$

где x_i – числовое значение i -й позиции; а N – общее число наблюдений (объем выборки).

Вычисление средней арифметической величины для переменных, значения которых измеряются не однозначно определенными числами, а изменяются вдоль непрерывного ряда значений, имеет свои особенности. Здесь рассчитывается не среднее арифметическое, а средневзвешенное:

$$\bar{x} = \frac{\sum x_i n_i}{N} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_i n_i}{N}, \quad (3)$$

где x_i – числовое значение i -й позиции; n_i – число респондентов, наблюдаемых по i -й позиции переменной; N – общее число наблюдений по всему массиву.

Показатели разброса для данных интервального или пропорционального уровня включают среднее отклонение, дисперсию и среднее квадратичное отклонение.

Среднее отклонение (MD) представляет собой меру разброса, основанную на отклонении каждого из значений от среднего.

Дисперсия (σ^2) представляет собой сумму квадратов отклонений от среднего, поделенную на число отметок:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}. \quad (4)$$

Среднее квадратическое отклонение (S) представляет собой корень квадратный из дисперсии (5):

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}. \quad (5)$$

Чем больше разброс точек данных вокруг среднего, тем выше значение σ^2 и S . Это означает, что если все данные одинаковы, то σ^2 и S равны нулю.

Для вычисления дисперсии и среднее квадратического отклонения надо пройти последовательно семь шагов:

- 1) вычислить среднее;
- 2) вычислить разности между средним и каждым из значений;
- 3) возвести в квадрат разности, вычисленные на этапе 2;
- 4) умножить квадраты разностей на частоты наблюдений каждого из значений;
- 5) просуммировать квадраты разностей, вычисленные на этапе 4;
- 6) разделить сумму квадратов из этапа 5 на N ; это равняется дисперсии;
- 7) извлечь квадратный корень из числа, вычисленного на этапе 6; это и будет среднеквадратическое отклонение.

В зависимости от того, насколько велика (или мала) дисперсия либо вычисленное с помощью ее среднеквадратическое отклонение, можно судить, насколько единодушны были в своих оценках респонденты (при меньшем значении дисперсии), или наоборот – насколько сильно они расходятся в своих мнениях (при большем значении дисперсии).

6.4. Проверка гипотез

В гипотезе, как правило, формулируется предположение о наличии связи между двумя и более переменными. Один из основных способов анализа, используемых для того, чтобы увидеть, какую связь переменные имеют друг с другом, называется *кросстабуляцией*. В результате такой операции получается таблица, по строкам которой размещены значения одной переменной, а по столбцам – значения другой. В каждой ячейке, образующейся на пересечении строки и столбца, указывается частота (процент) данного значения зависимой переменной, соответствующая данному значению независимой переменной. Такая таблица называется *кросстаб* (дословно – перекрестная таблица).

Анализ данных кросстаба позволяет сделать выводы: 1) о наличии связи между независимой и зависимой переменными; 2) о направлении этой связи, которая может быть либо прямой (или положительной), либо обратной (отрицательной); 3) о силе связи.

Наличие связи выявляется сравнением изменения значений зависимой переменной. При этом руководствуются правилом: нет изменения – нет связи.

Направление связи определяется следующим образом. Прямая связь: при увеличении значений независимой переменной одновременно возрастают значения зависимой переменной. Обратная связь: при увеличении значений независимой переменной одновременно уменьшаются значения зависимой переменной.

Сила связи определяется оценкой того, насколько резко или медленно изменяются значения зависимой переменной при изменении значений зависимой переменной: если степень различия мала, имеет место слабая связь, если велика, речь может идти о сильной связи.

В качестве показателя силы связи используют также различные коэффициенты корреляции. Одним из них является коэффициент ранговой корреляции Спирмена. Формула расчета его имеет следующий вид:

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{n^3 - n}, \quad (6)$$

где d – разность рангов; n – общее число рангов (т.е. вариантов ответов). По- нятно, что коэффициент ранговой корреляции Спирмена будет равен $+1$ (абсо- лютная положительная связь), если ответы всех респондентов обеих анализиру- емых групп будут в точности совпадать; он будет равен -1 (абсолютная отрица- тельная связь), если ответы всех респондентов обеих анализируемых групп бу- дут прямо противоположны; если $r_s = 0$, то это означает полное отсутствие вся- кой связи.

Табличное представление данных

Таблица – это перечень сведений, обычно числовых данных, приведенных в определенную систему и разнесенных по графам – строкам и столбцам. Ос- новное преимущество этой формы представления информации заключается в том, что в ней кратко и емко даются пояснения значений соответствующих группировок. Хорошо сконструированная таблица позволяет как самому иссле- дователю, так и заказчику более четко представить, описать и объяснить смысл и сущность изучаемого социального явления.

Любая маркетинговая статистическая таблица описывается с помощью следующих параметров.

Заголовок – название таблицы, раскрывающие структуру группировки описываемых переменных, либо характер связи (зависимости) между двумя и более переменными. Наряду с этим в названии (или чаще в подзаголовке) ино- гда указываются общие для всех переменных единицы измерения (число отве- тов, проценты, средний балл и т.д.).

Подлежащее – то, что подлежит описанию, то есть указание переменной, подвергаемой анализу, и тех конкретных значений, которые она может принимать.

Сказуемое – само описание, то есть числовые значения, разнесенные по графам – ячейкам или клеткам таблицы.

Частотные (линейные) распределения значений одной переменной отра- жаются в простой таблице. Простые таблицы представляют собой простой пе- речень (список) отдельных значений той или иной переменной с количествен- ной или качественной характеристикой каждой из них в отдельности. Поэтому иногда их называют также перечневыми.

Большинство комбинационных таблиц формируется в ходе операции, именуемой кросстабуляция. Получаемые в результате кросстабы должны быть, как и любые другие таблицы, наглядны, обозримы и, желательно, размещаться в пределах одной страницы.