

Н.С. МАРТЫШЕНКО

**МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ АНАЛИЗА ПОВЕДЕНИЯ ПОТРЕБИТЕЛЕЙ
НА РЕГИОНАЛЬНОМ ТУРИСТСКОМ РЫНКЕ**

Излагаются методические подходы при сборе и обработке первичных данных, характеризующих конъюнктуру рынка туристских услуг, а также методические рекомендации по использованию методов многомерного статистического анализа для сегментирования потребительского рынка услуг туристского комплекса.

В последнее десятилетие значимость Приморского края для экономики страны существенно повысилась, в том числе, как перспективной зоны для туризма. Многим жителям Дальнего Востока и Сибири стали недоступны черноморские и прибалтийские курорты, рекреационные объекты Кавказа и Карпат. Существенно возросли транспортные тарифы и их доля в структуре туристских расходов, что позволяет рассматривать Приморье как перспективную зону въездного туризма.

Одним из сдерживающих факторов развития туризма в крае является неразвитость туристской индустрии и туристской инфраструктуры, отсутствие навыков изучения рынка, отсутствие методик исследования поведения потребителей и механизмов воздействия на потребителей, слабое использование возможностей современных информационных технологий в решении этих задач. В настоящей работе рассматриваются некоторые пути преодоления проблем, сдерживающих развитие туристского комплекса региона.

Заинтересованность региона в развитии туристской отрасли зависит от того, какой вклад вносит и может внести эта отрасль в развитие края. Вклад отрасли оценивается, в первую очередь, вкладом в бюджет края, занятостью населения в отрасли и ростом их доходов. Специфика туристской отрасли такова, что основной доход в бюджет приносят не те туристские фирмы, которые формируют и реализуют туристский продукт потребителям, а предприятия социально-бытовой и туристской инфраструктуры региона, обслужива-

ющие туристов (транспорт, гостиницы, рестораны и т.п.). Тоже можно сказать и о занятости населения: если турфирмы Приморского края имеют всего менее двух тысяч сотрудников, то в сфере обслуживания занята значительная часть населения (50,4% в 2004г). Таким образом, задача анализа рынка туристских услуг усложнена нечеткостью границ сферы услуг, задействованной непосредственно на обслуживании туристов, тем более, что многие объекты можно расценивать как объекты двойного назначения.

Статистические данные, характеризующие сегодняшнее состояние рынка туристских услуг, по всем показателям свидетельствуют (в т.ч., по вкладу в бюджет) об устойчиво низкой активности потребителей. А при существующих потенциальных возможностях края по удовлетворению спроса на туристские услуги, вклад отрасли в экономику края нужно признать неадекватно малым (3% в ВРП). Отсюда вытекает цель анализа поведения потребителей туристских услуг: анализ должен обосновать действия и последовательность, в какой их необходимо предпринять, чтобы существенно повысить вклад отрасли туризма в экономику края.

В настоящее время в Приморье насчитывается более 250 туристских фирм и агентств. Коллективы фирм, как правило, очень малочисленны. Турфирмы с количеством сотрудников 3-5 человек составляют 37% от общего количества фирм, занимающихся организацией поездок, с количеством сотрудников более 20 человек – всего 5%.

Большое количество малых фирм-организаторов поездок объясняет и масштабы их деятельности. Около 63% туристских фирм принимают до 1 тыс. туристов в год и они обслуживают всего 7% от общего количества въездных туристов. С другой стороны 7% туристских фирм, принимающих более 22 тыс. туристов в год, обслуживают около 42% общего числа всех туристов.

Структура предложений туристского продукта пока очень упрощенная. Главным направлением поездок населения Приморского края является Китай – свыше 90% выездного потока, затем – США, Корея и Япония. Понятно, что

такой туризм, в основном, носит коммерческий характер, Китай – шоптуры, Япония – автомобильный бизнес. Наблюдается тенденция незначительного роста количества поездок российских граждан, проживающих в Приморском крае, в Тайланд, США, Сингапур и Австралию.

Перспективы повышения спроса на услуги туристской отрасли связаны с повышением уровня жизни населения, который согласно официальной статистике, в последние годы имеет устойчивую положительную тенденцию.

Потребителей рынка туристских услуг можно разбить на три большие группы или сегмента: иностранные туристы; туристы, прибывающие из других регионов страны; население края (потребление услуг во время отпусков и в свободное от работы время). Все они пользуются услугами предприятий, работающих на отрасль туризма.

Анализ статистических данных показал, что общая численность туристов, прибывающих в край в течение года из других стран, в последние годы сохраняется примерно на одном уровне, а по некоторым странам (в том числе Китай) даже падает. Иностранные туристы наиболее выгодны для края, но резко изменить ситуацию в ближайшие годы, по всей видимости, не удастся.

Главная причина этого состоит в том, что уровень требований зарубежных туристов намного выше, чем сегодня может предоставить туристская отрасль края. А поскольку у иностранного туриста есть большой выбор стран, где его требования будут удовлетворены, не стоит ожидать в ближайшие годы резкого изменения положения.

Вторая группа потребителей – это туристы, въезжающие из других регионов страны. Эта группа в настоящее время является наиболее перспективной, потому что очень многих жителей соседних регионов интересовал бы отдых у моря, традиционно считающийся в России наиболее популярным и привлекательным.

Несмотря на тот факт, что, оценивая уровень развития туризма, специалисты акцентируют свое внимание на природно-климатических ресурсах, перспективу развития внутреннего и въездного туризма в Приморском крае

мы связываем с развитием материально-технической базы предприятий туристской индустрии и туристской инфраструктуры и с повышением уровня обслуживания, а также с ведением разумной ценовой политики на товары и услуги, особенно туристского назначения.

Обращает на себя внимание отсутствие на сегодняшний день систематично собранных данных по туризму, что не дает возможности проведения комплексного анализа туристских ресурсов региона, а значит оценить его потенциал и экономический вклад в развитие территории края. Поэтому большое значение для оценки эффективности туризма и совершенствования методического обеспечения анализа поведения потребителей по результатам его эксплуатации имеет третья группа потребителей туристских услуг – это население края. Относительно этого сегмента все-таки имеется некоторая официальная информация, кроме того, автором настоящей работы были проведены собственные исследования с целью дополнительной информации.

В отличие от двух выше обозначенных групп въездных туристов, которые могут расширяться в огромных пределах за счет привлечения все новых и новых потребителей, масштаб третьей группы – жителей края – фиксирован. Возможности потребления услуг ограничены объемами доходов населения. Хотя возможности внутреннего туризма ограничены, он должен быть объектом первостепенного внимания, поскольку в дальнейшем должен привести к росту и въездного туризма. Значимость этой идеи возрастает, если принять во внимание то, что в последние три года наметились явно положительные тенденции в росте благосостояния населения, а значит и в возможности совершать поездки по краю и потреблять товары и услуги регионального туристского комплекса.

Еще одним аргументом в пользу более глубокого исследования населения края как потребителей туристских услуг, является отработка методик анализа поведения потребителей в этой специфической отрасли. Особенно это важно для методик, основанных на применении сложных математических моделей и многомерного анализа данных. При всем различии трех рассмот-

ренных потребительских групп отлаженные методики могут быть в дальнейшем распространены и на въездной иностранный туризм. Сбор данных по этой категории потребителей намного сложнее и соответственно требует больших финансовых и временных затрат. Поэтому использование испытанных методик и технологий позволит более эффективно организовать сбор и обработку данных, необходимых для исследования въездного иностранного туризма.

Первичный материал о поведении потребителей формируется в процессе анкетных опросов потребителей туристских услуг. Анализ данных анкет можно разбить на три этапа: анализ достоверности данных и устранения грубых ошибок; предварительный анализ данных; многомерный анализ данных.

Но прежде чем перейти к анализу данных необходимо разработать систему сбора первичного материала. Исследователь всегда ограничен в данных. Умение найти источник и собрать данные, которые возможно найти в условиях, сложившихся на момент проведения исследований, составляет основу практически любой научной работы. Наибольшую ценность представляют не те данные, которые удалось почерпнуть из внешних источников, отчетов и книг регистрации, а те, которые получены или изысканы самим исследователем путем целенаправленных действий. В этом случае исследователь самостоятельно определяет не только структуру и способ сбора информации, но и организует процесс сбора. При этом он представляет, каким образом эти данные будут представлены в компьютерной информационной системе, как будут использованы для проведения научного анализа. Все вместе, от определения информационной потребности и возможностей по сбору данных до системы представления данных на компьютере, можно определить как систему сбора первичных материалов.

Создание системы сбора информации, необходимой для решения задачи исследования поведения потребителей, - это сложный и трудоемкий процесс, предполагающий прохождение множества этапов на пути к поставлен-

ной цели. Разработку системы сбора первичных данных приходится производить в условиях неполной информации и ограниченности ресурсов. Поиск решений осуществляется для слабо структурированных и не полностью формализованных задач. Эффективность работы такой системы будет тем выше, чем более полно в ней будут реализованы принципы системного подхода [3 - 4].

Хотя сбор информации о поведении потребителей объединен единой целью (принцип конечной цели) достигнуть цели в рамках одного опроса невозможно. Каждый новый этап исследования (принцип поэтапности) дает новую информацию для организации работы на следующем этапе. Поэтому разрабатываемая система сбора должна отвечать принципу развития или открытости. Этот принцип предполагает учет изменяемости системы, способности к развитию, расширению, накоплению информации. С расширением спектра решаемых задач и появлением новых функций предполагается разработка новых модулей, совместимых с уже имеющимися.

Важнейшим принципом, которому должна отвечать как система сбора, так и система обработки информации является принцип модульности. В процессе накопления данных уточняются формулировки вопросов, а сами вопросы объединяются в некоторые блоки или модули. Повторяемостью модулей вопросов, используемых при анкетировании потребителей туристских услуг, позволяет выдержать основополагающее свойство системы, такое как целостность системы [6].

Построение системы с учетом состояния внешней среды, известно как удовлетворение принципа связности. Система сбора информации рассматривается как элемент (подсистема) более общей системы имеющей своей целью решение главной задачи - исследование поведения потребителей.

После сбора анкетных данных необходимо произвести анализ достоверности данных. Этот этап имеет исключительно важное значение в практических исследованиях, потому что от степени достоверности данных зависит обоснованность выводов на всех последующих этапах. Для решения этой за-

дачи используется такой метод многомерного статистического анализа как рабастное оценивание.

Рабастные методы оценки данных используются для выявления и подавления грубых ошибок в первичных данных многомерных наблюдений. Основы теории рабастности были разработаны академиками: А.Н. Колмогоровым, Н.В. Смирновым и Б.С. Ястремским. Свое дальнейшее развитие теория рабастности получила в работах зарубежных ученых – А. Тьюки, Дж. Хьюбера.

Суть классических методов сводится к проверке сложных статистических гипотез на основе столь же сложных критериев (критерий Хьюберта, Граббса, Титъена-Мура). Такие критерии не нашли очень широкого распространения в силу ряда причин и сложность их практического использования не является основной причиной. Из анализа примеров применения таких методов можно сделать выводы, что они применяются в задачах с крайне ограниченными объемами выборок. Цена включения или отбрасывания одного объекта (наблюдения) достаточно высока и поэтому для принятия решения недостаточно только интуиции исследователя и содержательного анализа. Решение должно быть подтверждено числом в вероятностном смысле. Соответственно, табулированные значения критических точек для специфических критериев рабастности, которые удалось обнаружить, в научной литературе приводятся для объемов выборки от 20 до 50. Программы расчета рабастных критериев не входят в состав распространенных пакетов по обработке статистических данных [2]. Тем не менее, проблема засорения первичных данных при обработке анкет существует и ее надо решать.

Для повышения достоверности анкетных данных можно было бы предложить более простые решения. Основная идея предложенного подхода состоит в построении некоторых фильтров, которые не дают точных результатов, а позволяют из большого количества наблюдений выделить «подозрительные» с точки зрения наличия грубой ошибки, далее принятие решения отводится исследователю. Специфика исследований, основанных на анкет-

ных данных, такова, что исследователь легко может расстаться с 5-ю – 10-ю наблюдениями из 2-х – 3-х тысяч.

Рассмотрим принципы построения процедур, предлагаемых для обнаружения и подавления грубых ошибок [1, 5].

Прежде чем производить поиск ошибок, необходимо определить возможные источники ошибок, затем порядок их диагностики. Диагностика производится с помощью специальных фильтров. Фильтры позволяют выделить многомерные наблюдения (объекты или анкеты), которые с точки зрения условий фильтра являются подозрительными или менее правдоподобными в ряду всех наблюдений. Принятие решения (например, отбрасывание, корректировка и т.п.) производится исследователем с позиций углубленного содержательного анализа конкретных данных.

При массовом сборе анкетных данных исследователь сталкивается с необходимостью привлечения большого количества интервьюеров. Каждый интервьюер производит опрос некоторой группы респондентов. Качество собранного материала каждым интервьюером может существенно отличаться в силу разных причин, - например, недобросовестной работы некоторых интервьюеров или их неумения контактировать с респондентом, или попытки интервьюера собрать данные в специфической контактной группе респондентов, негативно относящихся к опросу. Независимо от причины, исследователю часто выгоднее отказаться от всей серии анкет, существенно отличающихся от всех остальных, чем получить недостоверные данные. При нехватке статистического материала лучше организовать дополнительный опрос респондентов.

Производить анализ выбросов по одному признаку (вопросу анкеты) при большом количестве анкет крайне неэффективно, особенно при большом количестве вопросов в анкете. Практика показывает, что если анкета содержит недостоверные данные, то недостоверность не будет носить избирательный характер, а будет прослеживаться по большинству вопросов. Поэтому лучше производить фильтрацию по всем признакам. Однако поскольку отве-

ты на различные вопросы могут представлять собой признаки, измеряемые в различных шкалах, такой анализ целесообразно проводить по группе признаков, измеренных в одной шкале.

Предположим, что ответы на вопросы содержат m признаков, представленных в шкале отношений. Выборка представлена совокупностью k пакетов анкет, собранных различными интервьюерами. Одно наблюдение из r -го пакета (r – номер интервьюера) можно представить набором значений m признаков:

$$x_{i_r} = (x_{i_r,1}, x_{i_r,2}, \dots, x_{i_r,j}, \dots, x_{i_r,m})$$

где $j=1, 2, \dots, m$ – номер признака; m – количество признаков;

$r=1, 2, \dots, k$ – номер интервьюера, k – количество интервьюеров;

$i_r=1, 2, \dots, n_r$ – номер анкеты в пакете одного интервьюера;

n_r – объем пакета анкет интервьюера с номером r ;

Тогда, объем выборки, включающей все анкеты будет:

$$n_0 = \sum_{r=1}^k n_r.$$

Задача состоит в том, чтобы из k пакетов анкет выделить пакет, который имеет наибольшие отличия от остальных пакетов.

Для этого последовательно для каждого пакета r ($r=1, 2, \dots, k$) повторим следующую процедуру: рассчитаем средние значения m признаков по выборке за исключением пакета с номером r :

$$\bar{X}^{-r} = (\bar{x}_1^{-r}, \bar{x}_2^{-r}, \dots, \bar{x}_j^{-r}, \dots, \bar{x}_1^{-r}),$$

и средние значения признаков по пакету с номером r :

$$\bar{X}^r = (\bar{x}_1^r, \bar{x}_2^r, \dots, \bar{x}_j^r, \dots, \bar{x}_1^r).$$

Вычисляем поэлементные модули разностей двух векторов средних:

$$\lambda_r = |\bar{X}^{-r} - \bar{X}^r|.$$

Процедура, основанная на последовательном изъятии и восстановлении части выборки, называется скользящим экзаменом. Объединяем все отклонения λ_r в одну матрицу λ размерности $k \times m$. На основании матрицы λ рас-

считаем матрицу M той же размерности. Вычисления производятся по схеме: определяется максимум в каждом столбце матрицы λ , затем элементу матрицы M , соответствующему максимуму присваивается значение единицы, всем остальным элементам матрицы M присваивается значение ноль. В результате построчного суммирования элементов матрицы M получим вектор оценок для каждого интервьюера:

$$\mu = (\mu_1, \mu_2, \dots, \mu_r, \dots, \mu_k).$$

Интервьюер с наибольшим значением μ_r будет иметь максимальный штраф, и поэтому его данные могут быть поставлены под сомнение. Теперь исследователь может сосредоточить свое внимание на отдельном пакете первичных данных, подвергнуть их дополнительному содержательному анализу, в результате которого исследователь определяет, является отклонение допустимым или нет. Это достаточно грубый фильтр. Он основан на предположении о том, что если пакет анкет содержит недостоверную информацию, то большие отклонения от средних будут не только по одному признаку, но и по другим. Данную процедуру можно применить и в случае, когда пакеты содержат по одному наблюдению. В этом случае идентификатором интервьюера может выступить номер наблюдения в выборке.

Для признаков, измеренных в ранговых и номинальных шкалах, можно использовать фильтр, основанный на сравнении частотных рядов. К номинальным шкалам можно отнести и бинарные признаки. К номинальной шкале с помощью операции типологии могут быть преобразованы и ответы на открытые вопросы. Практика показывает их высокую информативность.

Предположим, что всего выделено m признаков. Тогда для расчета элемента вектор λ_r будем использовать формулу:

$$\lambda_{ij} = \sum_{i_r} \frac{(P^{-r}_{i_r j} - P^r_{i_r j})^2}{P^{-r}_{i_r j}},$$

где $i_r = 1, 2, \dots, f_r$ – номер интервала частотного ряда r -го признака;

f_r – количество интервалов в частотном ряду r -го признака;

$P^{-r}_{i,j}, P^r_{i,j}$ - относительные частоты, рассчитанные по частотным рядам выборки без r -го пакета и только по выборке r -го пакета.

Поскольку количество разрядов в частотных рядах различных признаков может отличаться, то в общей оценке интервьюера признаки должны браться с весовыми коэффициентами:

$$\mu_r = \sum_{j=1}^m \frac{f_j}{\sum_{j=1}^m f_j} \lambda_{rj}.$$

Необходимость весовых коэффициентов отпадает, если все признаки свести к бинарным шкалам. Соответственно можно использовать второй фильтр на всех признаках, если предварительно перейти от шкалы отношений к ранговой или бинарной шкале.

Третий фильтр является наиболее точным, поскольку максимально использует информацию, содержащуюся в выборке. Ввиду громоздкого формального описания этого фильтра, изложим только основную идею. Фильтр основан на использовании идей распознавания образов. Вначале все признаки приводятся к бинарному представлению. Выборка считается классифицированной на два класса. Первая выборка – анкеты одного пакета, вторая выборка – все остальные анкеты. Две выборки используются для построения эталонов классов (обучение). Затем производится распознавание наблюдений обеих выборок. По результатам распознавания рассчитывается количество ошибок распознавания. Далее процедура повторяется по схеме скользящего экзамена. Пакет анкет, обеспечивающий наименьшую ошибку распознавания, можно считать более обособленным и, следовательно, он рассматривается как аномалия, требующая содержательного анализа. При этом могут быть отброшены не только наблюдения (анкеты), принадлежащие этому пакету, но и прочие наблюдения, отнесенные к аномальной выборке в процессе распознавания.

При содержательном анализе подозрительных данных, как правило, обращаются к первичному материалу (анкетам на бумажном носителе). При

этом часто выбросы могут быть вызваны банальной ошибкой оператора или явной опiskeй респондента, автоматически перенесенной оператором в компьютер. Такие ошибки легко устранимы.

Рассмотренные выше фильтры предполагают, что все респонденты ответили на все вопросы анкет. На практике это выполняется крайне редко. Причин этому может быть несколько.

Первая причина - не все респонденты одинаково хорошо могут понимать вопрос анкеты. В этом случае респондент может просто отказаться отвечать на такой вопрос.

Вторая причина - на некоторые вопросы респондент может отказаться отвечать, потому что считает информацию конфиденциальной.

Третья причина - респондент может просто по невнимательности пропустить вопрос.

Четвертая причина - респондент может и понимать вопрос, но просто не знать требуемой информации.

Пятая причина - отсутствие данных в одном вопросе может быть вполне объяснимо ответом на другой вопрос. Например, если нужно указать сумму денежных средств, затрачиваемую в отпускное время, отсутствие данных вполне может быть объяснимо, если респондент вообще не использовал отпуск.

Шестая причина - респондент может не давать ответы на ряд вопросов при негативном отношении к опросу.

По крайней мере, пять причин отсутствия данных по одному признаку в наблюдении (ответ на один вопрос) не влияет на достоверность данных при ответах на другие вопросы. Все такие ситуации, хотя и доставляют дополнительные трудности исследователю, но вполне могут быть обработаны. Во-первых, в отдельных задачах анализа могут использоваться не все признаки и следовательно, отсутствие данных в неиспользуемых признаках вообще не отражается на таких задачах. Во-вторых, некоторые позиции могут быть заполнены на основе содержательного анализа соответствующей анкеты. В-

третьих, иногда отсутствие данных можно интерпретировать как значение, измеренное в порядковой шкале.

Шестая причина ставит под сомнение достоверность ответов и на все остальные вопросы. Такие анкеты должны быть по возможности выявлены с помощью специальных фильтров и удалены. Такой фильтр может быть построен на основе идей распознавания образов.

Для анализа многомерной выборки на отсутствие данных может быть очень полезен следующий простой фильтр.

Матрица данных наблюдений (за исключением столбцов, содержащих все данные) преобразуется к бинарному виду (0 – есть данные, 1 – нет данных). Обозначим такую матрицу через V . Она имеет размерность $n \times m$, где n – количество наблюдений, а m – количество признаков. Произвольный элемент матрицы обозначим b_{ij} . Тогда каждой строке матрицы (ассоциированной с наблюдением) можно сопоставить некоторую величину:

$$\lambda_i = \sum_{j=1}^m \frac{1}{v_j} b_{ij},$$

где v_j – количество единиц в столбце j .

Весовой коэффициент $1/v_j$ вводится для ранжирования признаков. Если признак содержит множество незаполненных позиций (отсутствие данных), то отсутствие данных признака в отдельном наблюдении событие не столь уж исключительное и имеет небольшой вес и наоборот. После расчета λ_i ($i=1, 2, \dots, n$) выборка может быть упорядочена по убыванию показателя λ_i . На первых позициях окажутся наблюдения, внушающие наибольшее беспокойство наличием отсутствующих данных. Анкеты, соответствующие таким наблюдениям, должны быть подвергнуты углубленному содержательному анализу. Если в процессе анализа возникает подозрение, что отсутствие данных вызвано шестой причиной, то такие анкеты должны быть исключены. Степень уверенности в необходимости исключения данных анкеты повышается, если в поле зрения исследователя попадает серия анкет, принадлежащих одному пакету.

Только после того, как исследователь убедился, что полученные данные заслуживают доверия и не содержат грубых ошибок, он может приступать к исследованию закономерностей структуры и закономерностей в поведении, присущего потребителям продуктов.

На втором этапе производится предварительный анализ данных, собранных в процессе анкетирования потребителей туристских услуг на региональном рынке.

Как правило, на этапе предварительного анализа производится построение частотных рядов отдельных признаков.

На этапе предварительного анализа кроме частотных рядов рассчитываются и многие другие характеристики одномерных признаков. Кроме этого производится расчет некоторых новых признаков на основе нескольких других и другие преобразования признаков. Наиболее интересным является формирование признаков на основании типологий создаваемых по открытым вопросам анкет.

Среди задач исследования поведения потребителей ключевой задачей является сегментирование потребителей или в терминологии методов статистического анализа (МСА) задача многомерной классификации. Поэтому предпосылки применения МСА содержатся как в самой постановке задачи, так и в данных, используемых для ее решения [5].

Хотя этот раздел МСА в последние два десятилетия очень бурно развивался, находя все новые и новые области применения, задачи, основанные на использовании методов кластерного анализа, до сих пор больше относятся к прерогативе научно-исследовательских, а не к повседневной практике. Особенно это относится к таким относительно новым объектам, как потребители туристских услуг. Однако, с распространением специализированных пакетов программ по статистике, появляется возможность приблизить сугубо научную задачу к практической работе. Именно такая цель и ставится в настоящей работе – не столько внести новое слово в теорию многомерной статистической классификации, а сколько расширить практическое использование

мощного средства на новый объект и способствовать еще более широкому использованию математического аппарата, доказать эффективность применения метода в этой области.

Для практического применения кластер-анализа для сегментирования потребителей экономисту недостаточно знания основ теории кластеризации и наличия соответствующих программных средств, а необходима конкретная методика, которая прослеживала бы все этапы от зарождения идеи поиска сегментов потребителей до получения результата. В методике должны быть определены возможные проблемы, с которыми может столкнуться исследователь, решивший воспользоваться новым инструментом и пути преодоления этих проблем.

На самом деле идея наличия классов однородных объектов возникает у исследователя еще до сбора первичных данных и разработки содержания вопросов анкет. Она основана на объективно существующем и наблюдаемом явлении в поведении потребителей, состоящем в предпочтении одними группами потребителей одних услуг, а другими группами потребителей – других услуг. Что объединяет одних, что разъединяет других, каковы закономерности формирования однородных групп потребителей? Самые общие идеи о причинах формирования групп потребителей должны зародиться у исследователя еще до сбора данных. При этом у него может быть несколько гипотез относительно формирования сегментов потребителей. Наличие нескольких гипотез даже желательно, поскольку если исходная гипотеза одна и, в конце концов, выяснится, что исходная гипотеза не подтвердилась, все усилия по поиску сегментов окажутся напрасными, что может обнаружиться на последних этапах предпринятого исследования. Если говорить о потребителях туристских услуг, то задача выделения однородных сегментов еще более усложняется, поскольку туристская услуга комплексная и не имеет четко очерченных границ. Этап формирования гипотез о формировании сегментов можно признать как **первый этап** на пути решения задачи сегментирования.

Чаще всего на первых этапах исследователь не задается вопросом о количестве сегментов.

На втором этапе исследователь определяет для себя, какими признаками он может охарактеризовать возможные сегменты, пока не задумываясь, как он сможет получить информацию о них. Это желаемые характеристики потребителей.

На третьем этапе он должен определить поле признаков или характеристик, которые он может реально получить в процессе анкетного опроса. На этом этапе он моделирует восприятие респондентами возможных вопросов. При этом он должен учитывать доступные ему средства организации опросов. Как правило, желаемые характеристики очень трудно получить, поэтому исследователь для оценки желаемой характеристики должен сформулировать целый ряд вопросов, которые, по его мнению, в комплексе отражают уровни желаемой характеристики. На этом же этапе исследователь определяет метод сбора информации.

На четвертом этапе происходит формирование содержания анкеты. Здесь опять исследователь стоит перед выбором, какие вопросы включать в анкету и в какой последовательности разместить. На этом этапе дорабатывается форма представления вопросов, уточняются шкалы измерения признаков, ассоциированных с вопросами. Результатом этапа является законченная форма анкеты и пробное анкетирование небольшой группы респондентов – потребителей продукта. Апробацию анкеты или пилотное исследование, в силу его важности, можно было бы выделить и в отдельный этап.

На этом же этапе исследователь обдумывает правила организации сбора анкет, т.е. ответы на вопросы: кто, где, когда, как, у кого будет проводить опрос. При организации опросов необходимо иметь в виду вопросы репрезентативности выборок. Разрабатывая систему сбора необходимой информации, исследователь должен учитывать доступные ему ресурсы.

На пятом этапе непосредственно происходит опрос потребителей продукта и заполнение анкет на бумажном носителе.

На шестом - производится разработка базы данных для представления данных в электронном виде. База данных должна разрабатываться с учетом требований к представлению данных, предъявляемых программными средствами, которые предполагается использовать для обработки данных.

На седьмом этапе осуществляется ввод данных. На этом же этапе производится оценка достоверности и подавления ошибок в данных. Этот этап, основанный на многомерном анализе выборок, уже был нами рассмотрен выше.

На восьмом этапе производится предварительный анализ данных, на котором исследуется структура одномерных признаков, оценивается степень зависимости пар признаков и т.п.

На девятом этапе производится выбор программного средства, с помощью которого можно выполнить кластерный анализ. Если исследователь не имеет опыта применения программ классификации, это не должно останавливать его. Для выбора программного продукта достаточно знания только основ теории классификации.

Распространенные программные продукты достаточно близки по своим характеристикам. Такие программы, как правило, предоставляют достаточно широкие возможности для эксперимента, намного превосходящие потребности пользователя, не искушенного в теории МСА. При всем обилии вариантов и возможностей по классификации данных многомерной выборки необходимо знать, что если выборка действительно содержит хорошо разделяемые классы, то большинство методов классификации дадут одни и те же или близкие результаты классификации. С другой стороны, если изучаемая выборка вообще не содержит классов, то даже самый изощренный алгоритм их не обнаружит. Эксперимент по использованию разных метрик сходства и правил классификации может дать эффект, но слишком рассчитывать на успех в задачах анализа анкетных данных не нужно. Другое дело, в технических системах, когда в отдельных случаях оправдан не только обширный эксперимент с алгоритмом, но и разработка специального алгоритма под за-

дачу. В случае с потребителями туристского продукта неэффективно искать алгоритм, который дал бы лучший результат классификации, эффективнее произвести поиски признаков, которые бы лучше работали на разделение выборки, тем более исследователь сам их формирует.

И только **на десятом этапе** можно приступить к выявлению однородных групп объектов (наблюдений) выборки. Собственно на этом этапе и производится сегментирование опрошенных потребителей. Тогда может возникнуть вопрос, зачем же были рассмотрены девять предшествующих этапов. Нами приведены эти этапы для того, чтобы показать, что на каждом из них исследователь должен учитывать, какую задачу он будет решать и соответственно ошибки, допущенные исследователем на любом из них, могут привести к отрицательному результату при решении задачи кластеризации. А получив отрицательный результат, исследователь должен проанализировать все предыдущие этапы, с тем, чтобы понять, где именно он допустил ошибку с тем, чтобы повторить исследование на новом уровне именно с этого этапа.

Этап выделения сегментов потребителей рассмотрим более подробно, разбив его на подэтапы, или действия.

Первое действие. Необходимо выбрать некоторый список признаков, который, по мнению исследователя, в совокупности может описать сегменты. Если задача решается впервые, следует включать в рассмотрение избыточное количество признаков. Однако важно, чтобы признаки были измерены в одинаковых шкалах. Иногда допускается включать в список признаков помимо признаков, измеренных в номинальной шкале отношений, признаки, измеренные в ранговой шкале.

Второе действие. Если в список признаков включены признаки, представленные в различных шкалах, необходимо произвести преобразование признаков от более богатой шкалы к более бедной. Часто при использовании признаков, заданных в шкале отношений, целесообразно произвести операцию нормировки. Преобразование ответов открытого вопроса путем типиза-

ции, тоже можно отнести к этому этапу. В результате преобразования признаков будет сформирована выборка, предназначенная для классификации.

Третье действие. Воспользоваться выбранной программой для классификации сформированной выборки. Для начала достаточно выбрать небольшое количество классов (2 – 3) и несложный вариант оценки меры сходства (например, Евклидову меру).

Четвертое действие. Производится предварительная оценка результатов классификации. Если результаты классификации покажут, что преобладающее количество объектов объединяется в один класс, а другие классы будут включать единичные объекты (до 5% выборки), то можно сделать вывод, что классификация для данной выборки и при данном способе классификации не удалась. В этом случае можно еще предпринять несколько попыток классификации выборки, меняя метрики сходства и другие параметры программы-классификатора. Это не требует больших затрат времени, поэтому такие попытки надо предпринять. При этом можно добиться положительного результата, но маловероятно, что картина резко изменится.

Если эксперимент с параметрами программы не даст результатов, то необходимо снова обратиться к выбору нового списка признаков (действие первое) или попробовать поработать над преобразованием уже выбранных (второе действие). Указанные действия необходимо повторять до тех пор, пока будет получен правдоподобный вариант классификации.

Если и эти действия не приведут к результату, необходимо рассмотреть процесс по этапам исследования в обратном порядке и постараться обнаружить ошибку на этих этапах или повторить исследования с первого этапа, исключив исходные гипотезы как ошибочные.

Но и в этом случае проделанная работа не бесполезна. По крайней мере, результаты предварительного анализа дают дополнительные знания об объекте, а соответственно – почву для генерации новых гипотез о структуре потребителей продукта.

Если получен удовлетворительный вариант классификации, необходимо перейти к пятому действию.

Пятое действие. Производится более детальный анализ полученного варианта классификации (первый уровень оценки качества классификации). Во-первых, анализируются средние значения признаков по классам; во-вторых, анализируется матрица сходства классов; в-третьих, предпринимается попытка содержательной интерпретации (объяснения) полученного разбиения.

Судить о качестве разбиения позволяют некоторые простейшие приемы. Например, сравнение средних значений признаков в отдельных классах со средними значениями в целом по всей совокупности объектов. Если отличие классовых средних от общего среднего значения существенное, то это может являться признаком хорошего разбиения. Оценка существенности различий может быть выполнена с помощью t-критерия Стьюдента.

После этого можно вернуться к третьему действию и попытаться увеличить число классов. Если это действие приведет к положительному результату, то это только повысит качество конечного результата.

Шестое действие. Производится второй уровень оценки качества классификации. Наиболее распространенные функционалы качества разбиения следующие:

1. Сумма квадратов расстояний до центров классов:
2. Сумма внутриклассовых расстояний между объектами:
3. Суммарная внутриклассовая дисперсия:
4. Среднее межклассовое расстояние:

Кроме названных функционалов качество классификации можно оценить и при помощи критерия Хотеллинга T^2 для проверки гипотезы о равенстве векторов средних для многомерной совокупности.

Перечень способов оценки качества классификации можно было бы продолжить. Однако, важнейшим критерием качества остается содержательная интерпретация классификации, производимая исследователем. В конечном

итоге, все десять этапов, приводящих к выделению сегментов, призваны подтвердить и оценить численно гипотезы исследователя о структуре потребителей, и никакая программа не сможет заменить самого исследователя.

Седьмое действие. Это действие связано со вторым действием выбора списка признаков. При выборе исходного списка признаков предполагалась некоторая избыточность их количества. Для того чтобы содержательная интерпретация сегментов стала более лаконичной желательно понизить размерность пространства признаков, исключив избыточные и неинформативные признаки.

После выполнения операции понижения размерности пространства признаков, целесообразно повторить все процедуры со второго действия.

Проделав все необходимые действия и получив сегменты потребителей продукта, можно оценить их мощность и далее использовать эту информацию для выработки решений по структуре предложений продуктов, более точно отвечающей потребностям, т.е. спросу потребителей.

Приведенная выше методика применения МСА для сегментирования потребителей продукта была выведена на основании многократного решения задачи классификации первичных данных анкетных опросов потребителей услуг регионального туристского комплекса. Она является главным элементом общей методики анализа поведения потребителей туристского комплекса.

Литература

1. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. – Финансы и статистика, 2000. – 352с.
2. Сигел Э. Практическая бизнес-статистика. / Пер. с англ. – М.: Издательский дом “Вильямс”, 2002. – 1056с.

3. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО “Речь”, 2000. – 350с.

4. Снапелев Ю.М, Старосельский В.А. Моделирование и управление в сложных системах. – М.: Сов. Радио, 1974. – 264с.

5. Сошникова Л.А. и др. Многомерный статистический анализ в экономике. Уч. пособие. – М.: ЮНИТИ, 1999. – 598с.

6. Токарев Б.Е. Методы сбора и использования маркетинговой информации: Учебно-практ. пособие / Б.Е. Токарев. – М.: Юристъ, 2003. – 254с.