**ФЕДЕРАЛЬНОЕ АГЕНСТВО ПО ОБРАЗОВАНИЮ**

Филиал государственного образовательного учреждения высшего профессионального образования казанского государственного университета имени В.И. Ульянова – Ленина в Г. набенежные Челны

**Факультет прикладной математики и информатики**

Специальность: 080116.65 :Математические методы в экономике

ДОКЛАД

**Применение кластерного анализа для сегментации рынка**

Выполнил:

студент III курса

дневного отделения

группы 40741

Ужогов А. А.

Набережны Челны

2009

Оглавление

1. Сегментация рынка

2. Сущность кластерного анализа

3. Выполнение кластерного анализа

3.1 Формулировка проблемы

3.2 Выбор способа измерениярасстояния или меры сходства

3.3 Выбор метода кластеризации

3.4 Принятие решения о количестве кластеров

3.5 Интерпретация и профилирование кластеров

3.6 Оценка надежности и достоверности

Вывод

**1. Сегментация рынка**

В наше время создание товара не является уже таким беспрецедентным и уникальным событием, как это было раньше. Еще буквально полвека назад создание чего-то нового производило переворот и приносило изобретателям сверхприбыли. Сейчас главным является не создание, а умелое и эффективное продвижение нового товара. Истории известны случаи, когда продукция, которая, по сути, не была уникальной или сверхновой становилась лидером продаж на многие годы, благодаря грамотной маркетинговой стратегии фирмы.

При разработке нового товара необходимо четко знать, на какую потребительскую аудиторию он рассчитан. Нужно знать о своем среднестатистическом потенциальном клиенте практически все, начиная с возраста и пола и заканчивая личными увлечениями. Именно эти сведения дают возможность разработать товар вплоть до мелочей и провести эффективную рекламную кампанию.

Одним из основных направлений маркетинговой деятельности является сегментация рынка, позволяющая аккумулировать средства предприятия на определенном направлении своего бизнеса. Рынок состоит из покупателей, а покупатели отличаются друг от друга по самым разным параметрам. И любой из этих переменных можно воспользоваться в качестве основы для сегментирования рынка.

К настоящему времени в экономической литературе достаточно четко обозначены понятия целевого рынка и целевого сегмента, выделение которых и является основной целью сегментации рынка. Целевой рынок – это потенциальный рынок фирмы, который определяется совокупностью людей со схожими потребностями в отношении конкретного товара или услуги, достаточными ресурсами, а также готовностью и возможностью покупать. Целевой сегмент – это однородная группа потребителей целевого рынка фирмы, обладающая схожими потребностями и покупательскими привычками по отношению к товару фирмы.

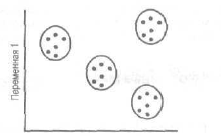
##### Таким образом, сегментирование рынка – это разбивка рынка на четкие группы покупателей, для каждой из которых могут потребоваться отдельные товары и/или комплексы маркетинга.

Целью сегментирования является выделение одной или нескольких целевых групп потребителей, под которых "затачивается" весь комплекс маркетинговых мероприятий - от разработки продуктов и брендинга до выбора тональности и носителей маркетинговых коммуникаций.

Разбивать рынок на сегменты можно разными способами. Можно использовать факторный, кластерный, дискриминантный анализы, можно разбивать и «на глазок», но этот метод пригоден только, для опытных и много знающих маркетологов. Мы рассмотрим только кластерный анализ.

# 2.Сущность кластерного анализа

Кластерный анализ представляет собой класс методов, используемых для классификации объектов или событий в относительно однородные группы, которые называют *кластерами* (clusters). Объекты в каждом кластере должны быть похожи между собой и отличаться от объектов в других кластерах. На рис. 20.1 показана идеальная ситуация кластеризации, когда кластеры четко отделены друг от друга на основании различий двух переменных: ориентация на качество (переменная 1), и чувствительность к цене (переменная 2),

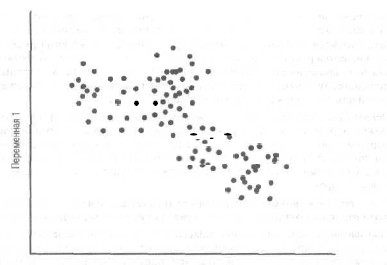


Переменная 2

*Рис. 20,1. Идеальная ситуация*

*кластеризации*

Следует отметить, что каждый потребитель попадает в один из кластеров, и перекрывающихся областей нет. С другой стороны, на рис. 20.2 представлена ситуация кластеризации, которая чаще всего встречается на практике



Переменная 2

*Рис. 20.2. Реальная ситуация кластеризации*

На рис. 20.2 границы некоторых кластеров очерчены нечетко, и отнесение некоторых потребителей к конкретному кластеру не очевидно, поскольку многие из них нельзя сгруппировать в тот или иной кластер. кластерном анализе нет необходимости в предварительной информации о кластерной принадлежности любого из объектов. Группы, или кластеры, определяют с помощью собранных данных, а не заранее. Кластерный анализ используют в маркетинге для различных целей.

# 3. Выполнение кластерного анализа

Этапы выполнения кластерного анализa:

1. Формулировка проблемы
2. Выбор меры расстояния
3. Выбор метода кластеризации
4. Принятие решения о качестве кластеров
5. Интерпритация и профелирование кластеров
6. Оценка достоверности кластеров

## 3.1Формулировка проблемы

Возможно, самая важная часть формулирования проблемы кластеризации — это выбор переменных, на основе которых проводят кластеризацию. Включение даже одной или двух посторонних (не имеющих отношение к группированию) переменных может исказить результаты кластеризации. Задача состоит в том, чтобы выбранный набор переменных смог описать сходство между объектами с точки зрения признаков, имеющих отношение к данной проблеме маркетингового исследования. Переменные следует выбирать, исходя из опыта прошлых исследований, теории или тестируемой гипотезы. Экспериментатор должен обладать интуицией и уметь делать выводы.

## 

## 3.2 Выбор способа измерения расстояния или меры сходства

Цель кластеризация — группирование схожих объектов. Поэтому для того чтобы оценить, насколько они похожи или непохожи, необходимо использовать некую единицу измерения. Наиболее распространенный метод заключается в том, чтобы в качестве такой меры использовать расстояния между двумя объектами. Объекты с меньшими расстояниями между собой больше похожи, чем объекты с большими расстояниями. Существует несколько способов вычисления расстояния между двумя объектами.

Наиболее часто используемая мера сходства— евклидово расстояние или его квадрат. Евклидова метрика это квадратный корень из суммы квадратов разностей в значениях для каждой переменной

Существуют и другие способы измерения расстояния. *Расстояние городских кварталов)* (city-block, или *манхэттенское расстояние* (Manhattan distance) между двумя объектами — это сумма абсолютных разностей в значениях для каждой переменной. *Расстояние Чебышева* (Chebychev distance) между двумя объектами — это максимальная абсолютная разность в значениях для любой переменной. Если переменные измерены в различных единицах, то единица измерения влияет на решение кластеризации. В этих случаях перед кластеризацией респондентов мы должны нормализовать данные, изменив шкалу измерения каждой переменной таким образом, чтобы среднее равнялось нулю, а стандартное отклонение — единице. Хотя нормализация может исключить влияние единицы измерения, она также уменьшает различия между группами по переменным, которые наилучшим образом дискриминируют (отличают) группы или кластеры. Кроме того, желательно удалить выбросы (т.е. случаи с нетипичными значениями). Использование различных способов измерения расстояния ведет к разным результатам кластеризации. Следовательно, целесообразно использовать различные меры сходства и затем сравнить результаты. Выбрав меру сходства, затем можно выбрать метод кластеризации.

## 3.3 Выбор метода кластеризации

Методы кластеризации могут быть иерархическими и неиерархическими. Иерархическая кластеризация (hierarchical clustering) характеризуется построением иерархической, или древовидной, структуры.

Иерархические методы могут быть агломеративными (объединительными) и дивизивными. Агломеративная кластеризация (agglomerative clustering) начинается с каждого объекта в отдельном кластере. Кластеры объединяют, группируя объекты каждый раз во все более и болеекрупные кластеры. Этот процесс продолжают до тех пор, пока все объекты не станут членами одного единственного кластера.

Разделяющая, или дивизивная, кластеризация (divisive clustering) начинается со всех объектов, сгруппированных в единственном кластере. Кластеры делят (расщепляют) до тех пор, пока каждый объект не окажется в отдельном кластере.

Обычно в маркетинговых исследованиях используют агломеративные методы, например методы связи, дисперсионные и центроидные методы. Методы связи (linkage methods) включают метод одиночной связи, метод полной связи и метод средней связи.

В основе **метода одиночной связи** (single method) лежит минимальное расстояние, или правило ближайшего соседа. При формировании кластера первыми объединяют два объекта, расстояние между которыми минимально. Далее определяют следующее по величине самое короткое расстояние, и в кластер с первыми двумя объектами вводят третий объект. На каждой стадии расстояние между двумя кластерами представляет собой расстояние между их ближайшими точками.

**Метод полной связи** (complete linkage) аналогичен методу одиночной связи, за исключением того, что в его основе лежит максимальное расстояние между объектами, или правило дальнего соседа. В методе полной связи расстояние между двумя кластерами вычисляют как расстояние между двумя их самыми удаленными точками.

Метод **средней связи** (average linkage) действует аналогично. Однако в этом методе расстояние между двумя кластерами определяют как среднее значение всех расстояний, измеренных между объектами двух кластеров, при этом в каждую пару входят объекты из разных кластеров.

Широко известным дисперсионным методом, используемым для этой цели, является **Метод Варда (Ward's procedure)** Дисперсионный метод, в котором кластеры формируют таким образом, чтобы минимизировать квадраты евклидовых расстояний до кластерных средних. Для каждого кластера вычисляют средние всех переменных. Затем для каждого объекта вычисляют квадраты евклидовых расстояний до кластерных средних. Эти квадраты расстояний суммируют для всех объектов. На каждой стадии объединяют два кластера с наименьшим приростом в полной внутрикластерной дисперсии.

В **центроидных** методах (centroid method) расстояние между двумя кластерами представляет собой расстояние между их центроидами (средними для всех переменных). **Центроидный метод (centroid method)** Дисперсионный метод иерархической кластеризации, в котором расстояние между двумя кластерами представляет собой расстояние между их центроидами (средними для всех переменных). Каждый раз объекты группируют и вычисляют новый центроид. Изо всех иерархических методов методы средней связи и Варда показывают наилучшие результаты по сравнению с другими методами.

К другому типу процедур кластеризации относятся **неиерахические методы кластеризации** (nonhierarchical clustering), часто называемые методом k-средних. Эти методы включают последовательный пороговый метод, параллельный пороговый метод и оптимизирующее распределение. В последовательном пороговом методе (sequential threshold method) выбирают центр кластера и все объекты, находящиеся в пределах заданного от центра порогового значения, группируют вместе. Затем выбирают новый кластерный центр, и процесс повторяют для несгруппированных точек. После того как объект помещен в кластер с этим новым центром, его уже не рассматривают как объект для дальнейшей кластеризации.

Аналогично работает параллельный пороговый метод (parallel threshold method), за исключением того, что одновременно выбирают несколько кластерных центров и объекты в пределах порогового уровня группируют с ближайшим центром.

Метод оптимизирующего распределения (optimizing partitioning method) отличается от двух изложенных выше пороговых методов тем, что объекты можно впоследствии поставить в соответствие другим кластерам (перераспределить), чтобы оптимизировать суммарный критерий, такой как среднее внутри кластерное расстояние для данного числа кластеров. Два главных недостатка неиерархических методов состоят в том, что число кластеров определяется заранее и выбор кластерных центров происходит независимо. Более того, результаты кластеризации могут зависеть от выбранных центров. Многие неиерархические процедуры выбирают первые *k* случаев *(k —* число кластеров), не пропуская никаких значений в качестве начальных кластерных центров. Таким образом, результаты кластеризации зависят от порядка наблюдений в данных. Неиерархическая кластеризация быстрее иерархических методов, и ее выгодно использовать при большом числе объектов или наблюдений, Высказано предположение о возможности использования иерархических и неиерархических методов в тандеме. Во-первых, первоначальное решение по кластеризации получают, используя такие иерархические методы, как метод средней связи или метод Варда. Полученное этими методами число кластеров и кластерных центроидов используют в качестве исходных данных в методе оптимизирующего распределения. Выбор метода кластеризации и выбор меры расстояния взаимосвязаны. Например, квадраты евклидовых расстояний используют наряду с методом Варда и центроидным методом. Некоторые из неиерархических методов также используют квадраты евклидовых расстояний.

## 

## 3.4 Принятие решения о количестве кластеров

Главный вопрос кластерного анализа — вопрос о количестве кластеров. Здесь нет твердых правил, позволяющих быстро принять решение, но можно руководствоваться следующим.

1. При определении количества кластеров руководствуются теоретическими и практическими соображениями. Например, если цель кластеризации — выявление сегментов рынка, то менеджмент может захотеть получить конкретное число кластеров.

2. В иерархической кластеризации в качестве критерия можно использовать расстояния, при которых объединяют кластеры/

3. В неиерархической кластеризации чертят график зависимости отношения суммарной внутригрупповой дисперсии к межгрупповой дисперсии от числа кластеров. Точка, в которой наблюдается изгиб или резкий поворот, указывает на приемлемое количество кластеров. Увеличение числа кластеров за эту точку обычно безрезультативно.

4. Относительные размеры кластеров должны быть достаточно выразительными.

## 3.5 Интерпретация и профилирование кластеров

Интерпретация и профилирование кластеров включает проверку кластерных центроидов. Центроиды представляют средние значения объектов, содержащиеся в кластере по каждой из переменных. Они позволяют описывать каждый кластер, если присвоить ему номер или метку. Если компьютерная программа кластеризации не выдаст такую информацию, ее можно получить через дискриминантный анализ.

Часто имеет смысл профилировать кластеры через переменные, которые не явились основанием для кластеризации. Эти переменные могут включать демографические, психографические характеристики, использование продукта или другие переменные. Например, кластеры можно вывести, исходя из искомых преимуществ. Дальнейшее профилирование осуществляют через демографические или психографические переменные, чтобы определить маркетинговую стратегию для каждого кластера. Переменные, существенно различающиеся между кластерами, можно идентифицировать дискриминантным анализом и однофакторным дисперсионным анализом.

## 3.6 Оценка надежности и достоверности

Имея несколько умозаключений, выведенных из кластерного анализа, не следует принимать никакого решения по кластеризации, не выполнив оценку надежности и достоверности этого решения. Формальные процедуры оценки надежности и достоверности решений кластеризации достаточно сложны и не всегда оправданы поэтому мы их опустим. Однако следующие процедуры обеспечат адекватную проверку качества кластерного анализа.

1. Выполняйте кластерный анализ на основании одних и тех же данных, но с использованием различных способов измерения расстояния. Сравните результаты, полученные на основе разных мер расстояния, чтобы определить, насколько совпадают полученные результаты.

2. Используйте разные методы кластерного анализа и сравните полученные результаты.

3. Разбейте данные на две равные части случайным образом. Выполните кластерный анализ отдельно для каждой половины. Сравните кластерные центроиды двух подвыборок.

4. Случайным образом удалите некоторые переменные. Выполните кластерный анализ по сокращенному набору переменных. Сравните результаты с полученными на основе полного набора переменных.

5. В неиерархической кластеризации решение может зависеть от порядка случаев в наборе данных. Выполните анализ несколько раз, меняя порядок случаев, до получения стабильного решения.

# Вывод

Кластерный анализ является очень удобным средством для выделения сегментов рынка. В особенности в наш век высоких технологий, когда на помощь человеку приходят машины, и столь трудоемкий процесс становиться буквально секундным делом.

Образование сегментов зависит от имеющихся данных, а не определяется заранее.

Переменные, которые являются основанием для кластеризации, следует выбирать, исходя из опыта предшествующих исследований, теоретических предпосылок, проверяемых гипотез, а также по усмотрению исследователя. Кроме того, следует выбрать соответствующую меру расстояния (сходства). Особенность иерархической кластеризации — разработка иерархической или древовидной структуры. Иерархические методы кластеризации могут быть агломеративными или дивизивными. Агломеративные методы включают: метод одиночной связи, метод полной связи и метод средней связи. Широко распространенным дисперсионным методом является метод Барда. Неиерархические методы кластеризации часто называют методами k-средних. Эти методы включают последовательный пороговый метод, параллельный пороговый метод и оптимизирующее распределение. Иерархические и неиерархические методы можно применять совместно. Выбор метода кластеризации и выбор меры расстояния взаимосвязаны.

Решение о числе кластеров принимают по теоретическим и практическим соображениям. В иерархической кластеризации важным критерием принятия решения о числе кластеров являются расстояния, при которых происходит объединение кластеров. Относительные размеры кластеров должны быть такими, чтобы имело смысл сохранить данный кластер, а не объединить его с другими. Кластеры интерпретируют с точки зрения кластерных центроидов. Часто интерпретировать кластеры помогает их профилирование через переменные, которые не лежали в основе кластеризации. Надежность и достоверность решений кластеризации оценивают разными способами.