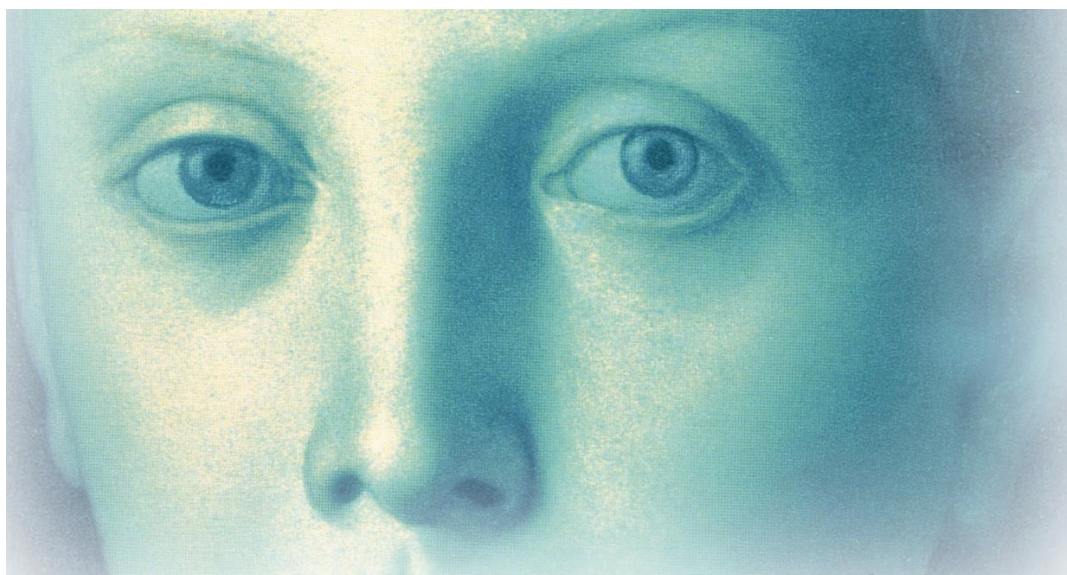




White Paper



Производная важность атрибутов и проблема коллинеарности

М.В.Бодунов
Старший аналитик
Ipsos Russia



В маркетинговых исследованиях часто применяется оценивание важности свойств объектов (продуктов, концептов и т.д.) путем анализа связей соответствующих показателей с некоторым «внешним» критерием (напр., общая оценка концепта или намерение покупать продукт). Если какой-либо показатель тесно связан с критерием, можно говорить о высокой важности этого показателя, и наоборот, если показатель слабо связан с критерием, это может свидетельствовать о его низком весе или важности. Важность, измеренную при помощи такого подхода, называют производной или аналитически выведенной [Цысарь А., 2001]. Считается, что получаемые оценки важности при таком косвенном подходе, оказываются менее зависимыми от субъективного контроля со стороны респондентов, чем прямые оценки важности исследуемых характеристик.

Как правило для оценки связей между изучаемыми показателями и критерием применяется корреляционный и регрессионный анализ. При этом результат сильно зависит от того, насколько тесно взаимосвязаны друг с другом независимые переменные, анализ важности которых проводится. В идеальной ситуации, редко встречающейся на практике, тестируемые показатели не связаны друг с другом (т.е. являются ортогональными по отношению друг к другу). В этом случае соответствующие корреляционные и регрессионные коэффициенты совпадают друг с другом, а сумма их квадратов будет равна объясняемой дисперсии. Квадрат корреляционного (регрессионного) коэффициента, отнесенный к объясняемой дисперсии, будет давать значение относительной (relative) дисперсионной важности соответствующего показателя.¹ Только в этой ситуации относительные важности разных показателей не будут зависеть друг от друга (естественно, при условии константности набора предикторов, поскольку им определяется полная объясняемая дисперсия).

Напротив, в ситуации, когда между изучаемыми показателями наблюдается умеренная или сильная связь, положение существенно образом меняется. Соответствующие корреляционные и регрессионные коэффициенты не будут совпадать и приобретают совершенно разный смысл. Коэффициент корреляции отражает специфический вклад каждого исследуемого показателя (и этот вклад состоит из двух основных компонентов: уникального вклада данного показателя, который никак не связан с другими показателями, включенными в анализ, и общего компонента, связанного с коррелирующими показателями). В отличие от коэффициентов корреляции, регрессионные коэффициенты показывают, какой дополнительный вклад в объяснение зависимой переменной вносит каждый предиктор в комбинации со всеми другими включенными в анализ предикторами. При наличии коллинеарности коэффициенты корреляции могут использоваться в качестве меры вклада в объяснение критерия, если исследователя интересует специфический вклад каждого отдельно взятого предиктора без рассмотрения того, в каких долях латентные переменные участвуют в детерминации изменчивости этих предикторов. Говорить об относительной важности предикторов, опираясь на корреляции, невозможно. На регрессионных коэффициентах ситуация с

¹ Сами коэффициенты (корреляционные или регрессионные) называют индикаторами теоретической важности. Используется также понятие уровневой важности, которое равно увеличению среднего значения критерия при увеличении соответствующего предиктора на единицу. Эта мера популярна в экономике [Johnson, J. W. и др., 2004].

коллинеарностью может сказаться еще более негативно. Коллинеарность может приводить к существенной недооценке вклада высоко коррелирующих показателей. Иногда коллинеарность может проявляться в появлении отрицательных регрессионных коэффициентов для тех показателей, связь которых с критерием может быть только положительная.

В настоящее время используется ряд методов и процедур для улучшения оценок важности путем устранения или компенсации эффекта коллинеарности.

Одним из способов (особенно эффективным в тех случаях, когда число независимых переменных велико) является применение факторного анализа к исследуемым переменным и преобразование исходного набора переменных в новые, взаимно не коррелирующие (ортогональные) индексы.

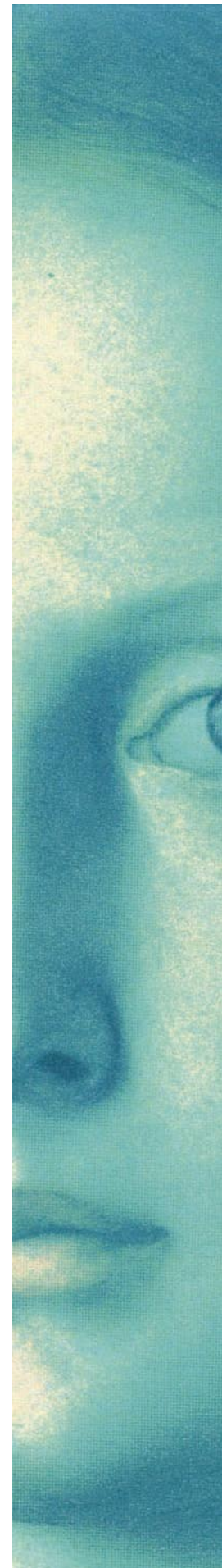
Следующие методы основаны на регрессионном анализе, применяемом к исходному набору исследуемых показателей без их предварительной факторизации:

1. Гребневая регрессия.
2. Методы, рассматривающие все возможные комбинации независимых переменных.
 - a. относительная важность Крускала;
 - b. показатель Шепли;
3. Относительная важность по Джонсону.

Относительная важность определяется как пропорциональный вклад каждой независимой переменной в R^2 , учитывающий как ее прямой эффект (т.е. ее корреляцию с критерием), так и эффект при комбинировании с другими переменными в регрессионном уравнении (Johnson, 2004, с. 240).

Если давать более строгую классификацию методов, то можно назвать три их группы:

1. Сокращение размерности при помощи методов группировки коррелирующих независимых переменных с образованием латентных переменных (напр., регрессионные модели на главных компонентах). К сожалению, эти методы не позволяют интерпретировать (изолировать, оценивать) эффект отдельных исходных переменных при прочих равных условиях. С другой стороны, этот подход позволяет оценить внутреннюю размерность исходных данных и перевести анализ на уровень более консистентных обобщенных переменных (в случае их приемлемой содержательной интерпретации).
2. Корректировка коэффициентов с использованием скалярного параметра пропорционального сжатия (масштабирования), например, гребневая регрессия. Недостатком этого подхода является то, что добавление скалярной величины к элементам главной диагонали ковариационной матрицы не улучшает интерпретируемость. Кроме того, отсутствуют достаточно четкие критерии выбора значения масштабирующего параметра.
3. Трансформация независимых переменных с целью устранения корреляции. Например, ортогонализация переменных-предикторов, которая производится при помощи последовательной регрессии (Kruskal (1987)). При таком подходе





сохраняется возможность интерпретации независимых переменных. С другой стороны, коэффициенты оказываются сильно зависящими от порядка их включения в модель. Для устранения этой зависимости используются приемы усреднения вклада каждой переменной при всех возможных комбинациях с остальными независимыми переменными. Однако при большом числе предикторов проделать это имеющимися программными средствами можно только теоретически. К существенному преимуществу относительных весов (как индикаторов важности) относится то, что они оказываются более стабильными от одного исследования к другому, чем стандартные регрессионные коэффициенты. Кроме того, в отличие от регрессионных коэффициентов, сумма относительных весов равна квадрату множественной корреляции (R^2).

Таким образом, коэффициенты регрессии при наличии корреляции между предикторами оказываются плохими индикаторами влияния (вклада, веса, важности). С другой стороны, альтернативных методов оценки их важности не существует.

Приемлемые результаты можно получить при использовании гребневой регрессии. Оптимальным же приемом следует назвать метод оценки относительного веса (усреднение по всем комбинациям). Имеется эвристическая аппроксимация метода усреднения по всем комбинациям, предложенная Дж.Джонсоном (Johnson, Jeff W, 2004).

Метод оценки относительных весов (важности) включает следующие этапы:

1. Выбирается переменная-критерий, наиболее полно отражающая интегральную привлекательность, успешность, эффективность изучаемого объекта.
2. Определяются потенциальные предикторы, или драйверы, наиболее важные для привлекательности объекта.
3. Оценивается относительная важность отобранного множества предикторов. Относительная важность может выражаться в процентах от общей дисперсии критерия, объясняемой каждым предиктором.

Оценка приближенных значений относительных весов (важности) осуществляется следующим образом:

1. Проводится множественный регрессионный анализ с выбранным критерием в качестве зависимой переменной, и отобранными предикторами в качестве независимых переменных. Регистрируется значение R^2 на этом шага.
2. К матрице корреляций между предикторами применяется метод главных компонент. Число выделяемых компонент равно числу предикторов (т.е. выделяется максимальное число компонент, объясняющих 100% дисперсии). Вращение компонент осуществляется при помощи процедур Varimax или Equamax. После вращения каждая компонента оказывается нагруженной большим весом только одной изучаемой переменной. Производится также оценка значений данных компонент.
3. На заключительном этапе производится вычисление «ортогонализированных» регрессионных коэффициентов. Вначале вычисляются корреляции между значениями факторов и

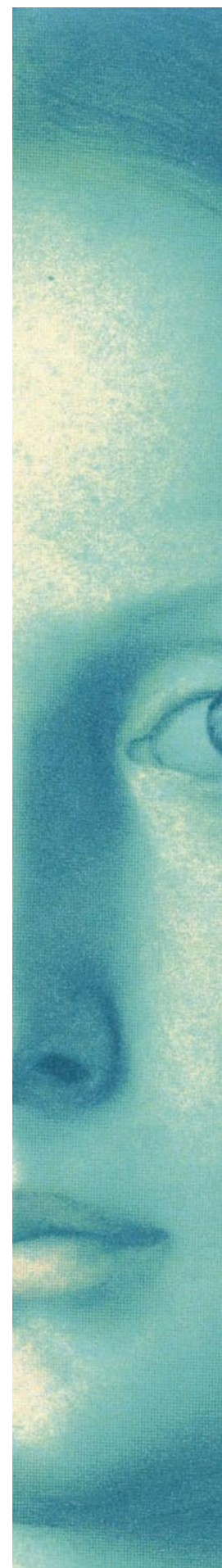
исходными предикторами, а также между значениями факторов и критерием. Затем эти две матрицы (квадратная матрица и вектор) перемножаются. Результирующий вектор представляет собой вектор стандартизованных относительных весов (коэффициентов важности) предикторов. Для удобства каждый коэффициент может быть выражен в процентах от R^2 (поскольку их полная сумма равна R^2).

Для иллюстрации соотношения между разными показателями важности ниже приводятся два примера их использования в типичных задачах маркетинговых исследований. Первый пример относится к тестированию пищевого продукта. В качестве критерия (зависимой переменной) использовалось заявленное респондентом намерение покупать продукт (5-балльная шкала). Оценка важности производилась для следующих интегральных характеристик восприятия респондентом продуктовых свойств (9-балльная шкала): вкус, текстура, качество. Свойства оценивались после того, как респондент пробовал продукт. Намерение покупать продукт оценивалось последним. В таблице 1 приведены корреляции (Пирсона) показателей свойств продукта с критерием (намерением покупать продукт), а также коэффициенты линейной регрессии, коэффициенты гребневой регрессии и относительные веса (важность по Джонсону) для данных показателей. По всем индикаторам важности максимальный вес получил вкус продукта, а минимальный – текстура продукта. Такое сходство рангового положения показателей по индикаторам важности встречается далеко не всегда (при большем количестве показателей очень часто наблюдаются несовпадения рангов, особенно в области средней важности). Большие различия между соответствующими коэффициентами корреляции и коэффициентами обычной линейной регрессии, а также нереалистичный отрицательный регрессионный коэффициент для текстуры говорят о наличии существенной мультиколлинеарности, обусловленной тесными связями между независимыми переменными (показателями свойств продукта). В данной ситуации (довольно типичной) использовать значения корреляций и регрессионных коэффициентов в качестве индикаторов важности нецелесообразно (эффекты разных показателей свойств продукта в них оказываются неразделенными). Гребневая регрессия улучшает соотношение между значениями весов для разных показателей, однако при этом она приводит к заметному недооцениванию этих значений. Относительные веса по Джонсону в качестве индикаторов важности являются оптимальными и с точки зрения их соотношения, и с точки зрения абсолютных значений.

Таблица 1

Показатели свойств продукта	Коэффициенты корреляции	Стандартизованные коэффициенты обычной регрессии	Стандартизованные коэффициенты гребневой регрессии	Относительные веса
Вкус (Аромат)	0.800	0.606	0.287	0.351
Текстура	0.544	-0.012	0.111	0.101
Качество	0.715	0.269	0.211	0.219

N = 850



$$R^2 = 0.671$$

Во втором примере представлены результаты тестирования упаковки для продукта питания. В качестве критерия использовалась оценка респондентом намерения покупать продукт (7-балльная шкала). Различные характеристики упаковки (5-балльная шкала), для которых производилась оценка производной потребительской важности, использовались как независимые переменные. Общее число предикторов было 19, т.е. значительно больше, чем в первом примере. Перед оценкой важности этих характеристик был проведен факторный анализ связей между ними (метод главных компонент, вращение Varimax). Факторизацию целесообразно выполнять во всех случаях, когда оценка выведенной важности производится для большого числа предикторов одновременно (более 10). Этим достигается следующее: выясняется подлинная внутренняя размерность исследуемого множества характеристик; получается важная дополнительная информация для интерпретации соотношений между ними. В ряде работ, рассматривающих способы устранения коллинеарности, факторизация предлагается как один из таких приемов ортогонализации предикторов. И хотя такой способ представляется вполне правомочным, отмечается, что его использование может приводить к потерям информации об уникальном вкладе предикторов в вариацию критерия (т.е. информации о важности предикторов с учетом уникального компонента). Если же исследователей в большей степени интересует именно вклад наиболее общих аспектов изучаемых показателей и в меньшей – уникальный вклад отдельных предикторов, то использование общих факторов при условии их хорошей интерпретируемости становится наилучшим решением проблемы мультиколлинеарности (за исключением случаев с малочисленными выборками). В этой ситуации коэффициенты корреляции предикторов с критерием будут тождественны соответствующим регрессионным коэффициентам, а их квадраты в сумме будут давать объясняемую дисперсию (R^2).

В случае наших 19 показателей методом главных компонент выделяется 2 общих фактора, суммарно объясняющих 62.4% общей дисперсии. Факторы после вращения (представлены в таблице 2) хорошо интерпретируются: первый фактор связан с мотивирующей функцией упаковки, тогда как второй фактор отражает главным образом перцептивные особенности упаковки как таковой, ее запоминаемость. Ряд предикторов оказались полиметричными, т.е. связанными с обоими факторами (напр., «Подходит для дорогого продукта», «Выглядит современно»), что может говорить о необходимости уточнения факторной структуры.² Однако для наших иллюстративных целей данная структура представляется исчерпывающей.

Таблица 2

Характеристики упаковки	F1 (34.3%)	F2 (28.1%)
Подходит для дорогого продукта	0.498	0.447
Вызывает желание попробовать этот продукт	0.620	0.480
Запоминающаяся упаковка		0.707

² В задачу данной статьи не входит детальное рассмотрение проблем, связанных с использованием факторного анализа в маркетинговых исследованиях. Этой теме будет посвящена специальная статья.

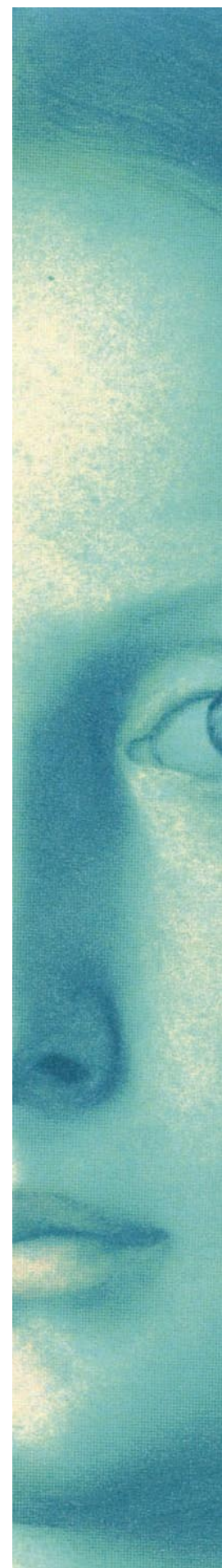
Имеет дешевый внешний вид		-0.619
Выглядит современно	0.475	0.527
Хорошо поставить на стол для гостей	0.661	
Действительно заметна на полке		0.757
Яркий и привлекающий внимание дизайн		0.747
Представляет марку, которой можно доверять	0.722	
Стильная	0.602	0.539
Скучная и унылая		-0.704
Создает впечатление продукта с отличным вкусом	0.660	0.493
Для таких людей как я	0.770	
Стоит потраченных денег	0.825	
Упаковка для натурального продукта	0.788	
Упаковка продукта, за который можно заплатить немного больше	0.653	
Создает впечатление продукта со вкусом свежеприготовленного продукта	0.740	
Упаковка имеет оригинальный, отличный от других дизайн		0.668
Соответствует имиджу премиального продукта	0.625	0.593

Коэффициенты корреляции данных факторов с критерием (намерением покупать продукт) оказались равными 0.566 и 0.418, соответственно (стандартизованные регрессионные коэффициенты имели те же значения, что обусловлено ортогональностью факторов). Квадраты коэффициентов давали следующие значения выведенной важности: 0.320 и 0.175 (при $R^2=0.495$ – 64.7% и 35.3%). Таким образом, при слегка большей мощности первый фактор («Мотивирующая роль упаковки») оказывался почти вдвое более важным с точки зрения влияния на намерение покупать, чем фактор визуально-перцептивных особенностей упаковки.

Рассмотрим индикаторы выведенной важности, полученные для всех 19 показателей (таблица 3). Хотелось бы сразу отметить, что учет уникальных вкладов исследуемых показателей позволил повысить объясняемую вариацию критерия всего на 10% - с 0.495 в случае факторной модели до 0.545 в развернутой модели. Как и в первом примере, из-за сильной мультиколлинеарности (в значительной степени отраженной в факторной структуре предикторов) наблюдаются существенные расхождения между коэффициентами корреляции и регрессионными коэффициентами, а также многочисленные инверсии знаков у последних (причем в обоих направлениях, см. коэффициенты для «Имеет дешевый внешний вид»). Применение гребневой регрессии значительно улучшает картину на фоне сильного подавления общего уровня коэффициентов (для их перевода в индикаторы важности коэффициенты должны быть возведены в квадрат; это не касается 4-го столбца таблицы – относительных весов). Оптимальными свойствами обладает показатель относительного веса по Джонсону.

Показатели, получившие наивысшие оценки важности (в данном примере - «Вызывает желание попробовать данный продукт», «Соответствует имиджу премиального продукта», «Запоминающаяся упаковка», «Стоит потраченных денег», «Для таких людей как я»), можно обозначить как ключевые драйверы.

В практике маркетинговых исследований относительная важность



свойств, характеристик и атрибутов продуктов очень часто рассматривается изолированно, вне какого-то контекста. В теоретических же исследованиях, напротив, оценка важности атрибутов всегда включается в ту или иную систему понятий, при помощи которых описывается поведение потребителей.

Рассмотрим некоторые аспекты современных представлений о соотношении выведенной важности атрибутов потребительских объектов, с одной стороны, и общих характеристик поведения потребителя как целостного субъекта, с другой. Понимание этих соотношений может способствовать разработке более эффективных стратегий позиционирования и продвижения продуктов/товаров/услуг в условиях рынка на современном этапе его развития.

Таблица 3

Показатели свойств продукта	Коэффициенты корреляции	Стандартизованные коэффициенты обычной регрессии	Стандартизованные коэффициенты гребневой регрессии	Относительные веса
Вызывает желание попробовать этот продукт	0.631	0.233	0.103	0.067
Соответствует имиджу премиального продукта	0.642	0.197	0.084	0.050
Запоминающаяся упаковка	0.568	0.165	0.080	0.045
Стоит потраченных денег	0.618	0.196	0.080	0.045
Для таких людей как я	0.615	0.099	0.073	0.041
Упаковка для натурального продукта	0.553	0.100	0.069	0.037
Стильная	0.554	0.048	0.046	0.030
Создает впечатление продукта со вкусом свежеприготовленного продукта	0.558	0.022	0.049	0.029
Упаковка имеет оригинальный, отличный от других дизайн	0.485	0.056	0.051	0.027
Подходит для дорогого продукта	0.463	0.056	0.041	0.023
Упаковка продукта, за который можно заплатить немного больше	0.490	-0.028	0.027	0.022
Создает впечатление продукта с отличным вкусом	0.526	-0.056	0.016	0.020
Представляет марку, которой можно доверять	0.467	-0.078	0.015	0.019
Действительно заметна на полке	0.489	-0.084	0.014	0.018
Яркий и привлекающий внимание дизайн	0.475	0.000	0.020	0.017

Хорошо поставить на стол для гостей	0.463	-0.026	0.016	0.016
Скучная и унылая	-0.391	-0.025	-0.021	0.015
Выглядит современно	0.433	-0.006	0.024	0.014
Имеет дешевый внешний вид	-0.359	0.055	0.001	0.010

N = 300

R² = 0.545

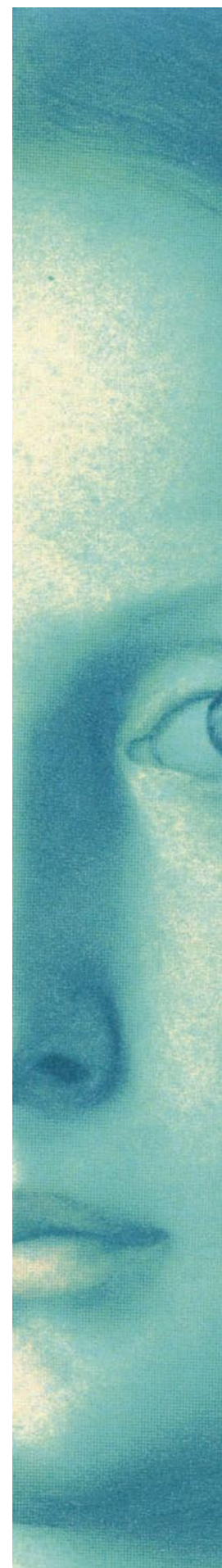
Данные в таблице представлены в порядке убывания относительного веса (столбец 4)

Приписываемая атрибутам (и другим характеристикам) продуктов важность может быть связана с личностными характеристиками разного уровня: от уникальных особенностей индивидуального опыта до таких обобщенных параметров, как ценностные ориентации личности. В общем случае такая связь устанавливается на неосознаваемом уровне благодаря механизму привычных действий. Эта связь выражается (и формируется) в различного вида предпочтениях при совершении актов выбора в континууме покупательского поведения. В случае, когда такая связь реализуется на осознаваемом уровне, ее действие может достигать высших личностных регуляторов индивидуального поведения – общеличных мотивов и ценностей. Благодаря этой связи личность удовлетворяет потребность во внутренней непротиворечивости, согласованности своего поведения. Более важным является то, что действия, согласующиеся с ценностной направленностью личности, могут не требовать какого-то внешнего подкрепления: эти действия сами дают личности то, что она хочет. В силу их абстрактности (идеальности) потенциальная недостижимость целей, обусловленных ценностными ориентациями, не оказывает деструктивное фрустрирующее воздействие на личность. Напротив, такие фрустрирующие (абстрактные) цели могут быть для нас более притягательными. Их надситуативность обеспечивает личности множество путей согласования поведенческих актов (в частности, актов потребительского выбора) со своей ценностной структурой. Все это делает ценности важным и перспективным объектом (и инструментом) в маркетинговых исследованиях.

В настоящее время разработано несколько подходов к диагностике структуры ценностей личности (потребителя).³ К наиболее известным относится Ценностный опросник Рокича (Rokeach Value Survey), Ценностный опросник Шварца (Schwartz Value Survey) и Перечень ценностей Кейла (Kahle's List of Values – LOV). Последний специально разрабатывался для исследования поведения потребителей.

В заключение рассмотрим гипотетические интерпретации представленных в примерах результатов (таблицы 1 и 3) с точки зрения ценностных ориентаций потребителя. В первом примере наибольшую важность получил атрибут «Вкус», вторым по важности оказался атрибут «Качество». Первый атрибут связан с такой ценностной ориентацией, как гедонизм, получение удовольствий от жизни. Второй атрибут имеет отношение к ценности «Самоуважение». По-видимому, данное соотношение важности двух атрибутов косвенно отражает доминирующее положение ценности «Гедонизм» по отношению к ценности «Самоуважение» у потребителей пищевых

³ В данной статье не ставится задача подробного освещения проблемы ценностей, их измерения и использования в маркетинговых исследованиях. Этой теме будет посвящена одна из ближайших статей.





продуктов массового спроса в ситуации индивидуального потребления. Во втором примере наивысшее значение важности выявлено у такой характеристики упаковки, как «Вызывает желание попробовать продукт», также связанной с гедонистической ценностной ориентацией. Второе место по важности заняла характеристика «Соответствует имиджу премиального продукта», релевантная ценности «Уважение со стороны окружающих». Можно предположить, что эти две универсальные ценности – гедонизм и уважение окружающих – являются ведущими при оценивании упаковки пищевого продукта массового спроса. Можно допустить существование двух сегментов с преобладанием одной из этих двух ценностных ориентаций.

Таким образом, рассмотрение значений важности атрибутов в контексте ценностных ориентаций потребителей (даже на гипотетическом уровне) существенно расширяет возможности интерпретации получаемых результатов и понимание механизмов выстраиваемой иерархии атрибутов по важности.

Литература

1. Цысарь А. Обзор существующих методик измерения значимости атрибутов товара/услуги. Маркетинг и маркетинговые исследования, №5, 2001
2. Beatty, S. E.;Kahle, L. R.; Homer, P.M. and Mirsa, S. Alternative Measurement Approaches to Consumer Values: The List of Values and the Rokeach Value Survey. Psychology and Marketing, Vol. 2 No. 3, 1985, 181-200.
3. Frost, Alan. A Consumer Choice Simulation Model for the Development of Product, their Packaging and Advertising. In Ian Greig (Ed.) Brand Choice Modelling, ESOMAR Monograph No 5, 1998, 7-26.
4. Johnson, Jeff W., LeBreton, James M. History and Use of Relative Importance Indices in Organizational Research. Organizational Research Methods, Vol. 7 No. 3, July 2004, 238-257.
5. Kruskal, W. Relative importance by averaging over orderings. The American Statistician, 41, 1987, 6–10.
6. LeBreton, James M., Binning, John F., Adorno, Anthony J., Melcher, Kathleen M. Importance of Personality and Job-Specific Affect for Predicting Job Attitudes and Withdrawal Behavior. Organizational Research Methods, Vol. 7 No. 3, July 2004, 300-325.
7. Overby, Jeffrey W., Woodruff, Robert B., Gardial, Sarah F. The influence of culture upon consumers' desired value perceptions: A research agenda. Marketing Theory, Vol. 5 No. 2, 139–163.
8. Weiner, Jay L., Tang, Jane. Multicollinearity in Customer Satisfaction Research. White Paper. Ipsos Loyalty, 2005.
9. Zeithaml, Valarie A. Consumer Perception of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. J. of Marketing, Vol. 52 No. 3, July 1988, 2-22.

