

# Статистические методы, используемые в маркетинговых исследованиях

**Маргарита  
АКУЛИЧ**  
Эксперт журнала

Статистические методы являются в маркетинговых исследованиях на предприятии наиболее часто используемыми среди количественных методов маркетингового анализа. Конечно, в большинстве случаев расчеты по реализации этих методов проводятся с помощью пакетов прикладных программ. Однако специалистам-исследователям необходимо понимать суть данных методов и область их применения. Краткое изложение большинства применяемых статистических методов приведено ниже.

## ОПИСАТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ

Под описательной статистикой или разведочным анализом данных принято понимание статистических методов обработки данных, осуществления их систематизирования, представления в наглядной форме в виде графиков и таблиц, а также количественного описания данных посредством системы статистических показателей.

В анализе описательного вида принято представление информации об анализируемом явлении в агрегированном (обобщенном) виде. Для этого прибегают к использованию методов: табличного; графического; исчисления статистических показателей.

Представление информации в рассматриваемом анализе осуществляется в виде графиков и таблиц (частотных и сопряженности). Количественная информация обобщается посредством графиков, а также показателей: среднего уровня (среднего значения, моды, медианы) процентилей; вариации (размаха, межквартильного размаха, дисперсии, стандарт-

ного отклонения, коэффициента вариации и др.); формы распределения (асимметрии и эксцесса).

Исследование данных ориентируется на такие оценки показателей, как интервальные и точечные.

Описательный анализ данных предусматривает, что при его реализации можно не только анализировать данные, но и выбирать аналитические методы их последующего анализа, к примеру, методы, которые позволяют проверять гипотезы и моделировать взаимосвязи.

В маркетинге описательный метод широко применяется для проведения презентаций и подготовки аналитических отчетов.

## ГРАФИЧЕСКИЙ МЕТОД

Под графическим методом принято понимание метода, содействующего агрегированию данных при первичном описательном анализе. График представляется в виде чертежа, показывающего соотношение конкретных данных на основе примене-

ния таких средств как изобразительные средства и геометрические образы. Графики содействуют представлению статистических данных в наглядном виде.

Графики могут быть представлены в виде диаграмм и статистических карт. Диаграммы принято подразделять на виды согласно задачам осуществляемого анализа. Они могут выступать в качестве диаграмм: динамики, взаимосвязи, сравнения и структуры. Благодаря диаграммам сравнения оказывается возможным изображение статистических данных, характеризующих разные территории (регионы, страны), объекты. Если применять диаграммы структуры, то можно наблюдать структуру и составные части изучаемой совокупности. Посредством диаграмм динамики анализируется развитие интересующих исследователя явлений во временном аспекте. На базе диаграмм взаимосвязи происходит изучение имеющих между данными зависимостей. С помощью статистических карт отображают статистические данные на географических территориях.

Особенности графиков зависят от тех задач, которые решаются благодаря им. Строить графики в настоящее время необходимо с применением пакетов программ для статистического профессионального анализа, обладающих огромными возможностями. Они позволяют строить сотни модификаций диаграмм, которые бывают: столбиковыми; секторными; линейными и т.д.

Нередко применение графического метода происходит, когда этот метод дополняет методы аналитические, к примеру, метод, с помощью которого анализируются взаимосвязи, или метод, выявляющий различия в группах. Но у графических средств имеются свои особенные преимущества, касающиеся выявления закономерностей, которые сложно описать количественно и проблематично обнаружить на базе прибегания к аналитическим процедурам.

### ТАБЛИЧНЫЙ МЕТОД

Под табличным методом принято понимание метода агрегирования имеющихся данных, которое проводится на этапе осуществления первичного описательного анализа. В качестве статистической таблицы выступает система столбцов и строк. В данной системе логично и последовательно излагают статистическую информацию об исследуемом процессе либо явлении. Табличное представление данных в маркетинге – это представление удобное и практичное, которое

обеспечивает наглядность информации. Но таблица должна отличаться высоким качеством, для этого важно ее правильное построение и оформление.

Табличный метод – это метод универсальный. Его можно применять в разных направлениях маркетинга. Прежде всего, он служит цели обеспечения удобства представления массивов данных об анализируемых объектах. Для этого прибегают к применению промежуточных таблиц (в которых представляют промежуточные итоги исследований) и сводных (в которых дается представление финальных данных). Представление посредством таблиц результатов маркетинговых исследований, опросов, группировок и сводок данных делают отчет об исследовании более наглядным, профессиональным, убедительным.

Для построения разных (сложных и простых) таблиц целесообразно прибегать к использованию: профессиональных статистических пакетов, например, таких как IBM SPSS Statistics; программы MS Excel. Благодаря пакетам и программам можно при минимуме временных затрат строить таблицы, осуществлять обработку данных, составлять отчеты по маркетинговым исследованиям.

### ВЫБОРОЧНЫЙ МЕТОД

В качестве выборочного метода в статистике принято понимание метода исследования общих свойств всей совокупности (генеральной совокупности или выборки) каких-то объектов на базе анализа свойств только части данных объектов (выборочной совокупности). Применение этого метода может быть целесообразно из-за: большой обширности объекта исследования (к примеру, когда изучаются потребительские предпочтения на рынке); необходимости в сборе первичной маркетинговой информации в рамках «пилотных» исследований.

При проведении выборочных обследований следует обеспечивать: количественную характеристику выборки либо определять минимальное количество объема выборки (наблюдений) для осуществления исследования; качественную характеристику выборки либо способов и методов, посредством которых происходит формирование выборочной совокупности.

Выборочное обследование предусматривает, что нужно добиваться минимальности объема выборки при максимуме точности описания генеральной совокупности на базе выборочных данных. Поэтому





выборке положено отличаться репрезентативностью (представительностью), чтобы объективно отражать свойства, имманентные генеральной совокупности.

Достижение точности итогов выборочных обследований возможно на основе прибегания к применению отличающихся сложностью методов, содействующих формированию выборок (это кластерный отбор, задание расслоения, использование отбора вероятностно-пропорционального вида, простой случайный либо неслучайный отбор, повторный или бесповторный отбор).

Значение показателя минимального объема выборки обуславливается многими параметрами, принадлежащими оцениваемому показателю или системе показателей (способ и методы формирования выборки, заданная надежность результатов, вариация исследуемых данных, максимально допустимая ошибка в оценке показателей). Данный показатель определяют на базе статистических методов либо экспертно.

### КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

В качестве корреляционного анализа принято понимание статистического метода анализа взаимосвязи между случайными переменными-величинами в количестве двух и более. Случайные переменные-величины – это свойства изучаемых объектов наблюдения, являющиеся измеряемыми.

При проведении корреляционного анализа предусматривается исчисление коэффициентов корреляции, принимающих обычно либо отрицательные, либо положительные значения. По знаку коэффициента судят о направлении имеющейся связи, а по абсолютному значению – силу имеющейся связи.

Для исчисления коэффициентов корреляции применяют способ расчета, зависящий от шкалы измерения тех из переменных, взаимосвязь между которыми исследуют.

Если переменные измерены в шкале количественного типа (шкала отношений либо шкала интерваль-



ная), то обеспечивается расчет ковариации или корреляционного момента, а на его базе – исчисление линейного коэффициента корреляции (коэффициента Пирсона).

Для оценивания направления связи между изучаемыми переменными, измерение которых производилось в шкале порядковой, прибегают к использованию непараметрических ранговых коэффициентов корреляции: коэффициента ранговой корреляции Кендалла и коэффициента корреляции Спирмена. Применение нередко находят и исчисления коэффициентов: Фехнера (корреляция знаков), конкордации (множественная ранговая корреляция). Между переменными дихотомическими также измеряют связи посредством соответствующих метрик.

Исчислять коэффициенты корреляции относительно нетрудно, к тому же их легко интерпретировать. Применять их могут даже люди специально не подготовленные. Однако у корреляционного анализа имеется своя специфика и методика. Необходимо соблюдать предпосылки исчисления каждого из коэффициентов корреляции, а также проверять их значимость, которая базируется на принципе проверки статистических гипотез, необходимо также правильно строить интервальные оценки коэффициентов. Необходимо также помнить, что иногда исследователям приходится сталкиваться с «ложными корреляциями», приводящими к ложным выводам. Поэтому нужно практиковать расчет не только общих коэффициентов корреляции, но и частных.

На базе корреляционного анализа невозможно определение формы связи между переменными и предсказание значения одной зависимой переменной по одной либо ряду переменных независимых. Если говорить о переменных количественных, то для решения этой задачи можно прибегнуть к применению линейного регрессионного анализа. В маркетинге корреляционный анализ в основном применяют в альянсе с анализом регрессионным. Посредством первого вида анализа определяют наличие связи и степень ее тесноты, а регрессионный анализ служит для математического моделирования формы связи.

### ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Регрессионный анализ понимается в качестве метода исследования статистической взаимосвязи между одной количественной зависимой переменной (результатирующей) и одной либо ряда количественных

независимых переменных-факторов (предикторов или объясняющих переменных).

Уравнение регрессии, по сути, является выражением взаимосвязи между усредненным значением переменной результирующей и усредненными значениями переменных объясняющих. Оно является математической функцией, подбираемой на базе статистических исходных данных зависимой переменной и предикторов. Чаще всего прибегают к использованию функции линейного вида и осуществляют линейный регрессионный анализ.

Имеет место очень сильная связь регрессионного анализа с анализом корреляционным, предусматривающим изучение направления и тесноты связи между количественно выраженными переменными. В анализе регрессионном исследуют форму зависимости между данными переменными. То есть при применении обоих методов осуществляют исследование одной и той же взаимосвязи, но с различных сторон, эти методы в принципе являются взаимодополняющими. Причем практикуется выполнение корреляционного анализа производить перед проведением анализа регрессионного. После того как посредством корреляционного анализа доказано наличие взаимосвязи, исследователь может обеспечить выражение форму данной связи, применив регрессионный анализ.

Целью регрессионного анализа является цель предсказания (прогнозирование) ожидаемого усредненного значения результирующей переменной посредством соответствующего уравнения. В маркетинге довольно часто необходимо прогнозировать разные важные показатели (к примеру, объема продаж или прибыли, или числа клиентов и т.д.).

### НЕЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Под нелинейной регрессией принято понимать регрессионную модель зависимости переменной результирующей от одной либо нескольких переменных объясняющих, выражаемую в виде нелинейной функции. Нелинейная модель (как и линейная) может быть парной и множественной.

Линейная регрессия, согласно ее целям и задачам, подобна регрессии линейной. Отличие обуславливается лишь формой связей и методами оценивания параметров. Выбрать форму связи зависимости нелинейного вида можно посредством: содержательного изучения исследуемого конкретного явления; опо-

ры на итоги изучения взаимосвязи между переменными, к примеру, с применением графического метода.

Оценка параметров нелинейной регрессии может базироваться: на линеаризации уравнения благодаря подходящим преобразованиям и оценки его параметров посредством применения метода наименьших квадратов; оценке параметров на базе метода максимального правдоподобия и применения процедур оптимизационных методов.

Нелинейные регрессии различают согласно: включаемым в эти регрессии предикторам (но они являются линейными по параметрам); включаемым в регрессии предикторам и подвергаемым оценке параметрам.

Если функции являются нелинейными по переменным объясняющим, возможно сведение их к линейным посредством замены переменных. Если функции являются нелинейными по переменным-факторам и подвергаемым оцениванию параметрам, их сведение к линейным моделям происходит благодаря логарифмированию и замене переменных. Если подобрать линеаризующее преобразование невозможно, для оценивания параметров прибегают к использованию методов нелинейной оптимизации на базе исходных данных.

Наилучшая нелинейная модель обычно выбирается на базе наименьшей стандартной остаточной ошибки, исчисленной для разных моделей. Если имеет место наличие ряда нелинейных моделей с сопоставимой точностью, рекомендуется останавливать выбор на модели, отличающейся большей простотой.

## КАТЕГОРИАЛЬНАЯ РЕГРЕССИЯ

В качестве категориальной регрессии принято рассмотрение статистического метода моделирования взаимосвязи между категориальными переменными: зависимой и независимыми. Для построения модели рассматриваемой регрессии прибегают к шкалированию либо оцифровке переменных путем присвоения категориям число-



Целью регрессионного анализа является цель предсказания (прогнозирования) ожидаемого усредненного значения результирующей переменной посредством соответствующего уравнения. В маркетинге довольно часто необходимо прогнозировать разные важные показатели.

вых значений. После этого идет построение оптимального уравнения линейной регрессии относительно преобразованных новых переменных.

В данной модели и переменная и ее предикторы – категориальные.

Применение категориальной регрессии в маркетинге происходит, когда нужно описать покупательскую удовлетворенность в зависимости от таких факторов, как простота совершения покупки, цена, качество товара. Посредством уравнения категориальной регрессии возможно прогнозирование уровня покупательской удовлетворенности в зависимости от всевозможных сочетаний значений категориальных независимых переменных.

## ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Под логистической регрессией или логит-регрессией понимают статистическую модель, используемую, чтобы предсказывать вероятности возникновения какого-то события посредством логистической функции. Эта регрессия относится к классу моделей бинарного выбора, в которых зависимая переменная является дихотомической (бинарной).





Зависимой переменной могут приниматься только значения в количестве двух и означать, к примеру, принадлежность к конкретной группе (скажем, к группе надежных клиентов либо ненадежных клиентов), совершаемое действие (покупка либо непокупка товара), ответы «да» либо «нет» (рекламный ролик нравится либо не нравится).

Построение обыкновенной регрессионной модели линейного типа с зависимыми бинарными переменными не допускается ввиду того, что если это построение будет иметь место, предсказанные значения зависимой переменной интерпретировать окажется практически невозможно.

Измерения значений факторов в моделях бинарного выбора – только количественные. В эти модели допускается включение категориальных переменных (выступающих в качестве факторов). В данных моделях обеспечивается построение регрессионной модели зависимости с принятием во внимание вероятности, что результативной дихотомической переменной будет принято

значение 0 или 1, если значение факторов заданное.

Для того чтобы смоделировать вероятность зависимой дихотомической переменной, нужно произвести подбор специальной монотонно возрастающей функции, значения которой могут варьироваться лишь от 0 до 1.

В моделях бинарного выбора в качестве специальной функции может быть выбрана функция: 1) логистическая; 2) стандартного нормального распределения.

Если модель бинарного выбора построена на базе логистической функции, то она рассматривается как логистическая регрессия или логит-модель. Если модель бинарного выбора построена на базе функции стандартного нормального распределения, то ее рассматривают как пробит-модель.

Посредством логистической регрессии осуществляется прогнозирование вероятности отклика для зависимой переменной от переменных независимых, которые включены в модель. Прогнозные значения вероятности можно использовать для разделения наблюдений на две группы.

При построении модели логистической регрессии можно осуществить отдельный анализ – анализ Receiver Operator Characteristic (ROC-кривых). Посредством данного анализа можно осуществить выбор оптимального порогового значения вероятности для классификации. ROC-кривую используют, чтобы представить результаты бинарной классификации и оценки уровня ее эффективности.

Использование логистической регрессии распространяется на решение задач, связанных с моделированием взаимосвязи и классификацией наблюдений. Она находит применение в скоринге: банковском (на ее основе возможно построение рейтинга заемщиков и управления кредитными рисками); потребительском (для моделирования потребительского поведения).

### МУЛЬТИНОМИАЛЬНАЯ ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

В качестве мультиномиальной логистической регрессии рассматривают общий случай модели логистической регрессии, в ней у зависимой переменной имеется более двух категорий.

Измерение зависимой переменной (ковариаты) в рассматриваемой регрессии возможно в таких шкалах, как порядковая и номинальная. В качестве нее может выступать переменная потребительского выбора торговой марки. Переменные независимые (факторы) могут быть количественными либо категориальными.

В данной модели для каждой из категорий переменной зависимой предусматривается построение уравнения логистической бинарной регрессии. Причем одной из категорий переменной зависимой отводится роль переменной опорной, и происходит сравнение с ней всех других категорий.

Посредством уравнения мультиномиальной логистической регрессии прогнозируется показатель вероятности принадлежности к каждой категории зависимой переменной согласно значениям переменных независимых.

### ПРОБИТ-МОДЕЛЬ РЕГРЕССИИ

Пробит-модель является статистической моделью бинарного выбора, используемой для того чтобы предсказывать вероятность возникновения какого-то события на базе функции нормального стандартного распределения.

Модель пробит-регрессии, подобно модели логистической регрессии, относят к виду моделей бинар-

ного выбора. По этой причине задачи ее построения и функции такие же, как в логит-модели.

В модели пробит-регрессии выражение расчетного значения зависимой переменной выступает в качестве значения функции нормального стандартного закона распределения. Пробит является значением, для которого исследователи вычисляют функцию нормального стандартного распределения. Имеет место зависимость значения пробита от комбинированных линейных значений факторных переменных. Для пробит-модели (также как и для логит-модели) зависимая переменная – дихотомическая. К факторам в пробит-модели предъявляется требование, чтобы они были количественно выраженными либо категориальными, но преобразованными в переменные дихотомические.

Применение пробит-модели относительно сферы аналогично применению логистической регрессии. Если осуществить моделирование и классификацию по пробит-модели и также по модели логистической регрессии, то результаты окажутся весьма сходными. Но в некоторых случаях результаты могут различаться.

### РЕГРЕССИЯ КОКСА

Регрессионную модель Кокса считают статистической моделью зависимости функции риска от переменных факторов независимого вида.

Регрессию Кокса рассматривают в качестве модели отличающихся пропорциональностью рисков. Благодаря ей прогнозируют риск наступления события для какого-то объекта и оценивают влияние определенных заранее независимых предикторов (переменных) на данный риск. Риск рассматривают в качестве зависящей от времени функции. Риск не является вероятностью, поэтому его значения могут превышать единицу.

Объектом может быть клиент, для которого в маркетинге практикуется прогнозирование риска наступления некоего события. Объект находится в поле зрения априори (т.е. его постоянно наблюдают), в любой временной отрезок возможно наступление события, приводящее к его выбытию из группы риска. К примеру, таким событием может оказаться отказ клиента от товара либо услуги компании или его неспособность оплаты кредита.

Выбор метода анализа с применением регрессии Кокса (анализа выживаемости) определяет наличие/отсутствие: цензурированных данных; независимых



предикаторов-переменных. Если независимых переменных нет, но имеются цензурированные данные, то оценить функцию дожития можно, применив таблицы дожития либо процедуры Каплана-Мейера. Если независимые переменные есть, но нет цензурированных данных, то оценить функцию дожития можно, прибегнув к применению нелинейной либо линейной регрессии (она применяется чрезвычайно редко). Если есть и независимые переменные, и цензурированные данные, то оценить функцию дожития и ее производные можно посредством регрессии, прибегнув к применению регрессии Кокса.

Применение модели пропорциональных рисков Кокса возможно, если для рисков выполняется условие пропорциональности. Это условие может считаться выполненным, когда для заданных исследователем двух наблюдений с разными по значениям независимыми переменными соотношение их функций интенсивности независимо от фактора времени. По значениям оценки функций риска не должны быть непостоянными во времени.

Использование модели Кокса с зависящими от фактора времени ковариатами распространяется на случаи, если условие пропорциональности не выполняется.

Регрессионную модель Кокса представляют в качестве зависимости функции риска в момент времени ( $t$ ) в виде сомножителей, которыми являются: базовая функция интенсивности, отражающая некий естественный риск, который не зависит от переменных независимого вида; значения функции риска, которую объясняют ковариатами.

При отсутствии в модели независимого вида переменных можно наблюдать совпадение базовой функции интенсивности с оценкой функции риска.

## АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Под анализом временных рядов принято понимание совокупности статистических методов, посредством которых выявляются составляющие временного ряда, и производится прогнозирование.

К временному ряду (или ряду динамики) относят последовательно размещенные статистические данные о каких-то параметрах изучаемого процесса, собранные в различные временные моменты. Каждое из значений временного ряда рассматривается в качестве его уровня. При изучении временных рядов для каждого из уровней указывают время и его номер согласно порядку.

У метода исследования временных рядов в сравнении с анализом данных, согласно простой выборке, имеются существенные отличия. Анализ временного ряда состоит не только в обеспечении статистических характеристик временного ряда, но и в учете взаимосвязи измерений с течением времени.

Необходимость в анализе временных рядов возникает обычно из-за изменения некоего показателя. В маркетинге зачастую таким показателем является объем продаж или товарных запасов, которые необходимо спрогнозировать на будущие временные периоды.

При анализе значений временного ряда необходимо прийти к пониманию, какие компоненты его формируют. Требуется построение математической модели для каждого из компонент либо их совокупности. К составляющим временного ряда принято причисление тренда и составляющих: сезонной, циклической и случайной. Тренд, сезонная и циклическая составляющие временного ряда образуют его составляющую неслучайную, присутствие которой имеет место во всех временных рядах. Присутствие же составляющей случайной не является обязательным.

Моделирование временных рядов возможно по направлениям: моделирования составляющей неслучайной в совокупности; разложения временного ряда на компоненты (составляющие) и моделирования значений каждой из компонент по отдельности.

Принято деление статистических методов прогнозирования на: аналитические (методы экстраполяции на базе кривых роста в виде временных функций); алгоритмические (метод скользящей средней – взвешенной и простой). Если во временном ряду имеется циклическая, либо сезонная компонента, практикуется проведение анализа периодических колебаний, либо спектрального анализа.

Временные ряды бывают стационарными и нестационарными. Если ряд стационарный, практикуется применение особых моделей: скользящего среднего (СС-модели), авторегрессии (AR-модели); смешанной (ARCC).

Для прогнозирования прибегают и к применению адаптивной модели. Для исследования временных рядов многофакторных и построения прогнозов возможно использование обычных регрессионных моделей.

Прогнозирование в маркетинге применяют в целях планирования, оно поэтому должно быть надежным.