



Министерство образования Республики Беларусь

Учреждение образования  
«Гомельский государственный технический  
университет имени П. О. Сухого»

Кафедра «Маркетинг»

**Н. В. Снопок**

# **МАРКЕТИНГОВЫЕ ИССЛЕДОВАНИЯ**

**КУРС ЛЕКЦИЙ**

**по одноименной дисциплине для студентов  
специальности 1-26 02 03 «Маркетинг»**

**дневной и заочной форм обучения**

**В двух частях**

**Часть 2**

**Электронный аналог печатного издания**

**Гомель 2009**

УДК 339.138(075.8)  
ББК 65.291.3я73  
С53

*Рекомендовано к изданию научно-методическим советом  
гуманитарно-экономического факультета ГГТУ им. П. О. Сухого  
(протокол № 4 от 26.12.2007 г.)*

Рецензент: зав. каф. менеджмента и экономики ГФ УО ФПБ МИТСО канд. экон. наук,  
доц. *Л. М. Латицкая*

**Снопко, Н. В.**  
С53 Маркетинговые исследования : курс лекций по одной дисциплине для студентов специальности 1-26 02 03 «Маркетинг» днев. и заоч. форм обучения. В 2 ч. Ч. 2 / Н. В. Снопко. – Гомель : ГГТУ им. П. О. Сухого, 2009. – 129 с. – Систем. требования: PC не ниже Intel Celeron 300 МГц; 32 Mb RAM; свободное место на HDD 16 Mb; Windows 98 и выше; Adobe Acrobat Reader. – Режим доступа: <http://lib.gstu.local>. – Загл. с титул. экрана.

ISBN 978-985-420-879-4.

Изложены теоретические сведения по разделам курса «Маркетинговые исследования». Представлены характеристики методов сбора данных, необходимые материалы для подготовки, проведения и анализа результатов маркетинговых исследований.

Для студентов специальности 1-26 02 03 «Маркетинг» дневной и заочной форм обучения.

УДК 339.138(075.8)  
ББК 65.291.3я73

ISBN 978-985-420-879-4 (ч. 2)  
ISBN 978-985-420-766-7

© Снопко Н. В., 2009  
© Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», 2009

## ПРЕДИСЛОВИЕ

Проведение маркетинговых исследований – достаточно сложная и трудоемкая процедура. Она основана не только на сборе необходимой информации для решения возникшей проблемы, но, как показывает опыт, в первую очередь на правильно разработанной гипотезе, выборе необходимого и целесообразного для данной проблемы метода исследования. Без учета теоретических наработок, изложенных в данном курсе, студент легко может запутаться во фрагментах информации и не сможет оценить общую картину в целом. Понимание целостности структуры процесса маркетинговых исследований, методов, применяемых при их проведении, необходимо будущим специалистам по маркетингу.

**В результате изучения дисциплины «Маркетинговые исследования» студент должен**

*иметь представление:*

- о сути и необходимости маркетинговых исследований;
- о корректной постановке проблемы;
- о методах сбора требуемой информации;
- о этапах сбора информации для решения проблемы;
- об анализе данных, и полученных результатах;

*знать:*

– основные понятия маркетинга, касающиеся маркетинговых исследований;

- основные элементы комплекса маркетинга;
- понятие рынка и его классификацию;
- понятие целевого рынка и методов его выбора;
- существующие методы определения потребностей и проблем;
- способы классификации маркетинговой информации;
- методы сбора необходимо маркетинговой информации;
- методы анализа и обработки полученной информации;

*владеть:*

– методами отбора и анализа необходимой для исследований вторичной информации;

- методологией сбора маркетинговой информации;
- технологией анализа маркетинговой информации;
- статистическими методами обработки полученных данных исследования;

*уметь использовать:*

– полученные результаты маркетинговых исследований для обоснования принимаемых к реализации управленческих решений по устранению выявленных проблем маркетинговой деятельности организации.

# Тема 1. ТИПЫ ВЫБОРОК И ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ

## 1.1. Этапы проектирования выборки

После того как точно определили задачу и заручились приемлемыми для ее решения схемой исследования и инструментами сбора данных, наступает следующий этап исследовательского проекта, который заключается в отборе тех элементов, которые будут подвергаться статистическому обследованию. Обследованию можно подвергнуть каждый элемент данной совокупности (популяции), произведя полную ее перепись. Полная перепись совокупности называется **цензом**.

**Совокупность (популяция)** – множество элементов, удовлетворяющих некоторым заданным условиям.

Существует и другая возможность – выборка, когда статистическому обследованию подвергается лишь некоторая часть совокупности.

**Выборка** – совокупность элементов подмножества большой группы объектов. По данным, полученным на этом подмножестве, делаются определенные выводы касательно всей группы. Возможность распространения выборочных данных на большую группу зависит от метода, с помощью которого была произведена выборка. Понятие совокупность (популяция) может относиться не только к людям, но и к фирмам, работающим в промышленности, к организациям розничной или оптовой торговли или даже к товарам, производимым на предприятии, и это понятие определяется как все множество элементов, удовлетворяющих некоторым заданным условиям.

**Основа (база) выборки** – перечень элементов, из которых будет производиться выборка. Она может состоять из территориальных единиц, организаций, лиц и других элементов.

Выборочный метод эффективнее переписи всей совокупности по следующим причинам. *Во-первых*, полное обследование совокупности даже сравнительно небольшого размера требует очень больших материальных и временных затрат. *Во-вторых*, к моменту завершения переписи и обработки данных, полученная информация часто устаревает. *В-третьих*, в некоторых случаях ценз просто невозможен. *Например*, если необходимо проверить соответствие реального срока службы электрических ламп расчетному, необходимо держать их во включенном состоянии до момента выхода из строя. Если исследо-

*вать таким образом весь запас ламп, будут получены достоверные сведения, но торговать уже будет нечем.*

*В-четвертых, проведение переписей требует привлечения большого штата сотрудников, что влечет за собой возрастание вероятности появления систематических ошибок.*

Существует следующая последовательность, которой можно придерживаться при составлении выборки:

I этап. Задание популяции.

II этап. Определение основы выборки.

III этап. Определение процедуры отбора.

IV этап. Определение объема выборки.

V этап. Отбор элементов выборки.

VI этап. Обследование отобранных элементов.

VII этап. Обследование отобранных элементов.

**Первый этап.** Прежде всего, необходимо задать совокупность или набор элементов, которые необходимо обследовать. *Например, при изучении предпочтений детей исследователям необходимо решить, будет ли обследуемая популяция состоять только из детей, только из родителей, или из тех и других.*

*Одна компания апробировала свои электрические беспроводные «гонки» только на детях. Детей они привели в полный восторг. Родители отнеслись к новинке иначе. Мамам не понравилось то обстоятельство, что аттракцион не приучает детей к бережному отношению к машинам, а пап не устраивало то, что серьезная вещь превратилась в пустую игрушку.*

*Возможна и обратная ситуация. Некая фирма приступила к производству нового продукта питания и развернула общенациональную рекламную кампанию, в которой основная роль была отведена не по годам развитому ребенку. Фирма проверяла действенность рекламных роликов только на матерях, которые млели от восторга. Дети же сочли этого «акселерата», а вместе с ним и сам рекламируемый продукт, противным. Продукту пришел конец.*

В начале исследования необходимо определиться с тем, из кого или из чего будет состоять соответствующая совокупность: из индивидов, семей, торговых фирм, определенных организаций и т. д. При этом необходимо также определиться и с элементами, которые должны быть исключены из популяции. Должна производиться как временная, так и географическая привязка элементов, на которую в ряде случаев могут налагаться дополнительные условия или ограничения.

*Например, если речь идет об индивидах, искомая популяция может состоять только из лиц старше 18 лет, или только из женщин, или только из лиц с высшим образованием.*

Задача определения географических границ для целевой популяции может представлять особую проблему при международных маркетинговых исследованиях, поскольку при этом возрастает неоднородность рассматриваемой системы. *Например, относительное соотношение городских и сельских территорий может существенно изменяться от страны к стране. Территориальный аспект может оказывать серьезное влияние на состав населения и в пределах одной страны.*

Чем проще определяется целевая популяция, тем выше ее охват и тем легче и дешевле процедура формирования выборки.

**Охват** – это выраженная в процентах доля элементов популяции или группы, которые удовлетворяют условиям включения в состав выборки.

Охват прямо влияет на временные и материальные затраты, необходимые для проведения обследования. Если охват большой (т. е. большая часть элементов популяции удовлетворяет одному или нескольким простым критериям, используемым для выявления потенциальных респондентов), временные и материальные затраты, потребные для сбора данных, сводятся к минимуму. И, наоборот, с увеличением количества критериев, которым должны удовлетворять потенциальные респонденты, возрастают и материальные, и временные издержки, потребные для их выявления.

**Второй этап** процесса отбора выборки состоит в определении основы выборки, которая является перечнем элементов, из которых будет производиться выборка. *Например, пусть целевой совокупностью какого-то исследования являются все семьи, проживающие в пределах Минска. Хорошей и легкодоступной основой выборки может стать телефонный справочник Минска. Тем не менее, при более внимательном рассмотрении становится очевидным, что содержащийся в справочнике список семей не вполне корректен, т. к. номера некоторых семей в нем опущены (разумеется, в него не входят и семьи, не имеющие телефона), некоторые же семьи имеют по несколько телефонных номеров. Лица, недавно поменявшие место жительства и, соответственно, номер своего телефона, также не присутствуют в справочнике.*

Поэтому точное соответствие между основой выборки и интересующей целевой совокупностью наблюдается очень редко. Один из

наиболее творческих этапов работы при выборочном контроле – это определение подходящей основы выборки в тех случаях, когда составление списка элементов совокупности вызывает затруднения. Подобные ситуации могут возникать при выборочном наблюдении территориальных зон или организаций с последующим взятием подвыборок, когда, например, целевой популяцией являются индивиды, но точного текущего их списка в наличии нет.

**Третий этап** процедуры составления выборки тесно связан с определением основы выборки. Выбор метода или процедуры составления выборки во многом зависит от принятой исследователем основы выборки. Различные типы выборок требуют различных типов основ выборки. *Об этом будет сказано в следующих вопросах.*

**Четвертый этап** процедуры составления выборки состоит в определении объема выборки. Этот этап будет рассмотрен в вопросе «Определение объема выборки».

На **пятом этапе** необходимо реально отобрать элементы, которые будут подвергнуты исследованию. Используемый для этого способ определяется избранным типом выборки. И, наконец, необходимо реально исследовать выделенных респондентов. На этом этапе существует большая вероятность совершения ряда ошибок, о которых будет сказано в теме «Реальные процедуры и систематические ошибки сбора данных».

## 1.2. Типы плана выборочного контроля

Все техники контроля выборки могут быть разделены на две категории: наблюдение за вероятностными и за детерминированными (квотными) выборками.

**Вероятностная выборка** – это выборка, в которую каждый член совокупности может включаться с некоторой *заданной ненулевой вероятностью*. Вероятности включения в выборку тех или иных членов совокупности могут отличаться друг от друга, но вероятность включения в нее каждого элемента известна. Эта вероятность определяется особой механической процедурой, используемой для отбора элементов выборки.

**Детерминированная выборка** – это выборка, основываемая на некоторых частных предпочтениях или суждениях, обуславливающих отбор тех или иных элементов, при этом оценка вероятности включения в выборку произвольного элемента совокупности становится невозможной. Гарантировать репрезентативность такой выборки нельзя.

*Например, за рубежом все зарегистрированные избиратели с какой-то вероятностью могут быть призваны для участия в суде присяжных. Если бы все заседатели отбирались случайным образом, жюри присяжных представляло бы собой пример вероятностной выборки. Однако там ни для кого не секрет, что подбор присяжных заседателей – вещь, мягко говоря, непростая. Ответчик платит своему поверенному лицу еще и за его умение подбирать потенциально «дружелюбное» жюри. Стало быть жюри является образчиком детерминированной выборки.*

Все детерминированные выборки основаны на частной позиции, суждении или предпочтении, а не на механической процедуре отбора элементов выборки. Эти предпочтения порой могут давать хорошие оценки характеристик совокупности, однако не существует способа объективного определения соответствия выборки поставленной задаче.

Выборки могут подразделяться также на выборки **фиксированного объема** и **последовательные выборки**.

**Выборка фиксированного объема (фиксированная выборка)** – выборка, определение объема которой производится априорно, требуемая информация определяется по отобраным элементам. При работе с данной выборкой объем определяется до начала обследования и анализу результатов предшествует сбор всех необходимых данных. При маркетинговых исследованиях обычно используется именно выборки фиксированного объема (*однако не следует забывать, что существуют и последовательные выборки, которые могут быть использованы с каждым из обсуждаемых ниже основных планов выборочного исследования*).

**Последовательная выборка** – выборка, формируемая на основании серии последовательных решений. Если после рассмотрения малой выборки результат представляется неубедительным, рассматривается выборка большего объема, если и этот шаг не приводит к результату, объем выборки вновь увеличивается и так далее. Таким образом, на каждом этапе принимается решение о том, можно ли считать полученный результат достаточно убедительным или же нет.

Работа с последовательной выборкой дает возможность оценить тренд данных по мере их сбора, что позволяет сократить расходы, связанные с дополнительными наблюдениями в тех случаях, когда их целесообразность невелика.

Как вероятностный, так и детерминированный план выборочного наблюдения делятся на ряд типов (рис. 1.1).



Рис. 1.1. Классификация техник выборочного контроля

Следует помнить о том, что основные типы выборок могут сочетаться, образуя более сложные планы выборочного наблюдения.

При отборе элементов **детерминированной выборки** определяющую роль играют частные оценки или решения. Иногда эти оценки исходят от исследователя, в некоторых же случаях отбор элементов совокупности определяется счетчиками. Так как элементы отбираются не механически, определение вероятности включения в выборку произвольного элемента и, соответственно, ошибки выборочного наблюдения становится невозможным, незнание ошибки, обусловленной избранной процедурой выборочного обследования, не позволяет оценить точность их оценок.

Нерепрезентативные выборки иногда называются **случайными**, т. к. отбор элементов выборки осуществляется «случайным» образом, — отбираются те элементы, которые являются или представляются наиболее доступными в период проведения отбора.

*Например, наша повседневная жизнь изобилует примерами подобных нерепрезентативных выборок. Мы беседуем с друзьями и на основании их реакций и позиций делаем выводы касательно царящих в обществе политических пристрастий; местная радиостанция призывает людей выразить свое отношение к какому-то спорному вопросу, выражаемое ими мнение интерпретируется как превали-*

рующее; мы призываем к сотрудничеству добровольцев и работаем с теми, кто вызывается помочь нам.

Проблема нерепрезентативных выборок очевидна – нельзя быть уверенным в том, что выборки такого рода действительно представляют целевую популяцию. И если мы еще можем усомниться в том, что мнение наших приятелей правильно отражает политические взгляды, преобладающие в обществе, нам зачастую хочется верить в то, что выборки большего объема, отобранные подобным же образом, репрезентативны. Покажем ошибочность подобного допущения на примере.

Несколько лет тому назад одна из локальных телевизионных станций города проводила ежедневный опрос общественного мнения по темам, представляющим интерес для местной общины. Опросы общественного мнения, носившие название «Пульс Мэдисона», проводились следующим образом. Каждый вечер во время шестичасовых новостей станция обращалась к зрителям с вопросом, касающимся определенной спорной проблемы, на который необходимо было дать положительный или отрицательный ответ. В случае положительного ответа надлежало звонить по одному, в случае отрицательного ответа – по другому телефону. Количество голосов «за» и «против» подсчитывалось автоматически. В десятичасовом выпуске новостей сообщались результаты телефонного опроса. Каждый вечер на студию звонило от 500 до 1000 человек, желавших выразить свою позицию по тому или иному вопросу; телевизионный комментатор интерпретировал результаты опроса как господствующее в обществе мнение.

В одном из шестичасовых выпусков зрителям был предложен следующий вопрос: «Не считаете ли вы, что возрастной ценз на употребление алкоголя может быть снижен до 18 лет?» Существовавший легальный ценз соответствовал 21 году. Аудитория отреагировала на этот вопрос необычайной активностью – в этот вечер на студию позвонили почти четыре тысячи человек, из которых за снижение возрастного ценза высказались 78 %. Представляется очевидным, что выборка из четырех тысяч человек «должна быть репрезентативной» для сообщества, состоящего из 180 000. Ничего подобного. Скорее всего определенная возрастная группа населения была заинтересована в известном исходе голосования куда сильнее прочих. Соответственно, не было ничего удивительного в том, что при обсуждении этого вопроса, проходившем несколькими неделями позже, выяснилось, что во

*время, отведенное для опроса, студенты действовали согласованно. Они звонили на телевидение по очереди, причем каждый из них совершал по несколько звонков. Таким образом, ни размер выборки, ни процент поборников либерализации закона не являлись чем-то удивительным. Выборка была нерепрезентативной.*

Простое увеличение объема выборки не делает ее репрезентативной. Репрезентативность выборки обеспечивается не ее объемом, а надлежащей процедурой отбора элементов. Когда участники опроса определяются добровольно или элементы выборки отбираются в силу их доступности, план контроля выборки не дает гарантии ее представительности. Эмпирические данные свидетельствуют о том, что выборки, формирование которых определялось соображениями удобства, вне зависимости от их размера редко оказываются репрезентативными.

Не рекомендуется использовать нерепрезентативные выборки при проведении описательных или каузальных исследований. Они допустимы лишь при поисковых исследованиях, имеющих целью отработку определенных идей или представлений, но даже и в этом случае предпочтительнее использовать преднамеренные выборки.

**Преднамеренные выборки (не вполне случайные)** – детерминированные выборки, элементы которых, отвечающие целям обследования, отбираются вручную. Основа отбора – мнение о том, что отбираемые элементы могут дать полноценное представление об изучаемой популяции. *Например, с характерным образчиком преднамеренной выборки мы сталкиваемся каждые четыре года во время президентских выборов, когда вниманию телезрителей представляется углубленный анализ общественного мнения. Эти мнения считаются репрезентативными, поскольку по опыту прошлых выборов выявляемый здесь победитель впоследствии становился президентом. Благодаря мониторингу анализа общественного мнения аналитики могут давать долговременные прогнозы о том, кто именно победит на выборах. Хотя подобный анализ и прогнозирование в последнее время претерпели существенные изменения, в их ходе по-прежнему используются преднамеренные выборки.*

В некоторых случаях элементы выборки отбираются не в силу их репрезентативности, а благодаря тому, что они могут предоставить интересующую информацию. *Например, когда суд руководствуется показаниями экспертизы, он, в известном смысле, прибегает к использованию преднамеренной выборки. Подобная же позиция может использоваться и при разработке исследовательских проектов.*

При первичной проработке вопроса интерес представлен, прежде всего, в определении перспектив исследования, что и обуславливает отбор элементов выборки.

Выборка по методу «**снежного кома**» является одним из типов преднамеренной выборки, используемой при работе с особыми видами совокупностей. Эта выборка зависит от умения исследователя задать начальное множество респондентов, которые обладают необходимыми характеристиками. Затем эти респонденты используются в качестве основы, которая определяет дальнейший отбор индивидов с нужными характеристиками.

*Например, компания хочет оценить потребность в некоем изделии, которое позволило бы глухим людям общаться по телефону. Исследователи могут начать разработку этой проблемы с идентификации ключевых фигур обследуемой совокупности; последние могли бы назвать имена других членов этой группы, которые согласились бы принять участие в обследовании. Выборка при подобной тактике растет подобно снежному кому.*

На начальных этапах проработки проблемы, когда определяются перспективы и возможные ограничения планируемого исследования, использование преднамеренной выборки может быть очень эффективным. Однако использование преднамеренных выборок при описательных или при казуальных исследованиях может повлиять на качество их результатов.

**Квотная выборка** – детерминированная выборка, отбираемая таким образом, что доля элементов выборки, обладающих определенными характеристиками, примерно соответствует доле таких же элементов в обследуемой совокупности, т. е. каждому задается квота, определяющая характеристики населения.

*В качестве примера можно рассмотреть попытку создания репрезентативной выборки студентов, проживающих на территории университета. Если в некой выборке, состоящей из 500 индивидов, не будет ни одного старшекурсника, то можно усомниться в ее репрезентативности и в правомерности применения полученных на этой выборке результатов к обследуемой совокупности. При работе с пропорциональной выборкой исследователь может проследить за тем, чтобы доля старшекурсников в выборке соответствовала их доле в общем количестве студентов.*

*Предположим, что исследователь проводит выборочное исследование студентов университета, при этом он заинтересован в том,*

чтобы выборка отражала не только их принадлежность к тому или иному полу, но и распределение их по курсам. Пусть общее число студентов составляет 10 000, из них: 3200 – первокурсники, 2600 – второкурсники, 2200 – студенты третьего курса и 2000 – выпускники; из них 7000 юношей и 3000 девушек. Для выборки объемом 1000 человек план пропорционального выборочного контроля требует наличия 320 первокурсников, 260 второкурсников, 220 третьекурсников и 200 выпускников; 700 юношей и 300 девушек. Исследователь может реализовать этот план, наделив каждого интервьюера определенной квотой, которая будет определять, с какими студентами он или она должны контактировать. Интервьюеру, которому надлежит провести 20 интервью, может быть дана инструкция опросить:

- шесть первокурсников – пять юношей и одну девушку;
- шесть второкурсников – четырех юношей и двух девушек;
- четырех третьекурсников – трех юношей и одну девушку;
- четырех выпускников – двух юношей и двух девушек.

Заметьте, что отбор конкретных элементов выборки определяется не исследовательским планом, а выбором интервьюера, призванного соблюдать только те условия, которые были заданы квотой: опросить пятерых первокурсников, одну первокурсницу и т. д.

Заметьте также, что данная квота точно отображает половое распределение студенческой популяции, но несколько искажает распределение студентов по курсам; 70 % (14 из 20) интервью приходится на долю юношей, но лишь 30 % (6 из 20) на долю первокурсников, в то время как те составляют 32 % от общего числа студентов. Квота, выделяемая каждому конкретному интервьюеру, может не отражать и обычно не отражает распределение контрольных характеристик в популяции – соответствующей пропорциональностью должна обладать только итоговая выборка.

Поэтому эти выборки зависят скорее от личных, субъективных позиций или суждений, чем от объективной процедуры отбора элементов выборки. Причем, в отличие от преднамеренной выборки, личное суждение здесь принадлежит не разработчику проекта, а интервьюеру.

Поэтому необходимо сделать три замечания. Во-первых, выборка может сильно отличаться от популяции по каким-то иным важным характеристикам, что может оказать серьезное влияние на результат. Например, если исследование будет посвящено проблеме бытующих в студенческой среде предрассудков, связанных с местом проживания,

небезразличным обстоятельством может оказаться то, откуда прибыли опрашиваемые – из города или из сельской местности. Поскольку квота для характеристики «выходец из города/села» не была означена, точное отображение этой характеристики становится маловероятным. Поэтому существует следующая альтернатива – определить квоты для всех потенциально значимых характеристик. Но увеличение количества контрольных характеристик приводит к усложнению спецификации. Это, в свою очередь, затрудняет (а иногда и делает невозможным) отбор элементов выборки и, следовательно, приводит к удорожанию. Если, например, принадлежность к городскому или сельскому населению и социоэкономический статус также окажутся значимыми для исследования, то интервьюеру возможно придется заняться поисками первокурсника, который был бы горожанином и принадлежал к высшему или к среднему классу. Иметь же дело с обычным первокурсником обычно проще.

Во-вторых, убедиться в том, что данная выборка действительно является репрезентативной, очень сложно. Можно проверить выборку на предмет соответствия распределения характеристик, которые не входят в число контрольных, их распределению в совокупности. Однако подобная проверка может приводить только к негативным выводам. Выявить можно только расхождение распределений. Если же распределения выборки и совокупности для каждой из этих характеристик и повторяют друг друга, существует вероятность того, что выборка отличается от совокупности по какому-то иному, не заданному явно признаку.

В-третьих, интервьюеры, предоставленные самим себе, слишком часто прибегают к опросу своих приятелей. А так как те зачастую оказываются подобными самим интервьюерам, возникает опасность ошибки. Например, при необходимости совершения домашних визитов интервьюеры зачастую оказываются движимыми соображениями удобства. Например, они могут проводить опросы только днем, что приводит к недооценке мнения работающих респондентов. Помимо прочего, они не заходят в обветшавшие дома и, как правило, не поднимаются на верхние этажи зданий, не имеющих лифтов.

В зависимости от специфики изучаемой проблемы, названные тенденции могут приводить к разного рода ошибкам, исправление же их на стадии анализа данных очень затруднено.

С другой стороны, при объективном отборе элементов выборки получают определенные средства, позволяющие упростить процедуру

оценки репрезентативности данной выборки. При анализе проблемы репрезентативности таких выборок рассматривается не столько состав выборки, сколько процедура отбора ее элементов.

**Вероятностные выборки** не всегда репрезентативнее детерминированных. Более репрезентативной может оказаться и детерминированная выборка. Преимущество вероятностных выборок состоит в том, что они позволяют оценить возможную ошибку выборочного обследования. Если же исследователь работает с детерминированной выборкой, он не имеет объективного метода оценки ее адекватности целям исследования.

Вероятностные выборки делятся на простую случайную, расслоненную и гнездовую.

**В простой случайной выборке** каждый элемент, включаемый в выборку, обладает одной и той же заданной вероятностью попадания в число исследуемых элементов, и любая комбинация элементов исходной популяции может потенциально стать выборкой. *Например, если мы захотим составить простую случайную выборку всех студентов, числящихся в определенном колледже, нам достаточно будет составить список всех студентов, присвоить каждой значащейся в нем фамилии свой номер и с помощью компьютера произвести случайный отбор заданного количества элементов.*

*Представьте, что исследуемой генеральной совокупностью является все взрослое население Гомеля. Для описания этой совокупности может быть использован ряд параметров: средний возраст, доля населения с высшим образованием, уровень доходов и т. д. Обратите внимание на то, что все эти показатели имеют определенное значение. Разумеется, мы можем рассчитать их, проведя полную перепись изучаемой совокупности. Обычно же мы опираемся не на цenz, а на отбираемую нами выборку и используем полученные при выборочном наблюдении значения для определения искомых параметров совокупности.*

**Параметр** – определенная характеристика или показатель генеральной или изучаемой совокупности.

*Приведем пример гипотетической совокупности, состоящей из 20 человек (табл. 1.1).*

## Данные по гипотетической совокупности

Элемент	Доход	Образование лет	Подписка на газету	Элемент	Доход	Образование лет	Подписка на газету
1A	5600	7	X	11K	9600	12	X
2B	6000	8	Y	12L	10000	12	Y
3C	6400	9	X	13M	10400	13	X
4D	6800	9	Y	14N	10800	13	Y
5E	7200	10	X	15O	11200	14	X
6F	7600	11	Y	16P	11600	15	Y
7G	8000	11	X	17Q	12000	15	X
8H	8400	11	Y	18R	12400	16	Y
9I	8800	11	X	19S	12800	17	X
10J	9200	11	Y	20T	13200	17	Y

Работа с небольшой гипотетической совокупностью имеет ряд преимуществ. *Во-первых*, небольшой объем выборки дает возможность легко вычислить параметры совокупности, которые могут использоваться для ее описания. *Во-вторых*, этот объем позволяет понять, что может произойти при принятии того или иного плана выборочного контроля. Обе эти особенности делают простым сравнение результатов выборки с «истинным» (известным) значением совокупности, чего нельзя сказать о типичной ситуации, при которой действительное значение совокупности неизвестно. Сравнение оценки с «истинным» значением приобретает в этом случае особую наглядность. Оценка параметров проводится по **средней** элементов, составляющих совокупность, и **дисперсии генеральной совокупности**.

**Производная совокупность** – совокупность всех возможных различных выборок, которые могут быть выделены из генеральной совокупности по заданному плану выборочного контроля. Значение статистики, используемое для оценки определенного параметра, зависит от выборки, определяемой планом.

**Статистика** – характеристика или показатель выборки.

Различные выборки дают различные статистики или оценки одного и того же параметра совокупности. *Например, рассмотрим произвольную совокупность всех возможных выборок, которые могут быть выделены из гипотетической генеральной совокупности, состоящей допустим из 20 индивидов, где у каждого будут свой средний доход и имя. По плану выборочного контроля, предполагающему, что выборка объемом  $n = 2$  может быть получена путем случайного*

бесповторного отбора. Данные индивидов записывают на диск, после чего все диски опускают в емкость и перемешивают. Затем извлекают диск из емкости, списывают с него информацию и откладывают его в сторону. Тоже проделывают и со вторым диском. Потом возвращают оба диска в емкость, опять перемешивают и повторяют ту же последовательность действий. Для 20 дисков возможны 190 таких парных комбинаций. Для каждой комбинации вычисляют среднюю величину дохода и величину ошибки (табл. 1.2).

Таблица 1.2

**Данные по произвольной совокупности всех возможных выборок**

Выборка			Выборка			Выборка		
К	Пара	Средний доход	К	пара	средний доход	К	Пара	Средний доход
1	AB	5800	101	G	8400	...	...	...
2	AC	6000	...	...	...	169	M	11800
...	...	...	112	G	10600	170	N	11000
19	AT	9400	113	H	8600	171	N	11200
20	BA	6200	114	H	8800	...	...	...
21	BC	6400	...	...	...	175	N	12000
...	...	...	124	H	10800	176	O	11400
37	BT	9600	125	I	9000	177	O	11600
38	C	6600	126	I	9200	...	...	...
39	C	6800	...	...	...	180	O	12200
...	...	...	135	I	11000	181	P	11800
54	C	9800	136	J	9400	182	P	12000
55	D	7000	137	J	9600	183	P	12200
56	D	7200	...	...	...	184	P	12400
...	...	...	145	J	11200	185	Q	12200
70	D	10000	146	K	9800	186	Q	12400
71	E	7400	147	K	10000	187	Q	12600
72	E	7600	...	...	...	188	R	12600
...	...	...	154	K	11400	189	R	12800
86	E	10200	155	L	10200	190	S	13000
86	F	7800	156	L	10400			
87	F	8000	...	...	...			
...	...	...	162	L	11600			
99	F	10400	163	M	10600			
100	G	8200	164	M	10800			

Недостатки производной совокупности. *Во-первых*, составление совокупностей такого рода требует слишком большой траты времени и сил. *Во-вторых*, совокупность определяется как совокупность всех возможных различных выборок, которые могут быть выделены из генеральной совокупности по заданному плану выборочного контроля. При изменении любой части выборочного контроля производная совокупность также меняется. *Например*, если возвращать в емкость первый диск прежде, чем вынуть второй, производная совокупность будет включать выборки  $AA$ ,  $BB$ ,  $CC$  и т. д. (а должно быть  $AB$ ,  $AC$ ,  $BC$ ). Если объем бесповторных выборок будет равен 3, а не 2, появятся выборки типа  $ABC$ , и их будет не 190, а 1140. *В-третьих*, при изменении простого случайного отбора на любой другой метод определения элементов выборки производная совокупность также изменится.

**Формирование простой случайной выборки.** *В приведенном выше примере отбор элементов выборки осуществлялся с помощью емкости, в которой находились все элементы исходной совокупности. Это позволило наглядно представить понятия производной совокупности и выборочного распределения. Применять же подобный метод на практике не рекомендуется, т. к. при этом повышается вероятность ошибки. Диски могут отличаться и размерами, и фактурой, что в известных случаях может приводить к предпочтению одних дисков другим.*

Предпочтительный метод формирования простой случайной выборки основан на использовании таблицы случайных чисел. Использование такой таблицы предполагает следующую последовательность шагов.

*Во-первых*, элементам генеральной совокупности должны быть присвоены последовательные номера от 1 до  $N$ . В нашей гипотетической совокупности элементу  $A$  был присвоен номер 1, элементу  $B$  – номер 2 и т. д.

*Во-вторых*, количество разрядов таблицы случайных чисел должно быть таким же, как у номера  $N$ . Для  $N = 20$  будут использоваться двузначные числа; для  $N$  между 100 и 999 – трехзначные числа и т. д.

*В-третьих*, начальная позиция должна определяться случайным образом. Можно раскрыть соответствующую таблицу случайных чисел и методом «тыка» формировать выборку. Поскольку числа в таблице случайных чисел следуют в случайном порядке, начальная позиция не имеет особого значения. И, наконец, можно двигаться

в любом произвольно выбранном направлении – вверх, вниз или поперек, отбирая те элементы, номера которых будут соответствовать случайным числам из таблицы.

### 1.3. Стратифицированная выборка

**Стратифицированная выборка** – это вероятностная выборка, для которой характерна следующая двухшаговая процедура:

1) генеральная (исходная) совокупность делится на ряд непересекающихся, исчерпывающих ее подмножеств;

2) в каждом подмножестве или группе производится независимый отбор элементов простых случайных выборок.

Подмножества, на которые подразделяется генеральная совокупность, называются **слоями** или **частными совокупностями**. Данное определение требует, чтобы выделяемые подмножества не пересекались и исчерпывали исходную совокупность. Это означает, что каждый элемент совокупности должен входить в один и только один из слоев; при этом процедура распределения должна охватывать все без исключения элементы генеральной совокупности.

*Вернемся к рассматривавшейся в предыдущем вопросе гипотетической совокупности, состоящей из 20 индивидов. Эта совокупность может быть описана несколькими параметрами, такими как средний уровень доходов, образовательный уровень, часть совокупности, подписавшаяся на то или иное издание. Предположим, мы хотим разделить генеральную совокупность на два слоя на основе образовательного уровня (табл. 1.1). Элементы А–J образуют первую страту или слой (уровень образования соответствует не более чем 12-летнему сроку обучения), элементы К–Т образуют вторую страту или слой (уровень образования соответствует более чем 12-летнему сроку обучения). Число страт не обязательно должно равняться двум. Генеральная совокупность может быть разделена на любое другое количество страт. Мы остановились на числе 2 только потому, что оно позволяет наглядно продемонстрировать технический аспект обсуждаемой процедуры (табл. 1.3).*

На втором этапе должен быть произведен отбор элементов простой случайной выборки **из каждой** страты. Пусть объем выборок и на сей раз будет равен 2. Это означает, что мы должны выбрать по одному элементу из каждой страты (в общем случае количество элементов из того или иного слоя не обязательно должно быть одинаковым).

Таблица 1.3

## Распределение данных по стратам

Элементы 1 страты		Элементы 2 страты	
<i>A</i>	<i>F</i>	<i>K</i>	<i>P</i>
<i>B</i>	<i>G</i>	<i>L</i>	<i>Q</i>
<i>C</i>	<i>H</i>	<i>M</i>	<i>R</i>
<i>D</i>	<i>I</i>	<i>N</i>	<i>S</i>
<i>E</i>	<i>J</i>	<i>O</i>	<i>T</i>

Процедура отбора элементов внутри стратифицированной выборки ничем не отличается от аналогичной процедуры для простой случайной выборки. Элементам генеральной совокупности каждой страты присваиваются порядковые номера от 1 до 10. Далее для отбора элементов может быть использована таблица случайных чисел. Первый элемент отбирается из 10 элементов первой страты, второй – из 10 элементов второй страты, при этом возможен как повторный «вход» в таблицу случайных чисел, так и продолжение движения по избранной ранее строке или столбцу, которое в любом случае должно продолжаться до появления первого числа от 1 до 10.

Рассмотрим **производную совокупность**.

Хотя реально может быть отобрана только одна выборка с объемом 2, рассмотрим производную совокупность всех возможных выборок с объемом 2, которые можно сформировать по заданному плану выборочного отбора (табл. 1.4).

Таблица 1.4

## Распределение данных по стратам

К	Пара	Среднее									
1	<i>AK</i>	7600	26	<i>CP</i>	9000	51	<i>FK</i>	8600	76	<i>HP</i>	10000
2	<i>AL</i>	7800	27	<i>CQ</i>	9200	52	<i>FL</i>	8800	77	<i>HQ</i>	10200
3	<i>AM</i>	8000	28	<i>CR</i>	9400	53	<i>FM</i>	9000	78	<i>HR</i>	10400
4	<i>AN</i>	8200	29	<i>CS</i>	9600	54	<i>FN</i>	9200	79	<i>HS</i>	10600
5	<i>AO</i>	8400	30	<i>CT</i>	9800	55	<i>FO</i>	9400	80	<i>HT</i>	10800
6	<i>AP</i>	8600	31	<i>DK</i>	8200	56	<i>FP</i>	9600	81	<i>IK</i>	9200
7	<i>AQ</i>	8800	32	<i>DL</i>	8400	57	<i>FQ</i>	9800	82	<i>IL</i>	9400
8	<i>AR</i>	9000	33	<i>DM</i>	8600	58	<i>FR</i>	10000	83	<i>IM</i>	9600
9	<i>AS</i>	9200	34	<i>DN</i>	8800	59	<i>FS</i>	10200	84	<i>IN</i>	9800
10	<i>AT</i>	9400	35	<i>DO</i>	9000	60	<i>FT</i>	10400	85	<i>IO</i>	10000
11	<i>BK</i>	7800	36	<i>DP</i>	9200	61	<i>GK</i>	8800	86	<i>IP</i>	10200

К	Пара	Среднее	К	Пара	Среднее	К	Пара	Среднее	К	Пара	Среднее
12	<i>BL</i>	8000	37	<i>DQ</i>	400	62	<i>GL</i>	9000	87	<i>IQ</i>	10400
13	<i>BM</i>	8200	38	<i>DR</i>	9600	63	<i>GM</i>	9200	88	<i>IR</i>	10600
14	<i>BN</i>	8400	39	<i>DS</i>	9800	64	<i>GN</i>	9400	89	<i>IS</i>	10800
15	<i>BO</i>	8600	40	<i>DT</i>	10000	65	<i>GO</i>	9600	90	<i>IT</i>	11000
16	<i>BP</i>	8800	41	<i>EK</i>	8400	66	<i>GP</i>	9800	91	<i>JK</i>	9400
17	<i>BQ</i>	9000	42	<i>EL</i>	8600	67	<i>GQ</i>	10000	92	<i>JL</i>	9600
18	<i>BR</i>	9200	43	<i>EM</i>	8800	68	<i>GR</i>	10200	93	<i>JM</i>	9800
19	<i>BS</i>	9400	44	<i>EN</i>	9000	69	<i>GS</i>	10400	94	<i>JN</i>	10000
20	<i>BT</i>	9600	45	<i>EO</i>	9200	70	<i>GT</i>	10600	95	<i>JO</i>	10200
21	<i>CK</i>	8000	46	<i>EP</i>	9400	71	<i>HK</i>	9000	96	<i>JP</i>	10400
22	<i>CL</i>	8200	47	<i>EQ</i>	9600	72	<i>HL</i>	9200	97	<i>JQ</i>	10600
23	<i>CM</i>	8400	48	<i>ER</i>	9800	73	<i>HM</i>	9400	98	<i>JR</i>	10800
24	<i>CN</i>	8600	49	<i>ES</i>	10000	74	<i>HN</i>	9600	99	<i>JS</i>	11000
25	<i>CO</i>	8800	50	<i>ET</i>	10200	75	<i>HO</i>	9800	100	<i>JT</i>	11200

При заданном плане выборочного отбора возможны только 100 парных комбинаций элементов, тогда как при отборе простой случайной выборки существовало 190 таких вариантов. Причина в том, что при таком виде выборочного наблюдения из каждой страты может быть выбран только один элемент. При формировании же простой случайной выборки из генеральной совокупности могли быть отобраны любые 2 элемента. В этом отношении стратифицированная выборка отличается от случайной большим числом ограничений. Каждый элемент имеет одинаковую вероятность включения в выборку, равную  $1/10$ , поскольку каждый из них может стать элементом, отбираемым из той или иной страты, т. е. в пределах страты речь идет о простой случайной выборке. Равновероятный отбор элементов может быть присущ и другим способам. Равновероятный отбор является необходимым, но не достаточным условием простого случайного выбора, его необходимо дополнить условием равной вероятности любой возможной комбинации из  $n$  элементов.

Одно из преимуществ стратифицированной выборки – такая выборка обеспечивает большую точность выборочных статистик, чем простая случайная выборка. *Например, если количественным признаком стратификации будет образование, то количество выборочных средних, сильно отклоняющихся от генерального среднего, существенно сократится.*

Второй довод в пользу стратифицированных выборок состоит в том, что разделение позволяет обследовать интересующие исследо-

вателя характеристики определенных подмножеств. Например, при стратификации можно гарантировать представление лиц с образованием не выше среднего и с образованием выше среднего. Эта возможность приобретает особую значимость при отборе элементов генеральной совокупности, включающей в себя редкие сегменты. Представим, например, что производитель колец с бриллиантами хочет изучить социальный состав потребителей его продукции. Если не будут приняты специальные меры, окажется, что высшие слои общества, составляющие всего около 3 % населения, либо вообще не будут представлены в выборке, либо окажутся представленными недостаточно полно. Тем не менее, производителя ювелирных изделий должен интересовать именно этот немногочисленный сегмент совокупности.

В маркетинге возможны ситуации, когда поведение совокупности, например, уровень потребления какой-то продукции, определяется ее небольшим подмножеством. В этих случаях становится важным адекватное представление этого подмножества в обследуемой выборке. Стратифицированное выборочное наблюдение является одним из вариантов обеспечения названного представления.

Предпочтение стратифицированной выборки по отношению к простой случайной, определяется стоимостью и точностью (хотя стратифицированные выборки дают более точные оценки, они имеют и большую стоимость). Поэтому при выборе стратифицированной выборки необходимо сделать выбор между пропорционально и непропорционально стратифицированными выборками.

**Пропорционально стратифицированная выборка** – стратифицированная выборка, в которой межслойное соотношение наблюдений пропорционально относительной доле элементов в каждом слое генеральной совокупности.

**Непропорционально стратифицированная выборка** – стратифицированная выборка, в которой объем отдельных слоев или подмножеств зависит от объема и изменчивости соответствующих слоев генеральной совокупности, т. е. слои с большей изменчивостью количественного признака получают в выборке большее представление, а слои, близкие к гомогенности, меньшее представление, чем в пропорционально стратифицированной выборке.

Преимущество **пропорционального распределения** состоит в том, что здесь достаточно знать только относительные размеры каждой страты для определения количества выборочных наблюдений, которые должны быть отобраны из каждого слоя для заданного объема выборки.

Однако непропорционально стратифицированная выборка может давать более точные результаты. При ее составлении одновременно учитывают два критерия: объем страты и ее изменчивость.

Очень часто путают стратифицированные выборки с квотными. У тех и других есть ряд сходств. В обоих случаях генеральная совокупность делится на сегменты, и элементы отбираются из каждого сегмента. Но между ними существует существенное различие. В стратифицированных выборках элементы выборки выбираются вероятностными методами; что касается квотных выборок, то их отбор обусловлен позицией исследователя.

#### 1.4. Групповая (гнездовая) выборка

Групповые выборки (кластерные) – еще один способ вероятностного выборочного исследования. Она в чем-то похожа, а в чем-то отличается от стратифицированного выборочного отбора.

**Групповая выборка** – вероятностная выборка, для которой характерна следующая двухступенчатая процедура:

- 1) генеральная совокупность делится на ряд непересекающихся исчерпывающих ее подмножеств;
- 2) производится случайный отбор подмножеств.

Если для выборки используют все элементы выбранных подмножеств, процедура называется *одноступенчатой*; если же выборка отбирается из этих подмножеств при помощи вероятностного метода, процедура называется *двухступенчатой*.

Существуют сходства и различия групповой и стратифицированной выборки. В каждом случае генеральная совокупность делится на ряд непересекающихся, исчерпывающих ее подмножеств, однако, в стратифицированной выборке производится отбор элементов *из каждого подмножества*, а при групповой выборке производится *отбор подмножеств*.

*Рассмотрим пример из вопроса 2 (данные табл. 1.1). Если все подписчики издания X будут рассматриваться в качестве первого, а все подписчики издания Y – в качестве второго подмножества, можно достаточно уверенно использовать в качестве контрольной выборки для оценки среднего уровня доходов ту или иную группу. Хотя распределение уровня доходов внутри каждого подмножества может отличаться от аналогичного распределения в генеральной совокупности, разброс значений уровня доходов таков, что при оценке среднего уровня доходов и дисперсии этого уровня по элементам*

любой из двух названных выборок можно допустить лишь незначительную ошибку.

**Систематическая выборка** – один из видов группового выборочного отбора, обеспечивающий возможность простейшего исследования многих генеральных совокупностей. Такая выборка предполагает включение в нее каждого  $k$ -го элемента генеральной совокупности, начиная с некоторого выбранного произвольно элемента. Рассмотрим старую совокупность из 20 индивидов и представим, что необходимо отобрать из них 5 элементов. Для 20 элементов совокупности и объема выборки, равного 5, выборочная доля равна  $f = n/N = 5/20 = 1/4$ , т. е. должен отбираться один элемент из четырех. Выборочный интервал  $I = 1/f$  будет равен 4. Это означает, что после произвольно выбранной начальной позиции должен отбираться каждый четвертый элемент. Если начальная позиция придется на 1, то в выборку будут отобраны 1-й, 5-й, 9-й, 13-й, 17-й элементы. Если позиция придется на 2, будут отобраны, соответственно, 2-й, 6-й, 10-й, 14-й, 18-й элементы и т. д.

Систематическая выборка относится к категории групповых одноступенчатых, т. к. используются все элементы выделенных групп, а не некоторая, возникающая в результате отбора их часть.

*Подмножества или группы в нашем случае:*

*Группа 1: A, E, I, M, Q.*

*Группа 2: B, F, J, N, R.*

*Группа 3: C, G, K, O, S.*

*Группа 4: D, H, L, P, T.*

*Для исследования производится случайный выбор одной из этих групп. В данном случае выбор делается один раз.*

*Достоинства систематической выборки:*

- простота формирования;*
- отсутствие проблемы дублирования элементов в отличие от случайной выборки;*
- большая репрезентативность, чем при простой случайной выборке.*

При использовании систематической выборки существует следующая опасность: если в списке элементов наблюдается естественная периодичность, оценка, производимая на основе такой выборки, связана с серьезными ошибками. Например, известен уровень продаж авиабилетов на каждый день года, и нужно проанализировать продажи с точки зрения продолжительности полета. Анализ всех 365 дней года будет весьма дорогостоящим. Предположим, что ис-

*следователяского бюджета хватит на исследование 52 дней. Систематическая выборка с выборочным интервалом равным 7 дней (365/52) скорее всего, приведет к ошибочным выводам, т. к. будет отражать уровень продаж авиабилетов на рейсы, совершаемые по понедельникам, средам или, например, по воскресеньям.*

Поэтому, правильное задание выборочного интервала возможно только на основе исследования причин периодичности.

**Территориальная выборка** – вид группового выборочного отбора, при котором территории или зоны (например, переписные районы, участки) выступают в роли первичных выборочных единиц. Генеральная совокупность делится (обычно с использованием карты) на ряд непересекающихся, исчерпывающих ее подмножеств, или территорий, после чего формируется случайная выборка этих территорий. Если в исследовании участвуют все семьи, живущие на выделенных территориях, то имеем дело с одноступенчатой территориальной выборкой; если же исследуются не все, а лишь отобранные из первичной выборки семьи, обследование называется двухступенчатым.

Принцип, который лежит в основе систематической выборки, задействован и в территориальном выборочном отборе.

*Составить точные исчерпывающие списки населения практически невозможно. Они либо просто отсутствуют, либо если и существуют, то содержат массу устаревшей недостоверной информации (люди переезжают, рождаются, умирают, женятся и разводятся). Несмотря на отсутствие списка семей, можно использовать в качестве относительно точных первичных выборочных единиц отдельные городские районы, представленные на карте.*

Одноступенчатая территориальная выборка проводится на основе следующих этапов:

- 1) производится простой случайный отбор  $n$  городских кварталов из совокупности  $N$  кварталов;
- 2) определяется потребление товара и доходы всех семей, живущих в выделенных кварталах, и распространяется выборочный показатель на всю совокупность.

Отличительной особенностью одноступенчатой территориальной выборки является регистрация и исследование всех семей, проживающих в выделенных территориальных единицах. Часто вместо сплошного исследования всех элементов выделенной группы проводится только выборочное обследование. Различают два типа двухступенчатой территориальной выборки:

- 1) простая двухступенчатая территориальная выборка;
- 2) территориальная выборка с вероятностями, пропорциональными объему (территориальная квотная выборка).

**Простая двухступенчатая территориальная выборка** – вид группового выборочного отбора, при котором из каждой первичной выборочной единицы (например, районов) отбирается определенная доля элементов выборки второго уровня (например, семей).

Рассмотрим совокупность, состоящую из 100 кварталов. Предположим, что в каждом квартале живут по 20 семей. Допустим, необходимо исследовать 80 семей из 2000. Таким образом, выборочная доля равна

$$\frac{80}{2000} = \frac{1}{25}.$$

Существует ряд способов формирования выборки:

- 1) отбор 10 кварталов и 8 семей в каждом квартале;
- 2) отбор 8 кварталов и 10 семей в каждом квартале;
- 3) отбор 20 кварталов и 4 семей в каждом квартале;
- 4) отбор 4 кварталов и 20 семей в каждом квартале.

Последний вариант представляет собой одноступенчатую территориальную выборку, а первые три относятся к категории двухступенчатых. Варианты, с которой будут отобраны кварталы, называются **первоуровневой выборочной долей**:

$$n_k / N_k, \quad (1.1)$$

где  $n_k$  и  $N_k$  – количество кварталов в выборке и в генеральной совокупности соответственно.

Для первых трех случаев первоуровневые выборочные доли равны, соответственно 1/10, 1/12,5 и 1/5.

Вероятность отбора семьи называется **выборочной долей второго уровня**. Так как выборка должна состоять из 80 семей, выборочная доля второго уровня отличается для каждого из приведенных выше вариантов:

$$n_{k/c} / N_{k/c}, \quad (1.2)$$

где  $n_{k/c}$  и  $N_{k/c}$  – количество семей на квартал в выборке и в генеральной совокупности.

Для первых трех схем отбора выборочная доля равна соответственно:  $8/20 = 2/5$ ,  $10/20 = 1/2$ ,  $4/20 = 1/5$ . При этом произведение выборочных долей первого и второго уровней в любом случае равно

общей выборочной доле  $1/25$ . Из соображения экономии при сборе данных выборочная доля второго уровня должна быть высокой (вторая схема), а из соображений статистической эффективности выборочная доля второго уровня должна быть небольшой. Поэтому статистические соображения свидетельствуют в пользу третьей схемы. Простая двухступенчатая территориальная выборка эффективна, если количество единиц второго уровня (семья) на единицу первого уровня (квартал) сохраняется примерно равным. Если количество единиц второго уровня существенно отличаются друг от друга, простая двухступенчатая выборка может привести к ошибочным оценкам. *Например, в некоторых кварталах могут находиться многоэтажные дома, где проживают семьи с низким доходом. В богатых кварталах могут находиться всего несколько домов, в каждом из которых будет проживать по одной семье. То есть количество единиц второго уровня на единицу первого уровня окажется совершенно различным. Данная проблема снимается путем комбинирования различных территорий, или прибегают к **территориальной квотной выборке**.*

**Территориальная квотная выборка** – вид группового отбора, при котором из каждой первичной выборочной единицы отбирается фиксированное количество элементов второго уровня. Вероятности, относящиеся к отбору каждой первичной единицы, могут изменяться, т. к. они напрямую связаны с относительными размерами первичных единиц.

*Приведем пример на основании предыдущего. Данные приведены в табл. 1.5.*

Таблица 1.5

**Данные по территориальной квотной выборке**

Квартал	Количество семей	Суммарное количество семей
1	800	800
2	400	1200
3	200	1400
4	200	1600
5	100	1700
6	100	1800
7	100	1900
8	50	1950
9	25	1975
10	25	2000

Пусть после анализа экономических и статистических критериев количество единиц второго уровня на единицу первого уровня было принято равным 10. Для получения выборки нужного объема (20 единиц) необходимо выбрать две единицы первого уровня. Вероятность выбора определенного элемента зависит от объема единицы первого уровня. В данном случае можно прибегнуть к таблицам трехзначных случайных чисел. Для этого можно воспользоваться двумя первыми числами от 1 до 2000. Числа от 1 до 800 относятся к первому кварталу; числа от 801 до 1200 – ко второму кварталу; числа от 1201 до 1400 – к третьему и т. д.

Вероятность того, что в выборку будет включена любая конкретная семья, остается постоянной, т. к. отличия вероятностей отбора элементов первого уровня компенсируются отличиями вероятностей отбора элементов второго уровня. Допустим, к примеру, рассмотрим два полюса – 1-й и 10-й кварталы. Вероятность выделения квартала 1 равна  $800/2000 = 2/5$ , т. к. 800 из возможных 2000 значений случайных чисел приходится именно на этот квартал. Кварталу 10 соответствует только 25 значений случайных чисел (от 1976 до 2000), соответственно выборочная доля первого уровня для 10 квартала равна  $25/2000 = 1/80$ . Так как из каждого квартала следует выбрать по 10 семей, выборочная доля второго уровня для первого квартала равна  $10/800 = 1/80$ , для 10 квартала  $10/25 = 1/2,5$ . При перемножении соответствующих вероятностей первого и второго уровней происходит их компенсация:

$$\frac{800}{2000} \times \frac{10}{800} = \frac{25}{2000} \times \frac{10}{25}$$

Подобное соотношение верно и для других кварталов.

Достоинства территориальной квотной выборки:

- возможность избежать ошибок, присущих простой двухступенчатой территориальной выборке;
- повышение точности оценок при увеличении изменчивости количества единиц второго уровня на единицу первого уровня.

### **1.5. Определение объема выборки при оценке среднего**

Выборочное распределение статистики является основой для определения выборки. Если известен размах выборочного распределения, можно определить величину ошибки, связанную с той или

иной оценкой. Ошибка, связанная с оценкой генерального среднего по выборочному среднему, определяется среднеквадратической ошибкой среднего:

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}, \quad (1.3)$$

где  $\sigma$  – среднеквадратическая ошибка средней;  $n$  – объем выборки.

Или отношением среднеквадратического отклонения выборки к квадратному корню из объема выборки при неизвестной генеральной дисперсии:

$$S_{\bar{x}} = S / \sqrt{n}. \quad (1.4)$$

Поэтому главным фактором оценки объема выборки должна стать среднеквадратическая ошибка оценки, определяемая из заданного выборочного распределения статистики.

При определении объема выборки необходимо учитывать точность и достоверность выборки.

**Точность** – степень ошибочности результатов обследования или размер доверительного интервала. *Абсолютная точность* задается определенным интервалом, в котором должно находиться оцениваемое значение. *Относительная точность* определяется относительно уровня оценки параметра.

**Достоверность** – степень уверенности в том, что оценка близка к истинному значению.

Для выборки фиксированного объема степень точности и степень достоверности являются связанными величинами. Можно определить либо одну, либо другую величины, но не обе одновременно.

#### **Определение объема выборки при оценке среднего.**

Рассмотрим случай, когда *генеральная дисперсия известна*.

Дисперсия генеральной совокупности  $\sigma^2$  может быть известна из предыдущих исследований. Это означает, что размах распределения  $\sigma_{\bar{x}}$ , заданного среднеквадратичной ошибкой оценки, известен с точностью до коэффициента пропорциональности квадратному корню из объема выборки, т. к.  $\sigma_{\bar{x}}$  – определяется по формуле (1.3).

Доверительный уровень того, что построенный интервал будет содержать в себе генеральное среднее, часто принимается равным 95 %. При этом коэффициент Стьюдента ( $t_c$ ) равен примерно 2. И так как размер заданного полуинтервала ( $H$ ) определяется по формуле

$$H = t_c \sigma_{\bar{x}} = t_c \frac{\sigma}{\sqrt{n}}, \quad (1.5)$$

то можно относительно определить объем выборки ( $n$ ), т. к.  $H$  и  $t_c$  заданы, а  $\sigma$  известна из предыдущих исследований:

$$n = \frac{t_c^2}{H^2} \sigma^2. \quad (1.6)$$

Например, предположим, что руководитель туристического агентства задался целью узнать средние годовые расходы рыболовов на питание и проживание во время их походов. При помощи простой случайной выборки нужно оценить среднегодовые расходы этих рыболовов, пользуясь списком лиц, получивших в течение года рыболовные лицензии. При этом заказчик хочет, чтобы оценка находилась в интервале  $\pm \$25$  от истинного генерального значения. Таким образом, общая точность будет составлять  $\$50$ , а половинная точность, которую будем обозначать  $H$ , составит  $\$25$ . Нужно определиться с доверительным уровнем результата. Пусть доверительный уровень того, что построенный интервал будет содержать в себе генеральное среднее, примем равным 95%. При этом  $t_{\bar{x}}$  равно примерно 2. Предположим, что определенная прежде дисперсия генеральной совокупности равна  $\$100$ . Тогда

$$n = \frac{2^2}{25^2} 100^2 = 64.$$

Таким образом, для оценки среднего уровня расходов при среднеквадратическом отклонении  $\$100$  и заданной точности плюс или минус  $\$25$  достаточно взять сравнительно небольшую выборку (64 рыболова).

Другой способ решения проблемы оценки заключается в построении номограммы уравнения и считывания с нее потребного объема выборки. Номограммы, которые по сути являются графическим решением уравнения, специально разработаны для наиболее часто используемых случаев и используются без предварительных расчетов.

Теперь посмотрим, что произойдет, если увеличить точность исследования в 2 раза: общая ширина требуемого интервала будет равняться  $\$25$ , половина его или  $H$ , соответственно,  $\$12,5$ .

$$n = \frac{2^2}{12,5^2} 100^2 = 256.$$

Те же данные получают и из номограммы. То есть для увеличения точности в 2 раза объем выборки должен быть увеличен в 4 раза.

Приведенное соотношение выражает зависимость точности и объема выборки. При увеличении точности в  $c$  раз объем выборки возрастает в  $c^2$  раз. Например, если бы заданная точность оценки должна была бы вырасти в 5 раз ( $c = 5$ ), то объем выборки увеличился бы с 64 до 1600 ( $c^2 = 25$ ). Увеличение доверительного интервала также приводит к увеличению объема выборки, а значит к удорожанию исследований. Например, задаем не 95 %, а 99 % доверительный интервал, тогда при  $H = 25$ , а  $\sigma = 100$ , получаем:

$$n = \frac{3^2}{25^2} 100^2 = 144.$$

В данном случае  $t_c$  при 99 % равен 3. Таким образом, при увеличении  $t_c$  в  $d$  раз (в нашем примере  $d = 3/2$ ), объем выборки увеличивается в  $d^2$  раз (в нашем случае  $d^2 = 9/4$ ). При всех вычислениях необходимо помнить о стоимости, связанной с увеличением точности и достоверности результатов.

Рассмотрим случай, когда *генеральная дисперсия неизвестна*.

Более типичным случаем является ситуация, когда генеральная дисперсия неизвестна. Однако процедура оценки объема выборки остается практически неизменной, просто вместо известного значения дисперсии использоваться будет *оценочное значение* генеральной дисперсии  $\sigma$ . Еще одна возможность заключается в учете того обстоятельства, что для величины с нормальным распределением область изменения примерно равна плюс-минус трем среднеквадратическим отклонениям. Таким образом, если можно определить область изменения, то можно путем деления на 6 определить и среднеквадратическое отклонение. Даже небольшой запас априорной информации может позволить определить область изменения. Ошибочная оценка скажется на точности доверительного интервала, которая может измениться и в большую, и в меньшую сторону.

Обратимся к примеру. *Некоторое число рыболовов не расходует на питание и проживание никаких средств, т. к. выезжают только на день. Другие совершают несколько недельных поездок в год. Предположим, что 15 дней в году – это типичный верхний предел пребывания на рыбалке. При этом ежедневные расходы составляют около \$30 в неделю, т. е. верхний денежный предел равен \$450. Область изменения этой величины также равна \$450 (т. к. потратить*

меньше \$0 невозможно) и расчетное среднеквадратическое отклонение составляет  $(450 - 0)/6 = 75$ . При заданной точности  $\pm\$25$  и 95 % доверительном интервале объем выборки составит

$$n = \frac{2^2}{25^2} 75^2 = 36.$$

Таким образом, отобранная выборка будет содержать 36 элементов. Предположим, что произведенные наблюдения дадут выборочное среднее  $\bar{X} = 35$  и среднеквадратическое выборочное отклонение  $\bar{S} = 60$ . Тогда доверительный интервал определяется:

$$35 \pm 2 \frac{S}{\sqrt{n}} = 35 \pm 2 \frac{60}{\sqrt{36}} = 35 \pm 20, \text{ или } 15 \leq \bar{X} \leq 55.$$

В результате получили следующее: заданная точность составляла  $\pm\$25$ ; полученная точность равна  $\pm\$20$ , т. е. интервал оказался уже, чем планировалось (выигрыш), т. к. выбранная оценка генерального среднеквадратического отклонения по выборочному была завышенной. Если бы эта оценка была заниженной, доверительный интервал оказался бы шире заданного.

При проведении исследования редко анализируется только один параметр. Обычно исследования носят комплексный многоцелевой характер. Приведем пояснения в виде примера. Предположим, что необходимо определить также годовые расходы рыбаков на снасти и снаряжение и расстояние, которое они преодолевают за время рыбалок. Теперь необходимо определить уже три величины. Допустим, каждая из них тоже должна иметь 95 % доверительный уровень; потребную абсолютную точность и среднеквадратическое отклонение сведем в таблицу с расчетными значениями объемов выборок, рассчитанных по вышеприведенным формулам (табл. 1.6)

Таблица 1.6

Данные расчета показателей

Показатель	Показатель		
	Расходы на питание и проживание	Расходы на снасти и снаряжение	Пройденное расстояние
Доверительный уровень	95 % ( $t_c = 2$ )	95 % ( $t_c = 2$ )	95 % ( $t_c = 2$ )
Заданная точность	$\pm 25$	$\pm 10$	$\pm 100$

Показатель	Показатель		
	Расходы на питание и проживание	Расходы на снасти и снаряжение	Пройденное расстояние
Среднеквадратическое отклонение	±75	±20	±500
Потребный объем выборки	36	16	100

Для каждой из трех оцениваемых величин получим свое значение объема выборки. В зависимости от величины объем выборки  $n$  должен быть равным 36, 16 или 100. Необходимо определенным образом согласовать эти три значения, так, чтобы принятый объем выборки отвечал сразу всем поставленным задачам. При консервативном подходе необходимо выбирать самое большое значение  $n = 100$ . Этим гарантируется нужная точность оценки каждой величины при условии, что оценки среднеквадратических отклонений были корректны. Теперь рассмотрим ситуацию, когда наименее критичной их трех оцениваемых величин будет расстояние, проходимое рыбаками. Этим можно снизить расходы на исследование, используя выборку меньшего объема. Оптимальный подход в подобных ситуациях состоит в выделении наиболее критичных величин и в соответствии заданной точности и достоверности определения объема их выборки. Величины, оценка которых требует большего объема выборки, в этом случае будут оцениваться с меньшей точностью или достоверностью, чем планировалось. Предположим, что наиболее критичным показателем является уровень расходов, и остановимся на значении объема выборки 36. Предположим также, что эта выборка (состоящая из 36 рыбаков) дает выборочное среднее 300 км и выборочное отклонение  $S = 500$  км. В данном случае результат выборки согласуется с первичной оценкой генерального среднеквадратического отклонения и, соответственно, неточность никак не сказывается на доверительном интервале:

$$\bar{X} \pm t_c \frac{S}{\sqrt{n}} = 300 \pm 2 \frac{500}{\sqrt{36}}, \text{ или } 133,3 \leq \bar{X} \leq 466,7.$$

В то время как заданная точность составляла  $\pm 100$  км, полученная точность равна  $\pm 166,7$  км. Для получения необходимой точности оценки доверительный уровень следует сделать меньше нынешнего 95%-го уровня.

## 1.6. Определение объема выборки при работе с выборочными долями

В приведенных выше примерах речь шла о средних. Однако многих заказчиков исследований интересуют другие параметры, например, выборочная доля  $p$ . В нашем примере может вызвать интерес определение доли лицензии рыболовов, прибывших из других мест или проживающих в сельской местности, или рыболовов, выезжающих с ночевкой. В рассмотренных методах речь шла о трех вещах, необходимых для определения объема выборки: определенном доверительном уровне, определенной точности и знании характера распределения выборочной статистики. Определение первых двух величин обусловлено спецификой постановки проблемы. Абсолютная точность, хоть и выраженная в процентном отношении, означает, что оценка будет находиться в интервале, определенном в долях истинного значения, например, в пределах  $\pm 5\%$  от истинного значения. *Теперь остается рассмотреть распределение выборочных долей.*

Среднеквадратическая ошибка доли определяется по формуле

$$\sigma_p = \sqrt{d(1-d)/n}, \quad (1.7)$$

где  $d$  – выборочная доля;  $\sigma_p$  – среднеквадратическая ошибка доли;  $n$  – объем выборки.

В этом случае средним значением является генеральная доля, а среднеквадратическим отклонением – среднеквадратическая ошибка доли. Отсюда объем выборки определяется по формуле

$$n = \frac{t_c^2}{H^2} d(1-d). \quad (1.8)$$

*Например, предположим, что необходимо узнать долю всех рыболовов, которые в течение года хотя бы раз выезжали на рыбалку с ночевкой. Предположим также, что нужно произвести оценку этой доли с точностью  $\pm 2\%$  и доверительным уровнем  $95\%$  ( $t_c = 2$ ).*

*Тогда*

$$n = \frac{2^2}{0,02^2} d(1-d).$$

*Данное уравнение содержит два неизвестных: оцениваемую долю признака в генеральной совокупности и объем выборки. Поэтому необходимо оценить первое значение и определить объем выборки,*

используя эту оценку. Для того чтобы провести предварительную оценку, необходимо обратиться к данным предыдущих исследований, либо провести предварительное обследование. Если в силу каких-то причин невозможно ни то, ни другое, можно провести оценку параметра на основе компетентного правдоподобного суждения. Следствием неточной оценки может стать увеличение или уменьшение точности доверительного интервала. Например, если исследователь принял долю рыболовов, выезжавших на рыбалку с ночевкой, равной 20 %, то в этом случае объем выборки составит:

$$n = \frac{2^2}{0,02^2} (0,20)(1 - 0,20) = 1600 .$$

Представим, что после обследования выборки, состоящей из 1600 рыболовов выборочная доля ( $p$ ) оказалась равной 0,40. Величина доверительного интервала будет зависеть от выборочной среднеквадратической доленой ошибки  $S_p$ , которая в данном случае подменяет собой неизвестную среднеквадратическую доленую ошибку генеральной совокупности  $\sigma_p$ :

$$S_p = \sqrt{\frac{pq}{n}}, \quad (1.9)$$

где  $p$  – доля единиц выборки;  $q = 1 - p$ .

В нашем примере:

$$S_p = \sqrt{\frac{0,40(0,60)}{1600}} = \sqrt{\frac{0,24}{1600}} = 0,012 .$$

Доверительный интервал для доли признака в генеральной совокупности определяется как выборочная доля  $p \pm ts_p = 0,40 \pm 2(0,012)$  или  $0,376 \leq d \leq 0,424$ .

Однако интервал оказался шире заданного. Причина состоит в том, что выборочная доля оказалась больше расчетной доли признака в генеральной совокупности.

Так как объем выборки прямо пропорционален произведению  $d(1 - d)$ , то произведение максимально при  $d = 0,5$ . Поэтому в отсутствии какой-либо информации о доле признака в генеральной совокупности можно принимать  $d = 0,5$ . При этом точность доверительного интервала возрастет в той же мере, в какой выборочная оценка будет отличаться от принятого значения, равного 0,5.

Необходимо отметить, что во всех выше приведенных расчетах по определению объема выборки генеральная совокупность не входит в формулу расчета, т. е. с оговоркой можно сказать, что объем выборки *не зависит* от генеральной совокупности.

Объясняется это тем, что если все элементы генеральной совокупности имеют одно и то же значение количественного признака (например, все рыболовы тратят на питание и проживание по \$74 в год), то для определения среднего достаточно будет отобрать выборку, состоящую из одного-единственного элемента. При этом количество элементов в исходной совокупности может равняться 1000, 10000 или 100000. Что действительно оказывает влияние на объем выборки, так это степень изменчивости количественного признака. Чем выше степень изменчивости признака, тем большим должен быть размер выборки, обеспечивающей заданную точность исследования. *Например, если наш район (область, страна) изобилует рыбой, привлекающей рыболовов и из других стран, то при определении среднего расстояния, преодолеваемого рыболовами во время их туров, можно столкнуться с огромным разбросом значений этой величины.*

Итак, объем генеральной совокупности оказывает на объем выборки лишь опосредованное влияние. В большинстве случаев с возрастанием объема исходной совокупности возрастает и **потенциальная возможность** увеличения изменчивости исследуемого количественного признака.

Объем выборки не зависит от объема генеральной совокупности и при оценке доли. Для доли определяющим фактором является расчетная доля генеральной совокупности, обладающая данным признаком: чем ближе эта доля к 0,5, тем большей должна быть выборка, при этом ее объем не зависит от объема всей совокупности. Значение 0,5 соответствует максимальной изменчивости, т. к. половина генеральной совокупности обладает данным признаком, а половина не обладает им.

Приведенные выше ситуации имеют отношение к тем, где целевая совокупность представляется, по сути, бесконечно большой (например, исследование потребительских товаров). В тех случаях, когда выборка представляет собой бóльшую часть генеральной совокупности, во избежание переоценки выборки требуется некоторое изменение формул расчета. И в первую очередь это связано с тем, что с увеличением объема выборки возрастает и стоимость исследования. В связи с этим в формулу вводят коэффициент поправки на конечность совокупности  $(N - n)/(N - 1)$ . Тогда формула расчета среднеквадратической ошибки среднего имеет следующий вид:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \quad (1.10)$$

где  $N$  – объем генеральной совокупности.

Когда требуемая выборка составляет более 5 % генеральной совокупности, ее расчетный объем должен быть уменьшен путем ввода в формулу поправки на размер совокупности:

$$n' = n \times \frac{N}{N+n-1}. \quad (1.11)$$

Например, генеральная совокупность состоит из 100 элементов, а оценка требуемого объема выборки составляет 20 элементов, при использовании указанной поправки исследованию должно подлежать меньшее число элементов. Требуемый объем выборки  $n' = nN/(N+n-1)$ , где  $n$  – первоначальное, а  $n'$  – пересмотренное значение объема выборки. Так, для  $N = 100$  и  $n = 20$  на деле можно обследовать только 17 элементов.

## Тема 2. СБОР ПЕРВИЧНОЙ ИНФОРМАЦИИ

### 2.1. Характеристика методов проведения опросов

Сбор первичной информации осуществляется с помощью различных методов сбора, к которым относят:

- опрос;
- наблюдение;
- эксперимент;
- имитационное моделирование.

**Опрос** позволяет выявить систему предпочтений, на которые ориентируется целевой рынок потребителей при выборе определенных товаров, оценке различных форм обслуживания, покупке изделий различных фирм. Это самый распространенный метод сбора данных в маркетинге (около 90 % всех исследований).

Опросы классифицируются по признакам (табл. 2.1).

Таблица 2.1

#### Классификация опросов

Признак классификации	Классификационные группировки
Источник первичной информации	Массовые Специализированные (экспертные)

Признак классификации	Классификационные группировки
Частота охвата	Точечные (разовые) Повторные
Степень охвата	Сплошные Выборочные
Форма опроса	Анкетирование Интервьюирование (прямое и опосредованное, например по телефону)

В **массовом** опросе основным источником информации выступают различные категории населения, профессиональная деятельность, которых не связана с предметом исследования.

В **специализированном** опросе главным источником информации являются компетентные лица, чья профессиональная деятельность тесно связана с предметом исследования.

**Повторные** опросы позволяют выявить жизненный цикл запросов и предпочтений, основные тенденции их изменений с течением времени для своевременного принятия решений о деятельности на рынке.

**Сплошные** исследования могут проводиться в отношении товаров производственного назначения, т. к. круг их потребителей относительно ограничен.

Сущность **выборочных** опросов заключается в том, что опрашивается часть всей изучаемой совокупности населения.

Рассмотрим более подробно классификацию опросов по форме. Основными, как отмечалось, являются личная беседа, телефонный опрос и опрос по почте.

**Личная беседа** представляет собой прямой диалог между интервьюером и респондентом. Беседа может происходить дома, в офисе или в каком-то общественном месте (например, торговый центр), где останавливают покупателя и просят ответить на ряд вопросов. Личная беседа позволяет реализовать гибкую тактику опроса, дополняя ответы наблюдениями интервьюера.

*Достоинство:* высокая достоверность полученных данных.

*Недостатки:*

- большая стоимость и продолжительность;
- вероятность влияния опрашивающего на мнение респондента;
- необходимость специальной подготовки интервьюера.

**Телефонный опрос** используется, когда возникает необходимость в кратчайшие сроки получить ответы на немногочисленные и

несложные вопросы. Применяется при осуществлении предварительных исследований для последующих личных опросов. *Достоинства:*

- быстрота и высокая результативность;
- незначительные затраты времени и сил.

В обоих случаях (личная беседа и телефонный опрос) интервьюер задает вопросы и записывает ответы респондента либо в процессе беседы, либо сразу после нее.

**Почтовый опрос** проводится посредством рассылки анкет выбранным респондентам с приложением оплаченного конверта для ответа. Респонденты заполняют анкеты в свободное время и отсылают ответы обратно.

*Достоинства:*

- возможность проведения на большой территории, в т. ч. труднодоступных районах;
- отсутствие психологического барьера и влияние опрашиваемого на ответ;
- отсутствие жестких временных рамок.

*Недостатки:*

- длительный период возврата ответов;
- неполный возврат анкет (число невозвратных анкет достигает более 90 %);
- трудности с определением достоверности полученной информации.

Разновидностью опроса по почте является потребительская панель, которая используется при опросе широкого круга потребителей и предприятий по большому числу вопросов. Потребительская панель позволяет получить необходимую информацию с помощью неоднократного опроса интересующей группы потребителей (чаще семей) через равные промежутки времени.

С помощью потребительской панели можно получить следующую информацию:

- количество товара, покупаемого семьей;
- размеры финансовых расходов;
- долю рынка основных производителей;
- предпочтительные цены, виды упаковки, виды товара, виды предприятий розничной торговли;
- различия в поведении потребителей, принадлежащих к различным социальным слоям, живущих в различных регионах, городах разной величины.

Однако использование потребительской панели связано с трудностью обеспечения репрезентативности:

– выбором необходимых объектов и получения их согласия на сотрудничество;

– «смертность панели», т. е. отказ участников от сотрудничества, смена места жительства, физическая смерть, переход в другую потребительскую категорию;

– «эффект панели», т. е. участники, чувствуя себя под контролем, сознательно или бессознательно меняют привычный образ поведения (например, снижается доля спонтанных покупок).

Для заинтересованности участники панели получают небольшое вознаграждение.

Еще одной разновидностью почтового опроса является **прессовый опрос**. В этом случае анкета печатается в газете или журнале.

*Недостаток* – относительно низкий возврат анкет.

Методы сбора информации по частоте использования делятся на два основных – *коммуникация* (опрос) и *наблюдение*. Классификация методов по признакам представлена в табл. 2.2.

Таблица 2.2

#### Методы классификации по признакам

Признак	Метод
<b>Коммуникация</b>	
Степень структурности	Структурированный Неструктурированный
Степень маскировки	Скрытый Нескрытый
Метод проведения	Личное интервью Телефонное интервью Анкета по почте
<b>Наблюдение</b>	
Степень структурности	Структурированный Неструктурированный
Степень маскировки	Скрытый Нескрытый
Параметры	Естественный Искусственный
Метод проведения	Человеком Механический

В коммуникацию входит опрос респондентов для подтверждения определенной информации с помощью опросного листа, в котором вопросы могут быть устными или письменными, и ответы могут быть даны в любой форме. Наблюдение же не включает в себя опросов.

Каждый метод сбора первичных данных имеет свои достоинства и недостатки. К достоинствам методов коммуникации относят многофакторность, скорость и цену, к достоинствам наблюдения – объективность и точность.

**Многофакторность** – это возможность сбора многих видов первичных данных, интересующих исследователя, таких как демографические или социо-экономические характеристики и образ жизни, взгляды и мнения индивидуума, мотивация, лежащая в основе действий, само поведение человека. Метод наблюдения может дать информацию лишь о поведении и некоторых демографических или социо-экономических характеристиках, при этом с определенными ограничениями (например, пол – очевидно, возраст и доход – опосредованно, социальный класс – неизвестно).

*Преимущества скорости и цены* тесно переплетаются. Даже если данные подтверждаются обоими методами, то коммуникация – более скоростной способ сбора информации, чем наблюдение, при котором можно долго ждать, чтобы что-либо увидеть.

Однако если наблюдения имеют ограничения в категориях масштаба, времени и денег, у них есть преимущества относительно объективности и точности. Это связано с тем, что наблюдение позволяет регистрировать поведение респондентов таким, какое оно есть, т. к. запись происходящего не зависит от памяти или настроения наблюдаемого.

## **2.2. Сбор информации посредством анкетирования**

При использовании метода опроса для сбора данных необходимо вначале решить, какой вид анкеты лучше поможет решить текущую задачу. Для этого важно определить характер **структуры (формализации)**, который требуется от анкеты, и **степень открытости** анкеты, которая соответствует исследуемой проблеме.

**Структура** – уровень стандартизации, предполагаемый инструментом сбора данных. *Высокостандартизованная анкета* – это такая, в которой задаваемые вопросы и ответы полностью предопределены темой опроса. *Предельно общая анкета* – это та, в которой

задаваемые вопросы четко не определены, и респонденты могут отвечать своими словами и в произвольной, подходящей для ответа форме.

**Закрытость** – объем информации о цели исследования, сообщаемый респонденту. *Открытая анкета* – это та, из которой цель исследования ясна.

Анкеты имеют классификацию, представленную в табл. 2.3.

В *стандартизированной открытой анкете* как ответы, так и вопросы являются стандартными, при этом используются **вопросы с многовариантным ответом**, в которых ответы ограничиваются перечисленными вариантами.

Основным преимуществом этих анкет является то, что они просты в заполнении и удобны для сведения ответов в таблицы и анализа результатов.

Таблица 2.3

Виды классификации анкет

Структура	Закрытость	Форма вопроса
Стандартизированные	Открытые	Вопрос с многовариантным ответом
	Закрытые	Вопрос, объединяющий скрытые подсознательные мотивы и формализованный вопрос
Нестандартизированные	Открытые	Вопрос с открытым ответом Глубокое интервью
	Закрытые	Словесная ассоциация Завершение предложения Составление рассказа Тематический апперцепционный тест (ТАТ)

*Например, если на вопрос «Как часто Вы смотрите телевизор?» не было предложено ни одной альтернативы, то один респондент мог бы ответить «каждый день», другой – «регулярно», а третий может указать количество часов в течение дня. Ответы в данном случае более сложно интерпретировать, чем результаты в случае с ограниченным набором ответов, допустим, по категориям «каждый день», «по крайней мере, три раза в неделю», «по меньшей мере, раз в неделю» или «меньше раза в неделю».*

Однако если эти вопросы дают более определенную информацию, они могут вызвать вводящие в заблуждение ответы. Например, если предлагаемые альтернативные ответы способны вызвать ответ на вопрос, о котором у респондента нет своего мнения, даже если существует категория «не знаю».

Следовательно, многовариантный ответ на вопросы анкеты является наиболее успешным в случае, когда возможные ответы являются хорошо известными, ограниченными по количеству и четкими. Они используются для получения мнений на вопросы, по которым люди имеют ясную точку зрения, и для извлечения реальной информации (возраст, образование, домашняя собственность и т. п.), не используются для сбора первичной информации в отношении мотивации.

*Нестандартизированные открытые анкеты* используются в случаях, когда цель исследования ясна, но ответ на вопрос остается открытым.

Например, вопрос «Что вы думаете о положении дел в частном предпринимательстве и необходимости расширения законодательства по данному вопросу?» ясен по своей цели. С его помощью пытаются вызвать респондента на свободный разговор о его соображениях по этой проблеме. Это **вопрос с открытым ответом**, т. к. он ведет к очень неформальному интервью (называемому **глубоким интервью**). Первоначальный ответ респондента, последующие вопросы и новые ответы определяют направление данного интервью.

**Вопрос с открытым ответом** – вопрос, характеризующийся тем, что респонденты могут отвечать своими собственными словами, а не ограничиваться возможностью выбора из определенного набора альтернатив.

**Глубокое интервью** – неформальная личная беседа, в которой интервьюер старается заставить собеседника говорить свободно и выразить свои истинные чувства.

*Достоинство* глубокого интервью – возможность получения более точного отношения респондента к некоторым проблемам.

*Недостатки:*

- наличие высококвалифицированных специалистов;
- более длительный период времени;
- меньший охват респондентов либо увеличение числа дополнительных интервьюеров;
- наличие квалифицированных психологов для интерпретации ответов;
- высокая стоимость услуг квалифицированных специалистов.

Таким образом, глубокое интервью лучше использовать для поисковых исследований, т. к. оно помогает решить практически все основные задачи, стоящие перед такого рода работой.

*Нестандартизированные закрытые анкеты* составляют основу исследования мотивов. В данном случае делается попытка преодолеть нежелание субъекта опроса обсуждать свои чувства, используя специальные методики, которые почти полностью исключают зависимость ответов от сознания опрашиваемого и его стремления не раскрыть себя. Данная техника опроса называется **проекционным методом**, и в ней основной упор делается на маскировке подлинного предмета исследования путем использования скрытых стимулов.

**Проекционный метод** – термин, используемый при описании анкеты, содержащей скрытые стимулы, которые заставляют опрашиваемого опираться на свои собственные эмоции, потребности, мотивации, предпочтения и ценности при формулировании ответа.

К наиболее часто используемым стимулам (побудителям) относятся *словесные ассоциации, завершение предложения, составление рассказа*.

**Словесная ассоциация** – вопросник, содержащий список слов, к которым респондент после их прочтения должен прибавить первое пришедшее в голову слово.

В словесной ассоциации слова настоящего текста смешиваются в списке с нейтральными словами, чтобы скрыть подлинную цель исследования.

*Например, если исследовать проблемы, связанные с загрязнением окружающей среды, то некоторые из ключевых слов могут быть такими: вода, воздух, озера, промышленность, заводская труба, город.*

Ответы на каждое ключевое слово записываются один за другим, а затем анализируются. Ответы обычно оцениваются по трем направлениям:

1) по *частоте, с которой слово упоминается в ходе ответов*. То есть все одинаковые ответы, которые были даны, группируются, чтобы выявить варианты интересов, скрытых мотивов или стереотипов. Часто ассоциации разбивают по таким категориям, как *предпочтительные/нежелательные, приятные/неприятные, современные/старомодные* и т. д.;

2) по *среднему промежутку времени*, который проходит перед получением ответа. В данном случае может использоваться секундомер или интервьюер может тихо считать до получения ответа. Рес-

понденты, которые колеблются (те, у кого более трех секунд уходит на ответ), рассматриваются как эмоционально связанные с данным словом, и здесь учитывается не столько их реакция на вопрос, а то, как они представляют себе подходящий ответ;

3) по **числу респондентов, которые не ответили на ключевые слова вообще** после истечения определенного времени. В данном случае эмоциональная заинтересованность респондента оценивается как настолько высокая, что способна привести к блокированию ответа.

Набор всех ответов отдельного человека вместе с деталями ответа на каждый вопрос затем используется для оценки личных предпочтений или ощущений в отношении предмета исследования.

**Завершение предложения** – вопросник, содержащий ряд предложений, которые участники опроса должны завершить первыми пришедшими в голову словами. Ответы дословно записываются и впоследствии анализируются.

*В одном исследовании опросили 179 курильщиков, которые верили, что сигареты приносят вред здоровью, почему они продолжают курить. Большинство ответили «Удовольствие важнее здоровья», «При умеренном курении все будет в порядке», «Мне нравится курить». При этом впечатление сложилось, что курильщики довольны своим положением. Но в той части исследования, где использовались тесты на завершение предложения, курильщики отвечали на вопрос «Люди, которые никогда не курят, ...» словами «более счастливы», «Более умные», «лучше выглядят». На вопрос «Подростки, которые курят», использовались ответы «глупые, дураки, сумасшедшие, выпендриваются» и т. п. Вывод, который можно сделать из теста на завершение предложений, состоит в том, что курильщики обеспокоены, испытывают чувство дискомфорта, внутреннего разлада и неудовлетворенности из-за своей привычки. Это в корне отличается от результатов опроса со свободным выбором ответов.*

**Достоинство:** предоставление более ясного побудительного мотива по сравнению со словесной ассоциацией.

**Недостаток:** наличие высококвалифицированных специалистов для разработки хороших тестов.

**Составление рассказа** – метод сбора данных в ходе опроса, опирающийся на образительные стимулы, такие, как карикатуры, фотографии или картинки, на основе которых участника опроса просят составить рассказ.

Определенный набор рисунков, с помощью которых опрашиваемый должен составить небольшой рассказ, называется **тематическим апперцепционным тестом (или тестом на восприятие)**. Некоторые из этих картинок изображают обычные события, а некоторые что-то непонятное. На некоторых люди или объекты нарисованы четко и ясно, а на других они относительно размыты. Способ, которым исследуемый реагирует на эти картинки, помогает, например, психологам интерпретировать его личность, т. е. импульсивен или контролирует себя, творческая натура или не обладает воображением и т. д.

При использовании данного метода в маркетинговых исследованиях работа проводится точно так же. Респондентам показывается изображение, а потом их просят рассказать об этой картинке, только ответы используются для оценки отношения к предмету исследования, а не для характеристики личности опрашиваемого.

*В примере с загрязнением окружающей среды побудительным стимулом может служить изображение большого города, а респондента попросят описать, на что будет похожа жизнь в подобном месте.*

*Стандартизированные закрытые анкеты* представляют собой вариант, объединяющий преимущества использования скрытых подсознательных мотивов и предпочтений с преимуществами формализованного подхода в вопросах расшифровки и оформления полученных ответов.

Например, известно, что индивидуальные знания, восприятие и память о каком-то предмете впоследствии определяет отношение к нему со стороны человека. При этом, чтобы не получить искаженную картину в случае прямого вопроса нужно просто спросить респондента о том, что он знает, а не выяснить его мнение о предмете исследования. Считается, что чем больше человек знает о предмете исследования, тем сильнее и определеннее будет его предварительное мнение о нем.

*От тех, кто голосует за демократическую партию, следует ожидать большей информированности о демократических кандидатах и их платформе, чем от голосующих за коммунистов.*

Главные преимущества данного метода проявляются в процессе анализа ответов. Они легко кодируются и классифицируются, из них можно быстро извлечь объективный показатель информированности о проблеме, однако в данном случае может возникнуть ситуация, при которой ответы респондента могут не показывать личных предпочтений, а просто говорят о том, что респондент находится в курсе текущих событий или проблемы.

### 2.3. Сбор информации посредством наблюдения

*Наблюдение – факт повседневной жизни. Мы постоянно наблюдаем за другими людьми и отдельными событиями для получения информации о мире вокруг нас. Соответственно, некоторые люди используют подобные наблюдения более продуктивно, чем остальные.*

*Одной из причин, часто приводимых в качестве объяснения успехов японцев в деле разработки новых продуктов, является то, что они являются мастерами наблюдения. Японские торговцы проводят много времени за тем, как покупатели используют купленные товары.*

Как и коммуникационные методы, методы наблюдения могут быть **стандартизированные** и **нестандартизированные**, **закрытые** или **открытые**. Наблюдение может проводиться естественным или искусственным образом.

**Наблюдение** – обычно используется в маркетинговых исследованиях поискового характера и представляет собой изучение поведения объекта исследования в естественной обстановке.

Преимущества наблюдения по сравнению с опросом:

- простота и низкие затраты на реализацию;
- независимость от желания объекта к сотрудничеству и его способности к словесному выражению сути дела;
- возможность обеспечивать более высокую объективность;
- возможность обеспечивать восприятие неосознанного поведения (выборов товара на полках магазинов);
- естественность обстановки исследования.

Недостатки:

- трудно обеспечивать репрезентативность (например, только тех, кто пришел в магазин, случайный выбор объектов невозможен);
- субъективность восприятия наблюдающего;
- не принимаются во внимание мотивы поведения потребителей, их интересы и факторы, определяющие поведение;
- невозможность однозначного поведения объекта наблюдения;
- поведение объектов может отличаться от естественного, если наблюдение ведется открытым способом (эффект наблюдения).

В маркетинге используются различные формы наблюдения, которые классифицируются по признакам:

- 1) по **характеру окружающей обстановки** наблюдение может быть **полевым** (в естественной обстановке) и **лабораторным** (в искусственно созданной ситуации).

**Естественные условия** это когда субъекты наблюдения исследуются в обычной среде, где им свойственно обычное поведение.

**Искусственные условия** это когда субъекты наблюдения изучаются в специально сконструированной обстановке, созданной для наблюдения за их поведением.

Преимущество первой заключается в естественности наблюдаемого поведения. Вторая позволяет поддерживать более стабильные условия наблюдения и испытывать технические средства;

2) по **способу осуществления** наблюдение может быть **скрытым** (с применением специальных камер, системы зеркал и пр.) и **открытым** (с непосредственным участием наблюдающего).

В случае **открытого наблюдения** субъекты наблюдения предупреждены о том, что за ними следят, при **скрытом наблюдении** участники не предупреждаются о том, что за ними идет наблюдение;

3) по **степени стандартизации** различают **стандартизированное** и **нестандартизированное (свободное)** наблюдение. Стандартизация подразумевает задание определенных категорий поведения, схем поступков.

**Стандартизированное наблюдение** – это когда проблема четко определена, поэтому виды поведения, подлежащие наблюдению, оговорены предварительно, так же, как и категории, которые будут использованы для записи и анализа ситуации.

**Нестандартизированное наблюдение** – это когда проблема не определена в конкретной постановке, поэтому допускается большая гибкость в использовании наблюдателями терминов, в которых они будут отражать сведения и результаты.

*Например, для наблюдения действенности рекламы в витрине магазина можно выработать следующие возможности поведения прохожих:*

- человек вошел в магазин, не взглянув на рекламу;
- человек вошел в магазин, предварительно взглянув на рекламу;
- человек взглянул на рекламу, но не вошел в магазин;
- человек прошел не взглянув на рекламу в витрине.

Особым видом наблюдения является системный сбор информации в определенном количестве торговых предприятий (панель розничной торговли). С помощью такой панели можно получить информацию:

- о развитии сбыта определенных групп товаров;
- об объемах сбыта в натуральном и стоимостном выражении;

- о средних запасах, сбыте, закупках по каждому предприятию, входящему в панель;
- о скорости сбыта.

Эта информация позволяет сделать вывод о путях распределения изделий, об эффективности работы сбытовых служб, о путях поступления товаров в розничную торговлю.

#### **2.4. Эксперимент и имитационное моделирование**

Под **экспериментальными исследованиями** понимается сбор первичной информации путем выбора однотипных групп обследуемых, выдачи им разных заданий, контроля за факторами, которые влияют на результаты, и сравнения различий в групповых реакциях (*например, по выявлению реакции на различные цены*).

**Экспериментом** называется манипулирование независимыми переменными с целью определения степени их влияния на зависимые переменные при сохранении контроля за влиянием других, не изучаемых параметров. Независимые переменные могут меняться по усмотрению экспериментатора (цены, затраты на рекламу и т. п.), в то время как зависимые переменные практически не находятся в сфере его непосредственного управления (объем продаж, показатель рыночной доли).

Выделяют два типа экспериментов:

- лабораторные;
- полевые.

К **лабораторным** относятся эксперименты, при проведении которых соблюдаются определенные искусственные условия с целью исключить влияние побочных факторов (*например, при оценке реакции покупателей на различные виды рекламы можно пригласить таких покупателей, чтобы они были репрезентативными с точки зрения пола, возраста, социального положения и т. п.*).

Помимо контроля над побочными факторами лабораторные эксперименты являются также более дешевыми и требуют меньше времени для своей реализации.

*В качестве примера можно привести результаты экспериментального изучения уровней запоминания и забывания рекламных сообщений.*

*Эксперименты показали, что доля людей, сохраняющих внушенное впечатление от рекламы, изменяется во времени по геометрическому закону (рис. 2.1).*

Однако скорость снижения способности вспомнить сильно варьируется в зависимости от того, что надо запомнить. Видно, что тема рекламного сообщения запоминается лучше, чем его источник. Аргументы, положенные в основу рекламного сообщения, запоминаются хуже всего. Способность вспомнить как тему, так и источник и аргументы рекламного сообщения резко падает в течение первой недели после появления рекламного сообщения. Следовательно, у рекламодателя имеется очень короткий срок для оправдания затрат на рекламу, или реклама требует повторения.

В результате исследования делается вывод о том, что рекламная телевизионная компания, состоящая из 6 повторений в первой волне и обеспечивающая уровень способности ее вспомнить, равный 60 %, не должна прерываться более чем на три месяца, если данный уровень запоминания нежелательно опускать ниже 20 %.

Здесь речь идет только об усредненных оценках. Реальные оценки в сильной мере зависят от ценности рекламы, творческого уровня ее исполнения.

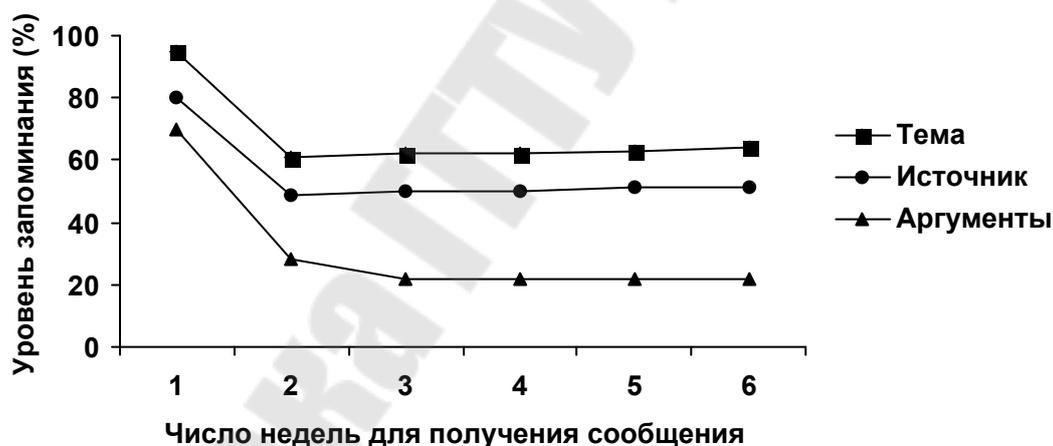


Рис. 2.1. Способность вспомнить рекламу как функцию времени

За последние годы при проведении лабораторных экспериментов все шире используется компьютерная техника. Существуют компьютеризованные программы-вопросники, которые позволяют потребителям «бродить» по супермаркету и выбирать товары для покупки. Компьютер регистрирует их покупки и измеряет их реакцию на применение определенных элементов комплекса маркетинга (цену, цвет и форму упаковки, внутримаркетинговые методы продвижения товара).

**Полевые эксперименты** проводятся в реальных условиях: в магазинах, на дому у потребителя и т. д. Результаты этих экспериментов заслуживают большего доверия, чем лабораторные, но при их проведении сложно точно учесть влияние побочных факторов, они требуют большего времени для реализации и больших затрат.

Полевые эксперименты чаще называют **пробным маркетингом**. Проводиться он может в одном или нескольких городах и обычно преследует две цели:

- определить потенциальный объем продаж нового товара;
- определить эффективность применения отдельных элементов комплекса маркетинга при реализации нового товара (как по отношению потребителей, так и посредников).

**Объектом** изучения пробного маркетинга являются различные рынки, поэтому такую деятельность часто называют **тестированием рынка**.

Тестирование рынка классифицируется:

- на стандартное;
- на контролируемое;
- на электронное;
- на имитационное.

При применении **стандартного тестирования** рынка предприятие испытывает продукты и другие переменные комплекса маркетинга через обычные каналы сбыта товара, используемые на данном предприятии. Этот метод является довольно дорогим, требует много времени и не является конфиденциальным.

*Например, было проведено тестирование влияния уровней рекламы и цен на объем продаж нового продукта питания в США. В четырех городах использовалась выборка, которая включала 30 магазинов. В двух городах проводилась очень интенсивная реклама, по объему превышающая в два раза рекламу, проводимую в двух других городах. Кроме того, все 30 магазинов были разбиты на три панели по 10 магазинов. При этом учитывались размеры магазинов, их местоположение в городе и другие факторы. Для каждой панели использовался свой уровень цен: базовая цена – ниже 50 центов; цена – на 10 центов выше базовой цены; цена на 20 центов выше базовой. Тестирование проводилось в течение 6 месяцев. Каждый месяц в каждом магазине фиксировался объем продаж. Было установлено, что более высокий уровень рекламы является более эффективным при использовании базовой цены и не оказывает влияния на объем продаж при использовании самой высокой цены (рис. 2.2).*

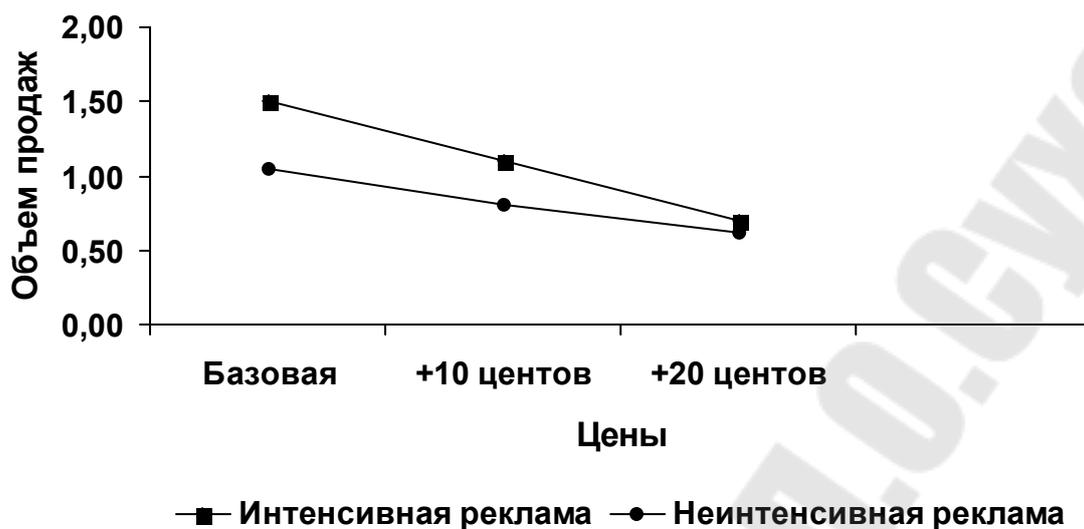


Рис. 2.2. Влияние уровней рекламы и цен на объем продаж

До проведения эксперимента была принята гипотеза, согласно которой при продаже (в данном случае пива) по всей стране целесообразно использовать две стратегии: 1) сочетание высоких затрат на рекламу с высокой ценой; 2) сочетание низких затрат на рекламу с низкой ценой. Предполагалось, что высокие цены покроют дополнительные затраты на рекламу. Тестирование рынка показало, что надо придерживаться другой стратегии: низкая цена – высокие затраты на рекламу. Из результатов эксперимента также обнаружилось, что проведение тестирования рынка при варьировании только одной переменной – затрат на рекламу или цены – дало бы неверные результаты.

**Контролируемое тестирование** рынка проводится специализированными исследовательскими фирмами, которые осуществляют реализацию товаров через определенных дистрибьюторов, поощряемых за участие в проведении эксперимента. Недостатком данного метода является то, что сбытовые каналы дистрибьюторов могут не соответствовать каналам, используемым предприятием в практической работе.

**Электронное тестирование** рынка заключается в том, что участники потребительской панели получают специальные идентификационные карточки, которые они предъявляют при покупке товара. При покупке тестируемых товаров или их отсутствии автоматически фиксируются демографические характеристики покупателя. (Такие испытания проводятся в городах, в которых предприятия торговли согласились принять участие в эксперименте).

Электронное тестирование рынка обеспечивает быстроту и низкую стоимость проведения эксперимента, конфиденциальность полученных результатов. Но тестируемый рынок может не соответствовать реальному рынку.

**Имитационное моделирование (имитационное тестирование)** заключается в построении математической, графической или иной модели контролируемых и неконтролируемых факторов, определяющих стратегию и тактику фирмы и в последующих экспериментах на модели с целью изучения влияния изменений этих факторов на объект исследования.

Наиболее широко применяется моделирование, обобщающее, объясняющее и прогнозирующее поведение потребителей. Зависимые и независимые переменные моделей покупательского поведения выражают особенности поведения покупателей в отношении конкретных товаров в рамках открытых сегментов рынка. К наиболее известным моделям конкретного поведения относятся:

– **стохастические (или вероятностные)**, которые основываются на предположении, что предыдущие, и особенно последняя, покупки определяют будущее поведение потребителя на рынке. Они используются при определении степени лояльности к маркам товаров, частоты смены привычных марок и прогнозами спроса на новые товары;

– **линейно-экспериментальные** – это модели, применяемые обычно для моделирования потенциального спроса на конкретные товары. Они отличаются формальной математической структурой и отражают тенденции поведения потребителей относительно отдельных товаров (например, доля той или иной торговой марки в совокупном объеме покупательского спроса). Такие модели носят обычно описательный характер и представляют собой обобщение движения спроса населения на всем рынке;

– **модели переработки информации**, строящиеся на том, что покупатель постоянно получает информацию из многих источников, перерабатывает ее и принимает решение о выборе того или иного товара.

*Достоинства* данного метода по сравнению со стандартным методом:

– на его реализацию требуется 18–24 недели (для стандартного – 12–18 месяцев);

– стоимость этого метода составляет около 5–10 % от стоимости стандартного;

- обеспечивает конфиденциальность;
- могут быть испытаны различные комплексы маркетинга.

*Недостатки* данного метода:

- не так полон, как полномасштабные испытания рынка;
- полученные результаты во многом зависят от предположений, заложенных в модели.

Недостатки методов тестирования рынка:

- высокая стоимость даже при проведении тестирования всего в нескольких городах и при ограниченной исследовательской программе;
- тестируемый продукт может стать известен конкурентам, которые могут повлиять на результаты (например, закупкой экспериментального товара в больших количествах для искажения результатов);
- наличие временной задержки между проведением эксперимента и принятием маркетинговых решений;
- не достаточно контролируемые.

Достоинство: возможность всестороннего изучения множества факторов, определяющих стратегию маркетинга.

*Рассмотрим пример поведения домохозяек на рынке потребительских товаров.*

*Сначала проводится сегментация домашних хозяек на несколько типов, которые характеризуются различными величинами вероятностей осуществления тех или иных покупок. Домохозяйки, включенные в каждый выявленный тип, образуют представительную выборку из населения, полученную на основе ряда переменных, характеризующих использование товаров населением. Поскольку система предпочтений осуществления различных покупок меняется во времени, то величины вероятностей, приписываемых домохозяйкам, характеризуют склонность к покупкам в пределах определенного диапазона времени.*

*Затем, для имитации первой группы покупок, в компьютер вводятся вероятности начальных покупок. Допустим, что продуктом первой группы покупок является кофе, и что для определенного типа домохозяек вероятность покупки кофе в течение недели равна 0,75, вероятность покупки быстрорастворимого кофе равна 0,35 и вероятность покупки быстрорастворимого кофе марки «А» равна 0,20. Это означает, что 75 % домохозяек данного типа покупают обычно кофе в течение данной недели; из тех, кто купил кофе, 35 % обычно покупают быстрорастворимый кофе; и из тех, кто купил быстрора-*

створимый кофе, 20 % обычно покупают кофе марки «А». (Данные вероятности получают, как правило, с помощью социологических и экспертных методов, а также на основе торговой статистики).

Процесс в компьютере осуществляется по типу рулетки, на колесе которой имеется 75 отделений с «да» и 35 отделений с «нет». Компьютер решает, купит ли домохозяйка на этой неделе кофе или нет. Если ответ «да», то вращается другое колесо, имеющее 35 отделений для быстрорастворимого кофе и 65 отделений для кофе в зернах. «Рулетка» решает, какой кофе купит эта домохозяйка. Если ответ – «быстрорастворимый», то вращается колесо 20–80, чтобы решить вопрос, будет ли купленный быстрорастворимый кофе марки «А».

После расчета имитируемого поведения при покупке первой домохозяйки компьютер фиксирует решение и устанавливает вероятности новых покупок кофе в течение рассматриваемого времени. Если покупка была сделана, вероятность другой покупки кофе на данной неделе уменьшится, т. к. имитируемая домохозяйка его уже имеет. Если покупка кофе не была сделана, вероятность его покупки на следующей неделе увеличится.

Компьютер повторяет этот процесс для всех моделируемых домохозяек, решая, какие из их числа сделают покупки, а какие – нет, изменяя вероятности и фиксируя полученные результаты. После того как подобным образом изучится поведение всех домохозяек, процесс имитирования для первой недели считается законченным.

Имитация покупок для второй недели осуществляется подобным же образом за важным исключением: берутся в расчет «события» первой недели. Домохозяйки, которые сделали покупки, возможно сделают новые, но с меньшей вероятностью, т. к. определенное количество нужного товара у них уже имеется. Домохозяйки, которые не сделали покупок в течение предыдущей недели, возможно сделают такие покупки в течение данной недели, вероятность чего является более высокой, т. к., скорее всего, запасы кофе закончились. Характер и количество этих изменений следует основывать на данных научных исследований о скорости использования продуктов домохозяйками различных типов. Повторяя подобные расчеты еженедельными циклами, можно перекрыть любой промежуток времени.

Когда компьютер рассчитывает циклы покупок для всего выбранного периода времени, он распечатывает данные о покупке кофе в целом, общих покупках быстрорастворимого кофе и общих покуп-

*как кофе марки «А». Эти цифры затем могут быть сравнены с отчетностью магазинов или другими внешними источниками информации. Таким образом определяется адекватность имитационной модели реальным условиям.*

Тестирование рынка осуществляется как для потребительских товаров, так и для товаров производственного назначения.

### **Тема 3. РАЗРАБОТКА АНКЕТЫ ДЛЯ СБОРА ПЕРВИЧНОЙ ИНФОРМАЦИИ**

#### **3.1. Этапы разработки вида анкеты и корректировка вопросов**

*Создание анкеты является творческим процессом. Большинство успехов относится, в первую, очередь, к определению того, чего нужно избегать в анкете, например, направляющих или неясных вопросов. Однако существует мало конкретных указаний на то, как сделать так, чтобы вопросы не оказались наводящими или нечеткими.*

Порядок разработки анкеты можно представить в виде следующих этапов:

- 1 этап. Определение необходимой информации.
- 2 этап. Определение вида анкеты и метода проведения опроса.
- 3 этап. Определение содержания вопросов.
- 4 этап. Определение формы ответов на вопросы.
- 5 этап. Определение формулировки вопросов.
- 6 этап. Определение последовательности вопросов.
- 7 этап. Определение физических характеристик анкеты.
- 8 этап. Перепроверка этапов 1–7 и при необходимости их корректировка.
- 9 этап. Проведение предварительного тестирования анкеты и при необходимости ее корректировка.

Данные этапы представляют собой не эталон, а примерный перечень процедур, которые могут выполняться при разработке анкеты.

Рассмотрим каждый этап более подробно.

##### **1 этап. Определение необходимой информации**

Первый шаг в разработке анкеты, состоящий в определении характера информации, которую надо получить, является несложным при условии, что предварительные стадии исследования были проведены достаточно тщательно и точно.

*Как описательные, так и казуальные исследования требуют определенных знаний по теме, чтобы сформулировать некоторые специальные гипотезы о направлении работы. Они и определяют характер анкеты, т. е. какая информация собирается и от кого. Если макеты таблиц для занесения результатов анализа уже разработаны, то работа по определению характера собираемой информации практически завершена.*

**Макет таблицы** – это таблица, созданная для регистрации информации, которая будет собираться. Она совпадает с таблицами, которые будут использованы в процессе работы, но на данной стадии еще не заполнена цифрами.

*Информация собирается о переменных, указанных в макетах таблиц, для проверки выдвинутых гипотез. Сведения необходимо получать от правильно выбранных людей и в правильном формате. Поэтому гипотезы не только являются указанием, какую информацию следует искать, но и определяют тип вопроса и форму ответа, используемую для ее получения.*

Подготовка анкеты может способствовать выдвижению новых гипотез, которые могли бы быть исследованы при небольших дополнительных усилиях и затратах. Но при этом следует учесть, что если эти новые гипотезы важны для понимания явления, то их следует использовать при разработке анкеты, а если они не являются принципиально необходимыми для данного исследования, то использовать их не следует. Это связано с тем, что они просто удлиняют анкету, вызывают дополнительные проблемы в процессе ее заполнения и анализа, а иногда и увеличивают уровень отказов отвечать на вопросы.

## **2 этап. Определение вида анкеты и метода проведения опроса**

После определения того, какая информация должна быть собрана, необходимо выяснить, как она будет собираться. Эти решения касаются степени стандартизации и открытости анкеты, а также методов, которыми она будет заполняться, – по почте, по телефону или при личной встрече.

*Например, если использовать закрытую нестандартизированную форму, при которой участникам опроса показывают картинку и просят составить по ней рассказ, то телефонное интервью отпадает, и даже опрос по почте может породить серьезные проблемы, поэтому использовать почтовый опрос для нестандартизированной закрытой анкеты, которая содержит вопросы со свободной формой, не рекомендуется.*

Так как тип собираемых данных имеет большое влияние на метод сбора информации, то необходимо определить, какая первичная информация нужна, как эти данные могут быть собраны, какую степень стандартизации и открытости следует использовать, и как будет проводиться опрос.

### **3 этап. Определение содержания вопросов**

Предварительные решения в отношении характера необходимой информации, стандартизации и открытости вопросов, используемых при составлении анкеты, о методе, которым будет проводиться анкетирование, во многом определяют решения в отношении конкретного содержания вопросов. При этом необходимо ответить на ряд вопросов.

#### **1. Является ли вопрос необходимым?**

Если данный момент не охватывается в достаточной степени остальными вопросами, то должен быть введен новый дополнительный вопрос. Вопрос должен быть сформулирован так, чтобы обеспечить достаточную подробность в ответе, но не больше, чем это необходимо.

*Например, очень часто в маркетинге используется концепция стадии жизненного цикла, чтобы объяснить потребительское поведение семьи. Стадия жизненного цикла – это переменная, которая состоит из нескольких элементов, включающих семейное положение, наличие детей, возраст детей. Наличие детей считается важным фактором, поскольку оно наиболее часто свидетельствует о наличии отношений зависимости. Это особенно справедливо, если младшему ребенку меньше 6 лет, что создает один тип ответственности, тогда как дети между 6-ю и 17-ю годами порождают другие отношения ответственности для родителей. В исследовании, использующем понятие стадии жизненного цикла как переменный фактор, нет никакой необходимости узнавать возраст каждого ребенка. Правильнее будет задать всего один вопрос относительно возраста самого младшего ребенка в семье, если в ней вообще есть дети.*

#### **2. Нужно ли задавать несколько вопросов вместо одного?**

Эта ситуация связана с тем, что при ответе на один и тот же вопрос респондент может дать несколько разных обоснования.

*Например, на вопрос «Почему вы используете зубную пасту «Жемчуг»?» один респондент может ответить: «Чтобы сократить количество дырок в зубах», а другой – «Потому что мне рекомендовал врач». Очевидно, что применены два разных обоснования при ответе на один и тот же вопрос. Первый человек отвечает на вопрос с точки зрения текущего использования пасты, в то время как*

другой – почему он начал использовать именно эту пасту. Поэтому лучше разбить этот вопрос на две части, которые отражают возможные направления ответа. Например:

– Почему вы в первый раз использовали пасту Жемчуг?

– В чем основная причина того, что вы пользуетесь ею?

### **3. Имеет ли респондент необходимую для ответа информацию?**

Для этого необходимо тщательно проверить каждый вопрос с точки зрения того, располагает ли типичный респондент той информацией, которую необходимо получить. Так как респонденты будут давать ответ, а вот будут ли иметь эти ответы какой-то смысл, это уже другое дело.

Поэтому, чтобы не получать бесполезные ответы, респонденту надо задавать те вопросы, которые имеют для него смысл. Во-первых, респондент должен быть проинформирован в отношении предмета, о котором проводится опрос, во-вторых, респондент должен помнить эту информацию.

Например, возьмем вопрос «Сколько денег ваша семья тратит в продовольственном магазине за обычную неделю?». Если респондент сам не делает покупки или семья имеет ограниченный бюджет, то он вряд ли знает такую информацию. В подобной ситуации полезно использовать «фильтрующие вопросы» перед тем как задать основной, чтобы определить имеется ли на самом деле у человека необходимая информация. Таким проверочным вопросом может стать вопрос «Кто делает продовольственные покупки в вашей семье?». Считается нормальным использовать вопросы такого рода: «У вас есть свое мнение о ...?» перед тем, как задать специальный вопрос о предмете проводимого опроса. Практика показывает, что использование фильтрующих вопросов повышает долю ответов «не знаю» на 20–25 %.

Респондент не только должен что-то знать о предмете исследования, но и помнить эту информацию. Способность человека запоминать отдельные события определяется не только самим событием, но и его важностью, временем, прошедшим с момента, когда оно произошло, присутствием или отсутствием стимулов, которые помогают его запомнить. Более важные события легче запоминаются, чем менее значимые. При этом также нужно учитывать тот факт, что способность человека запоминать какой-то факт зависит от того, сколько времени прошло с того момента, когда произошло событие.

*Например, мы можем вспомнить программу, которую смотрели вчера вечером, но гораздо труднее вспомнить то, что мы смотрели на прошлой неделе, и почти невозможно припомнить увиденные месяц назад программы. Суть всего этого состоит в том, что если событие рассматривается как относительно незначимое для большинства респондентов, то нужно интересоваться только самыми последними известиями о нем.*

Для более важных событий существуют две силы, действующие в противоположных направлениях, которые влияют на способность респондента давать достоверный ответ на вопросы, относящиеся к конкретному периоду времени. Одна из них – это **ошибка преувеличения**, которая представляет собой склонность помнить событие так, будто оно произошло значительно позднее, чем это было на самом деле. Другая – это **потеря воспоминания**, тенденция забывать относительно незначимые события целиком.

#### **4. Поделится ли респондент информацией?**

Нежелание респондентов поделиться информацией может зависеть от количества усилий, требующихся для ответов, их способности сформулировать ответ или чувствительности к данной тематике. Респонденты также могут отказаться от ответа, когда проблема смущает или, наоборот, представляется им угрожающей.

Подобные вопросы должны быть устранены везде, где это возможно. Если это невозможно из-за важности вопроса для исследования, то необходимо уделить особое внимание тому, как подать вопрос, особенно в плане месторасположения или формулировки вопроса.

В целом лучше поместить деликатный вопрос далее по ходу анкеты, чем в начале опроса.

При использовании деликатных вопросов, полезно использовать средства, которые помогают сделать их менее угрожающими:

- 1) поместить вопрос среди группы других, более нейтральных;
- 2) перед тем как задать особый вопрос, следует убедиться, что поведение или восприятие является самым обычным.

*Например, «Последние исследования показывают, что одно из четырех семейств сталкивается с ежемесячными трудностями при выполнении своих финансовых обязательств». Данная техника, известная как использование успокаивающих заявлений, делает более легким для респондентов преодолеть потенциальное замешательство в их поведении.*

3) формулировка вопроса в отношении других и о том, как они могут действовать или что чувствовать.

Например, «Не кажется ли вам, что большинство из нас мошенничает с налогами? Почему?» Когда вопрос задается в подобной манере, есть большая уверенность в том, что респондент может приоткрыть свое истинное отношение к вопросу о мошенничестве с налоговыми декларациями, по сравнению с вероятностью отказа, если бы его прямо спросили о том, мошенничает ли он со своими налогами и почему он это делает;

4) формулировка ответа в виде ряда категорий, чтобы респондент мог бы просто выбрать подходящий вариант. Вместо того чтобы спрашивать о возрасте респондента, например, можно просто спросить его по возрастным категориям. Например, 20–29; 50–59; 30–39; 40–49; 60 и более.

#### **4 этап. Определение формы ответов на вопросы**

После того как определено содержание каждого отдельного вопроса, необходимо решить, какие вопросы следует использовать в анкете. Чаще на практике используются следующие виды вопросов:

– **вопросы с произвольным вариантом ответа.** Это вопрос, на который респонденты отвечают на такие вопросы своими собственными словами, не ограничиваясь выбором из предлагаемого набора вариантов;

– **многовариантный вопрос.** Это вопрос с ограниченным набором ответов, из которого респондента просят выбрать вариант, который наиболее соответствует его собственной позиции по данному вопросу.

*Например, студенты отвечают на вопросы подобного многовариантного формата на экзаменах или при заполнении тестов (опрос по возрасту, доходу, причине покупки товара и т. п.).*

– **дихотомический вопрос.** Вопрос с фиксированным количеством вариантов ответа, при котором респондента просят указать тот из двух ответов, который наиболее полно соответствует его взглядам на предмет вопроса.

*Например, Вы собираетесь покупать автомобиль в этом году?*

*ДА      НЕТ*

– **шкалы.** Это другой тип вопросов с фиксированными вариантами ответов, который использует определенную шкалу, для того чтобы показать ответ.

Например, когда проводят опрос об использовании видеоманитофонов, может быть задан следующий вопрос: «Как часто вы записываете на видеоманитофон телепрограммы для последующего просмотра?»

НИКОГДА      ИЗРЕДКА      ИНОГДА      ЧАСТО

В данном случае использована форма многовариантного вопроса. Однако эти ответы могут быть также даны при помощи шкалы. Шкаловая природа данного вопроса возможно будет более очевидна, если при составлении ответа будет использована следующая форма, представленная на рис. 3.1.



Рис. 3.1. Схема ответов шкалирующих вопросов

Преимущество данной схемы состоит в том, что описания или категории могут быть представлены в верхней части страницы, а разновидности программ могут быть напечатаны по левой стороне листа – например, фильмы, спортивные передачи, специальные программы. Респондентов просят отметить на шкале, с какой частотой используется видеоманитофон для записи тех или иных передач. Инструкцию можно дать только один раз, в самом начале, и при этом большое количество информации от респондента может быть получено от участника опроса за небольшое время.

### **5 этап. Определение формулировки вопросов**

Формулирование каждого из вопросов – это сложная задача, т. к. неправильная формулировка может вызвать отказ респондентов отвечать на вопрос, даже если они и согласились принять участие в опросе. Плохая формулировка может также спровоцировать респондента ответить на вопрос умышленно или из-за недопонимания некорректно.

Поэтому следует избегать следующих моментов:

- состояние неответа;
- избегать сложных понятий и слов;
- избегать сомнительных слов и вопросов;
- избегать наводящих вопросов;
- избегать скрытых альтернатив;
- избегать скрытых допущений;

- избегать обобщений и оценок;
- избегать двухканальных вопросов.

### **Состояние неответа**

Может создать большие проблемы при анализе информации. Второе условие (все остальные) порождает ошибку измерения, состоящую в том, что записанная или полученная оценка не равняется настоящей оценке респондентом данного факта.

*Состояние неответа* – источник ошибки, которая вызывается не характером выборки, а возникает из-за того, что респондент соглашается участвовать в интервью, но отказывается или не способен ответить на отдельные вопросы.

*Например, в одном анекдоте по данному поводу рассказывается о двух священниках, доминиканце и иезуите, которые спорили, является ли грехом курение и молитва в одно и то же время. Не найдя окончательного ответа, они отправились проконсультироваться у своих наставников. На следующей неделе они снова встретились. Доминиканец спросил: «Ну и что сказал твой наставник?» Иезуит ответил: «Он сказал, что в данном случае греха нет». «Это забавно, поскольку мой куратор сказал, что это грех», – ответил доминиканец. Иезуит: «Как ты его спросил?» Ответ: «Я спросил, можно ли курить во время молитвы». Иезуит воскликнул: «О, а я спрашивал моего учителя, правильно ли молиться, когда куришь».*

### **Избегать сложных понятий и слов**

Так как большинство исследователей более образованны, чем типичные респонденты, поэтому они используют слова, которые привычны для них самих, но которые непонятны для многих респондентов. Трудность заключается в том, что нелегко забыть то, что знаешь, и поставить себя на место другого человека, когда пытаешься определить подходящий набор слов.

Поэтому необходимо, чтобы вопросы были составлены в более простом и понятном для респондентов виде.

### **Избегать сомнительных слов и вопросов**

Слова и вопросы должны не только быть простыми, но и не должны иметь двусмысленный характер.

*Например, возьмем многовариантный вопрос: «Как часто вы записываете на свой видеомаягнитофон телепрограммы для дальнейшего просмотра?»*

НИКОГДА

ИЗРЕДКА

ИНОГДА

ЧАСТО

Для любых практических целей ответы на это опрос будут бесполезны, т. к. слова «изредка» и «часто» являются неясными. Например, для одного респондента слово «часто» может означать «почти каждый день. Для другого это может означать «Да, я делаю это, когда у меня есть в этом потребность. Это случается один раз в неделю». Слова «изредка» и «иногда» также могут по-разному быть интерпретированы разными респондентами. Поэтому, хотя на вопрос и могут быть даны ответы, они будут давать мало представления о том, как обстоит дело с частотой использования видеомонофона для записи телепрограмм. Гораздо лучшей стратегией будет конкретизировать варианты выбора для респондента, например, так:

*Никогда не записываю.*

*Записываю примерно раз в месяц.*

*Записываю примерно раз в неделю.*

*Записываю почти каждый день.*

Другой вариант избежать двусмысленности состоит в том, чтобы спросить о том, когда записывали в последний раз. Например, «Вы записывали на видео какую-нибудь телепрограмму за последние два дня?»

*Да*

*Нет*

*Не могу вспомнить*

Процент ответивших положительно даст представление о частоте, с которой используется видеомонофон, а последующий вопрос к этой категории лиц – «С какой целью вы осуществляли запись?» – позволит определить, для записи чего конкретно его использовали. Среди ответивших будут люди, которые обычно используют монофон для записи программ, но не делали этого в течение последних двух дней. Будут и те, кто такого обычно не делает, но по каким-то причинам записал что-то за этот срок. Эти отклонения нейтрализуют одно другое при достаточно большом объеме выборки.

#### **Избегать наводящих вопросов**

*Наводящий вопрос* – вопрос, сформулированный таким образом, что, будучи заданным, указывает, как респондент должен отвечать.

#### **Избегать скрытых альтернатив**

*Скрытая альтернатива* – альтернативный ответ, который не содержится в составных частях вопроса.

*Например, при одном исследовании нужно было узнать мнения неработающих жен в отношении работы вне дома. Им задали похожие вопросы: «Хотели бы работать, если бы это было возмож-*

но?» и «Вы предпочитаете работать или предпочитаете делать только домашнюю работу?» Эти вопросы принесли разительно отличающиеся ответы. В первой версии 19 % домохозяек сказали, что они не хотели бы работать. При втором варианте 68 % заявили, что не хотели бы иметь какую-то работу, – в три с половиной раза больше, чем в первом случае. Разница в двух вопросах состоит в том, что вторая версия содержит явную альтернативу, которая только подразумевается в первом случае.

### **Скрытое допущение**

Проблема, которая возникает, когда вопрос не сформулирован так, чтобы четко были определены его последствия, и, таким образом, он порождает различные ответы от разных людей, которые допускают различные последствия.

Например, на два вопроса (вариант А «Нужно ли принять закон, требующий, чтобы люди пристегивались ремнями безопасности в автомобиле?» и вариант Б «Нужно ли принять закон, требующий, чтобы люди пристегивались ремнями безопасности в автомобиле или платили штраф?», т. е. когда последствия четко указаны в самом вопросе, то вариант Б делает скрытое последствие в версии А отчетливым; единственный способ добиться эффективности закона об обязательном использовании ремней безопасности состоял бы в том, чтобы ввести какие-то штрафы за его нарушение. В случае, когда нет никакого четкого заявления о том, что произойдет, если человек не будет соблюдать закон, 73 % высказались за принятие закона. Когда люди столкнулись с перспективой уплаты штрафа за несоблюдение закона, только 50 % высказались в пользу законодательного принуждения к использованию ремней безопасности.

### **Избегать обобщений и оценок**

Вопросы всегда должны задаваться в конкретных терминах, а не в общих.

Например, рассмотрим вопрос: «С каким количеством продавцов вы встречались за последний год?», который может быть задан специалисту по закупкам. Для ответа на данный вопрос агент должен приблизительно оценить, сколько продавцам он звонит в среднем за неделю, и умножить на 52. Задача не должна решаться самим респондентом. Более правильная оценка может быть получена, если спросить агента: «Сколько торговых представителям вы звонили за последнюю неделю?» и затем самим умножить полученный ответ на 52.

### **Избегать двухканальных вопросов**

Вопрос, который требует двух ответов и из-за этого создает затруднение для респондента, называется **двухканальным вопросом**. Например, вопрос «Какова ваша оценка цен и удобства расположения, предоставляемых демонстрационным залом фирмы XYZ?» требует от респондента реагировать на два разных параметра, которыми быть охарактеризован демонстрационный зал. Респондент может считать, что цены достаточно привлекательны, но местоположение неудобное, и, следовательно, сталкивается с дилеммой, как ему отвечать. Можно и нужно устранять такие двухканальные вопросы разделением исходного вопроса на два отдельных подвопроса. Полезным индикатором того, что должны быть использованы два вопроса, является наличие союза «и» в начальной формулировке вопроса.

## **3.2. Этапы определения последовательности вопросов и завершение разработки анкеты**

### **6 этап. Определение последовательности вопросов**

Когда определены форма ответа и отдельная формулировка для каждого вопроса, начинают объединение их в единую анкету. Строго определенного порядка не существует, но предлагается придерживаться следующих общих принципов:

– **использовать простые и интересные начальные вопросы.**

Например, если респондент не сможет легко ответить на первые вопросы или он находит их неинтересными или пугающими, он может отказаться заполнить остальную часть анкеты. Поэтому важно, чтобы несколько первых вопросов были простыми, интересными и не угрожали респонденту. Иногда полезно использовать такие вопросы для разогрева, даже если ответы на них не будут в дальнейшем подвергаться анализу, поскольку вопросы о мнении часто эффективны для расслабления респондентов и установления с ними контакта;

– **использовать туннельный подход.**

Один из подходов к построению последовательности вопросов называется «воронкообразным» подходом. А это подход к выстраиванию последовательности вопросов, который получил свое название от формы, начинающейся с общих вопросов и постепенно переходящей к вопросам с более узкой тематикой.

Например, если у респондента нужно спросить: «Какие изменения необходимы в сервисной политике компании?» и плюс к этому

*«Как вы оцениваете качество сервиса?», первый вопрос должен быть задан перед вторым. В противном случае качество обслуживания будет непропорционально выделено при ответе просто потому, что это свежее впечатление в памяти респондента.*

В вопросах должна присутствовать логика, т. е. резкие изменения в тематике и переход от одной темы к другой должны быть устранены. Иногда необходимы переходные средства для сглаживания процесса перехода, если появляются изменения в тематике, и для этого можно использовать фильтрующие вопросы или вставлять краткое объяснение как способ перехода от одного предмета к другому;

– **использовать разветвленные вопросы.**

**Разветвленный вопрос** – техника, используемая для того, чтобы отослать респондентов к различным местам в анкете, основываясь на их ответах по текущим вопросам.

*Например, начальный вопрос может формулироваться так: «Вы покупали автомобиль в течение последних 6-и месяцев?». Если респондент отвечает «да», то переходит к другому месту в анкете, где содержатся вопросы о конкретных деталях покупки. После отрицательного ответа на тот же вопрос рекомендуется пропускать вопросы, относящиеся к деталям покупки.*

Преимущество разветвленных вопросов заключается в том, что они требуют меньше альтернатив для индивидуальных вопросов, сохраняя при этом возможность получения необходимой информации от тех, кто может ее предоставить.

Разветвленные вопросы лучше использовать в телефонных и личных интервью, при почтовых опросах они должны быть сведены к минимуму:

– размещать сложные или деликатные вопросы в конечной части анкеты;

– задавать вопросы о классификационной информации в конце.

Типовая анкета содержит два типа информации: *основную информацию и классификационную информацию.*

**Основная информация** относится к предмету исследования (например, намерения или предпочтения респондента).

**Классификационная информация** относится к другим данным, которые собираются, чтобы классифицировать респондентов и извлечь дополнительную информацию. Демографические и социально-экономические характеристики респондентов часто используются как классификационные переменные.

### **7 этап. Определение физических характеристик анкеты**

Физические параметры анкеты могут повлиять на точность полученных ответов. При определении физического формата анкеты необходимо предусмотреть следующие моменты:

– *обеспечение привлекательности анкеты.* Физический вид анкеты может оказать влияние на склонность респондента к сотрудничеству. Это значит, что должна использоваться качественная бумага, анкета должна быть напечатана, а не размножена на ксероксе или представлена в виде фотоконии. При почтовых опросах должно быть представлено сопроводительное письмо, которое служит для представления исследования.

– *облегчение проведения и контроля.* Это подразумевает оптимальный размер анкеты, ее план и последовательность вопросов. Достоинства коротких анкет:

- проще заполнять;
- меньше времени на заполнение;
- меньше вероятность отказа со стороны респондента в вопросе участия;
- удобны при полевых исследованиях (небольшой вес при большом количестве);
- простота в сортировке, подсчитыванию и группировке полученной информации.

Переполненная анкета имеет нечеткое содержание, ведет к ошибкам при сборе информации, предоставляет более короткие и менее информативные ответы при самостоятельном заполнении, или при проведении устного опроса.

Нумерация вопросов облегчает кодирование, редактирование и табулирование ответов.

### **8 этап. Перепроверка этапов 1–7 и при необходимости их корректировка**

Анкета чаще всего не создается с первой попытки. Перепроверка и пересмотр анкеты выступают важной частью ее создания. Каждый из вопросов должен быть пересмотрен, для того чтобы убедиться, что этот вопрос легок для ответа и не вызовет затруднений, неясностей, потенциального недовольства респондента. Ни один из вопросов не должен быть направляющим или вызывающим какие-то определенные отклонения. Часто практикуется тестирование анкеты в небольшой группе участников опроса, по результатам которой выявленные недостатки и неточности устраняются.

## **9 этап. Проведение предварительного тестирования анкеты и при необходимости ее корректировка**

Предварительное тестирование выполняет ту же роль при разработке анкеты, что и рыночная проверка для вновь создаваемого продукта.

Предварительное тестирование помогает вскрыть отдельные вопросы, в которых могут быть улучшены формулировки или изменена последовательность вопросов. Если изменения существенны, переработанная анкета должна быть снова проверена.

## **Тема 4. ИЗМЕРЕНИЯ В МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЯХ**

### **4.1. Основы измерений**

*Большинство вещей, которые мы измеряем, предельно конкретны: килограммы на весах, чайные ложки, количество бензина в баке. Но как можно измерить отношение людей к жевательной резинке? Предпочтения подростков в отношении средства от прыщей? Социальное положение семьи? Исследователи рынка заинтересованы в измерении различных величин, о которых остальные люди редко думают в терминах количественных показателей.*

**Измерение** – правила для численной оценки в отношении объектов, которые обладают количественными характеристиками.

Существуют два момента в этом определении.

**Во-первых**, оно показывает, что измеряют характеристики объекта, а не сам объект. *(Мы не измеряем, например, самого человека, а выбираем для измерения его доход, принадлежность к социальной группе, образование, рост, вес, вкусы или что-то еще, что выступает как характеристика человека.)*

**Во-вторых**, определение широкое в том смысле, что оно не определяет как конкретно будут придаваться числовые значения.

*Например, мы часто присваиваем все свойства цифровой шкалы высчитанным величинам. Например, возьмем числа 1, 2, 3 и 4. Пусть цифра 1 обозначает один объект, цифра 2 – два объекта и т. д.*

*Эта шкала чисел обладает определенными свойствами. Например, можно сказать, что 2 больше чем 1, а 3 больше чем 2 и т. д. Также можно указать, что интервал между 1 и 2 такого же размера, как и между 3 и 4, который равен интервалу между 2 и 3. Дальше можно отметить, что 3 в три раза больше 1, а 4 в четыре раза больше 1 и в два раза больше 2.*

Поэтому необходимо определить свойства характеристик и присвоить им количественное выражение таким способом, чтобы они правильно отражали свойства этих характеристик.

Выделяют четыре типа шкал, при помощи которых характеристики могут быть измерены:

- номинальную;
- порядковую;
- интервальную;
- относительную.

Таблица 4.1

#### Измерительные шкалы

Шкала	Сравнительные признаки	Типовые примеры	Мера средних величин
Номинальная	Идентификация	Мужчина/женщина Пользуется /не пользуется Род занятий Количество служащих	Мода
Порядковая	Порядок	Предпочтение отдельных марок Социальная группа Твердость минералов Категория качества древесины	Медиана
Интервальная	Сравнение интервалов	Температурная шкала Шкала средних значений Отношение к торговой марке	Среднее значение
Относительная	Сравнение абсолютных величин	Количество проданных товаров Число покупателей Вероятность покупки Вес	Геометрическая средняя Гармоническая средняя

**Номинальная шкала** – измерение, при котором числа присваиваются объектам или классам объектов только с целью их **идентификации**.

Поэтому первое свойство шкалы чисел – это **идентификация**.

*Например, личный номер системы социального страхования – это **номинальная шкала**, наравне с номерами футболистов, ящиков*

и т. д. Эти числа просто выделяют индивидуальный предмет, присваивая ему конкретный номер. Соответственно, если при проведении исследования мужчины кодируются как 1, а женщины – как 2, то мы снова используем номинальную шкалу. Из этих цифр нельзя получить больше информации, чем информацию о поле человека. С таким же успехом можно провести кодировку в обратном порядке, присвоив мужчинам код 2, а женщинам – 1.

Второе свойство цифровой шкалы – порядок. Так, например, можно сказать, что число 2 больше числа 1, что число 3 больше числа 2, что 4 больше трех предыдущих чисел. Числа 1, 2, 3 и 4 упорядочены, и чем больше число, тем больше свойство.

Поэтому **порядковая шкала** включает в себя определенность, т. к. одно и то же число будет использоваться для всех одинаковых объектов. (Например, использование цифры 1 для обозначения первокурсника, цифры 2 – для второкурсника, 3 – для третьекурсника и 4 – для студента старшего курса. С таким же успехом можно было использовать числа 10, 20, 30 и 40. Эта нумерация будет просто означать уровень курса, на котором учится студент, и относительное положение двух человек с точки зрения сравнения того, насколько один из них ушел вперед в освоении учебной программы).

Третье свойство цифровой шкалы состоит в том, что интервалы между числами имеют определенное значение, и само число говорит о том, как далеки объекты (в отношении конкретной характеристики) друг от друга. Это означает, что разница этих объектов может быть измерена. (Например, разница между 1 и 2 равна разнице между 2 и 3, а разница между 2 и 4 в два раза больше разницы между 1 и 2.)

Классическим примером **интервальной шкалы** выступает температурная шкала, т. к. она показывает, что можно и что нельзя сказать, когда характеристика измеряется по интервальной шкале. Например, самая низкая температура за день составила 40, а самая высокая – 80 градусов по Фаренгейту. Можно ли утверждать, что самая высокая температура была в два раза горячее, чем низкая температура? Для того чтобы показать неправомочность утверждения о том, что 80 F в два раза теплее 40 F, надо просто преобразовать эти температуры в эквивалентные им по шкале Цельсия, где  $C = (5F - 160)/9$ . Теперь видно, что низшая температура составит 4,4 °C, а высокая – 26,6 °C, что представляет собой значительно большую разницу, чем в случае шкалы Фаренгейта.

*Этот пример показывает, что нельзя проводить сравнения абсолютных величин числовых показателей, когда проводится измерение на основе интервальной шкалы. Причина этого состоит в том, что на интервальной шкале точка отсчета выбирается произвольно. Например, одно и то же природное явление, замерзание воды, происходит при 0 °С, но одновременно и при 32 F. Следовательно, точка отсчета выбирается произвольно.*

При использовании интервальной шкалы среднее значение, медиана и мода могут выступать показателями среднего значения.

**Относительная шкала** отличается от интервальной шкалы тем, что предполагает наличие *естественного* или *абсолютного нуля*, в отношении которого существует единое мнение о месте его расположения. *Хорошие примеры – это рост и вес. Так как есть абсолютный нуль, правомерно использование сравнения абсолютных величин численных значений. Так, человек, весящий 100 кг, в два раза тяжелее человека с весом в 50 кг, а человек с весом 150 кг в три раза тяжелее последнего.*

В относительной шкале ноль имеет практическое значение, т. е. означает отсутствие измеряемого свойства. Поэтому при использовании относительной шкалы можно сравнивать интервалы, упорядочивать объекты в зависимости от их величины или использовать цифры для идентификации объектов (все, что позволяют делать интервальная, номинальная и порядковая шкалы). Геометрическая или чаще используемая арифметическая средняя, медиана и мода выступают как применимые показатели средних значений при измерении характеристик при помощи относительной шкалы.

## **4.2. Классификация и оценка ошибок**

Идеалом для всякой шкалы является создание такой системы баллов, которая отражала бы истинные различия в той характеристике, которую необходимо измерить, без вмешательства ненужных факторов. То, что на самом деле получаем, часто отличается от желаемого. Ошибка, которая может появиться при подсчете – **систематическая ошибка**, которая называется еще и **постоянной ошибкой** из-за того, что постоянно присутствует при измерениях. *Например, при измерении роста человека при помощи плохо размеченной линейки.*

Другой тип ошибки – **случайная ошибка** не является постоянной, а возникает из-за временных характеристик личности или ситуа-

ции, в которой проводится измерение и которая возникает при измерении нерегулярным образом. *Например, если, не поставив в известность исследователя, человек, которого измеряли, сменил туфли перед повторным измерением, то два результата могут не совпасть, даже если рост человека не изменился.*

Различие между систематической и случайной ошибкой является очень важным из-за способа оценки достоверности измеряемого свойства.

Существует два типа заключений, когда устанавливается надежность инструмента:

- прямая оценка достоверности;
- косвенная оценка, использующая надежность.

Существуют три методики **прямой оценки**, которые используются для выводов о достоверности:

- практическая достоверность;
- содержательная достоверность;
- конструкционная достоверность.

**Практическая достоверность** показывает, насколько хорошо показатель действительно предсказывает, что критерий является либо характеристикой, либо особенностью поведения человека. *Например, с большой долей уверенности можно утверждать, что студент, получающий в течении срока обучения высокие оценки, достаточно успешно сдаст государственный экзамен.*

**Содержательная достоверность** применима, если измерительный инструмент адекватно охватывает наиболее важные аспекты измеряемого понятия. *Например, оценивая уровень грамотности человека, с помощью некоторого текста нельзя с определенной долей уверенности утверждать, что человек написал его хорошо, только основываясь на полученные в прошлом знания. Здесь может сыграть роль фактор так называемой «врожденной грамотности» или то, что данный текст тесно связан с профессиональной либо другой сферой деятельности человека.*

**Конструкционная достоверность** является наиболее трудным типом достоверности с точки зрения ее определения. Не только сам инструмент должен быть внутренне последовательным, но измерять им нужно только то, для чего он предназначен. Так, инструмент измерения, предназначенный для измерения предпочтения, будет обладать конструкционной надежностью, если он в самом деле будет

измерять предпочтения в данном вопросе, а не какую-то другую базовую характеристику человека, которая влияет на его очки.

Поэтому **конструкционная достоверность** – оценка того, насколько хорошо инструмент охватывает теоретическое построение, концепцию или черту, которую предполагается измерить. *Например, вместо того чтобы говорить о том, что полученные 75 пунктов из 100 показывают лояльность человека к конкретной торговой марке, мы можем говорить о получении представления о лояльности к марке.*

**Косвенная оценка, использующая надежность**, основывается на понятии **надежность**. **Надежность** – одинаковость результатов, полученных независимыми, но сравнимыми измерениями одного и того же объекта, черты или понятия. *Например, если оценивать меткость стрельбы человека из винтовки, необходимо, чтобы на показатель надежности не могли повлиять временные факторы, а именно – старая винтовка не надежна, новая относительно надежна, но ее прицел может быть неправильно отрегулирован, и только из новой винтовки после корректировки прицела стрелок мог добиться регулярного попадания в центр мишени.*

### **4.3 Измерение ожиданий, восприятия и предпочтений**

Существует ряд способов, при помощи которых могут быть измерены ожидания, который включает в себя самоотчет, наблюдение обычного поведения, косвенные техники, метод решения реальных задач и метод психологических реакций. Наиболее простым подходом является **самоотчет**, при котором людей просят прямо высказать свои соображения или чувства в отношении объекта или класса объектов. *Например, информативность, привлекательность, надоедливость или раздражительность показатели, с помощью которых просят респондентов, чтобы они выбрали что-то одно, чтобы описать рекламу в каждом из средств массовой информации: газетах, журналах, радио, телевидении, рекламных щитов и прямых почтовых реклам.*

Наблюдательный подход к определению отношения основывается на предположении, что поведение объекта определяется его ожиданиями, и исследователи, следовательно, могут использовать наблюдаемое поведение для выводов об отношении человека к объекту. То поведение, которое исследователь хочет пронаблюдать, часто вызы-

вается созданием искусственной ситуации. Например, для оценки отношения человека к обязательному использованию ремней безопасности субъекта могут попросить подписать составленную в сильных выражениях петицию в защиту принятия закона о пользовании ремнями безопасности. Отношение индивида по интересующему вопросу было бы выведено, основываясь на том, согласился ли он подписать эту бумагу или нет.

**Косвенные техники** оценки ожиданий используют нестандартизированные или частично стандартизированные стимулы, которые обсуждались в теме «Сбор первичной информации», такие как словесно-ассоциативные техники, тесты на завершение предложений, составление рассказа по картинкам и т. д.

**Решение реальных задач** – метод оценки ожиданий, который опирается на предположение о том, что решение людьми отдельных задач будет отражать их личное отношение. *Например, для оценки отношения человека к использованию ремней безопасности можно попросить запомнить следующие факты: 1) количество жизней, которые были спасены благодаря тому, что они были пристегнуты; 2) количество людей, которые умерли при несчастных случаях, когда пассажиры не смогли отстегнуть ремни вовремя. Данные материалы должны отражать обе точки зрения на проблему. Затем исследователь должен определить, какие из фактов участник запомнил. Допущение состоит в том, что участники должны быть более склонны запомнить те аргументы, которые больше совпадают с собственной позицией человека.*

Другой подход к измерению ожиданий подразумевает использование **психологической реакции**. В этом случае при помощи специальных средств наблюдения исследователь отслеживает ответы респондента на контролируемое использование каких-либо стимулов. Проблема, которая возникает при использовании таких показателей для измерения отношения, состоит в том, что, за исключением частотного анализа голоса, индивидуальная психологическая реакция показывает только интенсивность индивидуальных ощущений, а не то, являются ли они негативными или позитивными.

Поскольку отношение является одной из наиболее распространенных концепций во всей социальной психологии, естественно, что исследователи разработали ряд методов для его измерения. Хотя многие из методов используют самоотчеты, каждый метод использует их различными способами.

**Шкала Ликерта**, также называемая шкалой **суммарных оценок** (рис. 4.1), является одной из наиболее широко используемых методик оценки ожиданий при помощи шкалы в маркетинговых исследованиях.

Характеристика	Безусловно не согласен	Не согласен	Нейтрален	Согласен	Безусловно согласен
Банк предоставляет хорошее обслуживание	—	—	—	—	—
Банк имеет удобное расположение	—	—	—	—	—
Банк имеет удобные часы работы	—	—	—	—	—
Банк предоставляет низкопроцентные кредиты	—	—	—	—	—

Рис. 4.1. Пример шкалы суммарных оценок

Она, в частности, полезна тем, что позволяет респондентам выразить интенсивность своих чувств. Рис. 4.1 представляет собой пример шкалы, которая может быть использована, например банком, заинтересованным в сравнении своего имиджа с имиджем конкурентов.

При создании шкалы Ликерта или шкалы суммарных оценок, исследователи разрабатывают ряд утверждений, которые относятся к проблеме или объекту исследования. Участников просят указать свою степень согласия или несогласия с каждым утверждением в серии.

Одной из наиболее популярных техник для измерения отношения в маркетинговых исследованиях является **семантическая дифференциальная шкала**.

Она признана особенно полезной при исследованиях имиджа корпорации, торговой марки или товара.

Эта шкала была разработана Чарльзом Осгудом и его коллегами.

Оригинальная семантическая дифференциальная шкала состояла из большого количества биполярных прилагательных, которые использовались для определения человеческих реакций в отношении интересующего объекта. Осгуд обнаружил, что большинство реакций может быть сгруппировано в три основных массива: 1) **массив оце-**

*нок*, представленный такими парами прилагательных, как плохой – хороший, сладкий – кислый, полезный – бесполезный; 2) *массив силы*, представленный такими парами прилагательных как могучий – беспомощный, сильный – слабый, глубокий – мелкий; 3) *массив активности*, представленный такими парами прилагательных, как быстрый – медленный, живой – мертвый, тихий – шумный.

Исследователи взяли общую идею Осгуда и адаптировали ее к своим собственным нуждам. Во-первых, вместо использования *основных* пар прилагательных к объектам своего интереса они разработали свои собственные. Эти пары не всегда были противоположны по значению и не всегда представляли собой только два слова. Кроме того, исследователи использовали отдельные фразы, чтобы обозначить концы шкалы, и некоторые из этих фраз содержали ожидания, приписываемые данному товару. *Например, один из концов шкалы мог быть обозначен как «стоит своих денег», а противоположный – «не стоит таких денег».* Во-вторых, вместо использования расчетов показателей оценки, силы и активности маркетологи были больше заинтересованы в создании профилей торговых марок, магазинов, компаний или чего-то еще, что нужно сравнить, и в общих оценках, при помощи которых объекты можно было бы сравнивать. В этом отношении использование семантического дифференциала в маркетинговых исследованиях последовало, скорее, за методом использования шкалы суммарных оценок в отношении конструирования самой шкалы.

Модификация семантической дифференциальной шкалы, которая получила отражение в литературе по маркетингу, это **шкала Стэпела**. Она отличается от семантической дифференциальной шкалы следующим: 1) прилагательные или описания как биполярные пары тестируются отдельно, а не одновременно; 2) пункты на шкале определяются числом; 3) выделяется 10 отдельных положений на шкале. Рис. 4.2 распределяет те же 4 характеристики, использованные ранее для измерения отношения применительно к разным банкам в формате шкалы Стэпела. Респондентов просили оценить, насколько точно каждое из ряда утверждений описывает интересующий объект, банк А.

	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5
Банк предоставляет хорошее обслуживание	<input type="checkbox"/>									
Банк имеет удобное расположение	<input type="checkbox"/>									
Банк имеет удобные часы работы	<input type="checkbox"/>									
Банк предоставляет низкопроцентные кредиты	<input type="checkbox"/>									

Рис. 4.2. Пример шкалы Стэпела

Респондентам даются примерно следующие инструкции: *Вы должны выбрать числа с плюсом для слов, которыми, на ваш взгляд, точно описывается банк А. Чем более точно слова описывают этот банк, тем большее число с плюсом вы должны выбрать. Вы должны выбрать число с минусом для слов, которые на ваш взгляд не характеризуют его правильно, Чем менее точно слова описывают его, тем большее число с минусом вы должны отбирать. Следовательно, вы можете выбрать любое число от +5, для слов, которые представляются вам точными, вплоть до -5 для слов, которые, по вашему мнению, совершенно неправильно отражают положение?*

Шкалы различаются по тонкости отличий, которые они позволяют измерить, и по процедурам, используемым при приписывании объектов к определенному месту. Три из наиболее часто используемых оценочных шкал – это **графическая шкала, пунктирная** и **сравнительная**.

При использовании **графической рейтинговой шкалы** люди показывают свои оценки при помощи отметки подходящего пункта на прямой, которая соединяет две крайние оценки какого-то параметра. Здесь возможны многочисленные варианты. Линия может быть вертикальной или горизонтальной, она может быть размечена или не размечена; если она размечена, то делений может быть мало или много, как в случае с термометрической шкалой, которая называется так из-за того, что похожа на шкалу обычного термометра. Рис. 4.3 представляет пример горизонтальной, имеющей только один фиксированный конец графической рейтинговой шкалы.

ПАРАМЕТР	ВАЖНО	НЕ ВАЖНО
Банк		предоставляет
<hr/>		
хорошее обслуживание		
Банк	имеет	хорошее
<hr/>		
расположение		
Банк	имеет	удобные
<hr/>		
часы работы		
Банк		предоставляет
<hr/>		
низкопроцентные кредиты		

Рис. 4.3. Пример графической рейтинговой шкалы

Каждого человека инструктируют, чтобы он указал важность параметра отметкой подходящей позиции на шкале. Показатель важности впоследствии может быть выведен измерением длины линии от левого края шкалы до отмеченного места.

Одно из больших преимуществ метода графической оценки состоит в легкости, с которой шкалы могут быть построены и использованы. Они предоставляют возможность установить тонкие различия и ограничиваются в этом отношении только возможностями дифференциации со стороны оценивающего. Для более эффективного использования исследователю рекомендуется избегать установления слишком крайних окончаний шкалы, поскольку крайние оценки подталкивают респондентов в центр шкалы, что способствует уменьшению степени полезности информации.

**Пунктирная оценочная шкала** в основном похожа на графическую шкалу, за тем лишь исключением, что респондент должен выбирать из конечного числа категорий вместо размещения оценки на нарисованной линии шкалы. В целом от 5 до 9 категорий сработают наилучшим образом, давая возможность провести достаточно тонкие различия и при этом, как представляется, оставаясь понятными для респондентов. Может использоваться и большее их количество.

Существует целый ряд возможных вариаций на тему пунктирной шкалы. Рис. 4.4 изображает две разные формы пунктирной рейтинговой шкалы, которые использовались для оценки степени удовлетворения потребителя.

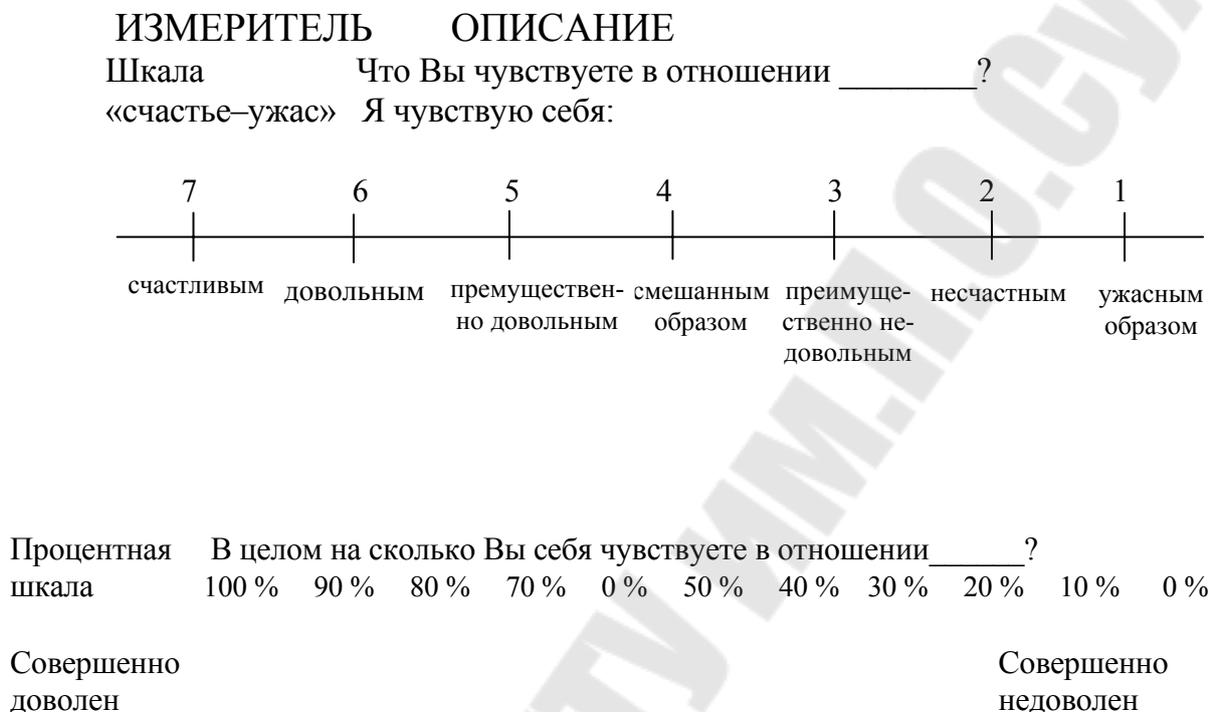


Рис. 4.4. Две различные формы пунктирной оценочной шкалы, используемой для измерения удовлетворения

Следует отметить, как категории расположены с точки зрения их положения на шкале и что в некоторых случаях они снабжаются словесным описанием, а в других нет. Описания категорий не являются абсолютно необходимыми в случае пунктирной шкалы, хотя их присутствие и характер все же влияют на ответы. Когда они используются, важно добиться, чтобы описания означали именно то, что они описывают.

В случае графической или пунктирной шкалы респондентов не просят сравнивать два параметра друг с другом или с некоторым стандартом, который предоставляет сам исследователь. *Например, респондентов могут попросить указать, насколько важно для них удобное месторасположение при выборе банка, но при этом не просят оценить, насколько более или менее важно для них удобное место по сравнению с удобными часами работы.* Однако при использовании **сравнительной рейтинговой шкалы** респондентов просят

оценить каждый параметр в связи с его отношением к другим оцениваемым характеристикам. Пример сравнительной рейтинговой шкалы представлен на рис. 4.5.

	Не важно	Относительно важно	Достаточно важно	Относительно важно
Банк предоставляет хорошее обслуживание	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Банк имеет удобное расположение	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Банк имеет удобные часы работы	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Банк предоставляет низкопроцентные кредиты	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Рис. 4.5. Сравнительная рейтинговая шкала, применяемая для оценки важности параметра

## Тема 5. АНАЛИЗ ДАННЫХ

### 5.1. Редактирование и кодирование данных

Следующим этапом после сбора данных идет определение значимости информации по сырым данным. Поиск значимости принимает много форм. Однако предварительные аналитические этапы редактирования, кодирования и табулирования данных являются общими для большинства исследований.

Основная цель редактирования состоит в установлении некоторых стандартов минимального качества сырых данных. Редактирование включает в себя просмотр и, если необходимо, исправление каждой анкеты или формы регистрации наблюдений. Просмотр и внесение исправлений часто выполняются в две стадии: *полевое редактирование* и *централизованное офисное редактирование*.

**Полевое редактирование** – это предварительное редактирование, которое строится таким образом, чтобы обнаружить наиболее бросающиеся в глаза пропуски и неточности данных.

Оно также полезно для контроля поведения персонала полевых сил и внесения ясности в любого рода недопонимание этими силами направлений их деятельности, методов, специфических вопросов и т. п. *Например, в одном опросе компании Roper, проводившемся на Украине, полевое редактирование позволило обнаружить, что один работник оставлял анкеты респондентам вместо того, чтобы брать у них интервью, как было определено инструкцией. Наводящая подсказка приводила к возникновению разных путей хождения ответов по кругу.*

В идеальных обстоятельствах полевое редактирование выполняется как можно скорее после того, как анкета или другая форма сбора данных заполнена. В этом случае проблемы могут быть устранены прежде, чем проводивший сбор информации штат будет расформирован, и пока конкретные контакты, явившиеся источником затруднения, еще свежи в памяти проводивших опрос или наблюдение людей. Полевое редактирование обычно выполняется руководителем полевых работ.

**Централизованное офисное редактирование** – точная всеобъемлющая проверка и коррекция заполненных форм сбора данных, включая принятие решения о том, что с этими данными делать.

Чтобы обеспечить логическую последовательность обработки материалов, лучше всего предоставить все носители собранных данных одному человеку. Если эту работу приходится делить по соображениям ее объема и имеющегося времени, подразделы должны определяться по частям анкеты, а не по респондентам. То есть один редактор должен редактировать часть «А» всех анкет, а другой – часть «В».

В отличие от полевого, *централизованное офисное редактирование* в меньшей степени зависит от последующих процедур, и в большей – от степени полноты данных. При анализе необходимо решить, каким образом будут обрабатываться носители собранных данных, содержащие неполные ответы, явно неправильные ответы и ответы, которые отражают утрату интереса.

Исследования, в которых все вернувшиеся анкеты заполнены целиком, довольно редки. В некоторых окажутся пропущенными целые разделы. Другие будут отражать оставленными без ответа отдельные позиции. Решение о том, каким образом обрабатывать эти неполные анкеты, зависит от серьезности пропусков. Анкеты, в которых пропущены целые разделы, не должны отбрасываться автоматически. *Может, например, быть, что пропущенный раздел относится к влиянию супруга*

на приобретение какой-то серьезной вещи длительного пользования, тогда как респондент в браке не состоит. Несмотря на незаполненный раздел, такой тип полученного ответа определенно полезен. Если нет никаких позитивных соображений относительно большого количества вопросов, оставшихся без ответов, весь подобный материал будет, отброшен, а это увеличивает долю оставшихся без ответов анкет по исследованию в целом. Анкеты, содержащие только изолированные безответные позиции, должны оставаться, а после кодирования они могут быть подвергнуты определенной очистке данных, о которой будем говорить позже.

Тщательное редактирование анкеты иногда показывает, что ответ на какой-то вопрос очевидно неправилен. Например, респондентам сначала может задаваться вопрос о типе магазина, в котором они приобретали фотоаппараты, а затем их просят назвать магазин. Если какое-то лицо на первый вопрос отвечает «в универмаге», а затем дает название магазина, торгующего по сниженным ценам, один из ответов неверен. Возможно, редактор сумеет определить, какой из двух ответов правильный, воспользовавшись другой информацией из той же анкеты.

В процессе анализа может потребоваться какая-то политика в подходе к обработке вопросов с точки зрения их правильности, когда возникают подобные несоответствия или неточности иного типа. Принятая политика должна отражать цель исследования. Например, директору Восточноевропейского маркетинга (Pepsi Cola International) были сданы результаты проводившегося в Венгрии опроса, в которых говорилось, что аптеки американского образца являются точками сбыта безалкогольных напитков. Директор не принял эту информацию, т. к. прекрасно знал, что в Венгрии нет американских аптек, и что эта информация была привнесена из структуры анкеты, разработанной на Западе.

При анализе необходимо стремиться не пропустить заполненные анкеты, которые неудачны с точки зрения пробуждения интереса респондента. Свидетельства отсутствия интереса могут быть и очевидными, и очень трудно распознаваемыми. Например, анкету субъекта, который проставил оценку «5» своего отношения к чему-то по пятибалльной шкале каждой из 40 позиций анкеты, даже не обратив внимания, что некоторые позиции выражали негативное, а некоторые позитивное отношение. Ясно, что этот субъект не принял исследование всерьез, и редактор, вероятно, должен отбросить такой ответ. Проницательный редактор может ухитриться разглядеть и

более трудно различимые указания на отсутствие интереса, такие как отметки, выходящие за пределы предусмотренных для них рамок, пометки, словоизлияния на анкете и т.п. Можно не отбрасывать такие ответы, но кодировать их необходимо таким образом, чтобы в последующем можно было провести отдельные табуляции для сомнительных носителей собранных данных и явно добротных анкет. Затем эти две группы можно сравнить, чтобы посмотреть, насколько значимо отсутствие интереса сказалось на результатах.

**Кодирование** – технический прием, с помощью которого данные распределяются по категориям. Прием связан со спецификацией альтернативных категорий или классов, в которые должны помещаться ответы, а самим классам должны назначаться кодовые номера. То есть посредством кодирования сырые данные превращаются в символы – обычно цифровые, которые можно табулировать и подсчитывать.

Первый этап кодирования заключается в специфицировании категорий или классов, к которым будут относиться ответы. Не существует какого-то идеального числа категорий. Скорее, это число будет зависеть от исследуемой проблемы и специальных позиций анкеты, используемых для генерирования информации. Выбор ответов должен быть взаимоисключающим и исчерпывающим, чтобы каждый ответ логически попадал в одну и только одну категорию. По ряду вопросов могут использоваться и множественные ответы. *Например, на вопрос: «Для каких целей вам служит йогурт?» ответы могут быть: «в качестве десерта», «как легкая вечерняя закуска», «как возможность перекусить после полудня» и т. п.* С другой стороны, если вопрос сосредоточен на возрасте лица, то приемлема, конечно, только одна категория возраста, и код должен четко указывать, о какой категории идет речь.

Кодирование закрытых вопросов и большинства средств балльной оценки не сложно, потому что оно устанавливается при конструировании самой анкеты.

Кодирование открытых вопросов более сложно и более дорогое, чем кодирование закрытых вопросов, т. к. приходится определять подходящие категории на базе ответов, которые не всегда предсказуемы. *Например, международные исследования могут создавать особые проблемы кодирования, поскольку разные понятия могут означать разные вещи. Например, консерватором на территории бывшего Советского Союза считается тот, кто желает оставить-*

*ся верным «старому коммунизму» или вернуться в него, тогда как в странах Запада люди тех же взглядов видятся принадлежащими к самому левому крылу. В свою очередь, либеральными русскими являются те, кто желает введения в экономику и политику рыночных перспектив, которые, как правило, поддерживаются на Западе консерваторами.*

Если анкет слишком много и необходимо использовать для кодирования анкет нескольких кодировщиков, дополнительной проблемой может стать возникновение несоответствия в самом кодировании. Поэтому для получения логической последовательности обработки данных эту работу необходимо разделять по задачам, а не в равных долях делить анкеты между кодировщиками.

Второй этап кодирования касается назначения кодовых номеров классов. Например, мужской пол может обозначаться буквой М, а женский – буквой Ж. Как альтернативный вариант, эти классы могут обозначаться: 1 – мужчина и 2 – женщина. Для обозначения классов принято использовать цифры, а не буквы. На этой стадии также лучше использовать цифры в том виде, как они зафиксировались в форме сбора данных, а не раскладывать их на более мелкие категории. *Например, если имеются данные о фактическом возрасте людей, не рекомендуется кодировать возраст как 1 = до 20 лет, 2 = 20–29 лет, 3 = 30–39 лет и т. д. Это привело бы к ненужной потере информации в ее исходном измерении, а если возникнет необходимость градации, это можно будет сделать с той же легкостью на более поздней стадии анализа.*

Когда для анализа данных предполагается использовать компьютер, кодирование необходимо выполнять таким образом, чтобы данные оказывались готовыми для ввода в машину, поэтому полезно обеспечить наглядность ввода посредством многоколонной записи. Кроме того, рекомендуется следовать установившимся традициям кодирования данных, т. е. располагать только один символ в каждой колонке. Когда вопрос допускает множество ответов, допускать отдельные колонки для кодирования каждого варианта ответа. *Так, в примере с йогуртом необходимо предусмотреть отдельные колонки для тех, кто употребляет этот продукт в качестве десерта, для тех, кому он служит вечерней легкой закуской, и т. д.*

*Использовать только числовые коды, а не буквы алфавита или специальные символы вроде @ или пробел. Для большинства компью-*

*терных программ при обработке статистических данных манипулирование чем-то иным, чем цифры, сопряжено с трудностями.*

Необходимо использовать ровно столько колонок поля, назначаемого для переменной, сколько необходимо для полного охвата всех ее возможных значений. Так, если переменная такова, что десяти кодов от 0 до 9 для охвата категории недостаточно, необходимо использовать две колонки, обеспечивающие 100 кодов от 00 до 99. Кроме того, любому полю должна назначаться не более, чем одна переменная.

Рекомендуется применять стандартные коды для «отсутствия информации». Так, все ответы «не знаю» должны кодироваться цифрой 8, «нет ответов» – цифрой 9, а «не применялось» обозначаться как 0. Лучше, если во всем исследовании для каждого из этих типов «нет информации» используется один и тот же код.

Желательно кодировать в каждой записи идентификационный номер респондента. Как правило, необходимости идентифицировать в этом номере имя респондента нет. Этот код просто связывает анкету с кодируемыми данными. Такая информация часто полезна на этапе очистки данных. Если анкета координируется не с одной записью, то в каждой записи кодируются идентификационный номер респондента и порядковый номер. Колонка 10 первой записи может указывать, как респондент ответил на вопрос 2, а в колонке 10 второй записи могут содержаться данные о том, мужского или женского пола эта персона.

Завершающий этап процесса кодирования состоит в подготовке **книги кодов**, которая содержит общие инструкции, указывающие, каким образом была закодирована каждая позиция данных. В ней перечисляются коды каждой переменной и категории, включенные в каждый код. Далее в ней указывается, где в компьютерной записи располагается переменная, и каким образом эта переменная читается – например, с десятичной точкой или как целое число. Последняя информация обеспечивается установлением формата.

## **5.2. Табулирование данных**

Табулирование заключается просто в подсчете количества событий, которые попадают в различные категории. Табулирование может принимать форму простой табуляции, или перекрестной табуляции.

**Простая табуляция** – подсчет количества событий, которые попадают в каждую категорию, когда категории базируются на одной переменной.

**Перекрестная табуляция** – подсчет количества событий, которые попадают в каждую из нескольких категорий, когда категории базируются на двух и более переменных, рассматриваемых одновременно.

Простая табуляция связана с подсчетом для единственной переменной. Она может повторяться для каждой из переменных исследования, но табуляция для каждой переменной не зависит от табуляции для других переменных. В перекрестной табуляции две или более переменных обрабатываются одновременно. *Например, количество людей, которые покупают суп «Галина Бланка» в супермаркете, является примером перекрестной табуляции, поскольку речь идет об измерении двух связанных характеристик.*

Табуляция может выполняться целиком от руки, целиком машинной или частично машинной и частично от руки. Какой из подходов более эффективен зависит и от числа необходимых табуляций, и от количества событий в каждой табуляции. Число табуляций является прямой функцией количества переменных, тогда как количество событий – это прямая функция размера выборки. Чем меньше число табуляций требуется и чем меньше выборка, применяются ручные методы. Однако применение любого подхода также в значительной степени зависит от сложности табуляций. Сложность возрастает по мере увеличения числа переменных, получаемых для одновременной обработки в перекрестной табуляции. Сложность также возрастает и с увеличением числа категорий на одну переменную.

В очень простых исследованиях ручная табуляция оказывается более полезной, особенно если вопросов немного и число возможных ответов ограничено, большинство же исследований полагается на компьютерную табуляцию, использующую пакеты программ.

**Одномерная табуляция** используется в следующих целях:

- 1) для определения степени безответности позиций анкеты;
- 2) для локализации грубых ошибок;
- 3) для локализации посторонних значений;
- 4) для определения эмпирического распределения рассматриваемой переменной;
- 5) для расчета итоговых статистик.

Первые три направления определяются как очистка данных. **Безответность позиций** является важной проблемой в большинстве исследований. Степень безответности позиций часто служит полезным индикатором качества исследования. Когда степень безответно-

сти большая, исследование в целом становится сомнительным и возникает необходимость пересмотреть его цели и методы. Даже когда безответность остается в определенных границах, необходимо принять решение относительно того, что делать с утраченными позициями, еще до анализа данных. Возможно использование нескольких стратегий.

1. Оставить позиции пустыми и описать их количество как отдельную категорию. *Хотя такой метод работает в случае простых одномерных и перекрестных табуляций, он неприемлем для ряда других статистических приемов.*

2. Исключать событие с утраченной позицией при анализе с использованием соответствующей переменной. При использовании такого подхода необходимо постоянно давать количество событий, на которых базируется анализ, т. к. размер выборки оказывается различным на протяжении анализа. Игнорируется также и тот факт, что степень безответности по какой-то конкретной позиции возможно показывает, насколько глубоко заботил респондентов адресовавшийся им вопрос.

3. Подставить значения утраченных позиций анкеты. Обычно такая подстановка имеет в своей основе меру срединной тенденции, медианы или моды. В порядке альтернативного варианта иногда можно оценить ответ с использованием другой информации, содержащейся в анкете. Однако этот подход требует большего объема работы и несет в себе потенциальную опасность смещения результатов.

Не существует «правильного» или простого ответа на вопрос о том, каким образом обрабатывать утраченные позиции. Все зависит от целей исследования, обстоятельств утраты информации и методов, которые используются для анализа данных.

Еще одна цель одномерной табуляции состоит в локализации *грубых ошибок*.

**Грубая ошибка** – ошибка, которая возникает при редактировании, кодировании, клавиатурном наборе или табулировании данных.

*Например, при одномерной табуляции данных по владению легковыми автомобилями на одну семью проверка исходной анкеты показала, что семья, сообщившая о нахождении у нее в собственности 9 машин, на самом деле имеет только одну. Цифра 9 является **грубой ошибкой**.*

Количество событий, используемых в качестве базы необходимо преобразовывать в проценты, причем обязательным является указа-

ние общего числа событий, на котором базируется процентное представление.

Следующее применение одномерной табуляции состоит в локализации посторонних значений.

**Постороннее значение** – наблюдение, настолько отличающееся по величине от остальных наблюдений, что возникает необходимость обрабатывать его как особое значение.

Это может означать исключение из анализа или определение особых факторов, которые ответственны за это наблюдение.

*Например, если семья действительно владеет 9 машинами, эта цифра должна рассматриваться как постороннее значение, т. к. слишком необычно, что у семьи так много легковых автомобилей.*

Четвертое направление использования одномерной табуляции состоит в определении **эмпирического распределения рассматриваемой характеристики**.

Часто распределение лучше всего представить в виде гистограммы.

**Гистограмма** – форма столбчатой диаграммы, на которой значения переменной размещаются по оси X, а частота или относительная частота появления значений указывается по оси Y.

Более глубокого понимания сути эмпирического распределения можно добиться, построив полигон частот.

**Полигон частот** – рисунок, получаемый из гистограммы посредством соединения верхних точек столбцов гистограммы прямыми линиями.

Одномерная табуляция применяется и для расчета **итоговых статистик** (таких, как мода, среднее значение, стандартное отклонение).

Мода, или наиболее часто происходящее событие, может быть определена непосредственно по одномерной табуляции.

Среднее значение рассчитывается посредством взвешивания каждого значения по частоте его появления, суммированием этих произведений и делением суммы на число событий.

Рассмотрим **перекрестную табуляцию**. Одномерная табуляция полезна для оценки переменных исследований по отдельности, перекрестная же табуляция является важным механизмом для изучения связей внутри и между переменными.

В перекрестной табуляции выборка делится на подгруппы таким образом, чтобы выяснить, как зависимые переменные изменяются от подгруппы к подгруппе.

Многие маркетинговые исследования не идут дальше перекрестной табуляции, и более того большинство исследований, использующих преимущества более сложных аналитических методов, тоже включают в себя перекрестную табуляцию в качестве важной составляющей.

Рассмотрим взаимосвязь между числом легковых автомобилей, которыми владеет семья, и семейным доходом. Чтобы сохранить выборку, предположим, что исследователя просто интересует определение того, верно ли, что семья с доходом выше среднего владеет двумя и более машинами, чем семья, доход которой ниже среднего. Предположим далее, что 37500 \$ – это медиана доходов всей совокупности, и что эту цифру правомерно использовать для разделения семей выборки на две группы – группу, имеющую доходы ниже среднего, и группу, в которой они выше среднего. Покажем в таблице двумерную классификацию семей выборки по доходу и числу легковых автомобилей (табл. 5.1).

Рассмотрим, зависит ли количество машин от семейного дохода? По табл. 5.1 это не вызывает сомнения. Однако в зависимости от того, что берется за базу при расчете процентов, результат может быть немного другой в плане процентного соотношения.

Таблица 5.1

**Количество легковых автомобилей по отношению к семейному доходу**

Доход, \$	Количество машин, шт.		
	1 или нет	2 или более	Всего
Менее 37500	48	6	54
Более 37500	27	19	46
<i>Всего</i>	75	25	100

Составим таблицу в процентном соотношении количества легковых автомобилей по отношению к семейному доходу.

Таблица 5.2

**Количество легковых автомобилей по отношению к семейному доходу**

Доход, \$	Количество машин			Число событий
	1 или нет	2 или более	Всего	
Менее 37500	89 %	11 %	100 %	54
Более 37500	59 %	41 %	100 %	46

Покажем влияние семейного дохода на владение количеством легковых автомобилей (табл. 5.3).

Таблица 5.3

**Семейный доход по отношению к количеству легковых автомобилей**

Доход	1 или нет	2 или более
Менее 37500	64 %	24 %
Более 37500	36 %	76 %
<i>Всего</i>	100 %	100 %
Число случаев	75	25

Из табл. 5.3 видно, что семейный доход оказывает влияние на владение более, чем одной машиной: 41 % семей с доходами выше среднего имеют два или более автомобиля и только 11 % семей с доходами ниже среднего могут позволить себе то же самое. С другой стороны, табл. 5.1 показывает, что 64 % семей, владеющих одной машиной, имеют доходы ниже среднего, а из владеющих двумя и более машинами только 24 % попадают в группу доходов ниже среднего. Этот пример показывает, что владение несколькими машинами не означает более высоких доходов. Это была иллюстрация правила процентных расчетов.

Проценты всегда рассчитываются в направлении причинного фактора или поперек результирующего фактора.

Поэтому в данном примере логично принимать доход за причинную или независимую переменную, а владение автомобилем зависимой переменной, т. е. правильный расчет в табл. 5.3.

Двумерная табуляция хотя и обеспечивает некоторое проникновение в суть зависимой взаимосвязи, но не дает окончательного ответа.

Например, то же самое можно сделать и с зависимостью количества машин и размером семьи. Поэтому правильный подход состоит в **одновременной обработке** дохода, размера семьи и владения не одним автомобилем.

Из табл. 5.5 видно, что 4 % составляют малые семьи с доходами ниже среднего, владеющие двумя или более автомобилями. Остальные 96 %, представляют семьи, во владении которых есть только одна машина или нет ни одной.

Отдельное влияние дохода на владение несколькими при постоянном размере семьи прочитывается по колонкам сверху вниз, а эффект размера семьи при постоянном доходе можно определить, считывая цифры по строкам (табл. 5.4).

Таблица 5.4

**Количество легковых автомобилей по отношению  
к доходу и размеру семьи**

Доход	4 человека или менее: число машин			5 человек или более: число машин			Всего: число машин		
	1 или нет	2 или более	всего	1 или нет	2 или более	всего	1 или нет	2 или более	всего
Менее 37500	44	2	46	4	4	8	48	6	54
Более 37500	26	6	32	1	13	14	27	19	46
<i>Всего</i>	70	8	78	5	17	22	75	25	100

Покажем процентное распределение количество легковых автомобилей по отношению к доходу и размеру семьи (табл. 4.5).

Таблица 5.5

**Количество легковых автомобилей по отношению  
к доходу и размеру семьи**

Доход	4 человека или менее: число машин			5 человек или более: число машин			Всего: число машин		
	1 или нет	2 или более	всего	1 или нет	2 или более	всего	1 или нет	2 или более	всего
Менее 37500	96 %	4 %	100 % (46)	50 %	50 %	100 % (8)	89 %	11 %	100 % (54)
Более 37500	81 %	19 %	100 % (32)	7 %	93 %	100 % (14)	59 %	41 %	100 % (46)

Исходная взаимосвязь, показанная в табл. 5.2, называется **полной (или нулевого порядка)** связью между переменными.

**Полная связь** – связь, существующая между переменными вне зависимости от уровней любых других переменных (еще называется связью нулевого порядка между переменными).

*Табл. 5.4, которая дает картину связи между двумя переменными (если брать только 2 или более машин) в пределах категорий размера семьи, называется **условной таблицей**, позволяющей обнаружить условную связь между переменными.*

**Условная связь** – связь, существующая между двумя переменными, когда в процессе анализа рассматриваются уровни одной или более других переменных (условных переменных).

*Размер семьи в табл. 5.4 является регулируемой переменной.*

Условные таблицы, построенные на основе одной регулируемой переменной, называются условными таблицами **первого порядка**.

Таблицы, составленные с использованием двух регулируемых переменных, называются условными таблицами второго порядка и т. д.

*Для определения того, доход или размер семьи имеют наибольшее влияние, необходимо рассмотреть их одновременно, используя метод разности долей.*

*Сначала рассмотрим влияние дохода и владение несколькими автомобилями для малых семей, а затем тоже для больших семей. Среди малых семей более высокий доход увеличивает вероятность владения несколькими автомобилями на 0,15 (0,19 – 0,04). Среди больших семей повышение вероятности владения несколькими автомобилями с ростом дохода составляет 0,43 (0,93–0,50). Эти разности являются мерами связи между доходом и владением несколькими автомобилями для семей разных размеров. Чтобы получить единственную оценку влияния дохода на владение несколькими автомобилями необходимо рассчитать среднее отдельных эффектов (средневзвешенное). Выборка объемом 100 семей состояла из 78 малых и 22 больших семей, поэтому весовой коэффициент для малых семей будет равен 0,78, а для больших – 0,22. Средневзвешенное рассчитывается как*

$$0,15 \times 0,78 + 0,43 \times 0,22 = 0,21.$$

*Это показывает, что переход из разряда низких доходов в разряд высоких в среднем повышает вероятность владения несколькими машинами на 0,21 (или 21 %).*

*Рассмотрим влияние размера семьи и владение несколькими автомобилями для семей с низкими доходами, а затем то же для семей с высокими доходами.*

*Среди семей с низкими доходами их большой размер повышает вероятность владения несколькими автомобилями на 0,46 (0,50 – 0,04) по сравнению с семьями малого размера. Среди семей с высокими доходами их большой размер повышает ту же вероятность на 0,74 (0,9 – 0,19) по сравнению с малочисленными семьями. Так как семей с низкими доходами в выборке было 54, а высокодоходных семей – 46, соответст-*

вующие веса для осреднения двух влияний равны 0,54 и 0,46 соответственно. Тогда средневзвешенное будет равно

$$0,46 \times 0,54 + 0,74 \times 0,46 = 0,59.$$

Таким образом, размер семьи обладает более существенным влиянием на владение не одним автомобилем, чем доход. Он повышает вероятность владения двумя и более автомобилями на 0,59, тогда как доход увеличивает ее только на 0,21.

Приведенный пример показывает важное приложение перекрестной табуляции – использование дополнительной переменной для очистки исходной перекрестной табуляции. В данном случае для очистки взаимосвязи между владением несколькими автомобилями и доходом был использован размер семьи. В действительности возможно появление ряда условий, при которых может возникнуть необходимость введения в перекрестную табуляцию дополнительных переменных. Исходная двумерная табуляция может указывать на существование или не существование взаимосвязи между переменными. Введение третьей переменной может либо оставить сделанное заключение неизменным, либо изменить его.

В настоящее время табулированные результаты чаще представляются в виде баннеров. **Баннер** – это последовательный ряд перекрестных табуляций между критерием и несколькими (часто большим числом) факторными переменными, оформленный в виде единой таблицы. Зависимая переменная или какое-то подлежащее объяснению явление определяет название строк таблицы. Заранее предсказанные или факторные переменные определяют колонки таблицы, а каждое значение этих переменных называется **баннерной точкой**. В табл. 5.6 показан **баннерный формат** для исследования владения легковыми автомобилями.

Достоинства баннерных таблиц:

– позволяют представить большой объем информации на небольшом носителе;

– формат представления удобен для восприятия.

Недостатки:

– имеют тенденцию к сокрытию взаимосвязей;

– затрудняют возможность зондирования возможных альтернативных объяснений полученных результатов;

– трудность в обнаружении ошибок в представляемых данных в результате неправильного кодирования или редактирования.

**Баннерный формат для данных о владении легковыми автомобилями**

Всего по выборке		Вопрос: Сколько легковых автомобилей имеет ваша семья?			
		Доход, \$		Размер семьи, чел.	
		Менее 37500	Более 37500	4 или менее	5 или более
Всего	100 (100)	54 (100)	46 (100)	78 (100)	22 (100)
1	75 (75)	48 (89)	27 (59)	70 (90)	5 (23)
2	20 (23)	6 (11)	17 (37)	8 (10)	15 (68)
3	5 (2)	– (0)	2 (4)	– (0)	2 (9)

*Примечание.* В скобках дано значение в процентах

**5.3. Методы анализа документов**

Выделяются два основных типа анализа:

- традиционный, классический;
- формализованный, количественный (контент-анализ)

**Традиционный анализ** – это цепочка умственных, логических построений, направленных на выявление сути анализируемого материала с определенной, интересующей исследователя точки зрения в каждом конкретном случае, т. к. интересующая информация, заложенная в документе, часто присутствует в неявном виде, в форме, отвечающей целям, с которыми документ был создан, но не всегда отвечающей целям маркетинговых исследований.

Поэтому проведение традиционного анализа означает преобразование первоначальной формы этой информации в необходимую исследователю форму.

Основным недостатком этого анализа является субъективность (т. к. проводится исследователем, а он человек).

В традиционном анализе различают внешний и внутренний анализ.

**Внешний анализ** – это анализ контекста документа и тех обстоятельств, которые сопутствовали его появлению.

Цель внешнего анализа – установить вид документа, его форму, время и место появления, автора и инициатора, какие цели преследовались при его составлении, степень надежности и достоверности, каков его контекст. *(Например, определенная политическая ситуация может способствовать предвзятому освещению событий в прессе,*

*внутренние трения между членами руководства могут отразиться на описании деятельности этой организации и т. п.)*

**Внутренний анализ** – это исследование содержания документа. То есть это работа, направленная на проведение внутреннего анализа документа и цифр, установление уровня компетенции автора документа, выяснение его личного отношения к описываемым в документе фактам (*искажения могут возникать не только в случае симпатии или антипатии автора, источником искажений могут выступать и методологические позиции автора*).

Отдельные виды документов из-за своей специфики, требуют специальных методов анализа и привлечения для их выполнения специалистов других областей знаний.

**Юридический анализ.** Он применяется для всех видов юридических документов. Его специфика заключается в том, что разработан особый словарь терминов, в котором значение каждого слова строго однозначно определено.

**Психологический анализ.** Он применяется при оценке отношения автора к какому-либо политическому, экономическому или социальному явлению. На основе таких исследований можно получить представление о формировании общественного мнения, общественных установок и т. д.

**Формализованный анализ.** Его называют часто количественный метод анализа документов (контент-анализ).

Суть этих методов сводится к тому, чтобы найти легко подсчитываемые признаки, черты, свойства документа (*например, такой признак, как частота употребления определенных терминов*), которые отражали бы определенные существенные стороны содержания. Тогда качественное содержание делается измеримым, становится доступным для точных вычислений.

Ограниченность формализованного анализа состоит в том, что не все содержание документа может быть измерено с помощью формализованных показателей.

**Контент-анализ** – это техника выделения заключения проводимого с помощью объективного и систематического выявления соответствующих характеристик текста задачам исследования.

Существуют общие принципы, когда применяется контент-анализ:

1. Когда требуется высокая степень точности или объективности анализа.

2. При наличии большого по объему и несистематизированного материала.

3. Когда категории, важные для целей исследования, характеризуются определенной частотой появления в изучаемых документах (*например, в работе с ответами на открытые вопросы анкет или глубинных интервью*).

Основными направлениями использования контент-анализа являются:

1. Выявление и оценка характеристик текста как индикаторов определенных сторон изучаемого объекта.

2. Выявление причин, породивших сообщение.

3. Оценка эффекта воздействия сообщения (*например, рекламного*).

Требование объективности анализа предполагает решение ряда проблем, связанных:

– с выработкой категорий анализа;

– с выделением единиц анализа;

– с выделением единиц счета.

**Категории анализа** – это понятие, в соответствии с которыми будут сортироваться единицы анализа (единицы содержания). (*Например, потребители с разным уровнем дохода, положительные отрицательные и нейтральные оценки товара определенной марки*).

Требования, предъявляемые к категориям:

– **должны быть исчерпывающими**, т. е. охватывать все части содержания, определяемые задачами исследования;

– **взаимоисключаемыми**, т. е. одни и те же части не должны принадлежать разным категориям;

– **надежность**, т. е. они должны быть сформулированы таким образом, чтобы у различных исследователей была высокая степень согласия по поводу отнесения какой-то части содержания к той или иной категории.

**Единицей анализа** (смысловой или качественной) является та часть содержания, которая выделяется как элемент, подводимый под ту или иную категорию. (Например, это может быть слово, устойчивое содержание слов, может не иметь явного терминологического выражения, а быть описательным, т.е. заголовок абзаца, раздела и т. п.)

Поэтому возникает задача выделения признаков (индикаторов), по которым определяется наличие в тексте темы или идеи, значимой с точки зрения целей исследования.

Индикаторами могут быть:

- относящиеся к теме слова и словосочетания;
- термины;
- имена людей;
- названия организаций;
- географические названия;
- пути решения экономических проблем,

*Например, при решении экономических проблем индикаторами могут быть отношение к событиям в терминах «за – против», «выгодно – невыгодно», «хорошо – плохо», в чьих это интересах, преследуемые интересы (политические, экономические, национальные, личные), социальные группы, общественные слои и т. п.*

Выбрав смысловую единицу и ее индикаторы, исследователь должен определить и единицу счета, которая является основанием для количественного анализа материала. Она может совпадать или не совпадать с единицей анализа или ее индикатора.

**Единицы счета** обладают разной степенью точности измерения, а также разным временем, уходящим на кодировку материала, попавшего в выборку.

Как правило, чем выше требование к точности, тем больше затраты времени на кодировку, с другой стороны, чем большая точность, тем более дифференцированы результаты и тем больше информации об объекте.

Поэтому, приступая к исследованию, необходимо решить:

- какая степень дифференциации нужна;
- как получить ее с минимальными затратами.

В практике методом контент-анализа были выделены общие единицы счета, отвечающие различным исследовательским требованиям.

**1. Время – пространство.** Пригоден для исследований сообщений, передаваемых СМИ. За единицы счета принимают: число строк, абзацев, см<sup>2</sup> площади, значков, колонок, посвященных какому-то вопросу, мнению, оценке. Для кино, радио и телевидения – это время, отведенное освещению определенного события или метраж пленки.

**2. Появление признаков в тексте,** т. е. наличие определенной характеристики содержания в любом признаке (*например, упоминание определенной марки продукта в каждой из единиц контекстов, т. е. частей, на которые разбит анализируемый текст. Так, если за единицу контекста принят абзац, то появление любого признака засчи-*

тывается 1 раз, независимо от количества повторений этих признаков в абзаце).

**3. Частота появления**, т. е. когда фиксируется каждое появление любого признака данной характеристики.

При разработке программы маркетинговых исследований необходимо четко определить, какого рода характеристики объекта подвергаются изучению, и в зависимости от этого оценивать документы с точки зрения их *адекватности, надежности, достоверности*.

**Адекватность** документа определяется как степень, в которой он отражает интересующие исследователя характеристики объекта, т. е. насколько он соответствует предмету исследования.

**Надежность** оценивается сопоставлением всех данных содержания с какими-то другими данными. Здесь возможны несколько вариантов проверки:

**1. Сравнение содержания документов**, исходящих из одного источника. Такое сравнение может проводиться:

а) *во времени*, т. е. когда сравниваются характеристики сообщений одного и того же источника в различные моменты времени;

б) *в различных ситуациях*, например, в условиях разной покупательной способности населения;

в) *в различных аудиториях*.

**2. Метод независимых источников**, т. е. выбираются характеристики из нескольких различных источников информации. Затем различия в характеристиках сравниваются с различиями в содержании сообщений.

Оценка **достоверности** данных документа проводится путем последовательного перебора источников встречающихся в документе ошибок.

Источники ошибок можно разделить на две категории:

– случайные (например, опечатки в статистических данных)

– систематические.

Систематические ошибки делятся на сознательные и несознательные

*Сознательные* ошибки часто определяются теми намерениями, которыми руководствуется автор при составлении документа (например, тенденциозное описание политической или экономической ситуации разными СМИ).

*Несознательные* ошибки представляют наиболее серьезную опасность. Чаще всего встречаются в статистических документах.

## Тема 6. РЕАЛЬНЫЕ ПРОЦЕДУРЫ И СИСТЕМАТИЧЕСКИЕ ОШИБКИ СБОРА ДАННЫХ

### 6.1. Влияние и значение систематических ошибок

Ошибки, возникающие при обследовании, можно разделить на два основных типа: *ошибки в выборке* и *систематические ошибки*.

**Ошибка в выборке** – разность между наблюдаемыми значениями количественного признака и их долгосрочным средним значением при повторении измерений.

Основой для ее рассмотрения была концепция выборочного распределения некоторой статистики, например, выборочного среднего, выборочной доли и т. п. Само понятие выборочного распределения неразрывно связано с понятием ошибки в выборке. Названное распределение существует благодаря тому, что различные выборки, сформированные в соответствии с принятым планом исследования, дают разные оценки параметра.

Статистика меняется от выборки к выборке в силу того, что в каждом случае выборочному отбору подвергается лишь часть генеральной совокупности. Соответственно можно определить ошибку в выборке как *разность между наблюдаемыми значениями количественного признака и их долгосрочным средним значением при повторении измерений*. Ошибка в выборке может быть уменьшена путем увеличения объема выборки.

Систематические ошибки являются отражением ошибок иного рода, которые могут возникать и не при выборочных исследованиях.

**Систематическая ошибка** – ошибка исследования, не связанная с выборкой. Она может быть вызвана концептуальными или логическими ошибками, неправильной интерпретацией ответов, а также статистическими, арифметическими, табуляционными, кодовыми или отчетными ошибками.

Систематическая ошибка подразделяется на *случайную* и *неслучайную*.

**Случайные ошибки** дают оценки, отличные от истинного значения; они могут приводить к отклонениям и в большую, и в меньшую сторону и имеют при этом случайный характер.

Неслучайные систематические ошибки имеют более тяжкие последствия. **Неслучайные систематические ошибки** приводят к односторонним отклонениям, Соответственно, для них характерна тен-

денция к смещению выборочного значения относительно параметра совокупности. Систематические ошибки могут являться следствием концептуальных или логических ошибок, неправильной интерпретации ответов, а также статистических, арифметических, табуляционных, кодовых или отчетных ошибок.

Один из авторов писал: «Перечень возможных бед и напастей с увеличением наших познаний только расширяется... Многолетняя работа в определенной области, позволяет приобрести известный методологический опыт, который, к сожалению, практически никогда не становится доступным другим. Подлинной уверенности в правильности выработанных подходов нет и быть не может».

Недостатки систематических ошибок:

– не так часты, но и не настолько подконтрольны, как ошибки в выборке. *При увеличении объема выборки ошибки в выборке уменьшаются. Сказать то же самое о систематических ошибках нельзя. В этом случае они могут как уменьшаться, так и возрасть;*

– в систематических ошибках как направление, так и величина ошибки могут оказаться совершенно непредсказуемыми, в отличие от выборок, где ошибки в выборке при использовании вероятностных методов могут быть оценены;

– приводят к смещению выборочного значения относительно параметра совокупности. *В ряде случаев нельзя судить даже о том, к чему именно приведут – к переоценке или к недооценке параметра;*

– влияют на достоверность выборочных оценок. *Вызванное ими смещение может увеличить ошибку оценки определенных статистик до такой степени, что оценка доверительного интервала окажется ошибочной.*

*Например, одно из исследований, отражающее финансовое положение потребителей, направленное оценить уровень накоплений потребителей, проведенное Иллинойским университетом, служит наглядным свидетельством сказанного. В ходе этого исследования сравнивали полученные от потребителей сведения об их финансовых средствах и долгах с известной информацией, отражающей финансовое положение потребителей.*

*Например, опытное обследование неожиданно показало, что систематические ошибки существуют не только в теории. Оказалось, что именно они являются основной причиной замеченной тенденции к занижению собранных показателей.*

*Эта ошибка не просто присутствовала в данных обследования, в ряде случаев роль систематических ошибок была настолько велика, что определение доверительных интервалов по известным формулам статистики теряло всяческий смысл. Следует отметить, что при увеличении объема выборки величина этой ошибки только возрастала.*

В некоторых ситуациях даже самые изоциренные выборки не могут избавить от систематических ошибок.

Особенно критичными ошибки становятся при работе с широкомасштабными, хорошо продуманными вероятностными выборками, т. к. при увеличении эффективности проектирования выборки и уменьшении выборочной дисперсии эффект систематических ошибок усиливается. В связи с тем что систематические отклонения практически не зависят от объема выборки, то сталкиваемся с парадоксальной ситуацией. Чем эффективнее составлена выборка, тем большую роль играют систематические ошибки и тем меньшим смыслом обладают вычисления по определению доверительного интервала, в основе которых лежат обычные формулы.

*В случае исследования, проведенного Иллинойским университетом, можно было определить систематическую ошибку, т. к. в распоряжении исследователей находились не только результаты опроса, но и реальные данные, отражающие финансовое положение потребителей. Предположим, что подобных данных нет. Исследователи могут предположить, что полученные ими ответы не совсем точны, но как они смогут определить хотя бы направление вызванного такими ошибками смещения? То ли респонденты сознательно завышали уровень своих сбережений, желая впечатлить интервьюера, то ли занижали их, боясь, что реальные цифры могут вызвать повышенный интерес у сотрудников Налогового управления. Предположим, что сам факт неточности приведенных сведений не вызывает у нас сомнений. Возникает еще один вопрос: какова величина этой «неточности»? Завышение реальной суммы на 10 000 долларов или ее занижение на 2000? Или наоборот?*

Часто проблема систематических ошибок оказывается центральной. Два типа систематических ошибок, отсутствие ответов одних и некорректные ответы других участников обследования, могут обратить результаты исследования на нет.

*Например, в результате специальных исследований, проведенных Бюро переписей, выяснилось, что такие систематические ошиб-*

ки могут в десять раз превышать ошибку выборки. Помимо прочего оказалось, что систематическая ошибка составляет большую часть ошибки исследования, в то время как случайная ошибка выборки сведена к минимуму.

Систематические ошибки могут быть уменьшены, но уменьшение их связано не столько с увеличением объема выборки, сколько с использованием специальных методов. А для этого необходимо, прежде всего, осознавать их причины.

На рис. 6.1 представлены основные типы систематических ошибок.

Систематические ошибки делятся на два основных типа: **ошибки, связанные с неполучением данных (ошибки ненаблюдения)**, и **ошибки наблюдения**.

**Ошибки ненаблюдения** возникают вследствие невозможности получения данных от части элементов обследуемой совокупности и быть вызваны тем, что часть обследуемой совокупности не была представлена в выборке, или же элементы, отобранные для включения в выборку, не представили данных.

**Ошибки наблюдений** возникают вследствие некорректной информации, полученной от элементов выборки, они могут возникнуть и на стадии обработки данных или формулирования итогового вывода.

По ряду характеристик они представляют еще более опасность, чем ошибки ненаблюдения. В случае последних ошибки этого вида обусловлены неполным охватом или неполучением данных. О существовании же ошибок наблюдения можно и не подозревать.

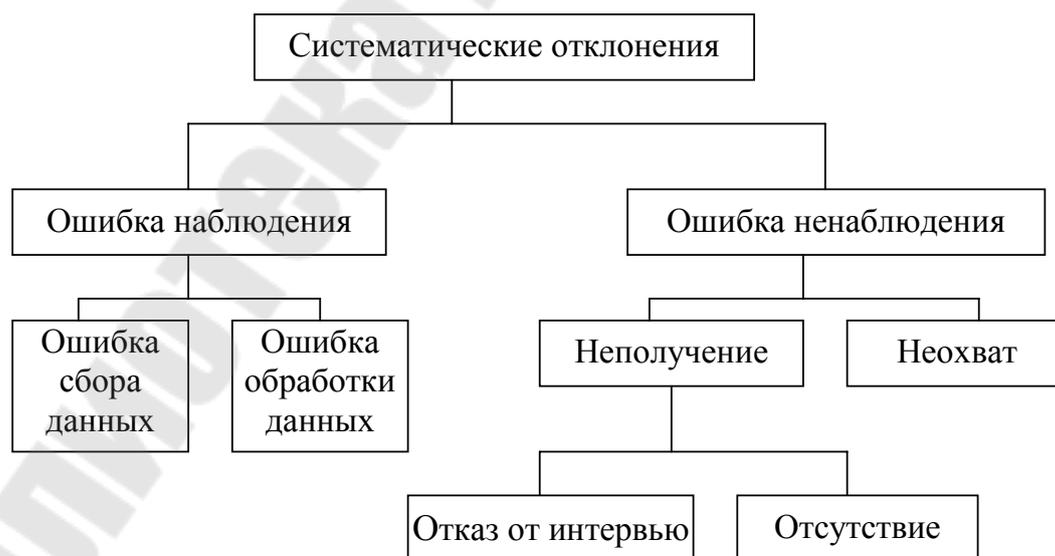


Рис. 6.1. Виды систематических ошибок

Само понятие ошибки наблюдения основывается на предположении о том, что для количественного признака или признаков существует некоторое «истинное» значение. Соответственно ошибка наблюдения является разностью объявленного и «истинного» значения. И естественно, что определение ошибки наблюдения создает трудности, т. к. при исследовании задаются той самой величиной, определение которой является целью исследования.

## 6.2. Понятие и сущность ошибок ненаблюдения

Рассмотрим ошибки ненаблюдения более подробно.

Существуют два типа ошибок ненаблюдения: *ошибки неохвата и ошибки неполучения данных*.

Любая из этих ошибок может привести к существенным неточностям, однако можно значительно уменьшить возможную ошибку.

**Ошибка неохвата** – систематическая ошибка, являющаяся следствием того, что определенные части или целые блоки генеральной совокупности не были включены в основу выборки.

Неохват может стать источником серьезных неточностей, при этом ошибка неохвата относится только к ошибочно выпавшим из рассмотрения частям совокупности, но никак не к частям, исключенным намеренно. Таким образом, проблема неохвата имеет отношение к основе выборки. *Например, при общем обследовании ошибка неохвата может возникнуть при использовании телефонного справочника в качестве основы выборки. Телефоны есть далеко не у всех семей, и не все номера телефонов включены в справочник. Помимо прочего, существует целый ряд демографических отличий между лицами, имеющими телефоны и не имеющими их.*

*При почтовом опросе, где основой выборки служит рассылочная ведомость, ошибка неохвата может быть следствием того, что рассылочная ведомость не дает адекватного представления о различных группах популяции.*

*В тех случаях, когда данные должны собираться методом обхода квартир, респондентов сводят в своеобразную территориальную выборку. В этом случае основой выборки становится скорее не список респондентов, а определенные районы, кварталы или дома. Тем не менее, это не снимает проблемы неполноты основы выборки. Городские карты могут устареть, вследствие чего районы новой застройки полностью выпадут из рассмотрения. Помимо прочего, инструкции, данные интервьюерам, могут оказаться недостаточно*

детальными. Указание «начните обследование с северо-западной оконечности означенного района, избрав начальный пункт случайным образом и подвергая обследованию каждый пятый жилой дом» может оказаться неполным, если в этом районе существуют многоквартирные дома. Практика показывает, что при проведении обследования интервьюеры предпочитают обходить стороной ветхие или запущенные строения. Вместе с тем интервьюеры предпочитают общаться с максимально доступными членами семей, что противоречит положению инструкции о случайном характере опроса. Все это приводит к недостаточному представлению определенной части популяции при одновременном избыточном представлении другой, наиболее доступной ее части.

Проблемы с основной выборки существуют и при проведении опросов в торговых центрах. С одной стороны, здесь отсутствует список элементов совокупности. С другой стороны, попасть в число обследуемых могут только те люди, которые привыкли совершать здесь покупки, при этом, чем чаще они это делают, тем выше вероятность того, что они станут респондентами.

Именно по этой причине при обследованиях такого рода часто используются пропорциональные выборки.

Тем не менее, использование пропорциональных выборок не снимает проблем, связанных с ошибками неохвата. Свобода же выбора респондентов интервьюером буквально распахивает двери для таких ошибок. «Недобор» характерен как для лиц с самыми низкими, так и для лиц с самыми высокими доходами. Руководитель проекта может не осознавать этого, поскольку исполнители зачастую склонны к фальсификации результатов, позволяющей скрыть это обстоятельство. Чем более сложной и развитой будет такая пропорциональная выборка, тем более критичным становится названный момент. Если некоторые элементы выборки будут задаваться набором, состоящим из трех или четырех признаков, интервьюер, испытывающий в этой связи определенные затруднения, может «немного слукавить», приписав требуемые характеристики тем, кто ими не обладает.

**Ошибка перебора** может возникать из-за возникновения повторов в сводке элементов выборки.

Единицы с множественными входами в основу выборки, например семьи, имеющие несколько телефонных номеров, имеют более высокую вероятность включения в выборку, чем единицы, соответствующие только одной позиции списка. Однако для большинства ис-

следований ошибки неохвата представляют значительно большую опасность. Ошибка неохвата представляет проблему не для любого исследования. В ряде случаев, при исследовании имеется в распоряжении четкая, ясная и полная основа выборки. *Например, если администрация универсама решит провести обследование покупателей, приобретающих товары в кредит, проблем с определением основы выборки у нее не будет. Основа выборки, очевидно, будет совпадать со списком таких кредитов по открытым счетам. Возможны какие-то затруднения при распознавании активных и неактивных счетов, но указанная проблема может быть разрешена уже на подготовительном этапе исследования.*

Ошибки неохвата ставят перед исследователем два вопроса:

1. Насколько они серьезны?
2. Каким образом можно их уменьшить?

Основная проблема состоит в том, что их величина может быть оценена только при сравнении результатов выборочного исследования с определенными независимыми внешними показателями.

Внешний же показатель в свою очередь может быть либо определен посредством дополнительной проверки качества части результатов, либо получен в ходе другого надежного обследования, результаты которого не утратили своей актуальности, *например, последней переписи населения.*

*Рассмотрим случай, когда ошибка неохвата очень велика. Самое действенное состоит в улучшении качества основы выборки. Улучшение это может заключаться в обновлении карт, в выборочной проверке качества и репрезентативности выборки и т. п. Например, проблема отсутствия ряда элементов, характерная для телефонных обследований, может решаться путем набора случайных цифр или дополнительными звонками, хотя такой способ не дает адекватного представления о лицах, не имеющих телефона.*

Несовершенная основа выборки может быть улучшена лишь до определенного предела. Когда этот предел будет достигнут, можно попытаться уменьшить ошибку неохвата посредством отбора элементов выборки или уточнением результатов.

Например, при формировании выборки по списку часто сталкиваются с тем, что в нем содержатся повторы, неподходящие элементы и пропуски. Первым шагом, направленным на исправление ситуации, должно стать обновление и уточнение списка, для чего могут использоваться дополнительные источники. Однако подобные меры могут

привести только к сокращению количества пропусков, но не к выявлению повторов или неприемлемых элементов. При отборе элементов выборки неприемлемые элементы должны исключаться из рассмотрения. Не рекомендуется заменить неприемлемый элемент элементом, следующим в списке за ним, т.к. использование подобного приема может оказать серьезное влияние на характер выборки. Если отбор осуществляется случайным образом, следует просто перейти к следующему, отобранному случайным образом элементу. Если же отбор производится систематически, то при появлении неприемлемых элементов следует, прежде всего, отрегулировать выборочный интервал.

Проблема повторов обычно решается методом корректировки. Обычно результаты получают вес, обратный, вероятности попадания элемента в выборку. *Например, при обследовании, базирующемся на регистрационной ведомости автомобилей, каждому респонденту будет задан вопрос: «Сколько у вас машин?» Владелец двух машин получит весовой коэффициент 1/2, владелец трех машин соответственно 1/3.*

Ошибки неохвата:

1) относятся к разряду систематических ошибок и потому не входят в стандартные статистические зависимости;

2) как правило, не могут быть устранены посредством увеличения объема выборки;

3) могут иметь существенный размер;

4) могут быть уменьшены (но не обязательно устранены) при осознании их наличия с помощью улучшения основы выборки и принятия ряда специальных мер, позволяющих до определенной степени компенсировать остаточное несовершенство основы.

Другой возможный источник систематического отклонения ненаблюдения – **ошибка неполучения данных**.

**Ошибка неполучения данных** – систематическая ошибка, порождаемая отсутствием информации о некоторых элементах, которые должны были войти в состав выборки.

Для того, чтобы оптимизировать и стандартизировать практику исследований предлагается следующее стандартное определение *доли ответивших*.

**Доля ответивших** – отношение количества проведенных с респондентами интервью к количеству приемлемых респондентов в выборке.

Различают две главные причины ошибки неполучения данных – это *отсутствие* и *отказ от интервью*. Ошибки неполучения дан-

ных могут возникать при обследованиях, использующих личный, телефонный или почтовый опросы. При проведении почтовых опросов проблема отсутствия подменяется проблемой неполучения опросного листа. *Опросный лист может просто затеряться на почте, и в этом случае систематическую ошибку можно считать случайной (если только этой потере не подлежат некие фундаментальные причины, такие как выбытие или смерть адресата; последние приведут к появлению систематической ошибки).*

Данные о некоторых элементах выборки могут отсутствовать, поскольку в момент звонка интервьюера респондентов может не оказаться дома. Опытные данные показывают, что процент **отсутствия** растет со временем, при этом многое зависит от респондента и времени суток, в которое совершается звонок. *Замужние женщины с маленькими детьми в течение дня (речь не идет о выходных) оказываются дома чаще, чем мужчины или бездетные и незамужние женщины.* Вероятность того, что ответят, выше для семей с низким достатком и для сельских семей. Этот показатель подвержен сезонным и недельным (рабочие дни или выходные) колебаниям.

Таким образом, **отсутствие** – систематическая ошибка, возникающая вследствие неполучения ответов от заранее определенных респондентов, отсутствующих дома в момент звонка регистратора.

Для снижения эффекта отсутствия может быть принят ряд мер при некоторых исследованиях интервьюер может заранее договориться с респондентом о времени своего звонка (этот подход особенно эффективен при исследовании административных работников, но может оказаться неоправданным при обследовании обычных потребителей, в последнем случае принято использовать повторный звонок или повторные звонки, которые должны производиться в другой час).

На деле неполучение данных, обусловленное отсутствием респондентов, имеет очень большое влияние на точность большинства обследований. *Например, один ведущий эксперт пришел к заключению, что небольшие выборки с 4–6 повторными звонками оказываются более эффективными, чем большие выборки без повторных звонков, если только процентное значение доли ответивших существенно не превышает нормальный уровень.* Некоторые данные свидетельствуют о том, что для осуществления контакта с тремя четвертями выборки семей необходимо совершить от 4 до 5 звонков.

Альтернативой **прямому повторному звонку** может являться **модифицированная попытка контакта**.

*Например, если после первой попытки контакта и нескольких повторных звонков интервьюеру так и не удастся вступить в контакт с нужным лицом, он может послать ему опросный лист с конвертом или оставить этот лист у его двери. Если же отсутствие обусловлено «отсутствием нужного лица», а не «отсутствием кого-либо дома», регистратор может узнать у домашних время, когда респондент обычно бывает дома.*

Нельзя пытаться решить проблему отсутствием посещением соседней квартиры или звонком по номеру, следующему в списке за нужным, т. к. подменяя «присутствующими» (которые могут существенно отличаться от отсутствующих по ряду характеристик) проблема не разрешается, а лишь усугубляется.

**Отказы от интервью.** Почти всегда при проведении обследования находят люди, которые отказываются принять в нем участие.

*Например, в одном из самых массированных исследований серьезности этой проблемы «Ваш голос» принимали участие 46 исследовательских фирм, проводивших почти 1,4 миллиона телефонных и персональных интервью. Обследование показало, что около 38 % опрошенных отказались от участия в опросе. Причем количество людей, отказавшихся участвовать в нем до проведения или в ходе предварительного собеседования, составляло 86 %. Остальные отказались от участия непосредственно в ходе обследования.*

**Отказы от интервью** – систематическая ошибка, возникающая вследствие того, что часть респондентов отказывается принимать участие в обследовании.

Доля отказов зависит часто от особенностей респондентов, организаций, осуществляющих финансовое обеспечение обследования, обстоятельств контакта, темы обследования и искусства интервьюера. На долю отказов может повлиять и культура данной территории. *Например, в некоторых государствах, таких, как Саудовская Аравия, обследовать женщин практически невозможно.*

Имеет значение и метод сбора информации. Опытные данные свидетельствуют о том, что наиболее эффективными являются персональные, а наименее эффективными – почтовые методы опроса. Телефонные интервью занимают промежуточное положение.

Перечислим субъективные и объективные причины отказа:

- наименее вовлекаемыми в процесс обследования являются женщины, лица с невысоким уровнем образования и невысоким уровнем дохода;
- характер организации, производящей обследование;

– внешние обстоятельства, например, респондент может быть занят, может испытывать усталость или чувствовать себя не лучшим образом;

– предмет исследования;

– сама личность интервьюера.

Стратегии, рекомендуемые для корректировки ошибки:

1) увеличение доли первичных ответов;

2) повторные попытки;

3) экстраполяция полученной информации.

**Увеличение доли первичных ответов.** Улучшение условий проведения интервью и углубленное обучение интервьюеров – основные пути увеличения доли ответивших. Целевая совокупность определяется поставленной задачей и может содержать семьи с различным образовательным уровнем и уровнем доходов, культурой и профессиональной принадлежностью и т. д.

Нужному сотрудничеству может способствовать и убеждение респондентов в ценности проводимого исследования и важности их участия в нем.

Чем больше информации о сути и цели исследования сообщают интервьюеры потенциальным респондентам, тем большей становится доля ответивших как при персональном, так и при телефонном исследовании. Гарантия конфиденциальности также способствует увеличению доли ответивших, т. к. многие участники опроса не хотят, чтобы их ответы ассоциировались с их именами. При проведении почтовых обследований позитивную роль может сыграть и материальный стимул. Однако подобный стимул не эффективен при персональных опросах, если только они не проводятся в торговых центрах.

**Повторные попытки.** В некоторых случаях причиной отказа участвовать в исследовании могут стать изменчивые или временные обстоятельства. Поэтому повторная попытка установления контакта может привести к успешному исходу и позитивно повлиять на общий показатель доли ответивших. *Если респондент отказался от участия в обследовании, сославшись на болезнь или на усталость, можно надеяться на успешный исход повторной (аналогичной или несколько видоизмененной) попытки. При почтовом опросе эта попытка выразится в отправке повторного почтового отправления. Успех подобных мероприятий во многом зависит от уровня компетентности персонала.*

Если же источником ошибки, обусловленной неполучением данных, является сам предмет исследования, то необходимо сделать так, чтобы заинтересовать респондента, например, можно ввести в опросный лист не относящиеся к делу вопросы.

**Экстраполяция полученной информации.** Третья стратегия состоит в оценке возможного эффекта, обусловленного неполучением данных, и в соответствующей коррекции результатов исследования.

До сих пор речь шла о *полном неполучении данных*. **Частичное неполучение данных**, которое также может приводить к возникновению серьезных проблем, состоит в том, что респондент, согласившийся принять участие в опросе, не хочет или не может ответить на некоторые вопросы из-за специфики их формы или содержания или вследствие нежелания обременять себя поиском нужной информации. Поэтому при разработке опросных листов необходимо сделать все для того, чтобы подобных проблем не возникало. Однако полностью исключить появление таких вопросов невозможно.

Возможность исправления ситуации во многом зависит от *размера частичного неполучения данных*. Здесь различают катастрофическое и случайное, частичное неполучение данных. Если слишком большое количество вопросов остается без ответов, последние теряют смысл, и возникает ситуация полного неполучения данных от ряда респондентов. Если же количество таких вопросов относительно невелико, ответ остается осмысленным.

### **6.3. Понятие и сущность ошибок наблюдения**

Наиболее распространенной ошибкой наблюдений является *ошибка сбора*.

**Ошибка сбора** – систематическая ошибка, возникающая при сборе данных.

Ошибка сбора возникает уже после того, как нужный индивид согласится принять участие в обследовании. Вместо того чтобы идти на полное сотрудничество, он отказывается отвечать на одни и дает неправильные ответы на другие вопросы интервьюера. Такие ошибки называются соответственно *ошибками пропуска* и *ошибками свидетельства*.

При рассмотрении ошибок в ответах следует понимать смысл процедуры опроса.

Во-первых, респонденту необходимо понять суть вопроса.

Во-вторых, ему необходимо обдумать свой ответ.

В-третьих, респондент понимает, что его ответ должен быть точным.

В-четвертых, он должен руководствоваться и иными соображениями, например, стремлением произвести на интервьюера должное впечатление.

В-пятых, он должен представить результаты процесса предоставления ответов на вопросы.

Целью опроса является реализация последней стадии. Нарушить же этот процесс возможно на любом этапе, следствием чего будет неточный ответ или ошибка в ответе.

Факторы, вызывающие ошибки в ответах, многочисленны и практически не поддаются классификации. Однако при работе с ошибками сбора данных можно воспользоваться схемой Кана–Кэннела (рис. 6.2)



Рис. 6.2. Модель возникновения ошибок (опрос)

Данная модель исходит из нескольких предположений. *Во-первых*, каждая личность имеет свойственные ей характеристики и психологические предпочтения, которые могут повлиять на отношение к опросу. Некоторые характеристики (такие, как возраст и пол) очевидны, другие (такие, как психологическое состояние) могут оставаться неизвестными интервьюеру.

*Во-вторых*, интервью является интерактивным процессом, детерминантами которого являются как интервьюер, так и респондент. Каждая из сторон отрабатывает специфическое поведение другой стороны.

Модель взаимодействия интервьюера-респондента имеет ряд положительных сторон. *Во-первых*, она соответствует эмпирическим данным. *Во-вторых*, она позволяет судить о том, как можно было бы свести к минимуму ошибки в ответах (а также ошибки неполучения данных, вызванные отказами). Данная модель может быть применима не только к персональному, но и к телефонному и почтовому методам опроса, что еще больше повышает ее ценность.

**Личные особенности (характеристики).** Опытные данные свидетельствуют о том, что личные особенности могут серьезно повлиять на ответы. Если интервьюер и респондент имеют много общего, их взаимодействие становится куда более легким и успешным. В первую очередь это относится к таким очевидным характеристикам, как национальность, возраст и пол, но это же правило приемлемо и к менее очевидным особенностям: принадлежности к определенному классу или группе населения. Имеет смысл подбирать интервьюера таким образом, чтобы он как можно меньше отличался от респондента, в этом случае вероятность позитивного исхода интервью существенно возрастает.

**Психологические факторы.** На практике результаты работы интервьюеров имеют явную обусловленность их взглядами, позициями и стремлениями. Естественно, подобную психологическую обусловленность имеет любой человек. Именно по этой причине большинство обследований проводится по жесткой фиксированной схеме, которой должны обязательно придерживаться все интервьюеры.

**Поведенческие факторы.** Биографические данные, мнения, позиции, намерения респондента также могут являться источниками ошибок. Их наличие или отсутствие зависит от характера взаимодействия интервьюер–респондент. То есть ошибки такого рода возникают непосредственно в ходе этого взаимодействия.

Существуют три модели поведения интервьюеров, которые приводят к появлению ошибок:

- 1) ошибки при формулировке вопросов и неумение задавать уточняющие вопросы;
- 2) ошибки при записи ответов;
- 3) подтасовка данных.

Ошибки при формулировке вопросов могут сопутствовать вопросам любого типа, наиболее острой эта проблема становится при работе с вопросами, допускающими различные толкования, предполагающими продолжение опроса.

Немаловажное значение имеет и то, как будет сформулирован исходный вопрос. Часто вопрос перефразируется, чтобы сделать суть вопроса более понятной для респондента. При этом интервьюеры могут «вчитывать» в вопрос собственное понимание или собственную позицию, тем самым предрасполагая респондента к тому или к иному ответу.

Альтернативные вопросы сопряжены с высокой вероятностью появления ошибки. Эта ошибка может возникать, например, из-за того, что при постановке вопроса интервьюер может сделать особый акцент на одной из альтернатив. Небольшого изменения тона достаточно для того, чтобы изменить смысл всего вопроса.

Одна из главных задач интервьюера – удержание внимания и интереса респондента. Бывает, что интервьюер одновременно должен фиксировать неоднозначные ответы респондента или перепроверять их соответствие друг другу. Одновременное решение двух этих задач в ряде случаев может приводить к ошибкам.

*Например, часто, занятый своим делом, интервьюер может «не услышать» респондента. Причиной этого может стать как неразборчивость сказанного, так и сосредоточенность интервьюера на чем-то ином. Последний может услышать именно то, что он хотел услышать и воспринять именно то, что он хотел воспринять.*

Подтасовка данных также может стать источником ошибки. Эта подтасовка может относиться как ко всему опросному листу, так и к отдельным его пунктам.

*Например, фонд рекламных исследований (АКР) порой проводит проверочные повторные обследования выборок, уже подвергавшихся обследованию. Сотрудники фонда хотят убедиться в том, что опрос действительно имел место и что были заданы все нужные вопросы. При проведении одного из таких обследований оказалось, что*

5,4 % опросных листов 33 различных обследований не находят подтверждения, а 7,9 % имеют не менее двух серьезных противоречий. Даже Бюро переписей, которое можно было бы назвать самым скрупулезным и аккуратным сборщиком информации в мире, постоянно сталкивается со сфабрикованными опросными листами.

Большинство коммерческих исследовательских фирм проверяет достоверность 10–20 % опросных листов путем проведения контрольного почтового или телефонного опроса.

Достоверность опросных листов может проверяться по следующим направлениям:

1) **метод опроса** – проверяется соответствие использованного метода заданному (например, действительно ли проводился персональный, а не телефонный опрос);

2) **поставленные вопросы** – проверка того, не были ли выпущены из рассмотрения важные вопросы (демографического или классификационного характера);

3) **демонстрация продукции** – проверка того, действительно ли была произведена необходимая для проведения опроса демонстрация продукта или информационного листа;

4) **знакомство респондента с интервьюером** – проверка того, не занимался ли интервьюер опросом своих знакомых или друзей;

5) **реакция на проведение опроса** – проверка «качества» работы интервьюера.

Другой разновидностью обмана, не являющейся ошибкой в ответах, но серьезно влияющей на все систематические ошибки, могут быть раздутые счета.

Интервьюер может указать завышенное расстояние или завышенную продолжительность обследования

**Офисные ошибки.** Систематические ошибки могут возникать не только при сборе информации. Они могут появляться при редактировании, кодировании, составлении таблиц и анализе данных. В большинстве случаев эти ошибки могут быть устранены частично или полностью введением надлежащего контроля над процессом обработки данных.

**Суммарные ошибки.** Частные ошибки, складываясь, приводят к ошибке суммарной, которая и должна интересовать исследователей в первую очередь. Прежде всего, это относится к ошибке выборки. При увеличении объема ошибка выборки уменьшается. Однако при этом могут возрасти систематические ошибки, т. к. исследование

большой выборки предполагает увеличение количества опросов, делает более затруднительным подбор кадров, обучение и контроль. Кроме того, систематическая ошибка куда опаснее ошибки выборки. Величину последней можно оценить. Многие же типы систематических ошибок не поддаются оценке. Величина ошибки выборки может быть снижена путем использования более сложных методов обследования или простым увеличением объема выборки. Эта процедура достаточно проста, и позволяет исследователю обеспечивать требуемую точность результатов. С систематическими ошибками все обстоит иначе. Очевидных методов повышения точности в этом случае нет. Обнаруживаются все новые и новые источники систематических ошибок, причем автоматическая коррекция последних оказывается затруднительной или невозможной.

## **Тема 7. ЭКСПЕРТНЫЕ ОЦЕНКИ**

### **7.1. Общая характеристика экспертных оценок**

Рассмотрим порядок проведения экспертных оценок.

Так как процедура формирования целей маркетинговых исследований, вариантов их реализации, моделей, критериев и т. п. не могут быть формализованы, поэтому оценки в этой области и в дальнейшем будут основываться на видении, интуиции, опыте того, кто их формирует.

Примерами традиционных эвристических процедур являются различные экспертизы, консилиумы, совещания и т. п. Основываются они чаще на опыте, однако в последнее время все чаще используются разнообразные математические методы обработки информации эвристического происхождения и в первую очередь статистические методы.

Методы экспертных оценок используются:

- 1) для прогнозирования будущего, если отсутствуют или недостаточно статистических данных;
- 2) для количественного измерения событий, для которых нет других способов измерения (например, при оценке важности целей и предпочтительности отдельных методов продвижения товара).

При проведении маркетинговых исследований широко применяются экспертные оценки, основанные на суждениях менеджеров, торговых или иных посредников, сотрудников сбытовых служб и потребителей.

Экспертные оценки имеют следующие недостатки:

- нет гарантий, что полученные оценки в действительности достоверны;
- имеются определенные трудности в проведении опроса экспертов и обработке полученных данных.

Если второй недостаток преодолим, то первый не всегда можно ликвидировать, т. к. даже существующие способы определения достоверности экспертных оценок не дают гарантий правильной оценки.

*Например, были случаи, когда отдельные эксперты, не согласные с мнением большинства, давали именно правильные оценки. Поэтому единодушие большинства не всегда является критерием достоверности.*

Одним из показателей, характеризующих пригодность эксперта, является степень его надежности. Под **степенью надежности** понимается относительная частота случаев, когда он приписывал более высокую вероятность тем событиям, которые в последствии реализовывались (поэтому считается, чем чаще эксперт оказывается прав, тем больше его авторитет).

Применяются *индивидуальные* и *групповые (коллективные)* экспертные опросы.

Основными целями **индивидуальных экспертных опросов** являются:

- прогнозирование хода развития событий и явлений в будущем, а также их оценка в настоящем (например, количественная и качественная оценка вероятности реализации определенного события в будущем; прогнозирования характеристик рынка и направлений их развития; определение целей и стратегий маркетинговой деятельности; ранжирование различных объектов и т. п.);

- анализ и обобщение результатов, представленных другими экспертами;

- составление сценариев;

- выдача заключений на работу других специалистов или организаций (рецензии, отзывы, экспертизы и т. п.).

Отдельный эксперт может быть полезен как источник информации и как советник для руководителей, принимающих решения.

К числу **групповых экспертных опросов** относятся следующие типы:

- открытое обсуждение поставленных вопросов с последующим открытым или закрытым голосованием;

- закрытое обсуждение с последующим закрытым голосованием или заполнением анкет экспертного опроса;
- свободное высказывание без обсуждения и голосования.

Групповые оценки могут быть получены при опросе экспертов как *одной специальности*, так и *разных* (в последнем случае оценка носит характер деловой игры).

**Первый тип** имеет существенные недостатки (например, взаимное влияние мнений экспертов, нежелание отказаться от ранее публично высказанных точек зрения), поэтому на практике чаще используют второй и третий типы групповых оценок.

**Второй тип** оценок можно подразделить на две категории:

- экспертный опрос, проводимый в один тур путем одноразового заполнения анкет;
- экспертный опрос, проводимый в несколько туров путем многократного заполнения анкет с целью последовательного уточнения оценок.

## 7.2. Типы экспертных опросов

Одним из наилучших методов использования суждений экспертов является метод «Дельфи», предусматривающий проведение экспертного опроса в несколько туров.

Метод «Дельфи» можно назвать «кибернетическим арбитражем» (кибернетическим потому, что процесс выработки суждений экспертами управляется рабочей группой через обратную связь). В данном методе делается попытка усовершенствовать групповой подход к решению задачи разработки прогноза или оценки путем взаимной критики субъективных взглядов, высказываемых отдельными специалистами без непосредственных контактов между ними и при сохранении анонимности мнений или аргументации в защиту этих мнений.

В одном из вариантов этого метода прямое обсуждение заменяется обменом информацией и мнениями с помощью тщательно разработанных вопросников.

К участникам обращаются с просьбой не только высказать свои мнения, но и привести их обоснование. А в каждом из последующих туров им выдается новая и уточненная информация по высказанным мнениям. Эта информация образуется в результате расчета совпадения точек зрения по ранее выполненным этапам работы. Этот процесс продолжается до тех пор, пока продвижение в направлении повыше-

ния совпадения точек зрения не станет незначительным. После этого фиксируются расходящиеся точки зрения.

Метод «Дельфи» применяется в маркетинге при решении таких задач, как экспертное определение сметной стоимости необычных проектов; выявление множества важнейших факторов; путей решения маркетинговых проблем (например, увеличение сбыта).

Метод коллективной генерации идей (метод «мозговой атаки») относится к третьему типу групповых экспертных оценок и направлен на получение большего количества идей, в том числе и от лиц, которые, обладая высокой степенью эрудиции, обычно воздерживаются от высказываний.

При проведении экспертных оценок с помощью данного метода проблема должна быть сформирована в основных терминах с выделением основного вопроса.

При этом предусматривается:

- отсутствие любой критики;
- свободная интерпретация идей в рамках данного вопроса;
- получение максимального количества идей;
- поощрение различных комбинаций идей и путей их усовершенствования.

Исследования эффективности этого метода показали, что групповое мышление производит на 70 % больше ценных новых идей, чем сумма индивидуальных мышлений. Наиболее продуктивными признаны группы в 10–15 человек, хотя есть примеры использования и более многочисленных групп – до 200 человек.

С помощью данного метода решаются следующие задачи маркетинговых исследований:

- определение возможных путей развития прогнозируемого процесса или объекта, один из которых оптимальный (при решении подобных задач данный метод применяется с целью определения полного набора возможных путей развития);
- определение наиболее широкого круга применяемых методов, если решение проблемы требует параллельного или последовательного использования ряда разнообразных методов;
- выявление круга факторов, которые необходимо принимать во внимание, определяя окончательный вариант решения задачи.

Основным недостатком данного метода является значительный уровень информационного шума, создаваемый тривиальными идеями (спонтанный и стихийный характер идей). Кстати, принципы данного метода просматриваются в «пятиминутках» и «оперативках».

### 7.3. Порядок проведения экспертных опросов

Выделяются следующие основные этапы проведения экспертных опросов:

1. Подбор экспертов и формирование групп.
2. Формирование вопросов и составление анкет.
3. Формирование правил определения суммарных оценок на основе оценок отдельных экспертов.
4. Работа с экспертами.
5. Анализ и обработка экспертных оценок.

**1 этап.** Здесь, исходя из целей экспертных оценок, решаются вопросы относительно структуры экспертной группы, количества экспертов и их индивидуальных качеств, т. е. определяются направления (разработка новых товаров, ценообразование, сбыт и т. п.), по которым необходимо привлечь экспертов (устанавливается их специализация). Далее по каждому направлению выделяются подгруппы экспертов, и устанавливается количество экспертов в каждой подгруппе, которое зависит от конкретной постановки задачи и должно обеспечить необходимый охват по каждому конкретному вопросу. Затем определяются требования к квалификации экспертов, стажу их работы в данной области опроса и общему стажу.

На практике чаще используется «прагматический подход» определения численности экспертной группы. По данному подходу **нижняя граница** численности ( $N_{\min}$ ) должна зависеть от числа оцениваемых событий ( $m$ ) (например, данное правило соблюдается при работе ученых советов, комиссий специалистов и т. п.). Поэтому принимается условие  $N_{\min} \geq m$ .

**Верхней границей** численности экспертной группы является потенциально возможное число экспертов ( $N_n$ ).

$N_{\max} \leq N_n$ , следовательно действительное значение численности группы  $N$  находится в пределах

$$m \leq N_{\min} \leq N \leq N_{\max} \leq N_n.$$

После этого определяется структура и состав группы, т. е. подбираются в группы эксперты требуемой специализации и квалификации, а для того чтобы событию была дана всесторонняя оценка, то число экспертов  $N_i$  каждого  $i$ -го направления желательно выбрать одинаковым:

$$N_i = \frac{N}{r}, \quad (7.1)$$

где  $r$  – число направлений;  $I = 1; 2; 3; \dots; r$ .

Дальше должно быть учтено ограничение, касающееся соответствия целей экспертов целям экспертных оценок, т. е. необходимо установить, есть ли тенденция у отдельных экспертов объективно оценивать рассматриваемые события. Для этого нужно выявить возможные потенциальные цели экспертов, противоречащие целям получения объективных результатов. Анализируя предшествующую деятельность экспертов, выясняют наличие причин, которые привели к завышению или занижению оценок таким образом, чтобы это повлияло на групповые оценки в направлении желательном для этого эксперта или других лиц.

Исходя из вышесказанного, последовательность формирования групп может быть следующей. Сначала определяется потенциально возможное число экспертов ( $N_n$ ), затем минимально допустимая численность ( $N_{\min}$ ). После этого среди потенциально возможных экспертов выделяют группу специалистов  $N_k$  ( $N_k \leq N_n$ ) требуемой квалификации. Затем из  $N_k$  исключают экспертов, у которых цели могут противоречить целям получения объективных результатов. Из оставшихся экспертов  $N_{\text{ц}}$  ( $N_{\text{ц}} \leq N_k$ ) формируют группы таким образом, чтобы обеспечить равное представительство. Для этого определяют вначале максимально возможное число представителей:

$$N_i^{\max} = \frac{N_{\text{ц}}}{r}. \quad (7.2)$$

Затем находят направление с наименьшим числом экспертов  $N_i^{\min}$ , которое входит в группу  $N_{\text{ц}}$  и затем уже численность группы будет определяться по формуле

$$N = N_i^{\min} \cdot r. \quad (7.3)$$

## **2 этап.** Формирование вопросов и составление анкет.

Правила опросов экспертов должны обеспечить соблюдение ряда условий, таких как:

- независимость формирования экспертами собственного мнения об оцениваемых событиях;
- удобство работы с предлагаемыми анкетами (вопросы должны быть в общепринятых терминах и не содержать смысловую неоднозначность);
- логическое соответствие вопросов структуре объекта опроса;
- приемлемые затраты времени по вопросам анкеты, удобное время получения вопросов и выдачи ответов;

- сохранение анонимности ответов;
- проведение коллективных обсуждений оцениваемых событий;
- представление экспертами требуемой информации.

Для обеспечения выполнения этих условий должны быть разработаны правила проведения опроса и организации работы экспертной группы.

**3 этап.** Формирование правил определения суммарных оценок на основе оценок отдельных экспертов.

В общем случае групповая оценка ( $a_i$ ) каждого события ( $C_i$ ) зависят от оценок этого события экспертами, а также степени компетентности ( $h_j$ ) экспертов. Степень компетентности ( $h_j$ ) эксперта учитывает его опыт и квалификацию и является основной характеристикой эксперта, которая используется при определении групповых оценок, т. е.  $a_i = f(a_{ij}, h_j)$ , где  $I = 1; 2; \dots; m$  – количество событий,  $j = 1; 2; \dots; p$  – количество экспертов.

Обычно данную функцию представляют в виде

$$a_i = \frac{\sum_{j=1}^p h_j \cdot a_{ij}}{\sum_{j=1}^p h_j}. \quad (7.4)$$

При равной компетентности экспертов эта формула принимает вид:

$$a_i = \frac{\sum_{j=1}^p a_{ij}}{p}. \quad (7.5)$$

В зависимости от специфики экспертных опросов, объекта исследования и используемой методики обработки экспертных данных оценки  $a_{ij}$  может иметь различную шкалу измерения: от 0 до 1; от 0 до 10; от 0 до 100. Выбор той или иной шкалы определяется удобством получения и обработки оценок, а также предпочтениями исследования.

В связи с тем, что оценка компетентности эксперта, как и оценка самого эксперта, носит субъективный характер, то на практике часто используют вторую формулу в качестве интегральной оценки. Только при этом на этапе отбора экспертов большее внимание уделяют неформальной оценке уровня квалификации экспертов.

#### 4 этап. Работа с экспертами.

Работа с экспертами в основном содержит 3 этапа:

На *первом этапе* эксперты привлекаются в индивидуальном порядке с целью:

- уточнить модель объекта, ее параметры и показатели, подлежащие оценке;
- уточнения формулировки вопросов и терминологии в анкетах;
- согласования целесообразности представления таблиц экспертных оценок в той или иной форме;
- уточнения состава группы экспертов.

На *втором этапе* экспертам направляются анкеты с пояснительным письмом, в котором описывается цель работы, структура и порядок заполнения анкет с примерами.

*Третий этап* работы с экспертами осуществляется *после* получения результатов опроса. На этом этапе от экспертов в форме консультаций получают всю недостающую информацию для уточнения полученных данных и окончательного анализа.

#### 5 этап. Анализ и обработка экспертных оценок.

При проведении анализа необходимо определить согласованность действий экспертов и достоверность экспертных оценок.

Пусть для каждого события ( $C_i$ ) на основании оценок  $a_{ij}$ , полученных группой из  $P$  экспертов образована матрица рангов важности –  $|| \alpha_{ij} ||$ , где  $i = 1; 2; \dots; m$  – число событий,  $j = 1; 2; \dots; p$  – число экспертов.

Матрица  $|| \alpha_{ij} ||$ , получается из матрицы  $|| a_{ij} ||$  исходя из рангов важности событий, т. е. при ранжировании события располагаются в порядке увеличения или уменьшения какого либо признака ( $X$ ) и ранг  $a_{ij}$  указывает то место, которое занимает  $i$ -е событие среди других ( $m$ ) событий, ранжированных в соответствии с признаком ( $X$ ).

Ранжирование применяется, когда события располагаются согласно неизмеримому и не подсчитываемому качеству (например, потребительские свойства товара, направления совершенствования товара и т. д.) или рассматриваются только относительно взаимного расположения во времени или пространстве.

Рассмотрим упорядоченную последовательность суммарных рангов  $m$  событий

$$a_1 \langle a_2 \langle \dots \langle a_i \langle \dots \langle a_m. \quad (7.6)$$

$\alpha_i$  находится из матрицы  $\|\alpha_{ij}\|$  по формуле

$$a_i = \sum_{j=1}^p a_{ij}. \quad (7.7)$$

Приведем условный пример определения рангов важности событий исходя из их коэффициентов относительной важности (более важное событие имеет меньший ранг важности)

Таблица 7.1

### Расчет рангов важности

Показатели	Эксперт	События С		
		1	2	3
Оценки экспертов, $\alpha_{ij}$	Эксперт $j = 1$	0,5	0,2	0,3
	Эксперт $j = 2$	0,3	0,6	0,1
Ранги важности, $\alpha_{ij}$	Эксперт $j = 1$	1	3	2
	Эксперт $j = 2$	2	1	3
Суммарный ранг важности, $\alpha_i$		3	4	5

Среднее значение для суммарных рангов рассматриваемого ряда равно

$$a = \frac{1}{2} p(m + 1). \quad (7.8)$$

Суммарное квадратическое отклонение ( $S$ ) суммарных событий от среднего значения  $\alpha$  равно

$$S = \sum_{i=1}^m \left( \sum_{j=1}^p a_{ij} - \frac{1}{2} p(m + 1) \right)^2. \quad (7.9)$$

Величина  $S$  будет максимальна, если все  $P$  экспертов дадут одинаковые оценки каждому событию ( $C_i$ ). Тогда

$$S_{\max} = \frac{1}{12} P^2 (m^3 - m). \quad (7.10)$$

Естественно, что в качестве меры согласованности экспертов принимается отношение  $S/S_{\max}$ , тогда имеем общую формулу

$$W = \frac{S}{S_{\max}} = \frac{\sum_{i=1}^m \left( \sum_{j=1}^p a_{ij} - \frac{1}{2} p(m+1) \right)^2}{\frac{1}{12} p^2 (m^3 - m)} = \frac{\sum_{i=1}^m \left( \sum_{j=1}^p a_{ij} - \frac{1}{2} p(m+1) \right)^2}{\frac{1}{12} m_p^2 (m^2 - 1)}, \quad (7.11)$$

где  $W$  – коэффициент конкордации. Его величина изменяется от 0 до 1. При  $W = 0$  согласованность мнений экспертов отсутствует, т. е. связи между оценками различных экспертов нет, и наоборот, при  $W = 1$  согласованность мнений полная (однако это может говорить о сговоре экспертов).

В том случае если последовательность (3) имеет еще и равенства, т. е. существует совпадение рангов, то формула для коэффициента конкордации имеет вид

$$W = \frac{S}{\frac{1}{12} m_p^2 (m^2 - 1) - p \sum_{j=1}^p T_j}, \quad (7.12)$$

где  $T_j$  определяется по формуле

$$T_j = \frac{1}{12} \sum_{K_j} (t_j^3 - t_j), \quad (7.13)$$

где  $t_j$  – число повторений каждого ряда в  $j$ -м ряду;  $K_j$  – число повторяющихся рангов в  $j$ -м ряду.

Когда ранги повторяются, то для получения нормальной ранжировки, необходимо приписать событиям, имеющим одинаковые ранги ранг, равный среднему значению мест, которые эти события поделили между собой.

Например, получена ранжировка событий (табл. 7.2).

Таблица 7.2

**Величина присвоенных рангов событиям**

События, $i$	1	2	3	4	5	6
Ранги, $\alpha_i$	1	2	3	3	2	3

События 2 и 5 поделили между собой 2-е и 3-е места. Значит, им приписывается ранг  $(2 + 3)/2 = 2,5$ .

События 3, 4 и 6 поделили между собой 4, 5 и 6-е места, и им приписывается ранг  $(4 + 5 + 6)/3 = 5$ . Таким образом, получаем нормальную ранжировку (табл. 7.3).

Таблица 7.3

**Нормальная ранжировка событий**

События, $i$	1	2	3	4	5	6
Ранги, $\alpha_i$	1	2,5	5	5	2,5	5

## Литература

1. Акулич, И. Л. Маркетинг : учебник / И. Л. Акулич. – 2-е изд., перераб. и доп. – Минск : Выш. шк., 2002. – 447 с.
2. Акулич, И. Л. Основы маркетинга : учеб. пособие / И. Л. Акулич, Е. В. Демченко. – Минск : Выш. шк., 1998. – 236 с.
3. Алексунин, В. А. Маркетинг: краткий курс : учеб. пособие / В. А. Алексунин. – Москва : Дашков и Ко, 2000. – 191 с.
4. Ансофф, И. Стратегическое управление / И. Ансофф ; сокр. пер. с англ. – Москва : Экономика, 1989. – 519 с.
5. Архипова, Л. В. Популярно о маркетинге / Л. В. Архипова, Б. В. Сребник. – Москва : Профиздат, 1991. – 80 с.
6. Багиев, Г. Л. Маркетинг : учеб. для вузов / Г. Л. Багиев, В. М. Тарасевич, Х. Анн ; под общ. ред. Г. Л. Багиева. – Москва : Экономика, 1999. – 703 с.
7. Балабанов, И. Т. Экономика туризма : учеб. пособие / И. Т. Балабанов, А. И. Балабанов. – Москва : Финансы и статистика, 1999. – 176 с.
8. Баркан, Д. И. Маркетинг для всех: Беседы для начинающих / Д. И. Баркан. – Ленинград, 1991. – 256 с.
9. Басовский, Л. Е. Маркетинг / Л. Е. Басовский. – Москва : ИНФРА-М, 1999. – 219 с.
10. Батра, Р. А. Рекламный менеджмент / Р. А. Батра, Д. Д. Майерс, Д. А. Аакер. – 5-е изд. – Москва : Вильямс, 1999. – 543 с.
11. Беляевский, И. К. Маркетинговое исследование: информация, анализ, прогноз : учеб. пособие / И. К. Беляевский. – Москва : Финансы и статистика, 2001. – 368 с.
12. Беляцкий, Н. П. Менеджмент: деловая карьера / Н. П. Беляцкий. – Минск : Выш. шк., 2001. – 302 с.
13. Березин, И. С. Маркетинг и исследование рынков / И. С. Березин. – Москва : Русская Деловая Литература, 1999. – 416 с.
14. Бороденя, В. А. Маркетинговые исследования : учеб.-метод. пособие / В. А. Бороденя. – Минск : БГЭУ, 2003. – 186 с.
15. Вайсман, А. Стратегия маркетинга: 10 шагов к успеху. Стратегия менеджмента: 5 факторов успеха / А. Вайсман. – Москва : Интерэксперт ; Экономика, 1995. – 246 с.
16. Васильева, Н. В. Реклама для всех / Н. В. Васильева. – Минск : Тесей, 2003. – 224 с.
17. Воробьева, Л. В. Теории конкуренции и современные факторы конкурентоспособности / Л. В. Воробьева. – Минск : Академия управления при Президенте Республики Беларусь, 2002. – 122 с.

18. Голубков, Е. П. Маркетинговые исследования. Теория, методология и практика / Е. П. Голубков. – Москва : Финпресс, 1998. – 416 с.
19. Голубкова, Е. Н. Маркетинговые коммуникации / Е. Н. Голубкова. – Москва : Финпресс, 2002. – 256 с.
20. Дурасов, А. С. Формирование маркетинговых коммуникаций / А. С. Дурасов, Е. В. Губанова. – Минск : Новое знание, 2000. – 120 с.
21. Дурович, А. П. Маркетинг в туризме : учеб. пособие / А. П. Дурович, А. С. Компанев. – Минск : Экономпресс, 1998. – 430 с.
22. Дурович, А. П. Основы маркетинга : учеб. пособие / А. П. Дурович. – Москва : Новое знание, 2004. – 512 с.
23. Уткин, Э. А. Рекламное дело / Э. А. Уткин. – Москва : Экмос, 1998. – 500 с.
24. Хруцкий, В. Е. Современный маркетинг: Настольная книга по исследованию рынка : учеб. пособие / В. Е. Хруцкий, И. В. Корнеева. – 3-е изд., перераб. и доп. – Москва : Финансы и статистика, 2003. – 560 с.
25. Черкасов, Г. И. Рынок: регулирование рынка : учеб. пособие для вузов / Г. И. Черкасов. – Москва : ЮНИТИ-ДАНА, 2004. – 222 с.
26. Черчилль, Г. А. Маркетинговые исследования / Г. А. Черчилль. – Санкт-Петербург : Питер, 2000. – 748 с.

## Содержание

Предисловие.....	3
Тема 1. Типы выборок и определение объема выборки.....	3
1.1. Этапы проектирования выборки .....	4
1.2. Типы плана выборочного контроля .....	7
1.3. Стратифицированная выборка.....	19
1.4. Групповая (гнездовая) выборка.....	23
1.5. Определение объема выборки при оценке среднего.....	28
1.6. Определение объема выборки при работе с выборочными долями .....	34
Тема 2. Сбор первичной информации.....	37
2.1. Характеристика методов проведения опросов .....	37
2.2. Сбор информации посредством анкетирования .....	41
2.3. Сбор информации посредством наблюдения.....	47
2.4. Эксперимент и имитационное моделирование.....	49
Тема 3. Разработка анкеты для сбора первичной информации.....	56
3.1. Этапы разработки вида анкеты и корректировка вопросов .....	56
3.2. Этапы определения последовательности вопросов и завершение разработки анкеты .....	66
Тема 4. Измерения в маркетинговых исследованиях .....	69
4.1. Основы измерений .....	69
4.2. Классификация и оценка ошибок.....	72
4.3. Измерение ожиданий, восприятия и предпочтений .....	74
Тема 5. Анализ данных .....	81
5.1. Редактирование и кодирование данных .....	81
5.2. Табулирование данных.....	86
5.3. Методы анализа документов.....	95
Тема 6. Реальные процедуры и систематические ошибки сбора данных.....	100
6.1. Влияние и значение систематических ошибок.....	100
6.2. Понятие и сущность ошибок ненаблюдения .....	104
6.3. Понятие и сущность ошибок наблюдения .....	111
Тема 7. Экспертные оценки.....	116
7.1. Общая характеристика экспертных оценок .....	116
7.2. Типы экспертных опросов .....	118
7.3. Порядок проведения экспертных опросов .....	120
Литература .....	127

Учебное электронное издание комбинированного распространения

Учебное издание

**Снопок Наталья Васильевна**

## **МАРКЕТИНГОВЫЕ ИССЛЕДОВАНИЯ**

**Курс лекций  
по одноименной дисциплине для студентов  
специальности 1-26 02 03 «Маркетинг»  
дневной и заочной форм обучения  
В двух частях  
Часть 2**

**Электронный аналог печатного издания**

Редактор *М. В. Аникеенко*  
Компьютерная верстка *Н. Б. Козловская*

Подписано в печать 18.11.09.

Формат 60x84/16. Бумага офсетная. Гарнитура «Таймс».

Ризография. Усл. печ. л. 7,67. Уч.-изд. л. 8,01.

Изд. № 194.

E-mail: [ic@gstu.gomel.by](mailto:ic@gstu.gomel.by)

<http://www.gstu.gomel.by>

Издатель и полиграфическое исполнение:  
Издательский центр учреждения образования  
«Гомельский государственный технический университет  
имени П. О. Сухого».

ЛИ № 02330/0549424 от 08.04.2009 г.

246746, г. Гомель, пр. Октября, 48.