

Методология маркетинговых исследований

Тема Процедура выборки

Вопросы лекции

- Основные понятия, используемые при проведении выборочных исследований. Генеральная совокупность. Выборка. Контур выборки. Единица и объем выборки. Ошибка выборки.
- Проблемы формирования выборки
- Репрезентативность выборки.
- Типы выборок. Виды вероятностной выборки. Виды невероятностной выборки.
- Комбинация типов выборки при проведении исследования. Многоступенчатая выборка.
- Этапы разработки выборочного плана. Подходы к определению объема выборки. Ошибка выборки.

Почему появляются выборки?

- Для маркетинговых целей в целях экономии времени и средств весьма типично подвергать исследованию лишь **небольшую часть** той группы людей или предприятий (генеральной совокупности), которая представляет интерес для исследователей.
- Она обычно составляет несколько сотен, иногда тысячу объектов.



Суть выборочного метода

- Суть выборочного метода заключается в том, что по определенным правилам из общей численности **генеральной совокупности** отбирается ограниченное **число единиц**, которое призвано в качестве своеобразной модели воспроизводить структуру всей генеральной совокупности
- **Генеральная совокупность** - та часть объекта исследования, которая локализована по времени и территориально, и на нее распространяются все выводы исследования

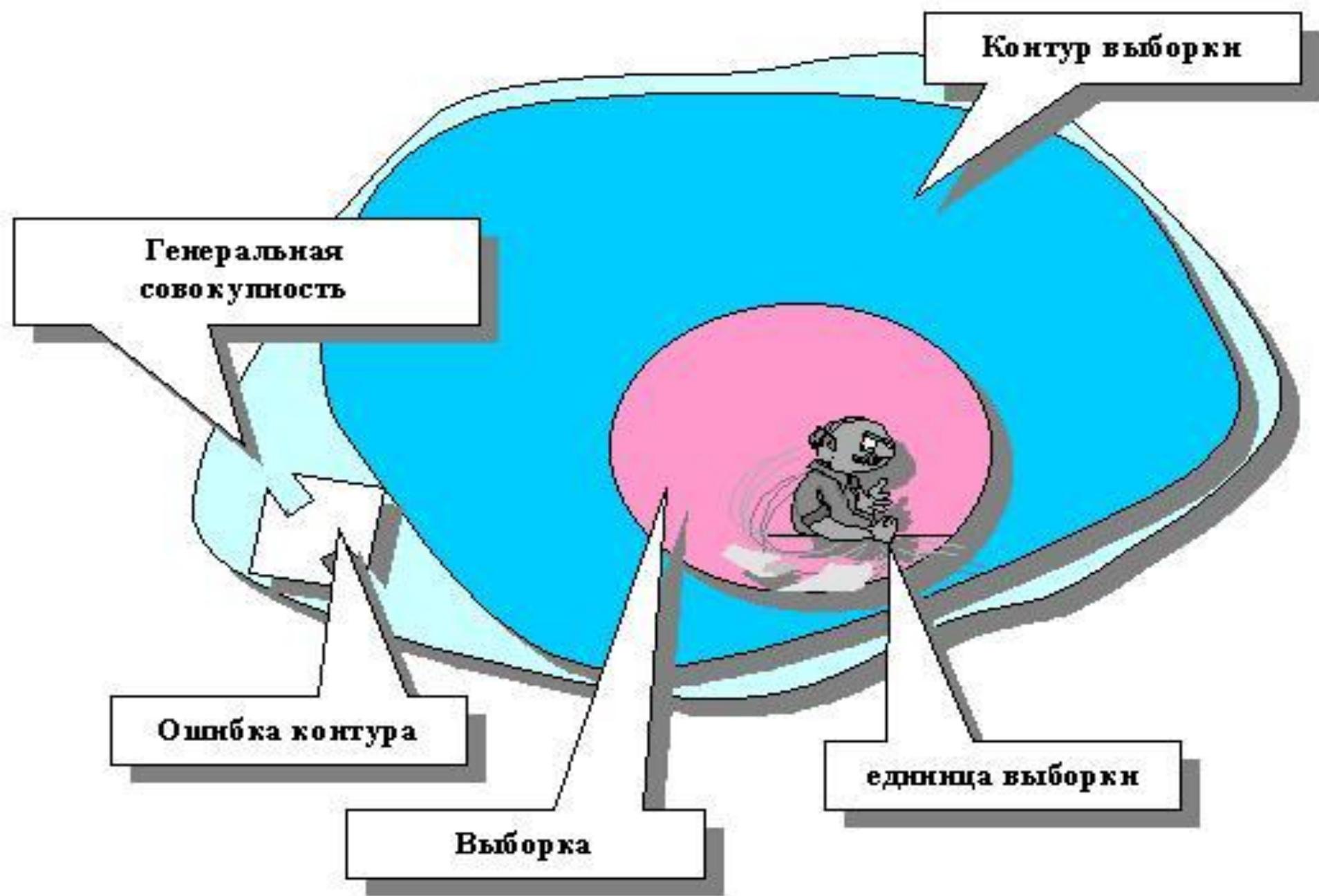
Примеры генеральных совокупностей:

- Все жители Нижнего Новгорода (1267760 человек по данным 2015 года)
- Юридические лица России (2,2 млн. на начало 2008 года)
- Розничные торговые точки, осуществляющие продажу продуктов питания (20 тысяч на начало 2008 года) и т.д.

Выборка. Контур выборки

- Ограниченное число единиц, выбранное из генеральной совокупности (равно как и процедура по их определению) именуется **выборкой**.
- **Контур выборки** – это наличие полного *списка* всех элементов генеральной совокупности.
- Контур выборки неизбежно содержит ошибку, называемую **ошибкой контура выборки** и характеризующую степень отклонения от истинных размеров совокупности.

Пример: Очевидно, что не существует полного официального списка всех автосервисных мастерских города Москвы, включая полулегальный и нелегальный бизнес в данной области. Исследователь должен информировать заказчика работы о размерах ошибки контура выборки.



Итак, основные понятия:

- **Генеральная совокупность** – вся группа, про которую необходимо собрать информацию;
- **Выборка** – часть совокупности, базовый уровень исследования;
- **Ошибка выборки** – разница между полученными данными и фактическими по совокупности;
- **Контур выборки** – все единицы совокупности, из которых будет формироваться выборка;
- **Ошибка контура выборки** – степень отклонения контура выборки от генеральной совокупности;
- **Единица выборки** – объект исследования

Три главные проблемы формирования выборки

- Прежде всего, исходя из сути рассматриваемой задачи, необходимо определить, кто является **единицей выборки**.
- Очень важным является также определение **контура выборки**.
- В целях выполнения правила репрезентативности проводимого исследования необходимо обратить внимание на **метод, с помощью которого выбираются единицы выборки**
- И наконец, необходимо решить вопрос об **объеме выборки**, который определяет число изучаемых единиц выборки.

Репрезентативность выборки

- Цель отбора объектов в состав выборки всегда состоит в получении такой совокупности, которая по своим качествам репрезентирует ту, которую мы намереваемся изучить
- **Под репрезентативностью в маркетинговых исследованиях понимают такие свойства выборки, которые позволяют ей выступать на момент опроса моделью, представителем генеральной совокупности.**
- Важно понимать, что репрезентативность выборки и ошибка выборки – разные явления.
Репрезентативность, в отличие от ошибки никак не зависит от размера выборки

Репрезентативность выборки

- Одна и та же выборка может быть репрезентативной и нерепрезентативной для разных генеральных совокупностей
- **Пример:**
 - Выборка, целиком состоящая из москвичей, владеющих автомобилем, не репрезентирует все население Москвы.
 - Выборка из российских предприятий численностью до 100 человек не репрезентирует все предприятия России.
 - Выборка из москвичей, совершающих покупки на рынке, не репрезентирует покупательское поведение всех москвичей.

Репрезентативность выборки

- В то же время, указанные выборки (при соблюдении прочих условий) могут отлично репрезентировать москвичей-автовладельцев, небольшие и средние российские предприятия и покупателей, совершающих покупки на рынках соответственно.
- Как бы мы не увеличивали количество опрошенных москвичей-автовладельцев, мы не сможем репрезентировать этой выборкой всех москвичей.

Группы выборочных методов (типы выборки)

- **Вероятностная выборка (случайная)** – это такая выборка, для которой каждый элемент генеральной совокупности имеет определенную, заранее заданную вероятность быть отобранным.
- Это позволяет исследователю рассчитать, насколько правильно выборка отражает популяцию, из которой она спроектирована (рассчитать ошибку выборки).

Неслучайная (невероятностная) выборка

Еще одно название -детерминированная

- **Неслучайная (невероятностная) выборка** – это такой способ отбора единиц, при котором мы не можем заранее рассчитать вероятность каждого элемента попасть в состав выборочной совокупности.
- Данный тип выборки не дает возможности рассчитать, насколько правильна (репрезентативна) выборка
- Иногда по условиям исследования оказывается единственно возможным провести неслучайную выборку.



Метод простого случайного отбора

- **Метод простого случайного отбора** предполагает, что вероятность быть избранным в выборку известна и является одинаковой для всех единиц совокупности. Для этого используется:

- Лотерейный метод (*жребия*)
- Таблицы случайных чисел

Главное условие осуществления вероятностной выборки – наличие полного *списка* (контура) всех элементов генеральной совокупности.

- Использование простого случайного отбора гарантирует, что каждая единица совокупности известна и имеет равные шансы быть включенной в выборку.

Лотерейный метод (*жребия*)

- При использовании лотерейного метода (или метода *жребия*) жетоны с номерами всех элементов помещают в урну, тщательно перемешивают и извлекают последовательно n жетонов, где n – число элементов выборочной совокупности.
- Элементы генеральной совокупности, имеющие номера оказавшиеся на извлеченных жетонах, будут составлять выборочную совокупность.

Фрагмент таблицы случайных цифр

37	52	35	15	04	80	44	83
21	28	31	42	46	72	13	20
07	99	95	64	13	06	27	29
10	16	69	93	39	08	09	38
00	84	65	56	09	29	67	19

Комментарий к использованию таблицы случайных цифр

- Если мы имеем, скажем, популяцию (то есть генеральную совокупность) из **1507 элементов** и хотим спроектировать выборку численностью **150 элементов**, мы можем выбирать любые *два смежных столбца* в таблице случайных чисел: цифры, стоящие в двух смежных ячейках будут образовывать четырехзначное число.
- Каждый раз, когда будет появляться число от 0001 до 1507, мы будем считать, что оно обозначает номер отбираемого элемента. Когда число появляется более чем один раз, этот номер игнорируется после первого раза.
- Если мы начнем с первых четырех столбцов в таблице, спускаясь по столбцам, то в выборку будут включены элементы под номерами **0799, 1016, 0084, 480, 1306, 929, 1320 и 938**.
- Поскольку мы не стремимся умышленно отыскать определенное число, мы можем начать с любого места таблицы и использовать любую систему для движения по таблице.
- С тем же успехом случайные числа могут генерироваться специальной программой компьютера.

Метод *систематической* (или *механической*) выборки

- На практике чаще всего используют метод *систематической* (или *механической*) выборки, когда из пронумерованного списка через равные интервалы k отбирается заданное число респондентов. При этом *шаг выборки k (интервал скачка)* рассчитывается по простой формуле:

$$k = \frac{N}{n}$$

где N – численность генеральной совокупности
 n – численность выборочной совокупности.

Пример *систематической* (или *механической*) выборки

- Предположим, что нам нужно спроектировать выборку численностью **100** из списка **5000 студентов** какого-то вуза.
- Если мы намерены использовать систематическую выборку, то должны вначале рассчитать интервал выборки делением числа элементов в списке на размер выборки.
- В данном случае, разделив общую численность студентов 5000 на требуемый размер выборки 100 единиц, мы получим **интервал (шаг) выборки 50**.
- Далее будем систематически двигаться по списку и отбирать каждого пятидесятого студента (отобрав таким образом 100 имен).
- Определение того места в списке, с которого мы начнем, проводится случайным образом, по **таблице случайных чисел** (это называется *случайный старт*).
- Таким образом, если случайно выбрана точка старта под номером 31, то в выборку будут включены студенты, стоящие под номерами 31, 81, 131, 181 и т.д.

Метод кластерного отбора. (гнездовая или серийная выборка)

- **Метод кластерного отбора основан на делении совокупности на подгруппы, каждая из которых представляет совокупность в целом.**
- Здесь отбираются не объекты исследования (например, люди), а группы. Группы отбираются случайным образом, а внутри них проводится сплошной опрос. Например, в ВУЗе с большим количеством студенческих групп отбор можно проводить путем случайного отбора этих групп и дальнейшего сплошного опроса в этих группах

Пример кластерного отбора

- Предположим, что исследуется мнение населения какого-то региона относительно марки какого-то товара. Регион разбивается на четко определяемые части (кластеры), например области.
- Исследователь может считать, что выделенные кластеры являются идентичными и мнение населения этих областей характерно для региона в целом.
- Далее одна из областей (один кластер) выбирается случайным образом, определяется совокупность для этой области, в ней проводится соответствующее исследование, а выводы обобщаются на совокупность всего региона

Причины появления стратифицированного отбора (типической или районированной выборки)

- В основе всех вероятностных методов лежит предположение, что любая совокупность характеризуется **симметричным распределением ее ключевых характеристик**. Другими словами, каждая выборка достаточно полно характеризует всю совокупность, различные крайности в выборке уравнивают друг друга.
- Но такая ситуация на практике встречается крайне редко. Скажем, исследуется рыночный потенциал определенного региона для какого-то товара. Население больших, средних и малых городов, сельской местности данного региона отличается по уровню образования, дохода, образу жизни и т.п. В таком случае необходим **стратифицированный отбор**

Понятие стратифицированной выборки

- При наличии в составе генеральной совокупности различных типов явлений с разными уровнями изучаемых признаков организуют стратифицированную выборку, чтобы обеспечить более равномерное представительство в выборочной совокупности различных частей (типов) явления.
- Генеральная совокупность разделяется на различные подгруппы (страты), например по уровню доходов, и выборки формируются из этих подгрупп, по сути дела являющихся сегментами рынка.

Условия осуществления стратифицированной выборки

- При использовании данного метода прежде всего следует выбрать некоторую наблюдаемую характеристику (признак), характеризующую каждую единицу совокупности, например уровень дохода. Далее для каждой страты с помощью **случайного отбора** формируется выборка.
- **Стратифицированная выборка может быть применена лишь при наличии дополнительной информации о генеральной совокупности** (например, нам необходимо процентное соотношение мужчин и женщин, в случае, если мы хотим стратифицировать выборку по полу).
- Отсутствие такой информации делает применение стратифицированной выборки невозможным

Методы (типы) невероятностной выборки

- Отбор в такой выборке осуществляется не по принципам случайности, а по субъективным критериям – доступности, типичности, равного представительства и т.д.
- При применении *невероятностных методов* отбора формирование выборки осуществляется без использования понятий теории вероятностей, вследствие чего невозможно рассчитать вероятность включения в выборку единицы совокупности.

Нерепрезентативный отбор.

Метод стихийной выборки: отбор на основе удобства и доступности

- **Смысл метода заключается в том, что формирование выборки осуществляется самым удобным с позиций исследователя образом, например, с позиций минимальных затрат времени и усилий, с позиции доступности респондентов.**
- Выбор места исследования и состава выборки производится субъективным образом, например, опрос покупателей осуществляется в магазине, ближайшем к месту жительства исследователя.
- Очевидно, что многие представители совокупности не принимают участия в опросе.

Метод экспертного отбора (на основе суждения)

- Метод основан на использовании мнения квалифицированных специалистов, экспертов относительно состава выборки. На основе такого подхода часто формируется состав фокус-группы.
- Например, изучается потребность в издании кулинарной книги для приготовления в микроволновой печи низкокалорийной пищи. На основе мнения исследователей в состав фокус-группы было включено 12 домохозяек; шестеро из них владело микроволновыми печами более трех лет, трое — менее трех лет и трое только собирались их купить.

Метод квотной выборки

- Формирование выборки на основе квот (квотный отбор) предполагает предварительное, исходя из целей исследования, определение состава и численности групп респондентов, отвечающих определенным требованиям (признакам).
- Например, мужчины в возрасте 20-30 лет, 31-45 лет и 46-60 лет; лица с доходом до 30 тысяч рублей, с доходом от 30 до 60 тысяч рублей и с доходом свыше 60 тысяч рублей)

Метод квотной выборки

- Для каждой группы задается количество объектов, которые должны попасть в выборку и быть обследованы.
- Количество объектов, которые должны попасть в каждую из групп, задается, чаще всего, либо пропорционально заранее известной доле группы в генеральной совокупности, либо одинаковым для каждой группы.
- Внутри групп объекты отбираются произвольно.
- Квотные выборки используются в маркетинговых исследованиях достаточно часто

Пример квотной выборки

- Приведем пример расчета выборки по таким параметрам квоты как пол и возраст.
- В одном из маркетинговых исследований при опросе населения, исследователи опирались на данные Областного комитета по статистике о половозрастной структуре населения районов города.
- Так, половозрастная структура населения Советского района Нижнего Новгорода выглядела следующим образом:

	Генеральная совокупность						Выборочная совокупность		
	численность			Доля, %			число		
	Всего	муж.	жен.	муж.	жен.	всего	муж.	жен.	всего
Основа выборки	128546	55757	72789	43,4	56,6	100,00	521	679	1200
Возраст (лет)									
20 - 24	11958	6009	5949	50,3	49,7	9,30	56	56	112
25 - 29	11858	6520	5338	55,0	45,0	9,22	61	50	111
30 - 34	12093	5861	6232	48,5	51,5	9,41	55	58	113
35 - 39	14360	6795	7565	47,3	52,7	11,17	63	71	134
40 - 44	14112	6598	7514	46,8	53,2	10,98	62	70	132
45-- 49	12779	5840	6939	45,7	54,3	9,94	55	65	119
50 - 54	7274	3222	4052	44,3	55,7	5,66	30	38	68
55 - 59	11981	5035	6946	42,0	58,0	9,32	47	65	112
60 - 69	17503	6828	10675	39,0	61,0	13,62	64	100	163
70 и старше	14628	3049	11579	20,8	79,2	11,38	28	108	137

Пример квотной выборки

- Изучая данные таблицы, нетрудно убедиться, что численность каждой из половозрастных групп в трех последних столбцах таблицы (объеме выборки) пропорциональна численности каждой из соответствующих групп в генеральной совокупности. Такая выборка называется **пропорциональной**.
- Иногда используются и **непропорциональные** выборки, если это необходимо исходя из целей исследования
- Таким образом на основе определенных заданных параметров создается **микромодель генеральной совокупности**

Метод снежного кома

- Формирование выборки в процессе опроса основано на расширении числа опрашиваемых на основе предложений респондентов, которые уже приняли участие в исследовании.
- У каждого респондента, начиная с первого, просят контакты его друзей, коллег, знакомых, которые подходили бы под условия отбора и могли бы принять участие в исследовании.
- Таким образом, за исключением первого шага, выборка формируется с участием самих объектов исследования.
- Метод часто применяется, когда необходимо найти и опросить труднодоступные группы респондентов (например, респондентов, имеющих высокий доход, респондентов, принадлежащих к одной профессиональной группе, респондентов, имеющих какие-либо схожие хобби/увлечения и т.д.)

Комбинация типов выборки при проведении исследования: **многоступенчатая выборка**

- **Многоступенчатая выборка** – строится с применением процедуры поэтапного отбора объектов опроса.
- При этом совокупность объектов, отобранных на предыдущем этапе (ступени), становится исходной для отбора на следующем.
- Промежуточные объекты, составляющие выборочную совокупность на высших ступенях, называют **единицами отбора**.
- Объекты самой нижней ступени, с которых ведется непосредственный сбор информации, называются **единицами наблюдения**.

Пример многоступенчатой выборки

Так, формирование выборки для оценки мнения населения России относительно различных марок какого-либо товара, например холодильника, может осуществляться по следующей схеме:

1. Россия, согласно ее административно - территориальному делению, делится на 89 субъектов Федерации.
2. Осуществляется **случайным образом** выбор девяти субъектов (исключая Москву и С.-Петербург).
3. Все населенные пункты (исключая Москву и С.-Петербург) в зависимости от численности населения делятся на шесть групп: до 1000 жителей, от 1000 до 10 000 (поселки), от 10 000 до 100 000, от 100 000 до 500 000, свыше 500 000 (города). Москва и С.-Петербург исследуются отдельно.
4. На основе **квотного метода** устанавливается, что для каждого из девяти выбранных субъектов Федерации исследуются три города и два поселка

Пример многоступенчатой выборки

Для этого берется каждый пятый населенный пункт из справочника Госкомстата «Численность населения Российской Федерации по городам, рабочим поселкам и районам» и определяется, в какую из категорий он попадет (**систематический отбор**). Так поступают до тех пор, пока не наберется нужное количество населенных пунктов. Если, дойдя до конца списка, не будет найдено квотное число населенных пунктов, процесс выбора населенных пунктов начинается сначала, но берется каждый четвертый населенный пункт, и т.д.

5. Выбирается **случайным образом** в каждом выбранном городе и поселке по меньшей мере 20 респондентов.

Таким образом, кроме Москвы и С.-Петербурга должно быть проинтервьюировано 900 респондентов. В Москве и С.-Петербурге опрашивается по 50 респондентов. В итоге должно быть проинтервьюировано **1000 респондентов**.

Этапы разработки выборочного плана

1. Определение соответствующей совокупности согласно поставленным целям исследования.
2. Получение «списка» совокупности для установления контура выборки.
3. Проектирование выборочного плана – определение состава и объема выборки.
4. Определение методов доступа к совокупности.
5. Достижение нужной численности выборки.
6. Проверка выборки на соответствие требованиям.
7. В случае необходимости формирование новой выборки.

Определение объема выборки

Исходным пунктом расчета объема выборки становится вопрос об определении размеров выборочной совокупности.

- Численность выборочной совокупности зависит от двух факторов:
 - (1) стоимости сбора информации
 - (2) стремления к определенной степени статистической достоверности результатов, которую надеется получить исследователь.
- Чем больше размеры выборки, то есть чем ближе они к размерам генеральной совокупности в целом, тем более надежны и достоверны полученные данные.

Определение объема выборки

- Самый спорный вопрос по выборочному плану.
- На практике, решение о объеме выборки всегда есть **компромисс между точностью результатов исследования и затратами на проведение исследования.**
- Способ определения объема выборки зависит от того, имеется ли у исследователя информация о составе генеральной совокупности

Связь между объемом выборки и репрезентативностью

- Следует отметить, что **объем выборки никак не влияет на репрезентативность** полученных результатов.
- Пример: Предположим, например, что в целях изучения степени использования в России персональных компьютеров в научной работе проводился опрос на основе принципа удобства на одном из московских перекрестков. И хотя было опрошено 5000 респондентов, полученные результаты не являются репрезентативными даже для Москвы.
- Это обусловлено тем, что был использован метод формирования выборки, который в данном случае применять было нельзя.

Методы определения объема выборки

Если исходная информация о генеральной совокупности отсутствует, то используется:

- **Правило большого пальца** – бездоказательно принимается, что для получения нужных результатов выборка должна составлять, например, 5% от генеральной совокупности
- Исходя из заранее оговоренных **с заказчиком условий** (он рекомендует цифры выборки)

Методы определения объема выборки

Если исходная информация о генеральной совокупности отсутствует, то используется:

- Многолетний **опыт социологов-практиков**. Он свидетельствует о том, что при массовых опросах, если величина генеральной совокупности составляет менее 5000, то достаточный объем выборочной совокупности должен быть не менее 500 человек.
- Если величина генеральной совокупности составляет более 5000, то для получения статистически значимых результатов нужно обследовать 10% ее состава, но не более 2000-2500 единиц.

Зависимость объема выборки от ее ошибки

(размер генеральной совокупности составляет 20 тыс. единиц)

Ошибка выборки, %	2	3	4	5	6	7	8	9	10	12	14	17	20
Объем выборки, ед.	2500	1100	620	400	280	200	160	110	100	67	50	30	25

Зависимость объема выборки от ее ошибки

В.И. Паниотто приводит следующие расчеты репрезентативной выборки с допущением 5-процентной ошибки:

Объем генеральной совокупности	500	1000	2000	3000	4000	5000	10000	100000
Объем выборки	222	286	333	350	360	370	385	358

Комментарии к таблице

- Суть данных, приведенных в этой таблице, состоит в следующем. Если вы, опросив, предположим, **400** человек в районном городе, где численность взрослого платежеспособного населения составляет 100 тысяч человек, выявили, что **33 процента** опрошенных покупателей предпочитают продукцию местного мясокомбината, то с **95-процентной степенью вероятности** можете утверждать, что **постоянными покупателями этой продукции являются $33 \pm 5\%$ (то есть от 28 до 38 процентов) жителей этого города.**
- Уровень уверенности в 95% – это условие, оно показывает 95% уверенности в том, что параметр популяции попадает в пределы допуска ошибки выборочной статистики.

Методы определения объема выборки

Если исходная информация о генеральной совокупности имеется, то:

- Наиболее теоретически обоснованный и корректный подход к определению объемов выборки основан на расчете **доверительных интервалов.**
- Метод требует понимания некоторых базовых понятий математической статистики (закон нормального распределения, меры вариации, среднего квадратического отклонения, доверительного интервала и т. д.)

Понятие вариации

- **Понятие вариации характеризует величину несхожести (схожести) ответов респондентов на определенный вопрос.**
- В более строгом плане вариацией значений какого-либо признака в совокупности называется различие его значений у разных единиц данной совокупности в один и тот же период или момент времени.
- Результаты ответов на вопросы опроса обычно представляются в форме **кривой распределения**.
- При высокой схожести ответов говорят о **малой вариации** (узкая кривая распределения) и при низкой схожести ответов — о **высокой вариации** (широкая кривая распределения).

Кривые распределения результатов ответа на вопрос: «Сколько миль за год проходит ваш автомобиль?»

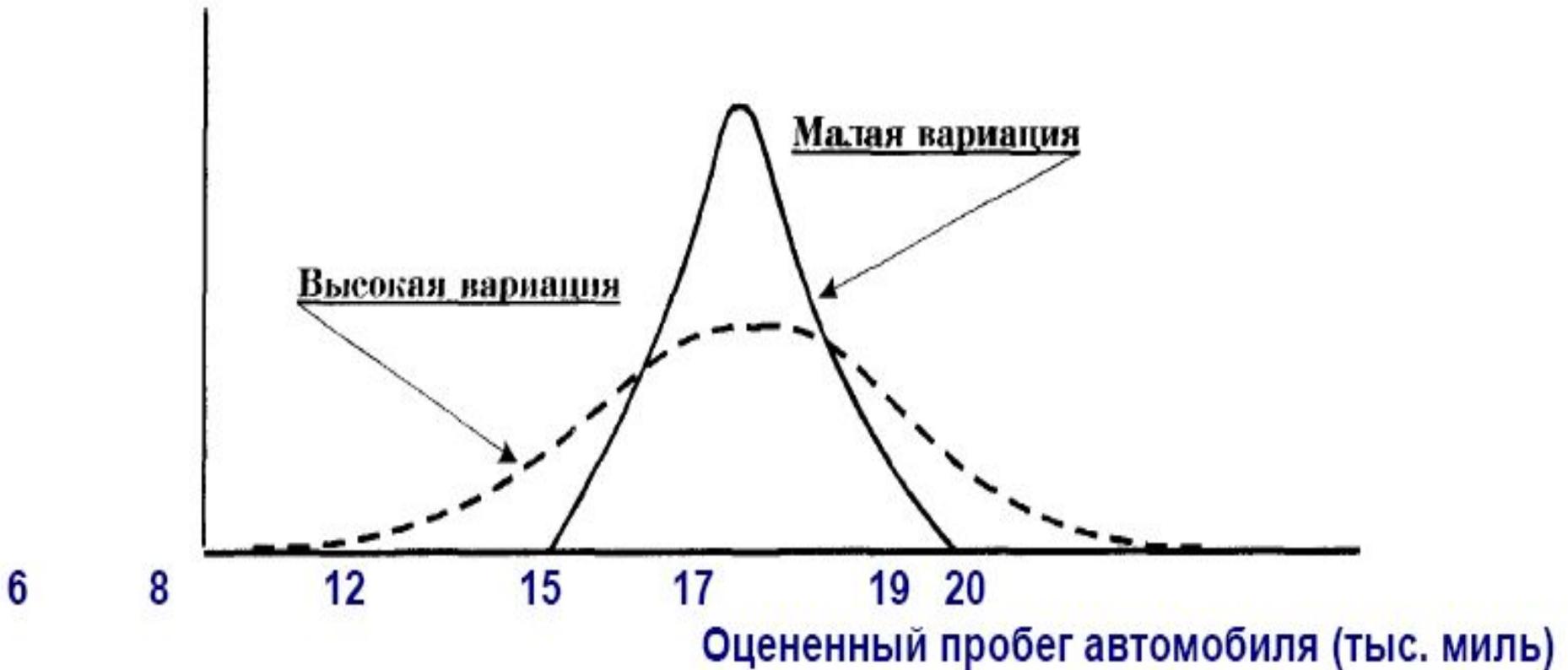


Рис. 4.5. Вариация и кривые распределения

Среднее квадратическое отклонение

- В качестве меры вариации обычно принимается **среднее квадратическое отклонение**, которое характеризует среднее расстояние от средней оценки ответов каждого респондента на **определенный вопрос**.
- Можно сравнить среднее квадратическое отклонения для двух выборок и определить, для какой из них вариация является меньшей.
- Поскольку все маркетинговые решения принимаются в условиях неопределенности, то это обстоятельство целесообразно учесть при определении объема выборки.

«Доверительный интервал»

- Так как определение исследуемых величин для совокупности в целом осуществляется на основе выборочной статистики, то следует установить диапазон (доверительный интервал), в который, как ожидается, попадут оценки для совокупности в целом, и ошибку их определения.
- **Понятие «доверительный интервал» — это диапазон, крайним точкам которого соответствует определенный процент определенных ответов на какой-то вопрос.**
- Данное понятие тесно связано с понятием «среднее квадратическое отклонение изучаемого признака в генеральной совокупности»: чем оно больше, тем шире должен быть доверительный интервал, чтобы включить в свой состав, например, 95% ответов.

Пример

- Проведено исследование числа визитов автовладельцев в сервисные мастерские за год. Доверительный интервал для среднего числа визитов был рассчитан равным 5—7 визитам при 99%-ном уровне доверительности (**надежности выборки**).
- Это означает, что если появится возможность провести независимо 100 раз выборочные исследования, то для 99 средних значений числа визитов попадут в диапазон от 5 до 7 визитов — другими словами, 99% автовладельцев попадут в доверительный интервал. (в 99 случаях результат будет находиться в указанных границах)

Важно понимать!

- Выборка всегда дает более относительный результат по сравнению с генеральной совокупностью.
- Всегда есть вероятность отклонения, которая считается ошибкой выборки (измеряется в процентах)

Чем больше объем выборки, тем меньше ошибка

Средняя квадратическая ошибка выборки

- Индикатором степени отличия оценки, истинной для совокупности в целом, от оценки, которая ожидается для типичной выборки, является **средняя квадратическая ошибка выборки**.
- Например, исследуется мнение потребителей о новом продукте и заказчик данного исследования указал, что его устроит **точность полученных результатов, равная $\pm 5\%$ (ошибка выборки)**
- Предположим, что 30% членов выборки высказалось за новый продукт. Это означает, что диапазон возможных оценок для всей совокупности составляет 25—35%.

Определение объема выборки на основе расчета доверительного интервала

Исходной информацией, необходимой для реализации данного подхода, является:

- Величина вариации, которой, как считается, обладает совокупность.
- Желаемая точность.
- Уровень доверительности, которому должны удовлетворять результаты проводимого обследования.

Когда на заданный в анкете вопрос существует только два варианта ответа, выраженные в процентах (используется процентная мера), объем выборки определяется по следующей формуле:

$$n = \frac{z^2 pq}{e^2}, \quad (4.1)$$

где n — объем выборки;

z — нормированное отклонение, определяемое исходя из выбранного уровня доверительности (табл. 4.15);

p — найденная вариация для выборки;

$q = (100 - p)$;

e — допустимая ошибка.

Таблица 4.15

Значение нормированного отклонения оценки (z) от среднего значения в зависимости от доверительной вероятности (α) полученного результата

$\alpha, \%$	60	70	80	85	90	95	97	99	99,7
z	0,84	1,03	1,29	1,44	1,65	1,96	2,18	2,58	3,0

Пример расчета объема выборки

- Фирмой, выпускающей покрышки, проводится опрос автолюбителей. Целью обследования является определение процента автолюбителей, использующих радиальные покрышки, поэтому на вопрос: «Используете ли вы радиальные покрышки?» возможно только два ответа: «Да» или «Нет» (шкала наименований).
- Если предположить, что совокупность автолюбителей обладает низким показателем вариации, то это означает, что почти каждый опрошенный использует радиальные покрышки. В этом случае может быть сформирована выборка достаточно малых размеров. В формуле (4.1) **произведение pq выражает вариацию, свойственную совокупности.**

Пример расчета объема выборки

- Предположим, что 90% единиц совокупности используют радиальные покрышки. Это означает, что $pq = 900$. Если принять, что показатель вариации выше ($p = 70\%$), то $pq = 2100$.
- Наибольшая вариация достигается в случае, когда половина совокупности (50%) используют радиальные покрышки, а другая (50%) — не использует. В этом случае произведение pq достигает наибольшего значения, равного 2500

Пример расчета объема выборки

- При проведении исследования следует указать **точность полученных оценок (точность выборки)**.
- Скажем, в результате исследования было установлено, что 44% респондентов используют радиальные покрышки.
- В этом случае результаты измерения желательно представить в виде: «Процент автолюбителей, использующих радиальные покрышки, составляет 44% плюс-минус ...%».

Пример расчета объема выборки

- Что касается **уровня доверительности**, то при проведении маркетинговых исследований, обычно рассматриваются только два его значения: **95%** или **99%**.
- Первому значению соответствует значение $z = 1,96$, второму — $z = 2,58$.
- Если выбирается уровень доверительности, равный 99%, то это говорит о том, что мы уверены на 99% (другими словами, доверительная вероятность равна 0,99) в том, что процент членов совокупности, попавших в диапазон $\pm e\%$, равен проценту членов выборки, попавших в тот же диапазон ошибки.

Пример расчета объема выборки

- Принимая вариацию равную 50%, точность, равную $\pm 10\%$ при 95%-ном уровне доверительности, рассчитаем размер выборки:
- При уровне доверительности, равном 99%, и $e = \pm 3\%$
 $n = 1067$.

$$n = \frac{1,96^2 (50 \cdot 50)}{10^2} = 96$$

Важно понимать!

- Приведенные формулы расчета объема выборки основаны на предположении, что все правила формирования выборки были соблюдены и единственной ошибкой выборки является ошибка, обусловленная ее объемом.
- Однако следует помнить, что объем выборки определяет точность полученных результатов, но не их **представительность**, которая определяется методом формирования выборки.
- Все формулы для расчета объема выборки предполагают, что репрезентативность гарантируется использованием корректных процедур формирования выборки.

Систематическая ошибка выборки

- Систематическая ошибка зависит от различных факторов, оказывающих постоянное воздействие на исследование и смещающих результаты исследования в определенную сторону.

Пример:

- Использование любых вероятностных выборок занижает долю людей с высоким доходом, ведущих активный образ жизни. Происходит это в силу того, что таких людей гораздо сложнее застать в каком-либо определенном месте (например, дома).
- Проблема респондентов, отказывающихся отвечать на вопросы анкеты (доля «отказников» в Москве, для разных опросов, колеблется от 50% до 80%)

Зависимость ошибки выборки от степени различий между единицами генеральной совокупности

- Чем больше различия (или гетерогенности) внутри генеральной совокупности, тем больше величина возможной ошибки выборки

Например, если нам нужно узнать, какое количество пива потребляется, то мы обнаружим, что внутри нашей генеральной совокупности нормы потребления у различных людей существенно различаются (*гетерогенная генеральная совокупность*).

- В другом случае мы будем изучать потребление хлеба и установим, что у разных людей оно различается гораздо менее существенно (*гомогенная генеральная совокупность*).

Важно понимать!

- Помимо четкого планирования репрезентативности выборки, **нельзя распространять полученные результаты за ее границы**. Так, результаты исследования мнения массового потребителя города Москвы о товарах определенной фирмы нельзя распространять на всю Россию.
- Можно быть поставленным в тупик разными результатами обследования степени лояльности потенциальных покупателей к определенной марке пылесоса (в одном исследовании была названа цифра 10%, в другом случае — 25%).
- Дело в том, что в первом случае цифра была получена от общего числа опрошенных, а во втором случае — только от числа тех покупателей, которые твердо решили приобрести пылесос.

Из этого следует:

- Для вдумчивого маркетолога очень важными являются те пояснения, которые сопровождают полученные данные (как минимум, формулировки вопросов и описание выборки)

Итак:

- Объем выборки не влияет на репрезентативность результатов (влияет метод ее формирования)
- Чем больше величина выборки, тем меньше ее возможная ошибка
- Чем больше различий (гетерогенность) внутри генеральной совокупности, тем больше величина возможной ошибки выборки
- Решение об объеме выборки является компромиссом между точностью результатов исследования и возможностями его практической реализации.

Спасибо за внимание!

