Основные понятия

Выборочный метод в социологии

Выборочное обследование представляет собой способ систематического сбора данных о поведении и установках людей посредством исследования специально подобранной группы респондентов, дающих информацию о себе и своем мнении.

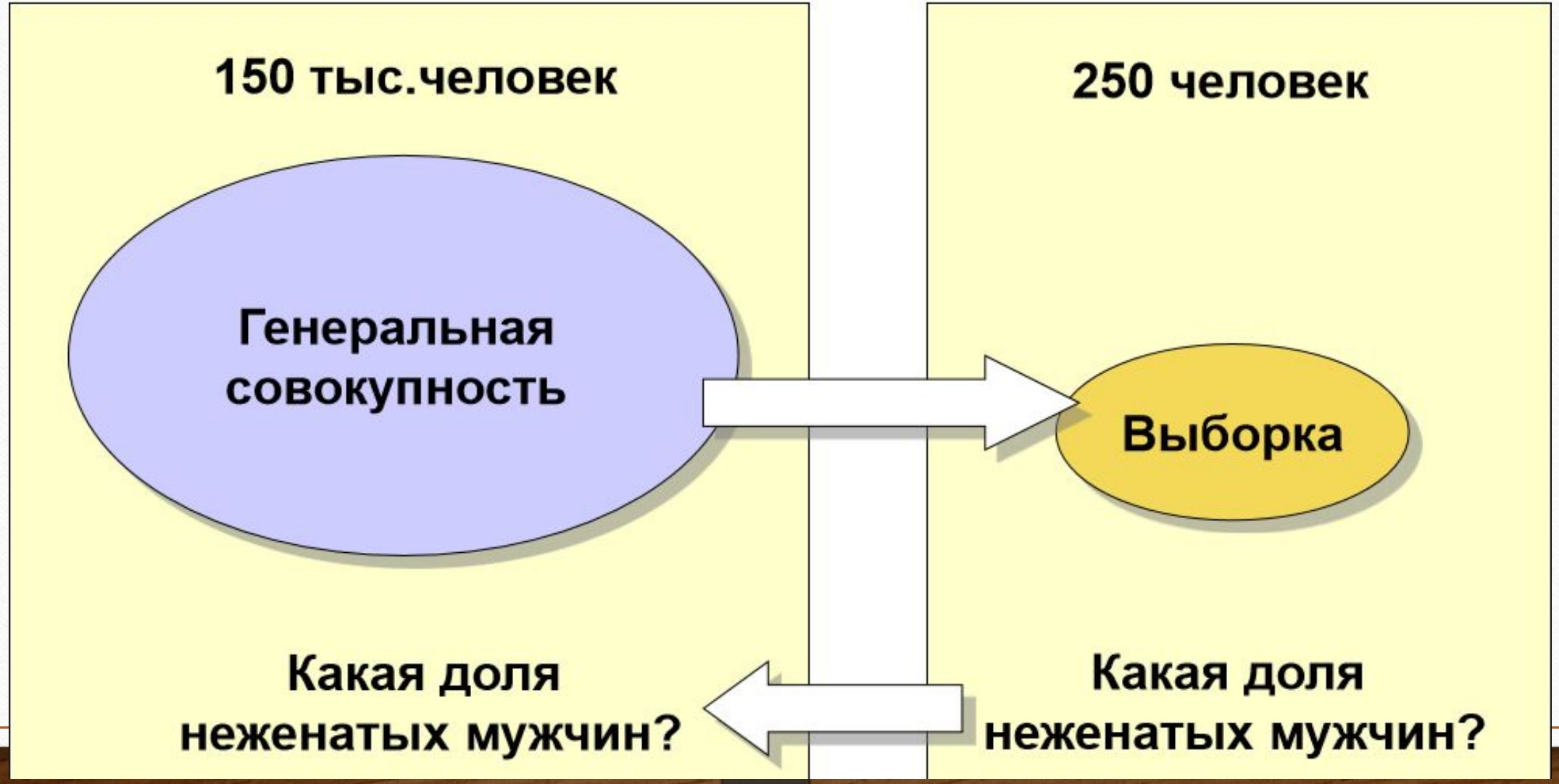
Проектирование выборки

Под данными имеются в виду значения переменных, приписанные единицам исследования – объектам.
Объекты – сообщества, институции, люди, тексты, вещи.
Проектируемый объект – это совокупность доступных исследователю единиц.

Генеральная совокупность и выборка

- Генеральная совокупность – вся интересующая исследователя совокупность изучаемых объектов.
- Выборка, **выборочная совокупность** – некоторая, обычно небольшая, часть генеральной совокупности, отбираемая специальным образом и исследуемая с целью получения выводов о свойствах генеральной совокупности.

Генеральная совокупность и выборка



Репрезентативная выборка

Репрезентативная выборка хорошо представляет генеральную совокупность.

Это означает, что каждое свойство (или комбинация свойств) наблюдается в выборке с той же частотой, что и в генеральной совокупности.
Под репрезентативностью в статистике понимают такие свойства выборки, которые позволяют ей выступать на момент опроса моделью, представителем генеральной совокупности.
Отсюда и само происхождение этого термина: *representative* (англ.) — представитель, делегат, уполномоченный.

Основное правило составления выборки гласит:

Каждый элемент генеральной совокупности должен иметь одинаковые шансы попасть в выборку

Типами выборки называются основные разновидности статистической выборки

Тип выборки говорит о том, как люди попадают в выборочную совокупность, **объем выборки** сообщает о том, какое их количество туда попало.

Основа выборки - полный и точный перечень единиц выборочной совокупности.

Элементы, предназначенные для отбора, называются **единицами отбора.**

Элемент генеральной совокупности, с которого непосредственно ведется сбор информации, называется **единицей наблюдения.** Обычно единица наблюдения — это отдельный человек.

Структура выборки — процентные пропорции признаков объекта, на основании которых составляется выборочная совокупность

- Если в генеральной совокупности, например, 30% молодежи, 50% людей среднего возраста и 20% пожилых, то и в выборочной совокупности должны соблюдаться те же самые процентные пропорции трех возрастов
- . Для каждого признака устанавливаются процентные пропорции в генеральной и выборочной совокупностях. Расхождение структуры двух совокупностей ведет к ошибке репрезентативности

Составление выборки

- — процесс выбора единиц (людей, организаций и т.п.) из интересующей исследователя генеральной совокупности с таким условием, чтобы, проведя свое исследование, ученый мог обобщить полученные результаты снова на всю генеральную совокупность, из которой осуществлялась выборка.

Репрезентативность выборки зависит от целого ряда условий и, прежде всего, от того, как она осуществляется :

- планомерно (т.е. по заранее намеченной схеме),
- путем непланомерного отбора вариантов из генеральной совокупности.

В любом случае выборка должна быть типичной и вполне объективной.

- Доброкачественная выборка должна быть объективной, т. е. производиться без предвзятых побуждений, при исключении субъективных влияний на ее состав. Выполнению этого условия репрезентативности отвечает принцип рандомизации (от англ. rendom—случай), или случайного отбора вариант из генеральной совокупности.
- Этот принцип положен в основу теории выборочного метода и должен соблюдаться во всех случаях образования репрезентативной выборочной совокупности, не исключая и случаев планомерного или преднамеренного отбора.

- **В статистической науке** в зависимости от способа отбора различают выборки следующих типов:

- случайная выборка с возвратом;
- случайная выборка без возврата;
- механическая;
- типическая;
- серийная.

Простой случайный отбор

Простой случайный отбор предполагает, что вероятность быть включенным в выборку известна и является одинаковой для всех единиц совокупности.

Простой случайный отбор реализуется двумя методами:

- отбор вслепую (другое название — метод лотереи или жребия),
- отбор не вслепую (происходит с помощью таблицы случайных чисел).

Отбор вслепую

- В этом случае выделяют:
- случайный повторный (с возвращением) отбор,
- случайный бесповторный (без возвращения) отбор.

Вероятностную выборку целесообразно применять только при наличии соответствующих условий

- **Первое условие** осуществления вероятностной выборки — наличие полного списка всех элементов генеральной совокупности от I до **N,** где **N—** общее число всех элементов.
- **Второе условие** вероятностной выборки — хорошая перемешанность элементов генеральной совокупности

- **Два различных способа образования выборочной совокупности**.
- **Первый—** вынутая карточка после фиксации ее номера возвращается в пачку, после чего карточки снова тщательно перемешиваются. Выборочная совокупность, образованная по такой схеме, получила название **случайной возвратной выборки.**
- **Второй—** каждая вынутая карточка после ее записи обратно не возвращается.. Выборочную совокупность, образованную по данной схеме называют **случайной безвозвратной выборкой.**

Систематическая выборка (отбор) — процедура отбора каждого k-го элемента из списка элементов генеральной совокупности

- Примером служат разного рода квартирные выборки: выбираются улицы, на которых интервьюер проводит квартирный опрос. Квартиры выбираются по определенной схеме (крайняя квартира справа от лестницы на последнем этаже первого подъезда и т.д.).
- Процедура систематического отбора проста:
- количество единиц генеральной совокупности, предположим 2000 единиц выборки, делится на количество единиц выборочной совокупности, скажем 200, и определяется **шаг выборки.** Он предполагает, что, начиная с любого номера из списка, опрашивается каждый десятый (2000:200 = 10).

Шаг выборки к рассчитывается по простой формуле: $$k = \frac{N}{n}$$

где ***N*** — численность генеральной совокупности, ***n*** — численность выборочной совокупности.

Шаг выборки показывает, сколько номеров в списке фамилий людей, вошедших в генеральную совокупность, надо пропустить (через сколько перешагнуть), чтобы в итоге получить список выборочной совокупности. Буквально шаг выборки означает расстояние между соседними фамилиями респондентов, измеренное количеством отбракованных фамилий из списка генеральной совокупности

Районированная и стратифицированная выборки

- Если генеральная совокупность велика, то обследуемую совокупность разделяют на более или менее однородные части, а затем осуществлять отбор единиц внутри этих частей. Такую раздробленную на части выборку правильнее всего было бы называть расслоенной («районированной» и «стратифицированной»).
- Сложилась практика различения двух разновидностей расслоенной выборки. Если деление происходит по стратам (социальным группам), то выборку именуют **стратифицированной**, если по экономико-географическим районам, то **—районированной.**
- **Стратифицированная выборка** — вероятностная выборка, обеспечивающая равномерное представительство в выборочной совокупности различных частей, типов, групп и слоев населения.

σ^2

Стратифицированная случайная выборка (ССВ)

ССВ основана на выборке по каждой страте отдельно.
Это повышает точность результатов либо уменьшает время исследования, допуская меньшие размеры выборки при заданном уровне точности.

Основная цель всякого расслоения — повышение точности выборочных оценок. Слои выделяются таким образом, чтобы *дисперсия*** изучаемых переменных внутри слоев была значительно меньше, чем между ними.

Различают стратификацию одномерную и многомерную в зависимости от того, один или несколько признаков положены в основу разделения совокупности.

****Дисперсия*** - представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины

Гнездовая выборка

- **Гнездовая выборка** — вид выборки, при котором отбираемые объекты представляют собой группы или гнезда (кластеры) более мелких единиц.
- **Гнездом** называют единицу отбора высшей ступени, состоящую из более мелких единиц низшей ступени. В выборку могут быть включены как все единицы низшего уровня, так и их часть.
- Число единиц, образующих **гнездо**, называют его **размером**.

- **Проблемы**, которые возникают здесь, связаны с определением величины гнезда, количеством гнезд, которые надо обследовать, их размещением в генеральной совокупности.
- **Основные рекомендации:**
- **1.** различия между гнездами были бы по возможности более неоднородными.
- **2.** большое число малых гнезд предпочтительнее малого числа крупных гнезд

МЕТОДЫ НЕВЕРОЯТНОСТНОЙ (НЕСЛУЧАЙНОЙ) ВЫБОРКИ

Неслучайная выборка — это способ отбора единиц, при котором мы не можем заранее рассчитать вероятность попадания каждого элемента в состав выборочной совокупности, что, разумеется, не дает возможности рассчитать, насколько правильна (репрезентативна) выборка

Разновидности неслучайной выборки:
квотная выборка,
метод снежного кома,
метод основного массива,
метод стихийного отбора.

Выделяют два основных вида неслучайного отбора: направленный и стихийный отбор

Направленный отбор

- **Целевые выборки** — группы физических или юридических лиц, отвечающие одному или нескольким определенным критериям, устанавливаемым ученым при организации исследования.
- **выбор типичных объектов** (методов типичных представителей),
- **Метод «снежного кома»** разновидность целенаправленного выбора, при котором предполагается, что отбор дополнительных (последующих) респондентов производится после ссылки на них первоначально отобранных

Направленный отбор

- **Стихийные выборки**

формируются произвольно и часто независимо от самого исследователя. Примерами стихийного отбора могут служить опросы с помощью средств массовой информации, выборка «первого встречного», опросы пассажиров на остановках и в общественном транспорте и т.д.

- **Метод основного массива**

представляет опрос 60—70% генеральной совокупности. Опрашивается простое большинство. В результате средние генеральной и выборочной совокупностей сближаются, а выборочная совокупность составляет преимущественную часть генеральной и перекрывает возможное смещение.

Квотная выборка - микромодель объекта социологического исследования, формируемая на основе статистических сведений (параметров квот) преимущественно о социально-демографических характеристиках элементов генеральной совокупности

Многоступенчатая выборка

многоступенчатая выборка, построена с применением процедуры поэтапного отбора объектов опроса. При этом совокупность объектов, отобранных на предыдущем этапе (ступени), становится исходной для отбора на следующем. Соответственно различают единицы отбора первой ступени (первичные единицы), единицы отбора второй ступени (вторичные единицы) и т.д. Объекты самой нижней ступени, с которых ведется непосредственный сбор информации, называются единицами наблюдения.

При этом не стоит забывать: чем больше ступеней в многоступенчатом отборе, тем больше ошибка выборки. В любом случае при многоступенчатом отборе ошибка всегда больше, чем при простом случайном. И еще: на каждой ступени все равно применяется случайный отбор

РАСЧЕТ ОБЪЕМА ВЫБОРКИ

Объем выборки — общее число единиц наблюдения, включенных в выборочную совокупность

Однородной считается такая совокупность, в которой контролируемый признак распределен равномерно, не образует пустот или сгущений.

Процедура расчета объема выборки — цепь бесконечных компромиссов между стремлением к точности и ограниченностью ресурсов, дефицитом времени и неполнотой сведений об изучаемом явлении.

Зависимость объема выборки от объема генеральной совокупности

Объем генеральной совокупности	500	1000	2000	3000	4000	5000	10000	100000	Бесконечная
Объем выборки	222	286	333	350	360	370	385	398	400

Для расчета необходимого размера выборки в количественном исследовании чаще всего используют два статистических понятия — доверительный интервал и доверительную вероятность

Доверительный интервал представляет собой заранее задаваемую вами погрешность выборки. Например, если вы задаете доверительный интервал в 3% и конкретный ответ на конкретный вопрос исследования составит 48%, это значит, что даже при проведении опроса всей генеральной совокупности реальное значение попадет в интервал между 45 (48-3) и 51% (48 + 3).

Доверительная вероятность показывает, насколько вы можете быть уверены в полученных результатах, в том, что характеристики выборки соответствуют характеристикам всей генеральной совокупности — или с какой вероятностью случайный ответ попадет в доверительный интервал.

Обычно используют доверительную вероятность 95 и 99%. Чаще всего используется 95% — этого вполне достаточно в подавляющем большинстве исследований. Если объединить доверительную вероятность и доверительный интервал, то можно сказать, что ответы на вопрос с 95%-ной вероятностью попадут в интервал между 45 и 51%.

Ошибка выборки

- Предположим, N – генеральная совокупность, обладает среднестатистической характеристикой $x_{ген}$. Это может быть средний возраст, доход, доля правонарушений и количество наркоманов.
- Предположим далее, что из генеральной совокупности N отобрана совокупность n – выборочная, которая обладает средней характеристикой $x_{выб}$. Ясно, что n<N. Тогда задача заключается в сопоставлении $x_{ген}$ и $x_{выб}$
- Отклонение выборочной средней от генеральной средней называется ошибкой выборки:
- $$\Delta = | x_{выб} - x_{ген} |$$
- Из формулы следует, что исчисление Δ – ошибки выборки является принципиально неосуществимым, если не известны значения переменной в выборочной и генеральной совокупностях. Так возникает замкнутый круг: чтобы установить значение генеральной средней, нужно построить выборку, но, даже реализовав выборку, мы не сможем определить ее ошибку, потому что не знаем генеральной средней

Ошибка репрезентативности — мера отклонения статистической структуры выборки от структуры соответствующей генеральной совокупности.

Ошибка выборки обусловливается двумя факторами:

1. методом формирования выборки
2. размером выборки.

Ошибки выборки

СЛУЧАЙНЫЕ ОШИБКИ

- **Случайные ошибки** уменьшаются при возрастании объема выборочной совокупности

СИСТЕМАТИЧЕСКИЕ ОШИБКИ

- Это неконтролируемые перекосы в распределении выборочных наблюдений, которые приводят к «утере» проектируемого объекта исследования.
- В отличие от случайных систематические ошибки распределяются вокруг средней неравномерно, при возрастании объема выборки не уменьшаются.
- Систематические ошибки в отличие от случайных не поддаются предварительному контролю

Зависимость объема выборки от ее ошибки

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

ТИПИЧНЫЕ СИСТЕМАТИЧЕСКИЕ ОШИБКИ ВЫБОРКИ И СПОСОБЫ ОТБОРА ЕДИНИЦ

Первый тип и наиболее часто встречающаяся ошибка заключается в давлении доступных объектов.

Второй тип систематической ошибки связан с иллюзией постоянства.

четвертый тип систематических ошибок - недостаточный учет отсутствующих в месте сбора данных

Пятый тип систематических ошибок – отказы от ответа, которые в зависимости от темы опроса могут составлять довольно значительный процент запланированных интервью.