



Методы получения и обработки информации в прикладном исследовании



План лекции и логика изложения

- **Цель лекции:** Дать представление об особенностях применения выборочного метода в социологии. Познакомиться с основными подходами к получению, хранению и обработке информации в рамках прикладного исследования.
- **План лекции:**
 - Выборочный метод в социологическом исследовании.
 - ✓ Основные виды выборки, области применения и особенности. Погрешности в выборочном исследовании.
 - Структура и общие принципы разработки инструментария исследования.
 - ✓ Типы вопросов и типы измерительных шкал. Особенности различных видов шкал и области их применения.
 - Методы обработки социальной информации и программные средства для анализа данных.
 - ✓ Основные группы методов обработки первичной социальной информации. Программные средства анализа данных – SPSS, Statistica, Vortex, MS Excel.

Программа прикладного исследования



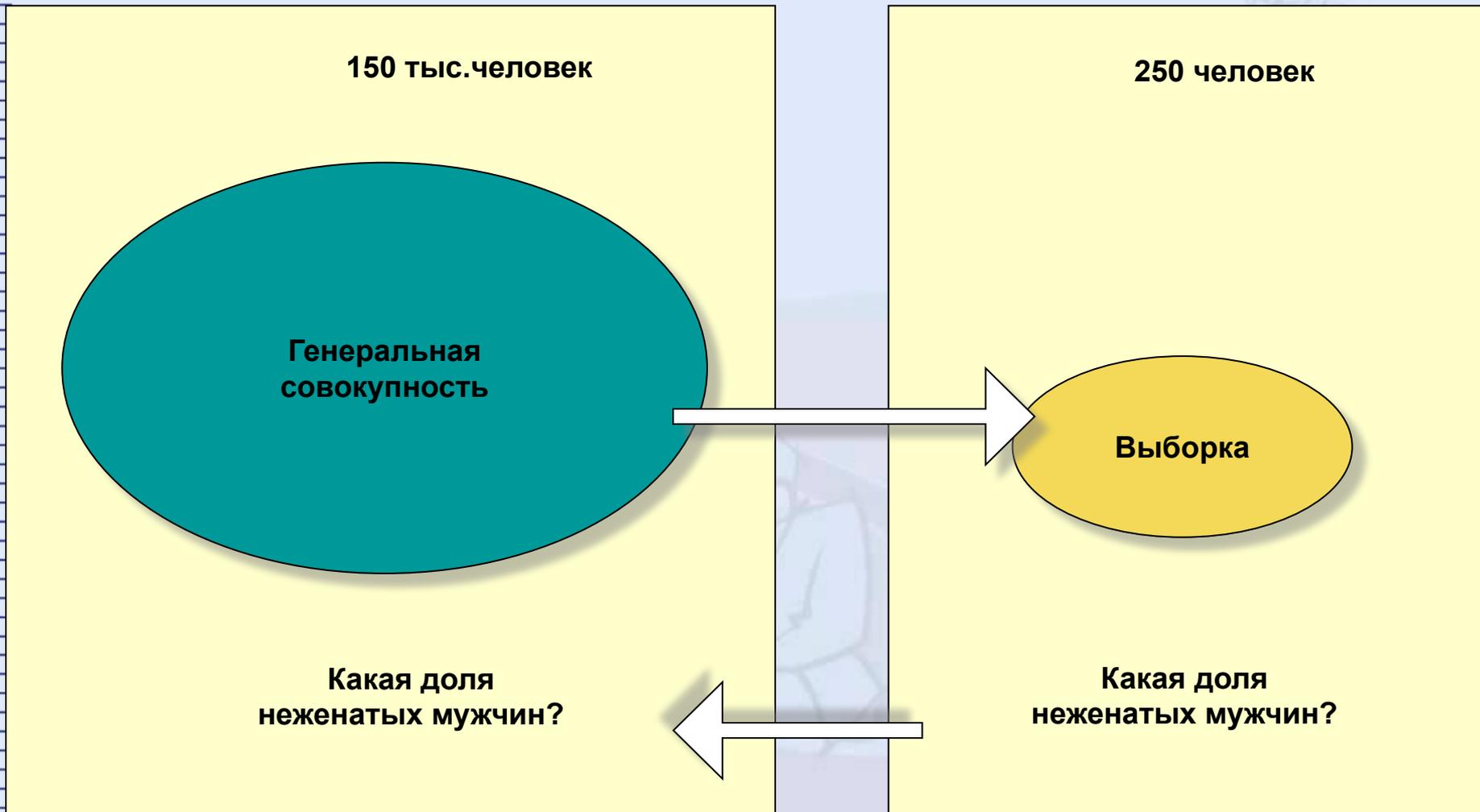
Понятия выборочного исследования

- Генеральная совокупность – вся социальная группа, про которую необходимо собрать информацию.
 - В большинстве случаев «генеральная совокупность» и «объект исследования» - синонимы (состоят из одних и тех же элементов).
- Выборка (выборочная совокупность).
 - Часть генеральной совокупности, которая непосредственно подвергается исследованию.
- Единица выборки (единицы отбора).
 - Элементы генеральной совокупности, из которых формируется выборочная совокупность.
- Ошибка выборки.
 - Степень рассогласования (невязки) между оценками важных с точки зрения исследования признаков, полученными на выборке и теми значениями, которые могли бы быть получены при проведении сплошного исследования (на всей генеральной совокупности).

Почему исследователи любят выборочные исследования

- **Меньше работы**
- **Меньше денег**
- **Меньше времени**
- **Меньше головной боли (?)**

Генеральная совокупность и выборка



Оптимальный размер выборки

- В интересах исследователя
- В интересах реально существующей ситуации в генеральной совокупности

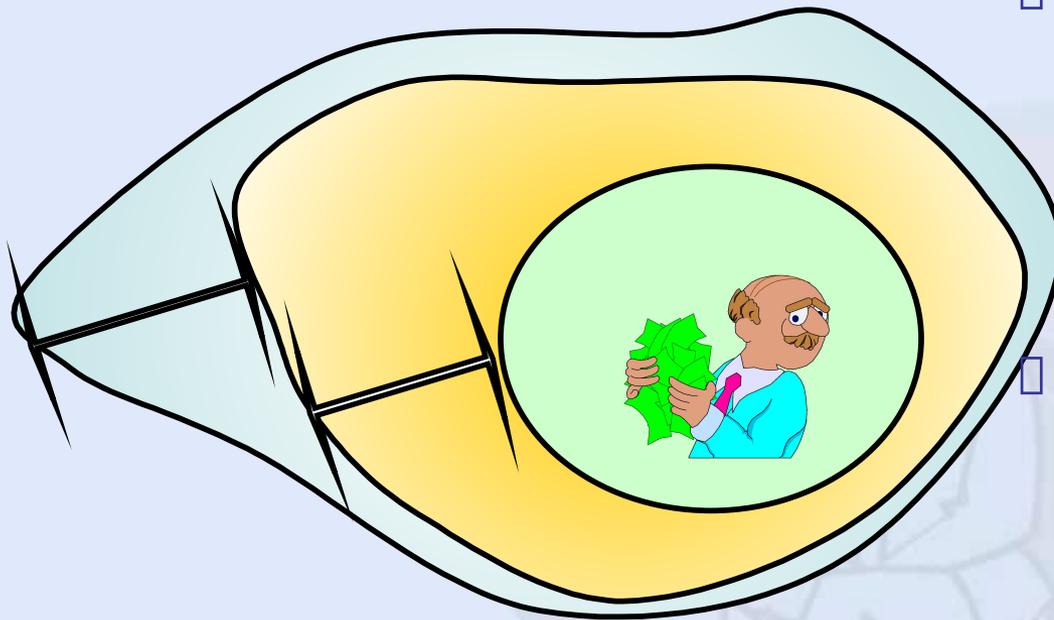
Ошибки при составлении основы выборки

- **Ошибки при формировании основы** выборочного наблюдения состоят в пропуске некоторых объектов или включении в основу объектов, не относящихся к изучаемой генеральной совокупности.
- Имеются специальные методы, позволяющие устранить такие ошибки или свести их действие к минимуму.

Ошибки в выборочном исследовании

Генеральная совокупность
(объект исследования)

«Достижимая»
совокупность
(контур выборки)



Ошибка модели

Ошибка модели

Выборка

- Социологическое исследование – это измерение.
- Измерение (всегда!!!) содержит ошибки:
 - Ошибки методики измерения.
 - Ошибки инструмента (прибора).
- Ошибка включает в себя два компонента:
 - Систематическая ошибка.
 - Случайная ошибка (статистическая).

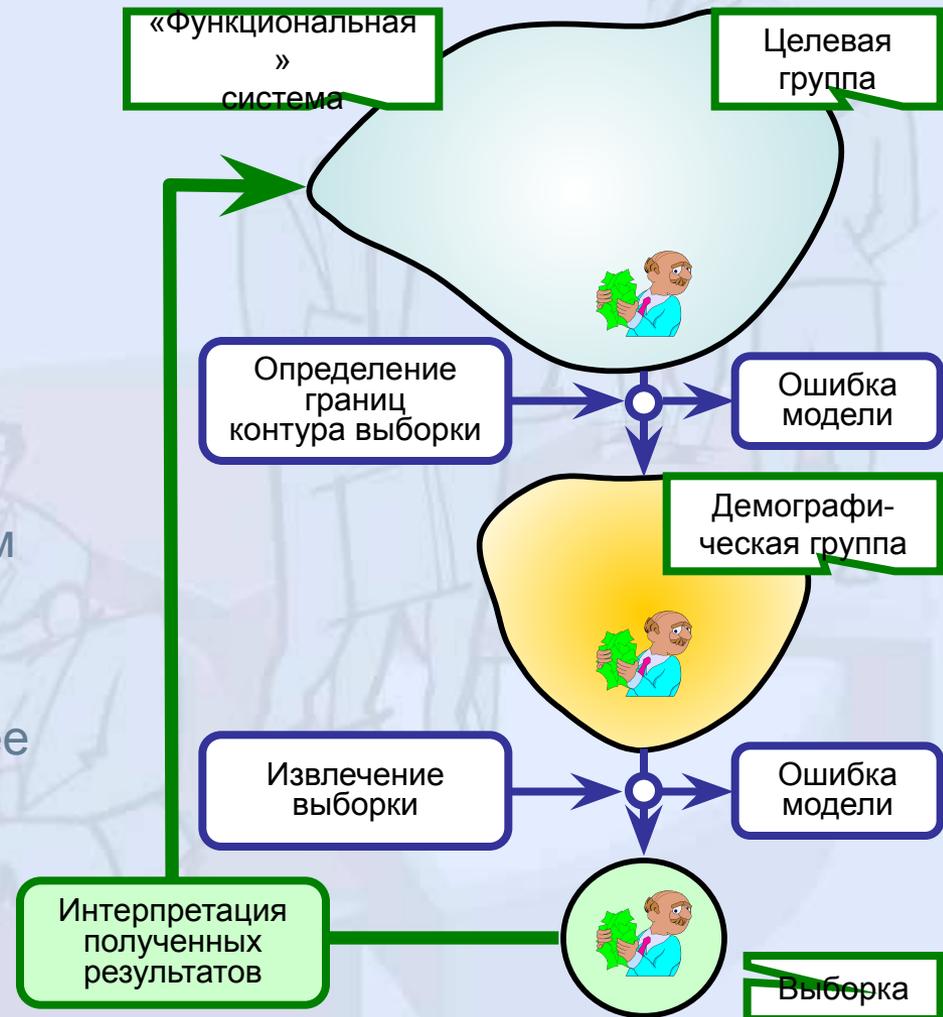
Систематическая и случайная ошибка

□ Систематическая ошибка.

- Источники ошибки – методика измерений и инструментарий.
- Может быть (хотя бы теоретически) компенсирована в процессе обработки.
- Не снижается с увеличением объема выборки.

□ Статистическая погрешность.

- Зависит от типа выборки и ее объема.
- Снижается (в некоторых случаях) с увеличением объема выборки.



Ошибки выборки:

Случайные: вероятность того, что выборочная средняя выйдет (или не выйдет) за пределы заданного интервала □ вероятностные, уменьшаются при возрастании объема выборочной совокупности

Эксперименты Дж.Гэллапа (1935 г. – отношение американцев к запрету спиртного)



Выборки	Одобрят	Не одобряют	Нет мнения
442 чел.	31	62	7
884 чел.	29	63	8
1327 чел.	30	63	7
2585 чел.	31	61	8
5255 чел.	33	59	8
8253 чел.	32	60	8
12494 чел.	32	61	7

Опрос лаборатории Х.Кентрила (1942 г., Нью-Йорк)

200 человек (распределение по плотности проживания на территориях штата+ раса+ эконом.положение + возраст)

Результаты: Ср.ошибка 3,3%

Кандидаты	Опрос Кентрила	Результаты голосования
Дью	58	53
Беннет	36	37
Альфанж	6	10



«Нью-Йорк дейли ньюс» - 48 тыс. человек – ср.ошибка 2,3%
Дж.Гэллап – 2500 человек – ср.ошибка 1,3 %

Выборочное исследование. Выборка

Систематические ошибки

- = неконтролируемые перекосы в распределении выборочных наблюдений («потеря» объекта). При возрастании объема выборки не уменьшаются
- 1936 г. выборы президента (Рузвельт – Лэндон):
«Литерэри Дайджест»: 2 376 523 человека

- - Прогноз: Лэндон – 57% , Рузвельт – 43%.
 - Результат: Рузвельт – 62,5%, Лэндон – 37,5%



Georg Gallup
(1901 –
1984)

Типичные систематические ошибки

- Давление доступных объектов
- Иллюзия постоянства (*пренебрежение группой неопределившихся респондентов*)
- Недостаточный учет аномальных и труднодоступных единиц исследования (*больные, личный состав ВС...*)
- Недостаточный учет отсутствующих в месте сбора данных (*нет дома*)
- Отказы от ответа (*нет мнения, негативная установка, внешние обстоятельства*)

Контроль и ремонт выборки

- Ремонт В. = процесс устранения погрешностей, т.е. расхождения ВС и ГС различными способами:
- Коррекция ВС (*напр., формирование повторной выборки или выборки большего размера, или следующих номер абонента...*)
- Коррекция распределений демографических характеристик респондентов (*взвешивание исходных данных...*)
- Коррекция резко выделяющихся ответов респондентов (брак?)
- Коррекция пропущенных ответов

Контроль и ремонт выборки

Контролем выборки будем называть процесс научного сравнения генеральной и выборочной совокупностей, выявление степени их расхождения, обнаружение причин отклонения и разработку возможных способов устранения погрешностей.

В узком смысле — это уравнивание выборочных и генеральных распределений социально-демографических характеристик респондентов.

Основная цель ремонта выборки — повышение качества уже собранной информации. Процедура ремонта выборки включает несколько операций.

Коррекция выборочной совокупности. Далеко не всегда отобранные респонденты, по самым разным обстоятельствам, могут или желают отвечать на вопросы. Кто-то заболел или уехал в срочную командировку, другой отказывается по идейным соображениям или не способен ответить в силу умственной недостаточности. Кого-то трудно застать дома, хотя анкетер приходил к нему не единожды.

Возникает проблема замены респондентов, которая может быть решена с помощью нескольких методов:

- выбор следующего по списку респондента (например, следующего номера в телефонном справочнике),
- использование первоначальной выборки больших размеров и формирование повторной выборки.

В последнем случае, если процент ответов оказался намного ниже, чем ожидалось, основа выборки расширяется за счет дополнительных имен, найденных, например, случайным образом. Самым эффективным способом считается поиск эквивалентной замены.

Коррекция распределений демографических характеристик респондентов

Если по окончании исследования в паспортичке вашего исследования получилось, что у вас, к примеру, перебор женщин, людей с высшим образованием или пожилых людей в сравнении с теми процентными долями, которые они имеют в генеральной совокупности, то можно применить **три способа**:

- 1) удалить те группы респондентов, которые оказались представлены в избыточном количестве;
- 2) доопросить те группы, которые оказались представленными в недостаточном количестве;
- 3) математически повысить значение ответов, представленных недостаточно, или снизить — представленных избыточно.

- В социологии еще не придумано единой и четкой формулы, используя которую можно рассчитать оптимальный объем выборочной совокупности, — такой формулы просто не существует в природе.
- Объем выборочной совокупности зависит от множества факторов, в том числе от целей и задач, теоретической модели, гипотез и методов исследования, степени однородности генеральной совокупности наконец, требующейся точности получаемой информации.

- Надо всегда помнить, что **каждый процент прироста точности информации в исследовании приводит к резкому увеличению расходов на его проведение.**
- Знаменитый институт Гэллапа, на протяжении многих десятилетий проводящий опросы в США, выявил, что при общенациональной выборке в 100 человек — ошибка выборки будет в пределах $\pm 11\%$; 200 человек - $\pm 8\%$; 400 - $\pm 6\%$; 600 - $\pm 5\%$; 750 — $\pm 4\%$; 1000 - $\pm 4\%$; 1500 - $\pm 3\%$; 4000 человек - $\pm 2\%$. Именно поэтому он проводит общенациональные опросы в США на выборке в 1500-2000 человек. Как видно, он предпочитает увеличение ошибки на 1% многократному увеличению стоимости исследования.

Для расчета объема случайной выборки надо знать

- желаемую точность оценивания,
- величину риска получаемого ответа
- степень изменчивости ответа.

Традиционно точность оценивания принимают за 5%, а величину риска — за 0,95.

Иными словами, если по данным выборочного исследования 60% опрошенных удовлетворены работой, то можно утверждать, что в генеральной совокупности доля удовлетворенных составит от 55 до 65% в 95% случаев, а в 5% случаев такая доля может выйти за этот интервал.

- С увеличением объема выборки точность возрастает, но до определенного порога.
- Уже при 600 опрошенных достигается желанный для всех 5%-ный уровень точности.
- 600 человек — приемлемый объем выборки.

- Между цифрами 400 и 600 человек **противоречия нет**.
- В первом случае объем выборки рассчитывался, исходя из положения о нормальном распределении ответов респондентов, а во втором — из практики.
- Расхождение между теорией и практикой обусловлено тем, что в реальной ситуации распределение оценок отличается от нормального, поэтому объем выборки надо рассчитывать с учетом именно этого обстоятельства;
- наиболее эффективным способом уменьшения объема выборки является снижение коэффициента вариации оценок.

Зависимость между объемом выборки и точностью оценивания

Количество интервью	Точность оценивания %
100	+/-11
200	+/-8
400	+/-6
600	+/-5
750	+/-4
1000	+/-4
1500	+/-3
4000	+/-2

- От чего действительно зависит размер выборки — так это от числа параметров, по которым мы желаем добиться репрезентативности.
- Если нас устраивает репрезентативность только по полу и возрасту, то выборки в 400 человек в одном населенном пункте будет более чем достаточно.
- Если параметров три, количество респондентов придется увеличить до 600.
- Добиться репрезентативности выборки одновременно по пяти параметрам: полу, возрасту, доходу, образованию, сфере профессиональной деятельности — можно лишь на выборке из 1 000-1200 человек в одном населенном пункте.

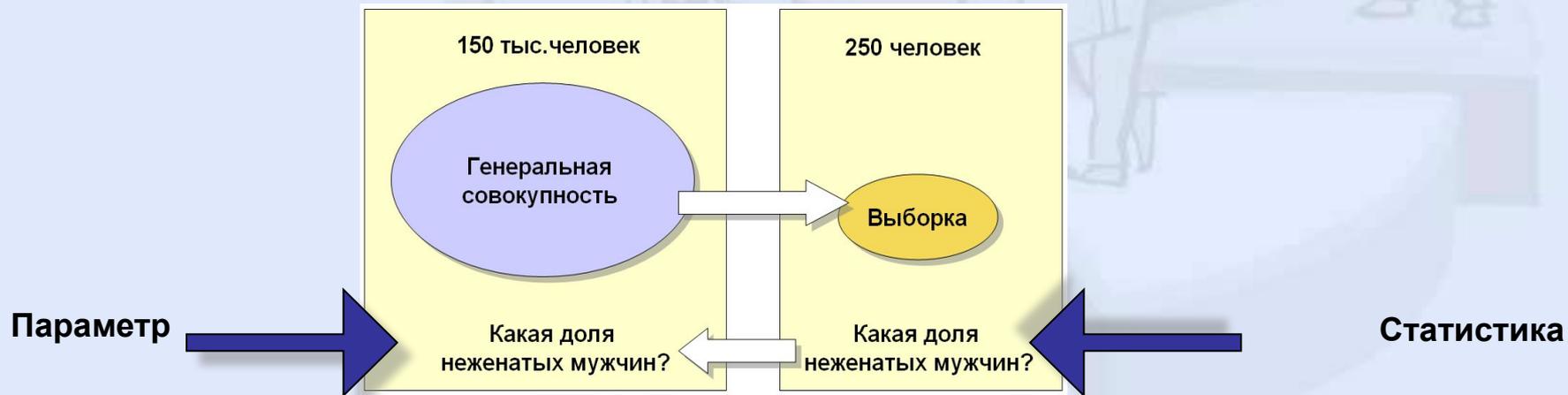
Определение объема выборочной совокупности и обеспечение ее репрезентативности

- предположим, что в генеральной совокупности 90% мужчин и 10% женщин.
- В этом случае, чтобы в выборочную совокупность попало 90 женщин, необходимо отобрать уже не менее 900 человек. Из примера видно, насколько велико влияние дисперсии (разброса признака) на объем выборочной совокупности. Поэтому вычислять его следует по признаку, дисперсия значений которого наибольшая.



Параметры и статистики

- **Параметры** - характеристики генеральной совокупности.
- **Статистики** - характеристики выборки.
- Мы будем использовать статистики для оценки параметров генеральной совокупности, которым они соответствуют.



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.

Репрезентативность выборки

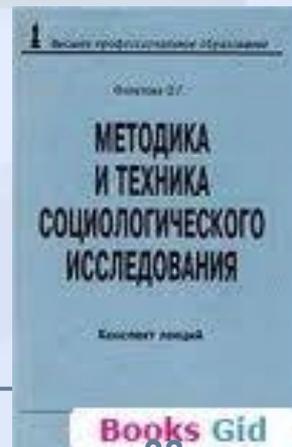
- Репрезентативность выборки.
 - Свойство выборки достаточно полно и точно представлять наиболее важные для исследователя и коррелирующие с ними признаки генеральной совокупности.
 - Результаты, полученные на нерепрезентативных выборка всегда содержат систематическую ошибку.
- Ограничения понятия «репрезентативность».
 - Не бывает «репрезентативности вообще». Репрезентативность возможна только по каким-либо определенным (важным для исследователя) признакам.
 - Репрезентативность не обязательно обеспечивает «надежность» или «точность» получаемых данных.
 - Для контроля репрезентативности необходимо привлечение «внешних» источников информации.
 - Репрезентативными могут быть (а могут и не быть) как «собственно случайные», так и направленные выборки.

Определение объема выборочной совокупности и обеспечение ее репрезентативности

- Когда информация о признаках элементов генеральной совокупности отсутствует, исключается возможность определения объема выборочной совокупности при помощи математических формул.
- В этом случае можно опереться на многолетний опыт социологов-практиков, свидетельствующий о том, что для пробных опросов достаточна выборочная совокупность объемом 100–250 человек.
- При массовых опросах, если величина генеральной совокупности составляет менее 5000 человек, достаточный объем выборочной совокупности – не менее 500 человек, если же 5000 человек и более, то – 10% ее состава, но не более 2000–2500 человек. Это гарантирует достаточно достоверные результаты исследования.

Определение объема выборочной совокупности и обеспечение ее репрезентативности

- Объемы выборочной совокупности, превышающие 3000 респондентов, используются лишь для комплексных исследований, а также при сложной структуре генеральной совокупности.
- Определение статистически значимого объема выборочной совокупности— важная, но не достаточная предпосылка правомерности распространения выводов исследования на всю генеральную совокупность.
- Дело в том, что из одной и той же генеральной совокупности можно отобрать большое число выборочных совокупностей.



Исследовательская практика подсказывает следующее правило: объем выборки должен обеспечивать не менее 100 наблюдений для каждой первостепенной и не менее 20—50 наблюдений для каждой второстепенной классификационной составляющей.

Первостепенные классификационные составляющие соответствуют наиболее критичным, а второстепенные — наименее критичным ячейкам перекрестной классификации, принятой в данном исследовании.

- Теоретические расчеты и практика доказывают, что для получения достоверных данных о мнении и предпочтениях населения такого крупного города, как Санкт-Петербург, достаточно опросить 700—800 человек.
- Однако большинство опросов населения здесь проходят на выборках объемом до 1,5 тыс. человек

Таким образом, как справедливо утверждает В. Ядов, «численность (объем) выборки зависит от уровня однородности или разнородности изучаемых объектов. Чем более они однородны, тем меньшая численность может обеспечить статистически достоверные выводы»

- Сегодня многие трудные расчеты берет на себя техника, а статистические программы можно получить по Интернету.
- Вот и с расчетом выборки ленивому социологу предоставили такую возможность на веб-сайте Аналитического центра «Бизнес и маркетинг» (<http://www.bma.ru/enter.htm>), где пользователю надо лишь внести необходимые данные, а затем нажать на кнопку «Рассчитать».

План выборочного наблюдения

1. Определение генеральной совокупности



2. Определение основы выборки



3. Определение метода получения выборки



4. Определение объема выборки



5. Процесс получения выборки

Определение метода получения выборки

- **Детерминированные методы:**
 - Нерепрезентативная выборка
 - Поверхностная выборка
 - Квотная выборка
 - Выборка по принципу «снежного кома»

- **Вероятностные методы:**
 - Простая случайная выборка
 - Систематическая выборка
 - Стратифицированная выборка
 - Кластерная выборка



За подробностями можно обратиться в эту книгу.

Выборка с возвращением и без возвращения

- После того, как объект извлечен из генеральной совокупности для включения в выборку, его либо возвращают в генеральную совокупность, либо нет. Если его возвратили, он может попасть в выборку повторно.
- **Выборка без возвращения** – любой объект не может попасть в выборку больше одного раза.
- **Выборка с возвращением** – любой объект может оказаться в выборке более одного раза.

Методы формирования выборки



Целевая генеральная совокупность

- На первом этапе определяют **целевую генеральную совокупность** (target population) – совокупность объектов, обладающих информацией, которую желает получить исследователь и о которой требуется сделать заключение.
- Какие объекты попадут в эту совокупность, где ее границы?
- **Пример.** Кто является целевой генеральной совокупностью при изучении зрительской аудитории кино?

Генеральная совокупность и объем выборки

- N — объем генеральной совокупности
- n — объем выборки
- Если изучается вся генеральная совокупность, то выборка называется **переписью** (census).
- Обычно мы будем считать, что целевая генеральная совокупность конечна, хотя ее объем велик:

$$\frac{n}{N} \approx \frac{n}{\infty}$$

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

- Чем больше объем выборки – тем выше точность ее результатов
- Чем более однородна генеральная совокупность – тем меньший объем выборки можно использовать
- **Однородная** совокупность – в которой контролируемый признак распределен равномерно (не образует пустот и сгущений)
- Объем выборки зависит от:
 - целей и задач, гипотез и методов исследования;
 - объема ГС;
 - степени однородности ГС;
 - требующейся точности полученной информации
 - характера предполагаемых распределений ответа...;

Стратегии расчета объема выборки

- *предварительная* – объем выборки определяется до проведения основного исследования;
- *последовательная* – объем выборки не рассчитывается заранее, а ставится в зависимость от конечных результатов исследования (например, в телефонном опросе постоянно проводится расчет распределений ответов на все вопросы и требуемого объема выборки);
- *комбинированная* – предварительно рассчитывают верхние допустимые значения для последовательной стратегии.

Стратегия предварительного расчета (случайная выборка)

Необходимо знать:

- Желаемую точность оценивания (5%)
- Величину риска получаемого ответа (0,95)
- Степень изменчивости ответа (50% на 50% при дихотомическом вопросе, т.е. макс. разброс берем, если нет информации о распределении ответов на вопрос)



$V_{ГС}$	500	1000	2000	3000	4000	5000	10 т.	100 т.	более
$V_{ВС}$	222	286	333	350	360	370	385	398	400

Зависимость объема выборки от распределения дихотомического ответа

Распределе ние ответов, %	50	40	30	20	10
	50	60	70	80	90
$V_{выборки}$	384	369	323	246	139

Определение основы выборки

- **Основа выборочного наблюдения** (sampling frame) есть представление элементов изучаемой генеральной совокупности. Обычно это список всех объектов или перечень инструкций для определения границ и объектов изучаемой совокупности.
- **Примеры.** Телефонная книга, отраслевые справочники предприятий, список адресатов.

Определение основы выборки

- Если исследователь не может составить подробный перечень элементов, то следует, по крайней мере, установить **правила для отбора** объектов изучаемой генеральной совокупности.
- **Пример.** Случайный компьютерный набор номеров при проведении опроса по телефону.

Вероятностный отбор

- Вероятностный отбор - это способ формирования выборки, при котором каждая единица генеральной совокупности имеет равную вероятность быть включенной в выборку.
 - При вероятностном отборе имеется возможность количественной оценки статистической погрешности выборки.
 - Является единственным приемлемым методом формирования выборки в тех случаях, когда отсутствуют или недостаточны предварительные данные о распределении объектов, входящих в генеральную совокупность, по интересующим исследователя характеристикам.

Вероятностные выборки строятся с использованием таблиц или генераторов случайных чисел.

10	09	73	25	3	09	37	67	07	15
37	54	20	48	0	80	15	73	61	47
08	42	26	89	5					
99	01	90	25	29					
12	80	79	99	70					
66	06	57	47	17	34	07	27	68	50
31	06	01	08	05	45	57	18	24	06
85	26	97	76	02	02	05	16	56	92
63	57	33	21	35	05	32	54	70	48
73	79	64	57	53	03	52	96	47	78

Фрагмент таблицы равномерно распределенных случайных чисел

Вероятностный отбор

А	В	С	Д	Е	Ж	З	И	К
4791								
4792	ЗАВОДОУКОВСКИЙ РАЙОН							
4793								
4794	(101) Заводоуковский район							
4795	2							
4796								
4797	Регион	Город	Код города	Станция	Число номеров	Минимальный	Максимальный	FIN
4798	(72) Тюменская	(101) Заводоуковский	34542	2	17	1064	6110	FIN
4799								
4800	Число инт.	5		Диапазон	5046			
4801	Кратность	50		Число номеров	250			
4802								
4803								
4804	##	Полный номер	Краткий номер	FIN				
4805	1	83454224480	24480	FIN				
4806	2	83454223073	23073	FIN				
4807	3	83454223561	23561	FIN				
4808	4	83454220283	20283	FIN				
4809	5	83454224451	24451	FIN				
4810	6	83454222733	22733	FIN				
4811	7	83454221451	21451	FIN				
4812	8	83454221939	21939	FIN				
4813	9	83454224316	24316	FIN				
4814	10	83454220844	20844	FIN				
4815	11	83454220432	20432	FIN				
4816	12	83454223236	23236	FIN				

Номер телефонной станции (в пределах города)

Диапазон разрешенных номеров

Телефонные номера, сгенерированные случайным образом

Телефонные номера с учетом кода города

Вероятностный отбор практически всегда используется в телефонных опросах, когда стоимость контакта практически не зависит от местонахождения респондента

Систематический отбор

- Систематический отбор - способ формирования выборочной совокупности, при котором выбор первого объекта репрезентации (как правило, случайный) полностью определяет набор объектов репрезентации, включаемых в выборку.
 - Отбор единиц осуществляется через один и тот же интервал (шаг) в исходном списке. Номер первого объекта в списке, включаемого в выборку, определяется случайным образом или принимается равным половине шага.

МАРШРУТНОЕ ЗАДАНИЕ

Организатор опроса выдает Вам маршрутное задание, в котором указаны адреса отобранных к опросу домов. Выполнять маршрутное задание можно в любом направлении. Не так важно, в каком порядке Вы проходите дома, важно, чтобы дом был указан в маршрутном задании.

ОТБОР КВАРТИРЫ

В ПЕРВОМ ВЫБРАННОМ многоквартирном доме или общежитии интервьюер обращается в квартиру №5.

ЕСЛИ В КВАРТИРЕ ИНТЕРВЬЮ БЫЛО ВЗЯТО, то следующая квартира, куда надо обращаться, вычисляется согласно описываемому ниже шагу. Так, в одно, двух и трехэтажных домах следующая квартира вычисляется прибавлением цифры 3 (№5, №8, №11...), в четырех и пятиэтажных домах шаг равен 5 (№5, №10, №15...) и в домах свыше 5 этажей шаг равен 7 (т.е. №5, №12, №19...). Когда дом закончится, следующий дом надо рассматривать как продолжение предыдущего.

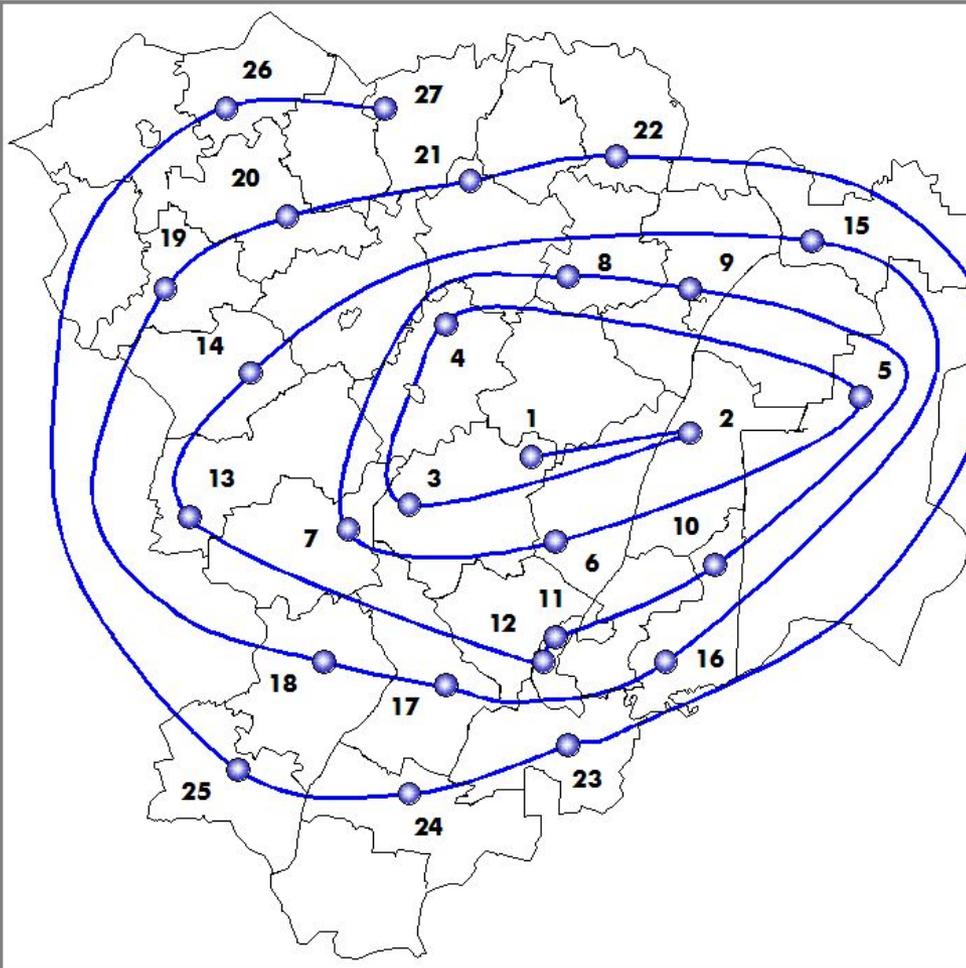
Требование строгой случайности нарушается. Выбор первого объекта и шага отбора полностью определяет выборку

Систематический отбор

На территории области расположено «N» объектов.
Объекты упорядочиваются в списке «раскручивающейся спиралью», вычерченной на карте.

В выборку включается «n» объектов. Объем выборки определяется целями исследования, его бюджетом и сроками проведения

Объекты, включаемые в выборку, отбираются с шагом $\text{Step} = \text{ОКРУГЛВНИЗ}(N/n)$. Номер первого объекта, включаемого в выборку, определяется случайным образом.



Систематическая выборка

- **Систематическая выборка** получается путем нумерации каждого члена генеральной совокупности и затем выбором каждого k -ого номера.
- **Пример.** Генеральная совокупность включает 2000 единиц, требуется отобрать 50. Поскольку $2000/50=40$, то будем выбирать каждый 40-й элемент. Для начала случайным образом выберем первый элемент выборки среди первых сорока элементов генеральной совокупности. Если первым оказался номер 12, тогда выборка будет включать объекты с номерами 12, 52, 92 и так далее, всего 50 объектов.

Простая случайная выборка

- **Простая случайная выборка** отбирается при помощи методов случайного отбора или случайных чисел.
- Один из таких методов заключается в нумерации каждого объекта генеральной совокупности и выборе номеров объектов при помощи генератора случайных чисел в компьютере или калькуляторе.
- До применения компьютеров случайные числа были получены и сведены в таблицу случайных чисел.

Как составить случайную выборку

- В аудитории находится 120 студентов. Составьте случайную выборку из 15 человек.
- Следует воспользоваться таблицей случайных чисел.

Таблица случайных чисел

79	41	71	93	60	35	04	67	96	04	79	10	86
26	52	53	13	43	50	92	09	87	21	83	75	17
18	13	41	30	56	20	37	74	49	56	45	46	83
19	82	02	69	34	27	77	34	24	93	16	77	00
14	57	44	30	93	76	32	13	55	29	49	30	77
29	12	18	50	65	33	15	79	50	28	50	45	45
01	27	92	67	62	31	97	55	29	21	64	27	29
55	75	65	68	82	73	07	95	66	43	43	92	16
84	95	95	96	13	30	91	64	74	83	47	89	71
62	62	21	37	29	62	19	44	08	64	34	50	11
66	57	28	69	75	99	74	31	58	19	47	66	89
48	13	69	97	01	01	75	58	05	40	40	18	29
94	31	73	19	80	76	33	18	05	53	04	51	41
00	06	53	98	62	55	08	38	49	42	10	44	38
46	16	44	27	39	15	28	01	64	27	89	03	27
77	49	85	95	23	93	25	39	63	74	54	82	85

Квотная выборка

- *Квотная выборка* — микромодель объекта социологического исследования, формируемая на основе статистических сведений (параметров квот) преимущественно о социально-демографических характеристиках элементов генеральной совокупности. Нужные данные обычно берут из статистических справочников.
- Квотный метод выборки предполагает предварительное наличие статистических сведений по ряду существенных либо коррелирующих с ними характеристик генеральной совокупности. Однако эти сведения не используются для определения объема выборки, так как в последующем отбор респондентов осуществляется не случайно, а целенаправленно, при помощи интервьюеров.

Таблица 2

Структура выборки

Возрастной сегмент	Доля ГС, %	Численность, чел.
17-25	26,9	322
26-40	35,8	430
41-55	37,3	448
Итого:	100	1200

Колебания выборки в зависимости от распределения на примере выборки в 1000 респондентов

Распределение	Точность (стандартная ошибка)	Интервал, в пределах которого находится истинный результат
50%:50%	3,1%	46,9%-53,1%
40%:60%	3,0	37%-43%
30%:70%	2,8	27,2%-32,8%
20%:80%	2,5%	17,5%-22,5%
10%:90%	1,9%	8,1%-11,9%
5%:95%	1,4%	3,6%-6,4%

Квотная выборка

- Приведем пример расчета выборки по таким параметрам квоты, как пол и возраст, в реальном социологическом исследовании.
- Обосновывая репрезентативность выборки при опросе избирателей в ходе избирательной кампании по выборам депутатов Городской думы Череповца, мы опирались на данные Облкомстата о половозрастной структуре населения районов города.
- Учитывая, что в состав выборки должны были войти избиратели, т.е. лица в возрасте не моложе 18 лет, мы взяли в качестве основы выборки ту часть таблицы, где представлено половозрастное распределение населения от 20 лет (разделить предыдущую строку, выделив из нее лиц 18—19 лет, не представляясь возможным)

Стратифицированный отбор

- Процедура районированного (стратифицированного, расслоенного) отбора.
 - До извлечения выборки производится разделение генеральной совокупности на «страты», различающиеся по контролируемым (важным с точки зрения цели исследования) признакам.
 - Выборка (случайная или направленная) извлекается независимо из отдельных страт.
 - Объем выборки из каждой страты определяется исследователем исходя из специфики исследования (как правило, из соображений минимизации статистической погрешности выборки).
 - На этапе обработки результатов исследования проводится перевзвешивание полученных данных (различным стратам присваиваются различные «весовые коэффициенты»).
- Стратифицированный отбор – единственный вид отбора, который при определенных условиях может обеспечить ошибку выборки ниже, чем при вероятностном отборе.

Стратифицированная выборка

- **Стратифицированная выборка** получается путем разбиения генеральной совокупности на группы или страты в зависимости от характеристик, важных для изучения.
- Преимуществом стратифицированной выборки является наличие представителей каждой страты в выборке в соотношении, сходном с генеральной совокупностью. Недостатком является сложность организации процесса при наличии нескольких признаков, скажем, возраста, дохода и социального статуса.

Пример стратифицированной выборки

- **Пример.** На двух факультетах обучается 2000 студентов, среди которых 60% филологов и 40% биологов. Соотношение мужчин и женщин 30/70. Тогда все они могут быть разделены на страты филологи-биологи и мужчины-женщины.

- Генеральная совокупность

- **Филологи Биологи**

- мужчины 360 240

- женщины 840 560

- Всего 2000 человек

- Выборочная совокупность

- **Филологи Биологи**

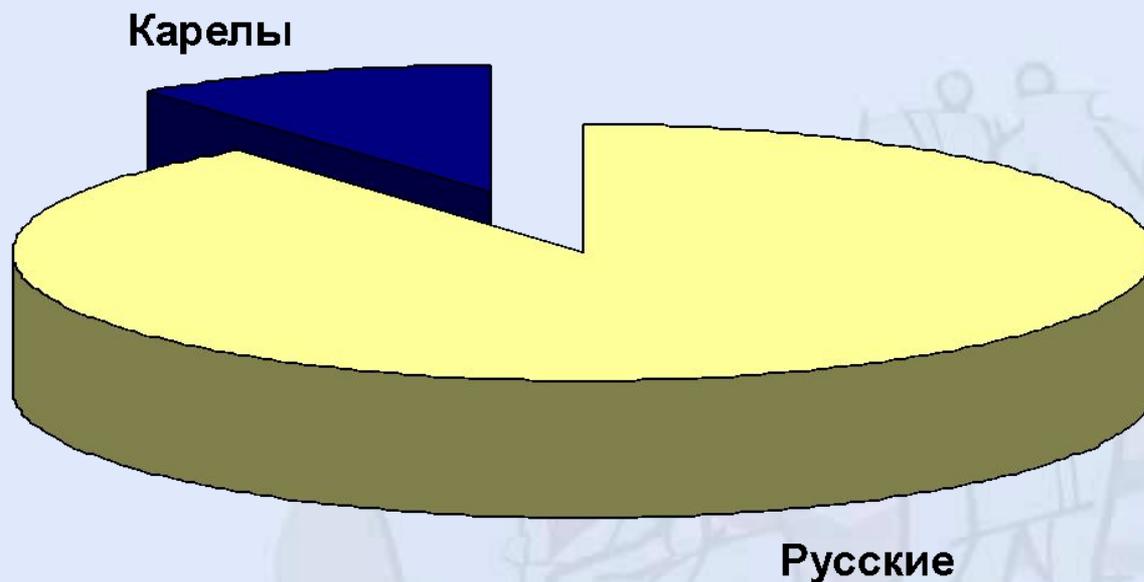
- мужчины 12 18

- женщины 28 42

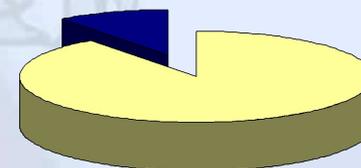
- Всего 100 человек



Стратифицированная выборка

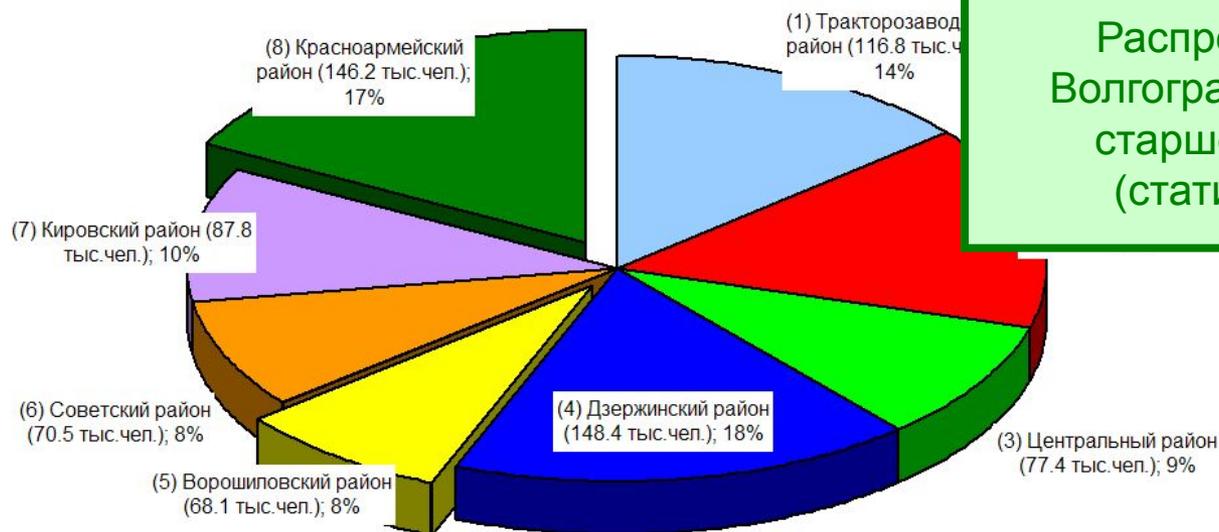


**Выборка соответствует
генеральной совокупности
по структуре**



Пример стратифицированного отбора

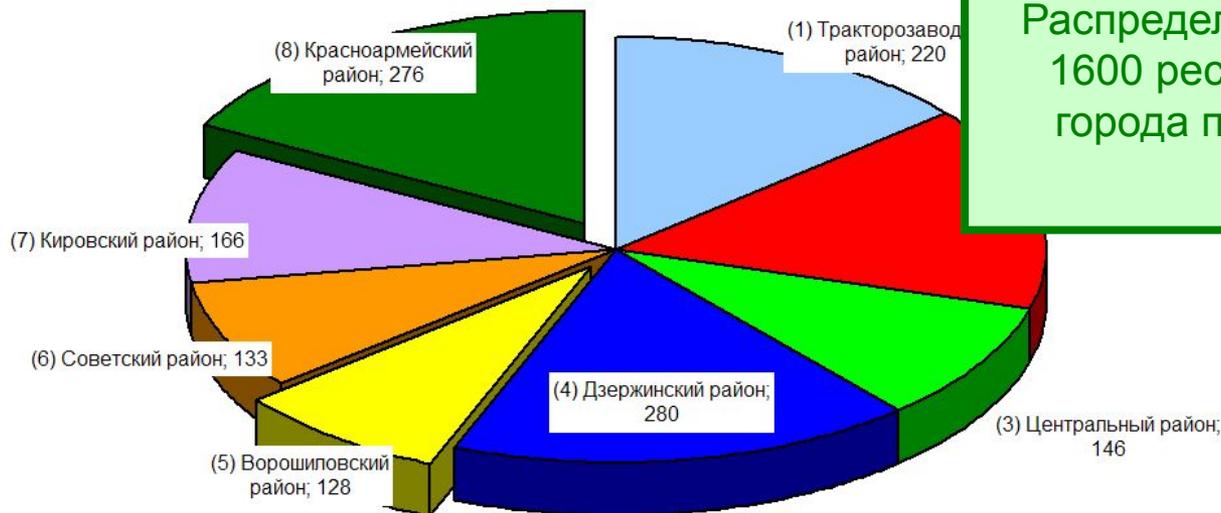
Распределение населения Волгограда в возрасте 16 лет и старше по районам города (статистические данные)



Район	Всего в возрасте 16 лет и старше, тыс. человек	Доля в общей совокупности, %	Объем выборки при пропорциональном отборе	Максимальная статистическая погрешность	Объем выборки при стратифицированном отборе	Максимальная статистическая погрешность выборки
(1) Тракторозаводский район	116.8	14%	220	6.6%	200	6.9%
(2) Краснооктябрьский район	133.4	16%	252	6.2%	200	6.9%
(3) Центральный район	77.4	9%	146	8.1%	200	6.9%
(4) Дзержинский район	148.4	17%	280	5.9%	200	6.9%
(5) Ворошиловский район	68.1	8%	128	8.7%	200	6.9%
(6) Советский район	70.5	8%	133	8.5%	200	6.9%
(7) Кировский район	87.8	10%	166	7.6%	200	6.9%
(8) Красноармейский район	146.2	17%	276	5.9%	200	6.9%
Общий итог	848.6	100%	1600	2.5%	1600	2.5%

Пример стратифицированного отбора

Распределение выборки объемом 1600 респондентов по районам города при пропорциональном отборе

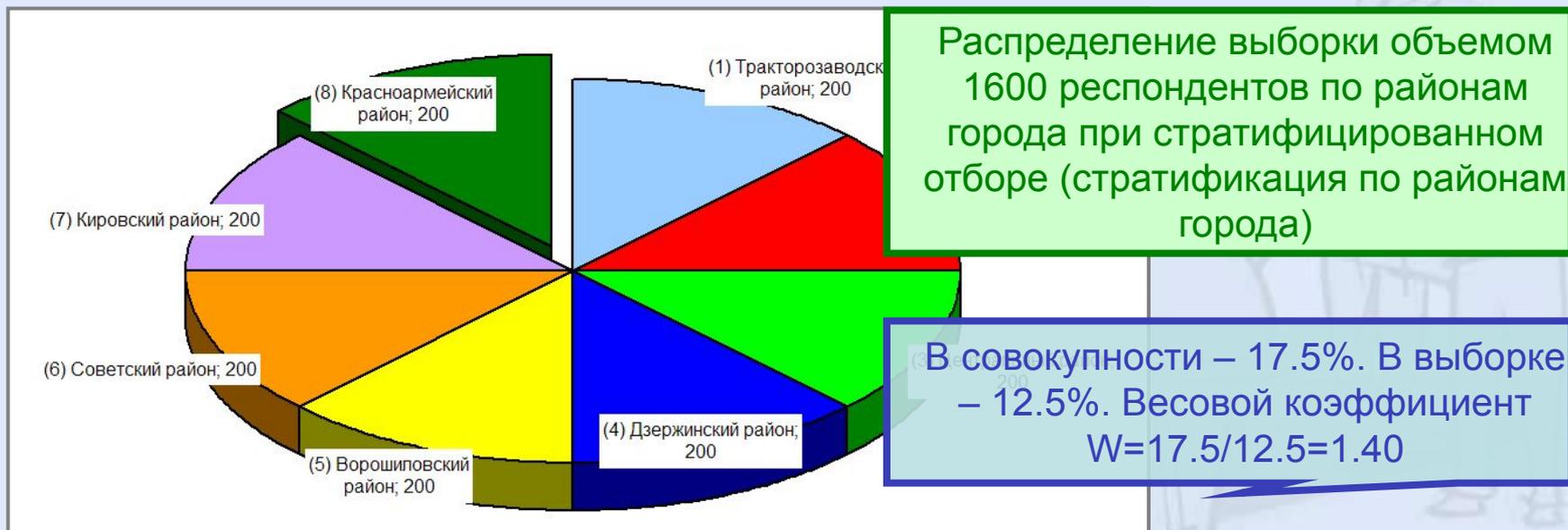


Район	Всего в возрасте 16 лет и старше, тыс. человек	Доля в общей совокупности, %	Объем выборки при пропорциональном отборе	Максимальная статистическая погрешность	Объем выборки при стратифицированном отборе	Максимальная статистическая погрешность выборки
(1) Тракторозавод район	116.8	14%	220	6.9%	200	6.9%
(2) Красноармейский район	134.4	16%	276	6.9%	200	6.9%
(3) Центральный район	71.4	9%	146	8.1%	200	6.9%
(4) Дзержинский район	148.4	17%	280	5.9%	200	6.9%
(5) Ворошиловский район	68.1	8%	128	8.7%	200	6.9%
(6) Советский район	70.5	8%	133	8.5%	200	6.9%
(7) Кировский район	87.8	10%	166	7.6%	200	6.9%
(8) Красноармейский район	146.2	17%	276	5.9%	200	6.9%
Общий итог	848.6	100%	1600	2.5%	1600	2.5%

Максимальная статистическая погрешность – 8.7%

Максимальная статистическая погрешность – 5.9%

Пример стратифицированного отбора



Максимальная
статистическая
погрешность – 6.9%

Максимальная
статистическая
погрешность – 6.9%

Район	Доля в генеральной совокупности, %	Доля в выборке	Весовой коэффициент
(1) Тракторозаводский район	13.8%	12.5%	1.10
(2) Краснооктябрьский район	15.7%	12.5%	1.26
(3) Центральный район	9.1%	12.5%	0.73
(4) Дзержинский район	17.5%	12.5%	1.40
(5) Ворошиловский район	8.0%	12.5%	0.64
(6) Советский район	8.3%	12.5%	0.66
(7) Кировский район	10.3%	12.5%	0.83
(8) Красноармейский район	17.2%	12.5%	1.38
Общий итог	100.0%	100.0%	1.00

Кластерная выборка



Классы школы

Цеха завода

Дома микрорайона

Кластерная выборка

- **Кластерная выборка** образуется при выделении отдельных групп, которые называются кластерами.
- **Пример.** Исследователю необходимо опросить жителей, проживающих в квартирах небольшого города. Если в городе 100 жилых домов, исследователь может выбрать любые 10 и опросить всех жителей этих домов.

Направленный отбор

- «Направленным отбором» называют такой способ отбора, при котором при формировании выборки не соблюдаются условия случайности.
 - Случайный отбор возможен лишь в том случае, если имеется «основа выборки» - полный список единиц отбора, входящих в генеральную совокупность. В противном случае использование направленного отбора неизбежно.
 - Часто в исследованиях используется «функциональная выборка» («целевая выборка»), которая формируется из объектов, удовлетворяющих некоторым критериям («молодежь», «студенчество»). Такой вид отбора также является направленным.
- Оценка погрешности при направленном отборе.
 - Для направленных выборок принципиально невозможно оценить случайную погрешность оценки. На практике для оценки погрешности используются те же формулы, что и для одноступенчатой вероятностной выборки.

Метод основного массива

- К стихийному отбору тесно примыкает *метод основного массива*. Метод основного массива представляет опрос **60—70%** генеральной совокупности.
- Процедура его крайне проста: из жителей данного района или работников предприятия опрашивается простое большинство. В результате средние генеральной и выборочной совокупностей сближаются, а выборочная совокупность составляет преимущественную часть генеральной и перекрывает возможное смещение.
- К подобному методу в прошлом часто прибегали заводские социологи, не искушенные в математических процедурах составления сложной выборки, зато располагающие материальными и временными ресурсами для опросов.
- Кроме того, к невероятным методам отбора относятся также:
 - ♦ отбор на основе принципа удобства;
 - ♦ отбор на основе суждений.

Отбор на принципе удобства

- Смысл *метода отбора на основе принципа удобства* 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.
- Формирование *выборки на основе суждений* основано на использовании мнений квалифицированных специалистов, экспертов относительно состава выборки. На основе такого подхода часто формируется состав фокус-группы



Расчет числа интервьюеров

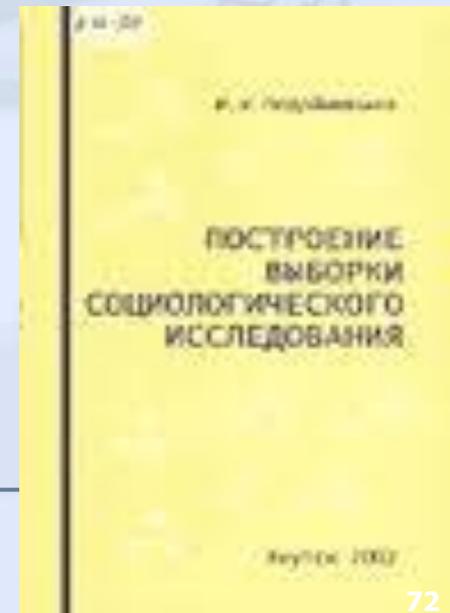
$$\checkmark I = n/a*t$$

- где I – оптимальное число интервьюеров;
- t – сроки (число дней), за которые предполагается провести исследование;
- a – норма опроса респондентов в день (при индивидуальном интервью – не более пяти-семи человек);
- n – объем выборочной совокупности (число подлежащих
- опросу).



Расчет числа интервьюеров

- Предположим, что опросу подлежит 1000 человек, а норма интервью – пять респондентов в день. Тогда для проведения
- опроса за 10 дней потребуется привлечь:
- $1000/5 * 10 = 20$ интервьюеров
- Можно также рассчитать
- количество кодировщиков



Пример плана выборочного наблюдения

- Требовалось изучить поведение в туристических поездках жителей Вологды. Опишем, как был составлен план выборочного наблюдения.
- Семьи стратифицировали на три района Вологды : северный, центральный и южный. Для отбора семей воспользовались случайным компьютерным набором телефонных номеров.

Составление плана

- Из каждой семьи отбирались кандидаты, соответствующие четырем критериям:
 - 1. Возраст 25 лет и старше.
 - 2. Проживает в Вологде как минимум 7 месяцев в году.
 - 3. Прожил в Вологде как минимум 2 года.
 - 4. Получал паспорт в Вологде.

- Для отбора респондента из каждой семьи требуется перечислить всех членов семьи, удовлетворяющих четырем критериям и из них выбрать того, кто следующим отпразднует свой день рождения.

План выборочного наблюдения

- **Исследуемая совокупность:**
 - Элементы совокупности - люди, отвечающие критериям
 - Единицы выборки - в семье с телефоном
 - Территория - в Вологде
 - Время - в период проведения опроса

- **2. Основа выборки:** Компьютерная программа, случайным образом генерирующая номера телефонов.

- **3. Метод получения выборки:** Стратифицированная выборка. Три района **ВОЛОГДЫ** : северный, центральный и южный.

План выборочного наблюдения

- 4. Единица выборки: Номера работающих телефонов.
- 5. Объем выборки: 868.
- 6. Инструкции по получению выборки:
 - Поделите выборку на страты.
 - С помощью компьютера наберите произвольные телефонные номера.
 - Перечислите всех членов семьи, отвечающих четырем критериям.
 - Выберите одного члена семьи методом следующего дня рождения.

Методы формирования выборки



Литература по теме занятия

- Дулина Н. В., Токарев В. В. Элементы прикладной социологии. Часть I: Учебное пособие / ВолгГТУ - Волгоград, 1996. – 142 с.
- Дулина, Н. В. Основы прикладной социологии: учеб. пособие / Н. В. Дулина, И. А. Небыков, В. В. Токарев / ВолгГТУ. – Волгоград, 2006. –155 с.
- Дулина Н. В., Небыков И. А., Ситникова О. И. Практикум по социологии: Учебное пособие / ВолгГТУ. – Волгоград, 2004. – 132 с.
- Рабочая книга социолога / отв. ред. Г. В. Осипов. – М.: КомКнига, 2006. – 480 с.
- Российская социологическая энциклопедия. Под общей редакцией Г. В. Осипова. – М.: Издательская группа НОРМА ИНФРА М, 1999. – 672 с.
- Ядов, В. А. Социологическое исследование: методология, программа, методы / В. А. Ядов; изд. 2-е, перераб и доп. – М.: Наука, 1987. – 248 с.
- Ядов, В. А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В. А. Ядов – М.: «Добросвет», Книжный дом «Университет», 1998. – 596 с.