Случайные выборки

12-1. Выборочное наблюдение

Генеральная совокупность и выборка

Генеральная совокупность (population) – вся интересующая исследователя совокупность изучаемых объектов.

Выборка (sample) – некоторая, обычно небольшая, часть генеральной совокупности, отбираемая специальным образом и исследуемая с целью получения выводов о свойствах генеральной совокупности.

Основные требования к выборке

- репрезентативность (отображение основных параметров генеральной совокупности – ее структуры и состояния),
- типичность
- объективность (исключение всех субъективных влияний на состав выборки)

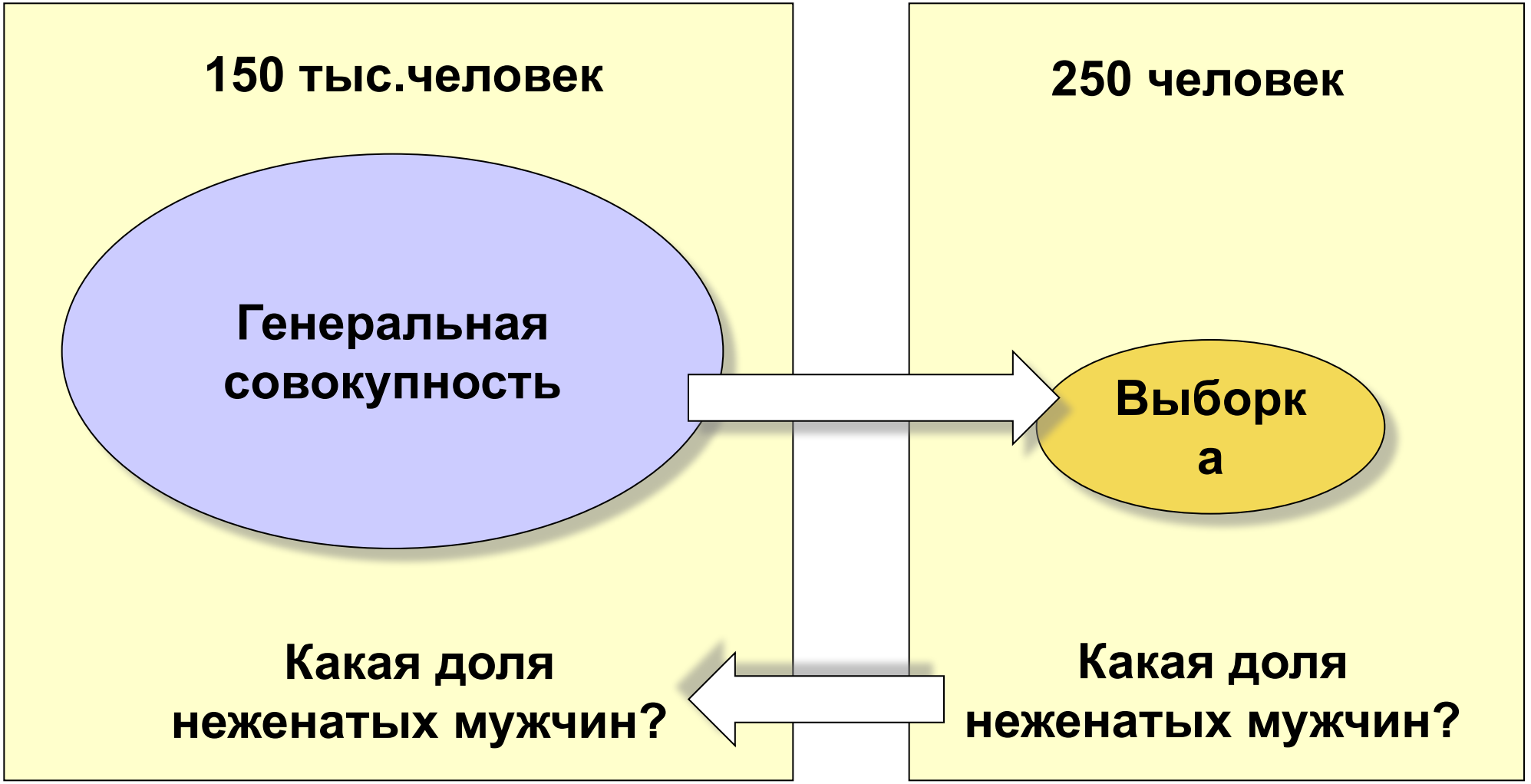
Основа выборки – полный и точный перечень единиц выборочной совокупности;

объем – количество единиц в выборочной совокупности;

метод – способ построения выборочной совокупности;

структура – процентные пропорции признаков объекта, на основе которых формируется выборочная совокупность.

Генеральная совокупность и выборка



Репрезентативная выборка

Репрезентативная выборка *хорошо* представляет генеральную совокупность.

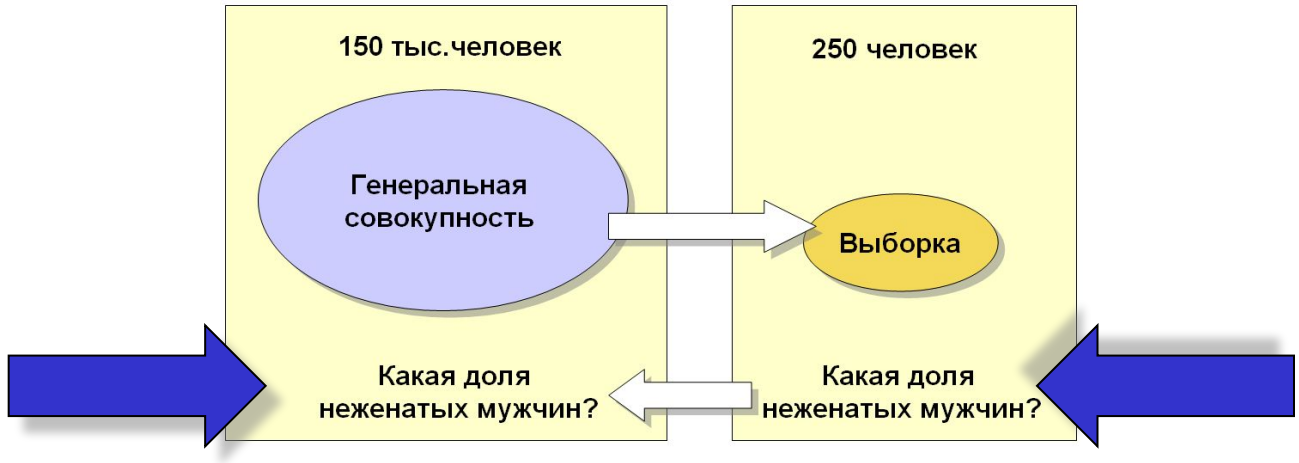
Это означает, что каждое свойство (или комбинация свойств) наблюдается в выборке с той же частотой, что и в генеральной совокупности.

Параметры и статистики

Параметры - характеристики генеральной совокупности.

Статистики - характеристики выборки.

Мы будем использовать статистики для оценки параметров генеральной совокупности, которым они соответствуют.



Параметр **Статистика**

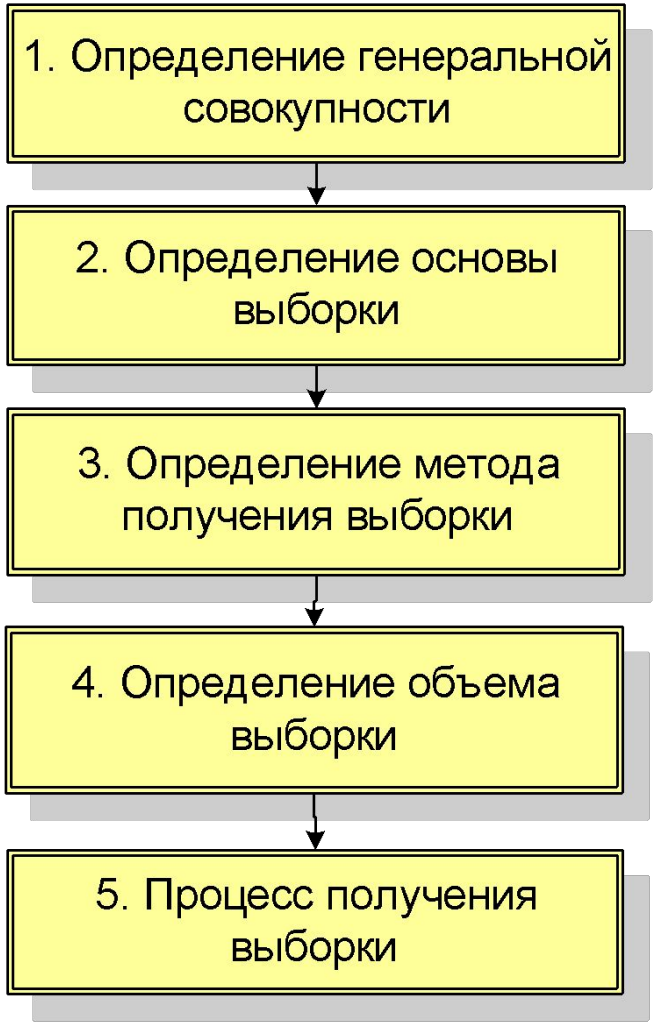
7.1. Выборочное наблюдение

План выборочного наблюдения

Методы получения выборок

Простая случайная выборка

План выборочного наблюдения



Целевая генеральная совокупность

На первом этапе следует 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:

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

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.

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.

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.

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

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.

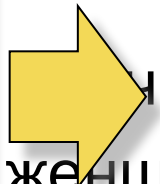
Пример стратифицированной выборки

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	18	12
женщины	42	28

Всего 100 человек

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.

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.

Пример плана выборочного наблюдения

Требовалось изучить поведение в туристических поездках жителей штата Флорида. Опишем, как был составлен план выборочного наблюдения.

Семьи стратифицировали на три района Флориды: 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 во Флориде как минимум 7 месяцев в году.
3. Прожил во Флориде как минимум 2 года.
4. Получал водительские права во Флориде.

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

1. Изучаемая совокупность:

Элементы совокупности - люди, отвечающие критериям

Единицы выборки - в семье с телефоном

Территория - в штате Флорида

Время - в период проведения опроса

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

3. Метод получения выборки: Стратифицированная выборка. Три района Флориды: 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.