

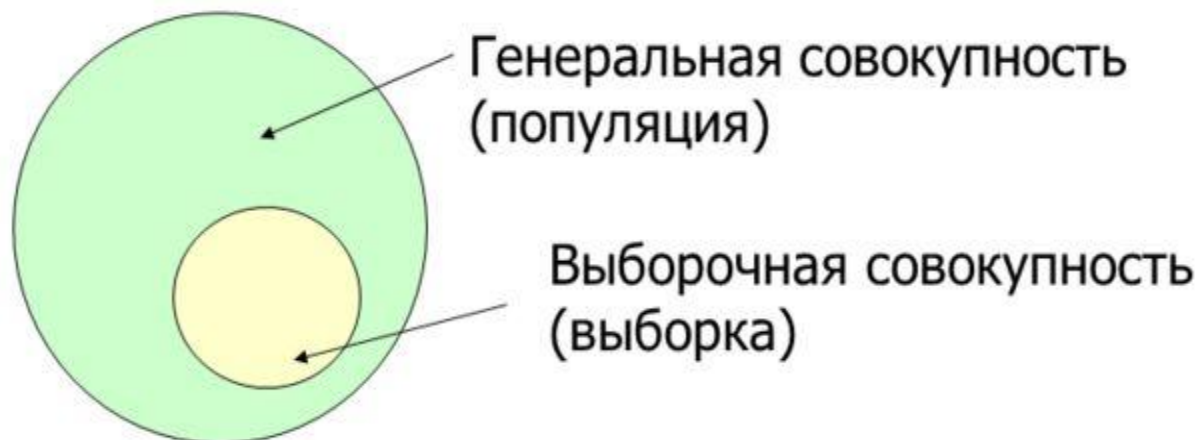
Выборки

# Целевая генеральная совокупность

- совокупность объектов, обладающих информацией, которую желает получить исследователь и о которой требуется сделать заключение.

Какие объекты попадут в эту совокупность, где ее границы?

# Генеральная совокупность и выборка



## **Наилучший подход:**

каждый участник генеральной совокупности имеет равную вероятность быть включенным в выборку

# Перепись

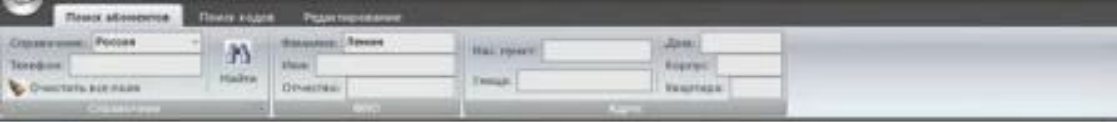
Если изучается вся  
генеральная  
совокупность, то  
выборка называется  
переписью (census)



# Определение основы выборки

Основа выборочного наблюдения (sampling frame) – некоторое представление элементов изучаемой генеральной совокупности. Это список всех объектов или перечень инструкций для определения границ и объектов изучаемой совокупности

Примеры: телефонная книга, список адресатов, отраслевые справочники предприятий



The screenshot shows a web application with a search bar at the top. Below the search bar, there is a table with columns: Телефон, Полный номер, Полное имя, Нас. пункт, Улица, Дом, Кв., Адрес, Страна, Город. The table contains multiple rows of data, including addresses in Russia and other countries.

| Телефон | Полный номер     | Полное имя                | Нас. пункт | Улица                    | Дом     | Кв. | Адрес                                       | Страна | Город           |
|---------|------------------|---------------------------|------------|--------------------------|---------|-----|---------------------------------------------|--------|-----------------|
| 495306  | +7 (495) 335306  | Левин С. И.               |            | Ильинский ул.            | 18      | 37  | Ильинский ул., д. 18, кв. 37                | Россия | Екатеринбург    |
| 495264  | +7 (2462) 495264 | Левин Ф. Ал. Фокс         |            | Ильинский ул.            | 113     | 0   | Ильинский ул., д. 113                       | Россия | Томск           |
| 363008  | +7 (3962) 363008 | Левин С. И.               |            | Синафонов ул.            | 235-А   | 29  | Синафонов ул., д. 235-А, кв. 29             | Россия | Красноярск      |
| 338503  | +7 (3953) 338503 | Левин Владимир Михайлович | Энергия    | Генерал ул.              | 60      | 27  | Энергия, Генерал ул., д. 18, кв. 27         | Россия | Братск          |
| 517284  | +7 (4842) 517284 | Левин А. И.               |            |                          | 6       | 0   | д. 6                                        | Россия | Калуга          |
| 518625  | +7 (4842) 518625 | Левин С. И.               |            |                          | 7       | 0   | д. 7                                        | Россия | Калуга          |
| 519753  | +7 (4842) 519753 | Левин С. И.               |            |                          | 12      | 0   | д. 12                                       | Россия | Калуга          |
| 504359  | +7 (4842) 504359 | Левин С. И.               |            |                          | 102     | 0   | д. 102                                      | Россия | Калуга          |
| 320462  | +7 (4842) 320462 | Левин М. В.               |            |                          | 38      | 0   | д. 38                                       | Россия | Калуга          |
| 451318  | +7 (4842) 451318 | Левин С. И.               |            | Щенная ул.               | 66      | 8   | Щенная ул., д. 66, кв. 8                    | Россия | Орен            |
| 1486089 | +7 (495) 1486089 | Левин А. Н.               |            | Душина Ольга ул.         | 13 к. 2 | 476 | Душина Ольга ул., д. 13 к. 2, кв. 476       | Россия | Москва          |
| 3364361 | +7 (495) 3364361 | Левин С. И.               |            |                          | 0       | 0   |                                             | Россия | Москва          |
| 1281154 | +7 (495) 1281154 | Левин С. И.               |            | Васкова архитектора ул.  | 11 к. 2 | 41  | Васкова архитектора ул., д. 11 к. 2, кв. 41 | Россия | Москва          |
| 1179806 | +7 (495) 1179806 | Левин С. И.               |            | Андропова просп.         | 19      | 102 | Андропова просп., д. 19, кв. 102            | Россия | Москва          |
| 9198894 | +7 (495) 9198894 | Левин С. И.               |            | Текстильщиков 11-й ул.   | 9       | 32  | Текстильщиков 11-й ул., д. 9, кв. 32        | Россия | Москва          |
| 3364438 | +7 (495) 3364438 | Левин С. И.               |            | Антонова генерала ул.    | 5 к. 2  | 55  | Антонова генерала ул., д. 5 к. 2, кв. 55    | Россия | Москва          |
| 72985   | +7 (8153) 72985  | Левин А. Н.               |            | Зеленая ул.              | 22      | 86  | Зеленая ул., д. 22, кв. 86                  | Россия | Великие Луки    |
| 5200836 | +7 (812) 5200836 | Левин С. Г.               |            | Зеленая пр.              | 54/31   | 156 | Зеленая пр., д. 54/31, кв. 156              | Россия | Санкт-Петербург |
| 5482941 | +7 (812) 5482941 | Левин С. И.               |            | Металлистов пр.          | 132/39  | 69  | Металлистов пр., д. 132/39, кв. 69          | Россия | Санкт-Петербург |
| 2544551 | +7 (812) 2544551 | Левин С. И.               |            | Новаторов б-р            | 50/39   | 61  | Новаторов б-р, д. 50/39, кв. 61             | Россия | Санкт-Петербург |
| 761823  | +7 (8142) 761823 | Левин Владимир Сергеевич  |            | Ленина пр.               | 37      | 63  | Ленина пр., д. 37, кв. 63                   | Россия | Петрозаводск    |
| 761825  | +7 (8142) 761825 | Левин Владимир Сергеевич  |            | Мирная пр.               | 12      | 25  | Мирная пр., д. 12, кв. 25                   | Россия | Петрозаводск    |
| 505771  | +7 (8202) 505771 | Левин О. Л.               |            | Вологодская ул.          | 56      | 51  | Вологодская ул., д. 56, кв. 51              | Россия | Череповец       |
| 949616  | +7 (8312) 949616 | Левин С. И.               |            | Первомайская ул.         | 22      | 0   | Первомайская ул., д. 22                     | Россия | Новый Новгород  |
| 439581  | +7 (8312) 439581 | Левин М. А.               |            | Защита ул.               | 10      | 18  | Защита ул., д. 10, кв. 18                   | Россия | Новый Новгород  |
| 318063  | +7 (8312) 318063 | Левин М. В.               |            | Мирная ул.               | 11      | 14  | Мирная ул., д. 11, кв. 14                   | Россия | Новый Новгород  |
| 5177686 | +7 (843) 5177686 | Левин А. А.               |            | Якушев К. пр.            | 51      | 183 | Якушев К. пр., д. 51, кв. 183               | Россия | Новый Новгород  |
| 5213923 | +7 (843) 5213923 | Левин А. М.               |            | Лаврентьева академик ул. | 2       | 11  | Лаврентьева академик ул., д. 2, кв. 11      | Россия | Новый Новгород  |
| 2255605 | +7 (847) 2255605 | Левин С. И.               |            | Мирная ул.               | 47      | 45  | Мирная ул., д. 47, кв. 45                   | Россия | Курган          |

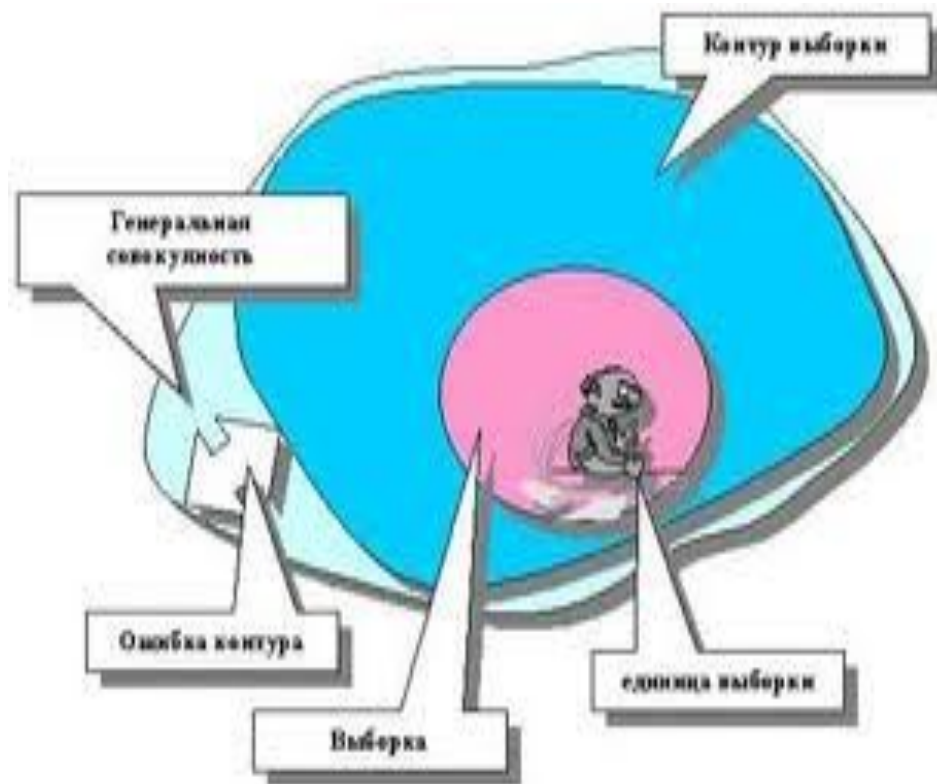
# Определение основы выборки

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



# Ошибки при составлении основы выборки

- Пропуск некоторых объектов
- Включение в основу объектов, не относящихся к изучаемой генеральной совокупности
- Имеются методы устранения или минимизации ошибок



# Методы получения выборки

## Детерминированные методы:

- Нерепрезентативная выборка
- Поверхностная выборка
- Квотная выборка
- Выборка по принципу «снежного кома»

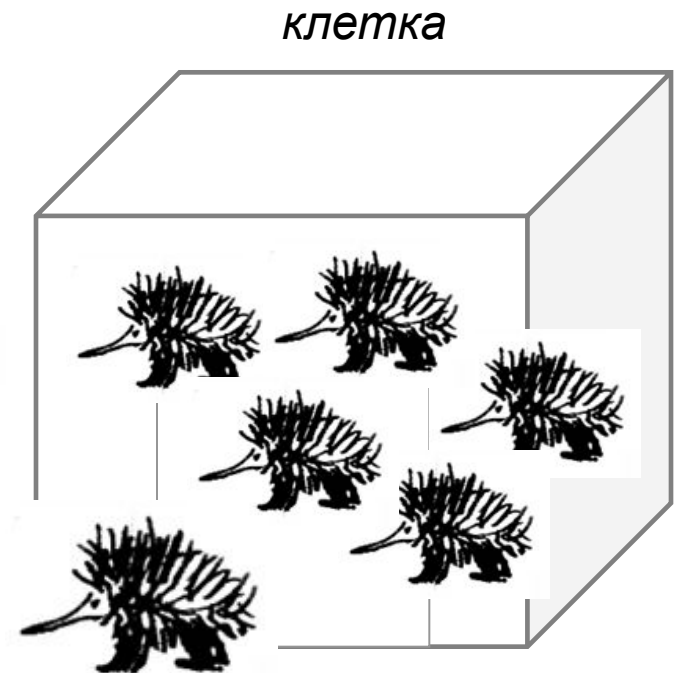
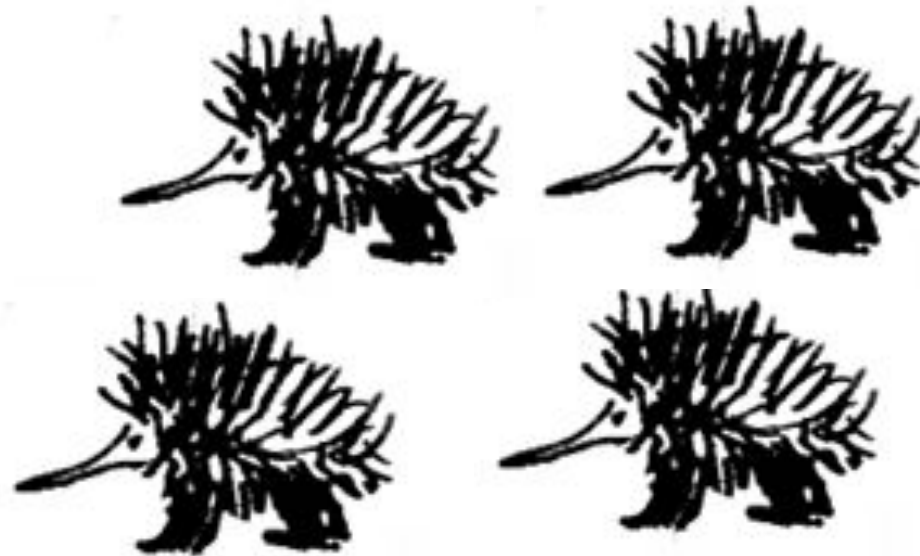
## Вероятностные методы:

- Простая случайная выборка
- Систематическая выборка
- Стратифицированная выборка
- Кластерная выборка



Выборка должна быть **РЕПРЕЗЕНТАТИВНОЙ**, т.е. её свойства должны отражать свойства популяции.

Для этого она должна быть **СЛУЧАЙНОЙ** (random) – т.е., все особи в популяции должны иметь одинаковые шансы попасть в неё, и попадание в выборку одного элемента не должно влиять на попадание другого элемента.



# Фиксированная выборка

- **Фиксированная выборка имеет фиксированный объем,**

подразумевает априорное определение ее размера и получение информации только от выбранных элементов.

- **Последовательная выборка подразумевает возможность**

принятия дополнительных решений в ходе ее формирования. Если при использовании небольшой выборки полученные результаты не позволяют сделать надежных выводов, то сбор информации продолжается.

# Удобная выборка

Удобная выборка  
(произвольная,  
акцидентная) включает  
элементы, которые  
встретились в  
соответствующий момент  
времени в месте сбора  
информации.

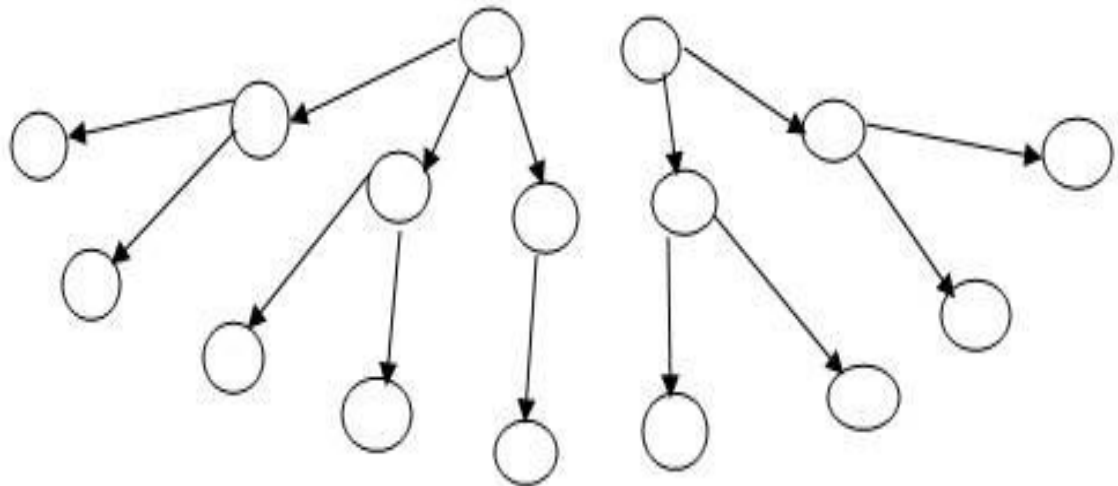
Не является  
репрезентативной!



# Типовая выборка

- **Типовая выборка (преднамеренная)** подразумевает целенаправленный выбор элементов, поскольку ожидается, что эти элементы смогут помочь выполнению задачи исследования.

Разновидностью типовой выборки является выборка **мето**



# Квотированные выборки

Квотированная выборка - это типовая выборка, на которую накладываются ограничения: оговаривается минимальное число каждой группы совокупности. Если необходимо опросить выборку из 1000 жителей города, можно оговорить квоты для северной и южной части этого города.

Квотированные выборки чаще всего основываются на демографических критериях: пол, возраст, регион, доход, образование и прочих.

# Простая случайная выборка

каждый элемент исходной совокупности имеет равную вероятность отбора, а любая комбинация элементов выборки столь же вероятна, как и любое другое сочетание ее  $n$  элементов.

*Пример: простая случайная выборка из списка*

|    | А         |  |
|----|-----------|--|
| 1  | Елена     |  |
| 2  | Петр      |  |
| 3  | Иван      |  |
| 4  | Дарья     |  |
| 5  | Оксана    |  |
| 6  | Николай   |  |
| 7  | Александр |  |
| 8  | Дмитрий   |  |
| 9  | Михаил    |  |
| 10 | Аристарх  |  |
| 11 | Мария     |  |
| 12 | Василий   |  |
| 13 | Светлана  |  |
| 14 |           |  |

# Получение простой случайной выборки

- 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.

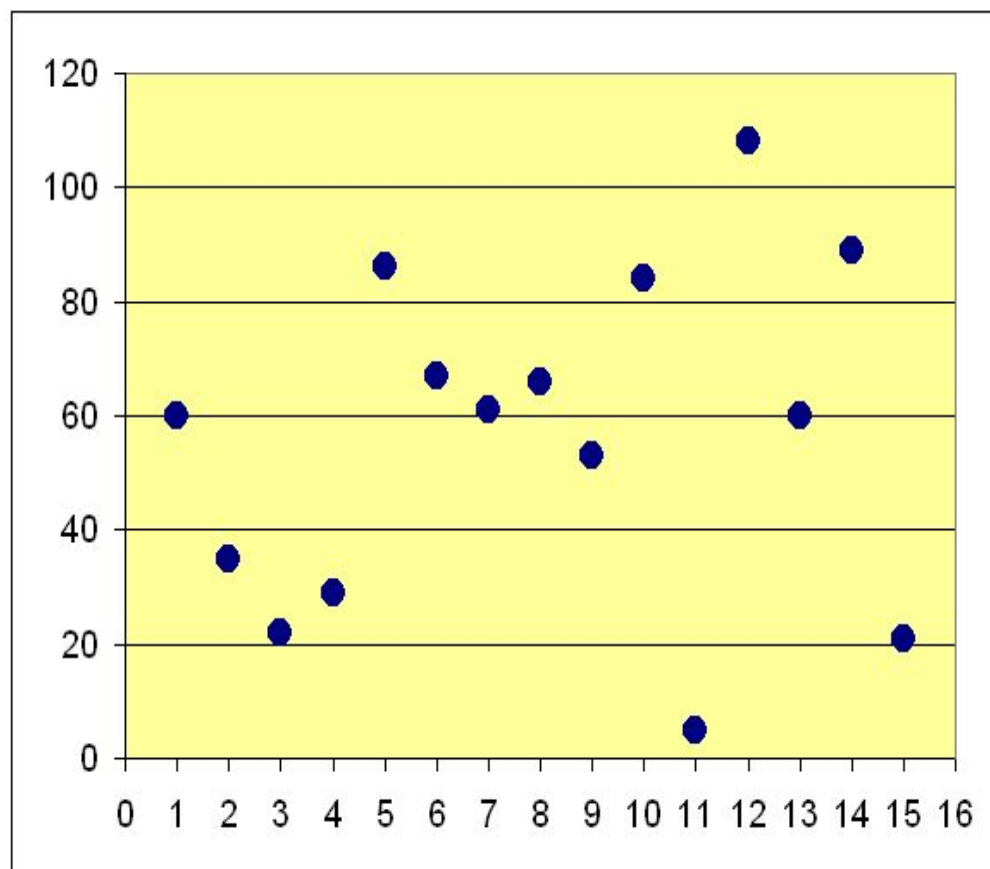
# Таблица случайных чисел

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

# Составление случайной выборки

Случайная выборка составлена в таблице Excel при помощи функции СЛЧИС().

| Номер | Случайное число | 120   | Округление |
|-------|-----------------|-------|------------|
| 1     | 0,4951          | 59,9  | 60         |
| 2     | 0,2842          | 34,8  | 35         |
| 3     | 0,1800          | 22,4  | 22         |
| 4     | 0,2329          | 28,7  | 29         |
| 5     | 0,7121          | 85,7  | 86         |
| 6     | 0,5560          | 67,2  | 67         |
| 7     | 0,5050          | 61,1  | 61         |
| 8     | 0,5442          | 65,8  | 66         |
| 9     | 0,4363          | 52,9  | 53         |
| 10    | 0,6933          | 83,5  | 84         |
| 11    | 0,0315          | 4,7   | 5          |
| 12    | 0,8985          | 107,9 | 108        |
| 13    | 0,4922          | 59,6  | 60         |
| 14    | 0,7401          | 89,1  | 89         |
| 15    | 0,1713          | 21,4  | 21         |



# Стратифицированная выборка

**Стратифицированная выборка** получается путем разбиения генеральной совокупности на подгруппы (или страты) в зависимости от характеристик, важных для изучения. Затем для каждой страты надо провести простой случайный отбор элементов.

**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

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  
совокупность

**Филологи Биологи**  
**Биологи**

|         |              |     |
|---------|--------------|-----|
| мужчины | 360          | 240 |
| женщины | 840          | 560 |
| Всего   | 2000 человек |     |

Выборочная

**Филологи**

|         |     |    |
|---------|-----|----|
| мужчины | 18  | 12 |
| женщины | 42  | 28 |
| Всего   | 100 |    |

# 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:

- случайным образом выберем первый из первых сорока элементов генеральной совокупности

- Если первым оказался номер 12, то тогда выборка будет включать объекты с номерами

12, 52, 92 и так далее,

всего 50 объектов

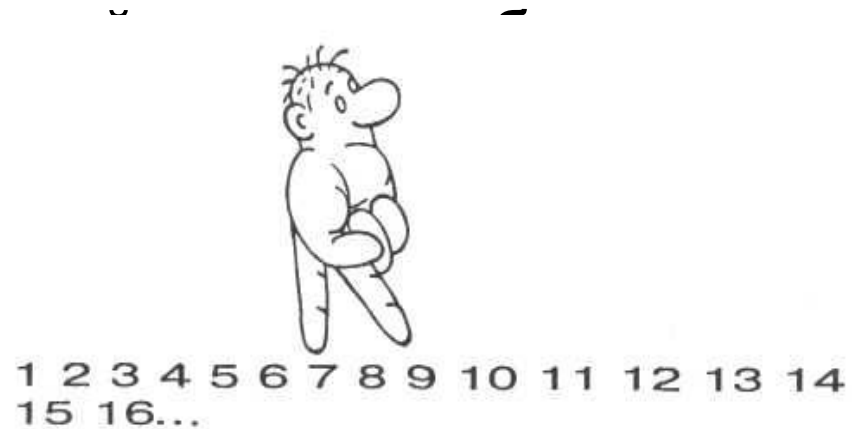


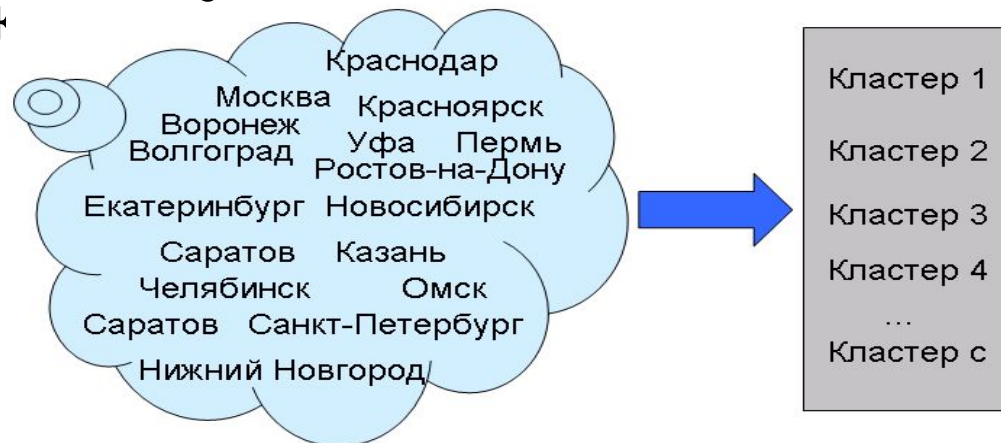
Рис. 2.3. Шаг выборки

# Кластерная выборка

- выборка, в которой исходная совокупность разделяется на непересекающиеся подгруппы (называемые кластерами), а затем из этих подгрупп формируется случайная выборка.

Пример. Исследователю необходимо опросить жителей, проживающих в квартирах небольшого города. Если в городе 100 жилых домов, исследователь может выбрать любые 10 и опросить всех ж

ДОМОВ.



# Пример плана выборочного наблюдения

Требовалось изучить поведение в туристических поездках жителей штата Флорида.

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.

# Какая выборка потребуется?

1. Почему посетители крупного торгового центра мало ходят в кинотеатр, который расположен на территории?
2. Почему крупный торговый центр в одном из районов Москвы мало посещают?
3. Почему иностранцы не посещают определенный ресторан?
4. Почему иностранцы не пользуются услугами определенного центра пластической хирургии?

# Размер выборки

Размер выборки зависит от величин допустимых  $\alpha$  и  $\beta$  ошибок

# Понятие об альфа- и бета-ошибках

**Вероятность упустить и вероятность обознаться**

В жизни, а также при проведение статистических тестов возможны два типа ошибок:

- отвергнуть правильную нулевую гипотезу
- принять неправильную нулевую гипотезу

**Нулевая гипотеза** – обычно предположение об отсутствии различий, например, 2 выборки взяты из одной генеральной совокупности

**«Различий нет, все всегда случайно»**

## Ошибка I рода ( $\alpha$ )

Вероятность отвергнуть правильную нулевую гипотезу

Вероятность обнаружить различия там,

где их нет = **Вероятность совершить фальшивое открытие**



## Ошибка II рода ( $\beta$ )

Вероятность принять неправильную нулевую гипотезу

Вероятность не обнаружить существующие различия

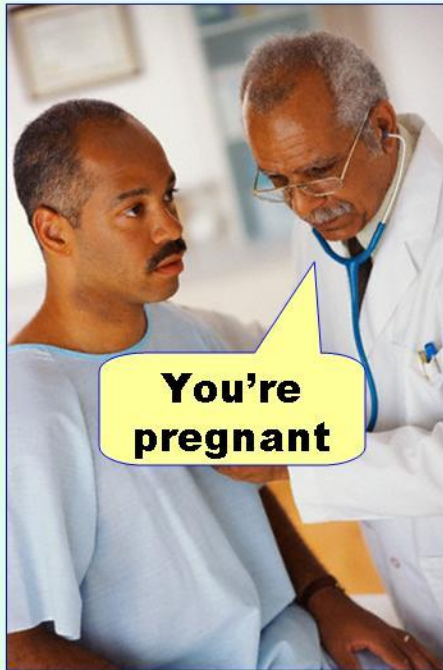
**Вероятность упустить открытие**



# Вероятность обознаться и вероятность упустить

$H_0$  – беременности  
нет

**Type I error**  
(false positive)



Отвергнута правильная  
нулевая гипотеза.  
Сделано  
ложноположительное  
открытие

**Type II error**  
(false negative)



Принята неправильная  
нулевая гипотеза.  
ложноотрицательный  
вывод. Открытие  
упущено

# Вероятность упустить и вероятность обозначаться

$\alpha$  vs.  $\beta$ :  
противоборство показателей теста

Всегда  
принимаяем  $H_0$   
 $\alpha=0, \beta=1$

Тест

Всегда  
отвергаем  $H_0$   
 $\alpha=1, \beta=0$

Ошибка I рода

Ошибка II рода

Уменьшая ошибку I рода, увеличиваем ошибку II рода,  
т.е. теряем мощность теста (*et converso*)

# Вероятность упустить и вероятность обознаться

$$\text{Мощность теста} = 1 - \beta$$

т.е. вероятность правильно отвергнуть нулевую гипотезу

- ❑ или вероятность не упустить открытие
- ❑ Мощность 80% считается приемлемой
- ❑ Консервативный тест - это тест с низкой мощностью
- ❑ Мощностью теста резко возрастает при увеличении объемов выборок
- ❑ При планировании экспериментов имеет смысл прикинуть возможную мощность

WinPepi

PORTAL

Copyright J.H. Abramson, Jan. 8, 2010. Version 10.0

WIN  
PEPI

Например, Compare2/ Power/ Comparison of proportions

Size A - 100    Size B - 100

a/A - 0.2    b/B - 0.1

Мощность = 44%

... и необходимый объем выборок

Например, Compare2/ Sample size/ Proportions

Size A/ Size B = 1

a/A - 0.2    b/B - 0.1

Общий объем выборок = 398

# $\alpha$ - и $\beta$ ошибки

- $\alpha$ -ошибка – отклонение верной нулевой гипотезы
- $\beta$ -ошибка – принятие ложной нулевой гипотезы
- $\alpha$  и  $\beta$  в биомедицинских исследованиях принимаются обычно за 0.05 и 0.20, соответственно
- Статистическая мощность ( $1-\beta$ ) – вероятность отклонения ложной нулевой гипотезы (вероятность обнаружить различия если они есть)

# Таблица для определения индекса МОЩНОСТИ

| $\alpha$ | $\beta=0.10$ | $\beta=0.20$ | $\beta=0.50$ |
|----------|--------------|--------------|--------------|
| 0.10     | 8.6          | 6.2          | 2.7          |
| 0.05     | 10.5         | 7.9          | 3.8          |
| 0.01     | 14.9         | 11.7         | 6.6          |

## Формулы для приблизительного расчета необходимого объема выборки

$$N \approx \frac{4 \cdot p(1-p)}{\Delta^2}$$

Для доли

$$N \approx 4 \cdot \left( \frac{s}{\Delta} \right)^2$$

Для средней величины

# Формулы для приблизительного расчета необходимого объема выборки

$$N \approx \frac{2 \cdot s^2 \cdot (z_\alpha + z_\beta)^2}{\Delta^2}$$

$$(z_\alpha + z_\beta)^2 = \textit{PowerIndex}(PI)$$

$\Delta^2$  – Искомая разница

PI – Можно найти в таблице

Формула предназначена только для сравнения средних арифметических  
(N – количество наблюдений в одной группе)

# Формулы для приблизительного расчета необходимого объема выборки

$$N \approx \frac{\bar{p} \cdot (1 - \bar{p}) \cdot PI}{\Delta^2}$$

$\bar{p}$  Среднее для двух изучаемых долей

PI – Можно найти в таблице

Формула предназначена только для сравнения долей  
(N – количество наблюдений в одной группе)

# Практическое задание 1

- Вы готовите заявку на грант на исследование нового препарата, снижающего АД в среднем на 10 мм рт. ст., зная, что  $\sigma = 10$  мм рт. ст. Представьте доказательства того, что Вам достаточно всего 32 испытуемых для выявления различий при уровнях  $\alpha = 0.05$  и  $\beta = 0.20$ .

# Практическое задание 2

- Сколько человек необходимо обследовать для определения средне-популяционного уровня АД с точностью до 3% в Архангельске, если  $\sigma = 15$  мм рт. ст.? При каком условии?