

Типы шкал измерения

- ▣ Шкала наименований (номинальная) – только описывает объект, количественные показатели не используются.
 - ▣ Оцениваемому объекту приписывается значение, отражающей его принадлежность к группе.
- ▣ Шкала порядка (ранговая) – размечает объекты по степени выраженности у них того или иного признака.
 - ▣ Позволяет упорядочить любую пару объектов относительно друг друга на шкале.
- ▣ Интервальная шкала не только упорядочивает объекты, но и количественно описывает их отличия.
 - ▣ Результаты, представленные в интервальной шкале, позволяют указывать «на сколько» степень выраженности признака у одного объекта больше, чем у другого.
- ▣ Шкала отношений это интервальная шкала с «нулевой точкой».
 - ▣ Значения, измеренные в шкале отношений, позволяют указать «во сколько раз» различаются показатели.



Типы шкал измерения

- Номера бегунов (номер на майке) - номинальная шкала.
 - Используется только для различения спортсменов и не для чего больше.
- Очередность прихода к финишу - порядковая шкала.
 - Порядок мест победителей на пьедестале почета.
- Рейтинг спортсмена - интервальная шкала.
 - Рейтинг спортсмена по 10-балльной шкале (от 1 до 10).
- Время прохождения дистанции в секундах - шкала отношений.
 - Можно говорить о том, «на сколько» и «во сколько» раз быстрее прошел дистанцию каждый из бегунов.



Свойства шкал измерения

ВИДЫ ШКАЛЫ	Отношения, определенные на множестве шкальных значений	Примеры
Номинальная	Только отношение эквивалентности	Национальность, семейное положение, род занятий
Порядковая	Дополнительно – транзитивность: если $A > B$ и $B > C$, то $A > C$	Уровень образования, военные ранги, разряды тарифной сетки
Интервальная	Дополнительно: симметричность, коммутативность, субстантивность, ассоциативность	Год рождения, температура по Цельсию
Шкала отношений	Любые числовые операции	Доход, температура по Кельвину

«Простые» шкалы. Минимальное количество «разрешенных» методов обработки

- «Пол» - признак, представленный в номинальной шкале.
- «Доля женщин» - признак, представленный в шкале отношений.

«Хорошие» шкалы. Допускают применение разнообразных статистических методов обработки



Свойства шкал измерения

- ☞ Чем более «грубой» является шкала, тем меньше методов можно использовать при обработке данных, но тем ниже вероятность ошибиться при измерениях.

Вид шкалы	Примеры	Описательные статистики	Способы обработки, статистические критерии
Номинальная шкала (шкала наименований)	Номера торговых марок, коды магазинов, классификация по полу и району проживания	Процентные отношения (доля категории в совокупности), мода	Критерий хи-квадрат, биномиальный критерий
Порядковая (ординальная) шкала	Сорт товара (ранг качества), ранги предпочтений, место на рынке, принадлежность к социальной группе	Дополнительно: процентиля, медиана	Ранговая корреляция, дисперсионный анализ
Интервальная шкала	Числа обозначают степень различия между объектами. Нулевая точка шкалы выбирается произвольно	Дополнительно: диапазоны (размах), среднее значение, стандартное отклонение	Коэффициент корреляции Пирсона, t-критерий, регрессионный анализ, факторный анализ, дисперсионный анализ
Шкала отношений	Числа обозначают степень различия между объектами. Точка начала отсчета фиксирована и соответствует «отсутствию признака»	Дополнительно: среднее геометрическое, среднее гармоническое	Дополнительно: коэффициент вариации

«Простые» шкалы.
Минимальная
вероятность допустить
ошибку при измерении

То, что можно сделать
просто – должно
быть сделано
просто.



Стратегии создания выборки

Вероятностные

- Величина и характер выборочной ошибки

предсказуемы:

- Можно обобщать выводы на генеральную совокупность
- Можно оценивать достоверность наших выводов

- Обычно используются в **социологии** (и иногда в образовании).

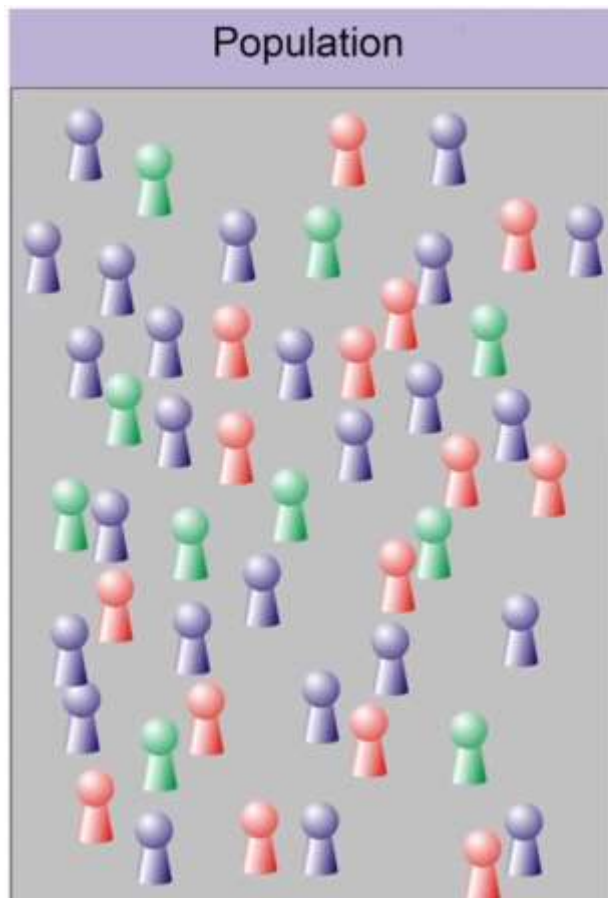
Невероятностные

- Отклонения статистик на выборке от их истинных значений

непредсказуемы:

- Невозможно понять, насколько картина, наблюдаемая на выборке, характерна для генеральной совокупности
- Не очень хорошо!

- Часто используются в **психологии** (и иногда в образовании).

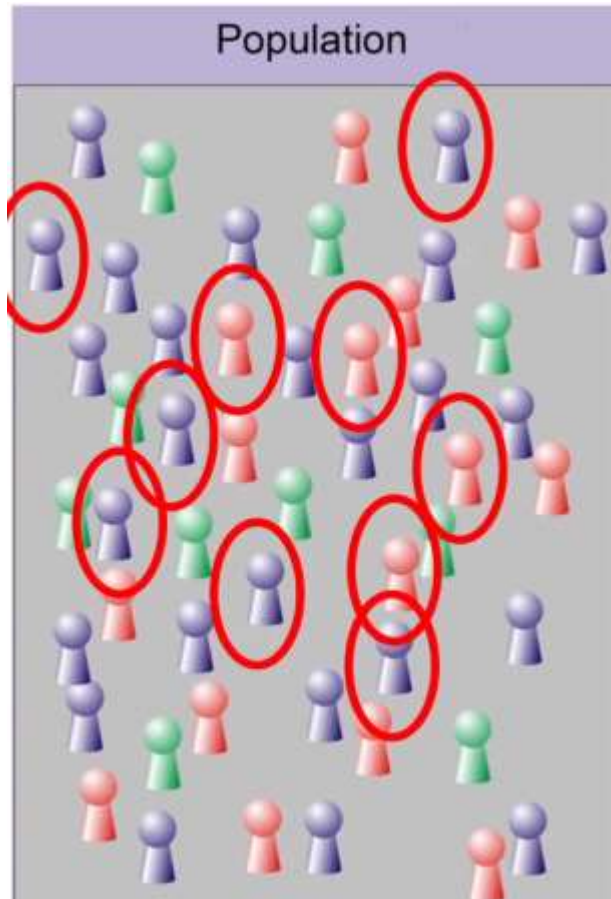


Генеральная совокупность
($N = 50$):

Синие:	$N = 25$ (50%)
Красные:	$N = 15$ (30%)
Зеленые:	$N = 10$ (20%)

Сделаем выборку объёма
 $N = 10$.

Случайная выборка



Генеральная совокупность:

Синие: $N = 25$ (50%)

Красные: $N = 15$ (30%)

Зеленые: $N = 10$ (20%)

Простая случайная выборка:

Синие: $N = 6$ (60%)

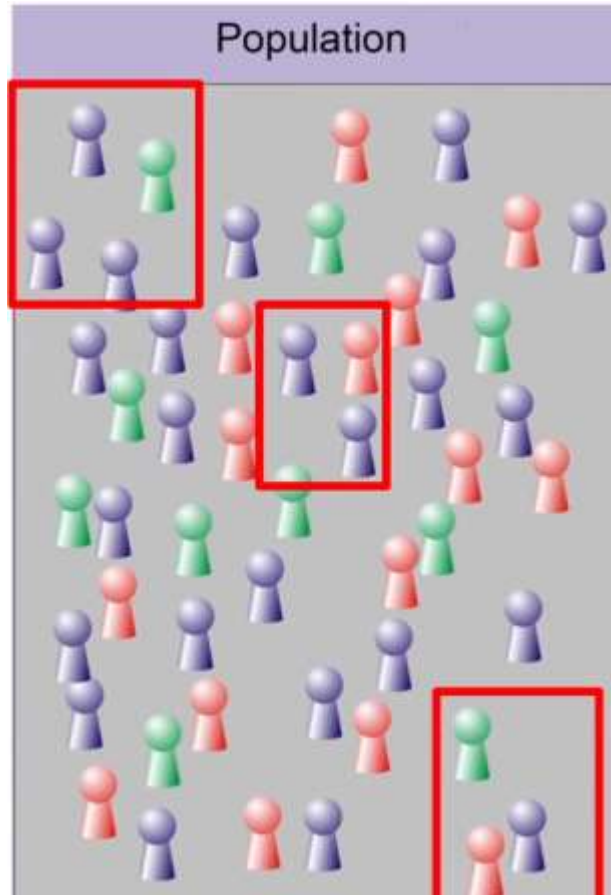
Красные: $N = 4$ (40%)

Зелёные: $N = 0$ (0%)

Мы выбрали 10 индивидов
совершенно случайно...

...и ни один зелёный в нашу
выборку не попал!

Кластерная выборка



Генеральная совокупность:

Синие: $N = 25$ (50%)

Красные: $N = 15$ (30%)

Зелёные: $N = 10$ (20%)

Кластерная выборка

Синие: $N = 6$ (60%)

Красные: $N = 2$ (20%)

Зелёные: $N = 2$ (20%)

Мы вытаскиваем группу
индивидов по соседству
друг с другом:

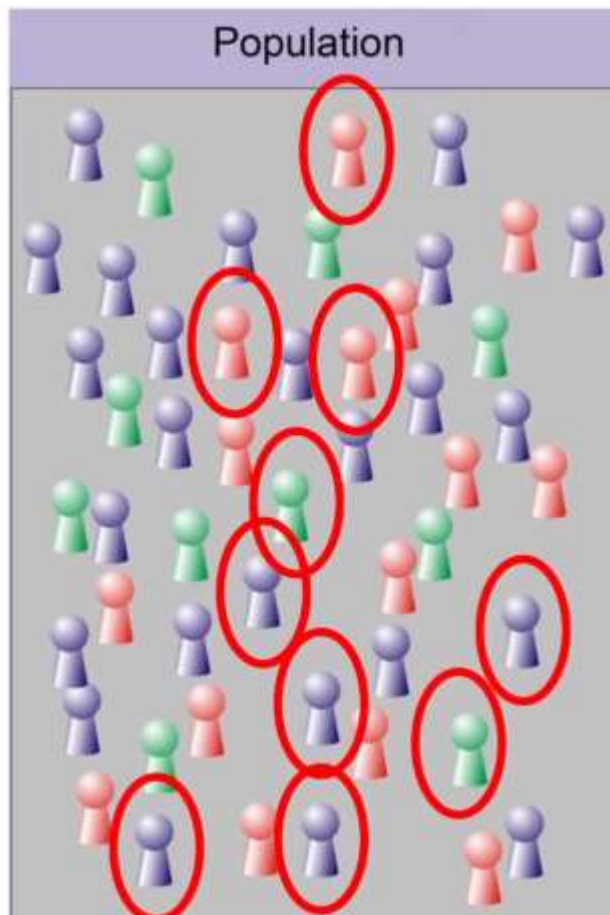
(+) Экономим время и силы

(-) Но выборочная ошибка
растёт, по сравнению со
случайной выборкой

Вероятностные стратегии

- **Стратифицированный отбор:**

- мы делим генеральную совокупность на страты (типы) и **случайно выбираем** людей из каждой страты (= равная вероятность) в нужной нам пропорции;
- на практике позволяет учесть ограниченный набор характеристик;
- **гарантирует репрезентативность** относительно тех характеристик, которые мы выбрали при создании страт;
- **выборочная ошибка ниже**, чем для случайной и тем более кластерной выборки.



Генеральная совокупность:

Синие: $N = 25$ (50%)

Красные: $N = 15$ (30%)

Зелёные: $N = 10$ (20%)

Стратифицированная случайная
выборка:

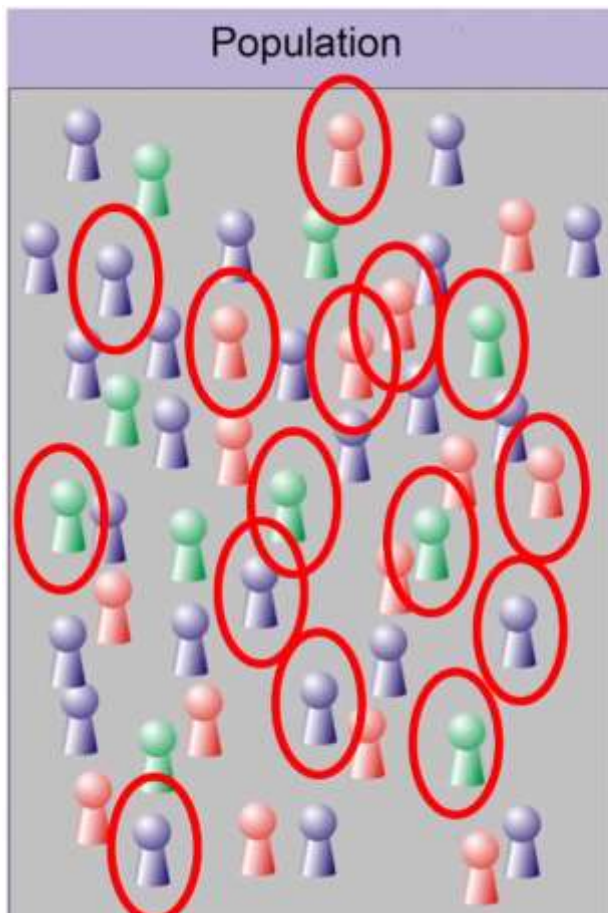
Синие: $N = 5$ (50%)

Красные: $N = 3$ (30%)

Зелёные: $N = 2$ (20%)

(Случайно выбираем нужную долю
индивидов из каждой страты →
репрезентативная выборка)

Если зелёных очень мало, мы можем
сказать: «а возьму-ка я их побольше,
чтобы точнее оценить для них среднее».
Это диспропорциональная
стратификация = для расчёта
показателей генеральной совокупности
нужно использовать **весовые
коэффициенты**. Этот подход часто
используется в масштабных
исследованиях.



Генеральная совокупность:

Синие: $N = 25$ (50%)

Красные: $N = 15$ (30%)

Зелёные: $N = 10$ (20%)

**Диспропорциональная
стратификация:**

Синие: $N = 5$ (33.3%)

Красные: $N = 5$ (33.3%)

Зелёные: $N = 5$ (33.3%)

Диспропорциональная

стратификация: я беру поровну из всех трёх видов, чтобы сравнить их, в результате:

- Каждый синий в нашей выборке представляет $25/5 = 5$ индивидов
- Каждый красный в нашей выборке представляет $15/5 = 3$ индивида
- Каждый зелёные в нашей выборке представляет $10/5 = 2$ индивида

Эти веса задают коэффициенты, на которые надо умножить средние синих, красных и зелёных в нашей выборке, чтобы получить оценку среднего для генеральной совокупности.

Невероятностные стратегии

- **Целевая выборка (purposive sample):**
 - подбираем испытуемых на основе теории.
- **Метод «снежного кома»:**
 - начинаем с нескольких ключевых испытуемых и через них выходим на других.
- **Удобная выборка (convenience sample):**
 - используем доступную выборку (группу людей, которых нам удобно опросить).
- **«Самоотбор» (self-selecting sample):**
 - используем тех, кто соглашается (или вызывается) принять участие в исследовании;
 - «эффект добровольца».

«Эффект добровольца»

- Как правило, у добровольцев:
 - более высокий уровень образования;
 - более высокий социальный статус;
 - более высокий интеллект;
 - более выраженная потребность в одобрении;
 - ...гендер: чаще женщины?
- Особенности выборки зависят от **мотивации участников** исследования (можем её задавать):
 - ситуация помощи → альтруистическая мотивация;
 - ситуация соревнования → мотивация достижения;
 - ситуация отбора → мотивация одобрения;
 - вознаграждение → материалистическая мотивация;
 - есть обратная связь → мотивация самопознания; и т.д.

Стратегии выборок: обзор

Вероятностные

- Выборочная ошибка **предсказуема**
 - Можно обобщать выборочные данные на генеральную совокупность и оценивать точность прогнозов
- **Обязательны**, если наш исследовательский вопрос касается распределения или параметров признака в популяции (генеральной совокупности).

Невероятностные

- Выборочная ошибка **непредсказуема**
 - Нельзя надёжно обобщать
- **Возможны**, если:
 - на вероятностные стратегии нет ресурсов;
 - обобщение не так важно для исследователя;
 - предполагается что изучаемый феномен одинаков у всех индивидов.