Методы уменьшения дисперсии

$$n \geq \frac{2\left(F^{-1}\left(1-\frac{\alpha}{2}\right)-F^{-1}(\beta)\right)^{2} s^{2}}{MDE^{2}}$$

s^2 – дисперсия

Сокращение дисперсии —> меньше выборка, чувствительнее метрика —> ускорение эксперимента

Стратификация

Стратификация

Стратификация

(стратифицированное семплирование + стратифицированная оценка)

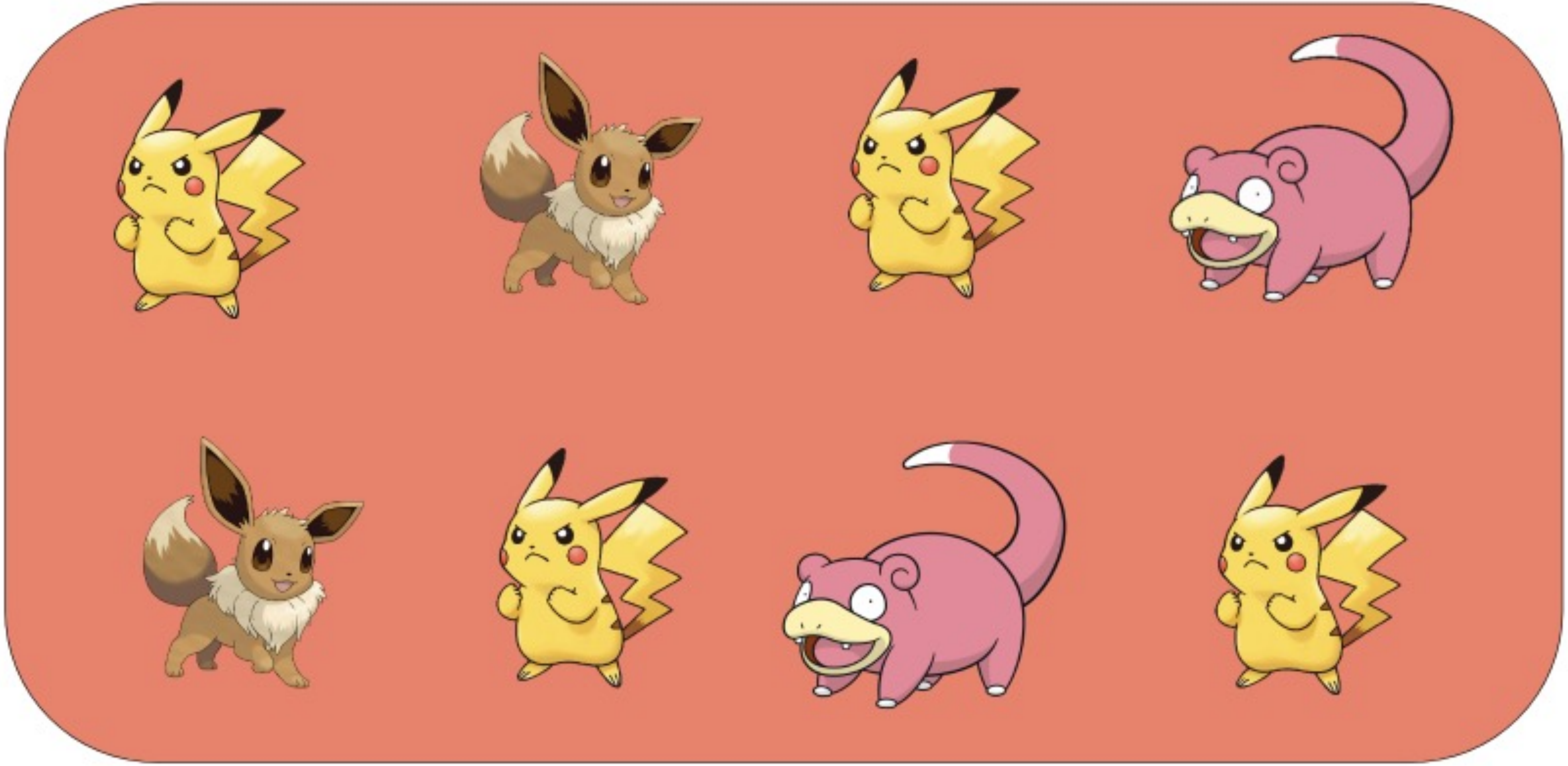
До начала эксперимента разбиваем пользователей на k страт (групп), определяем вероятность попадания в страту

Пост-стратификация

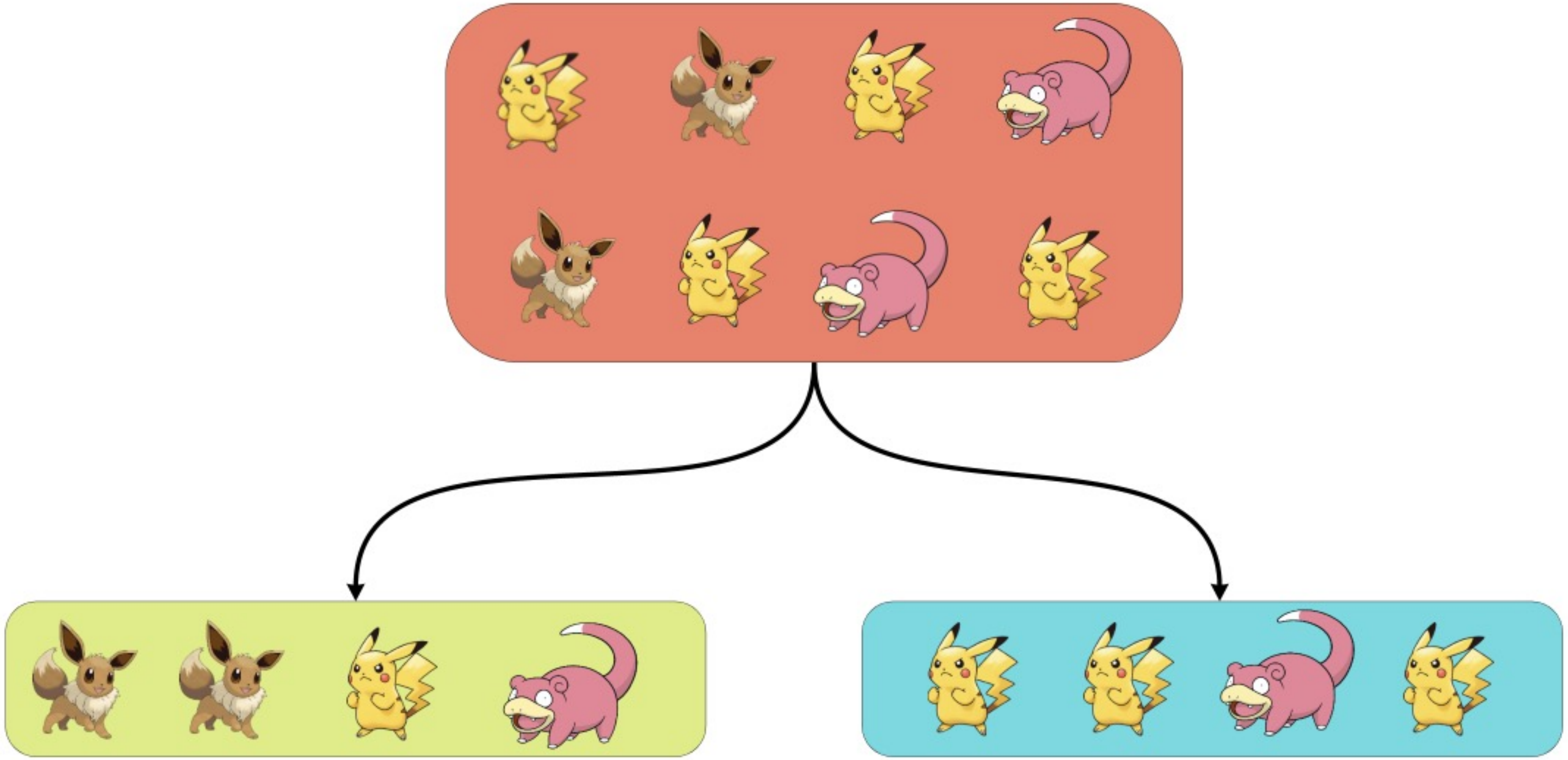
(стратифицированная оценка)

Случайная выборка + стратифицированная оценка на основе заранее известных вероятностей попадания в группу

Пусть у нас есть генеральная совокупность пользователей-покемонов

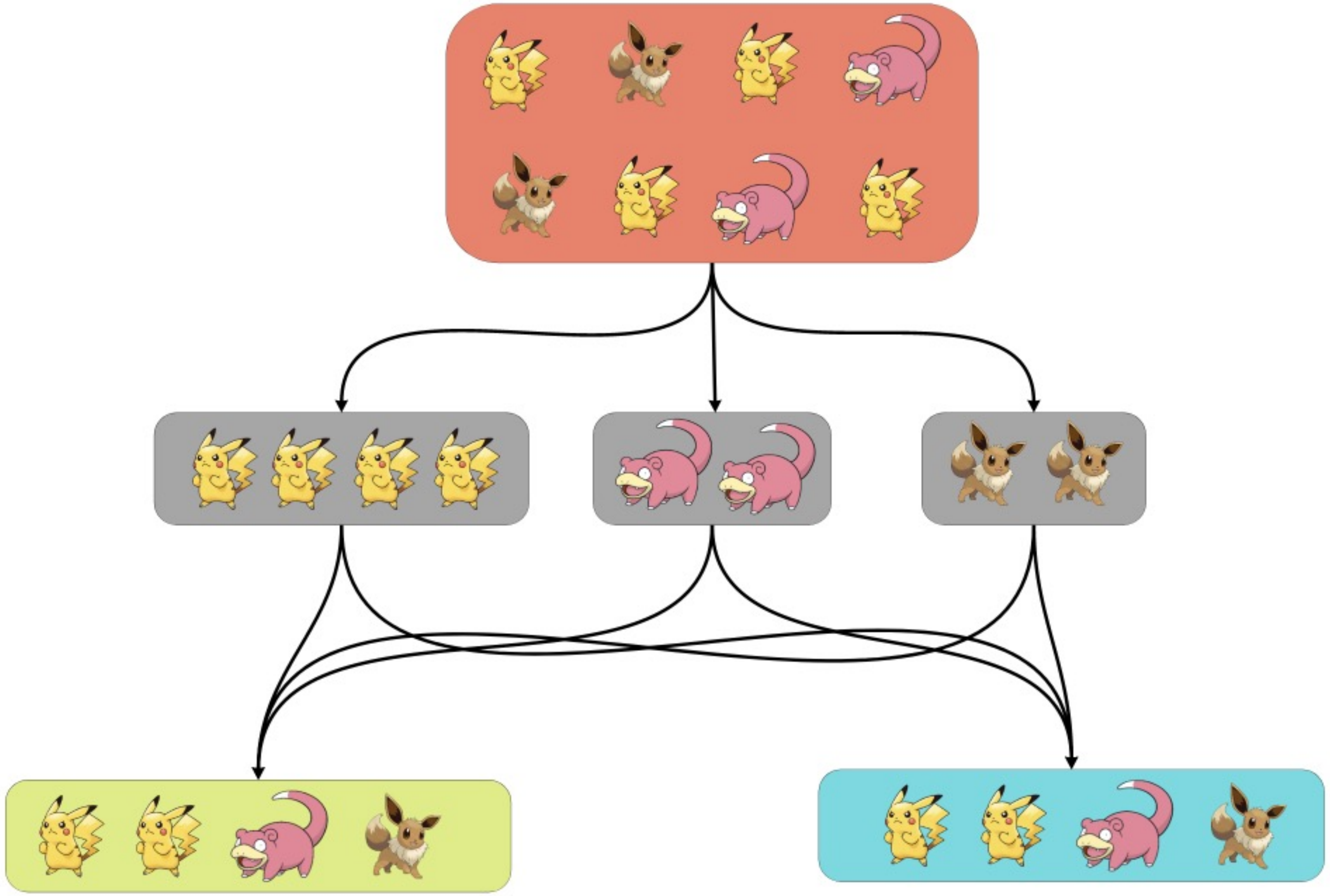


Случайное разбиение в A/B тесте по пользователям



Что если поведение покемонов разное?
Во второй выборке 3 пикачу, а в первой один. Это добавляет дополнительный шум

Стратифицированное семплирование



Чем это лучше?

Рассмотрим оценки простого среднего и стратифицированного среднего

Обозначения:

- $\mu = \mathbb{E}Y$ — популяционное среднее;
- $\sigma^2 = \mathbb{V}Y$ — популяционная дисперсия;
- μ_k, σ_k^2 — среднее значение и дисперсия бизнес метрики для k-й страты;
- w_k — доля k-й страты в популяции;
- n_k — число пользователей из k-й страты в рассматриваемой группе;
- $n = \sum_{k=1}^{K} n_k$ — общий размер группы.
- $Y_{11}, \ldots Y_{1n_1}, \ldots Y_{K1}, \ldots Y_{Kn_K}$ — выборка из г.с., где Y_{kj} — метрика для j-го пользователя k-й страты.

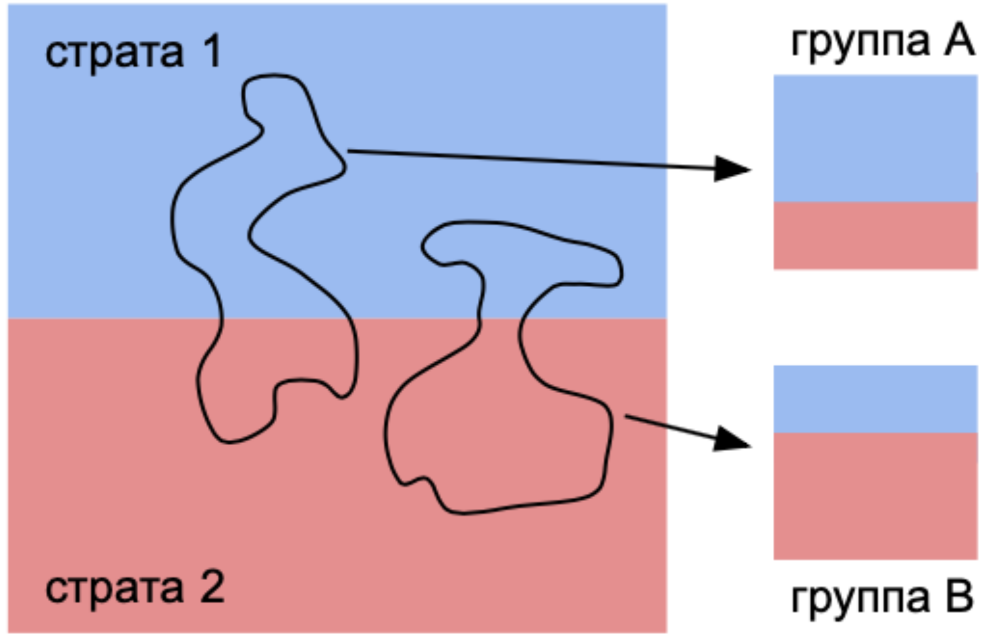
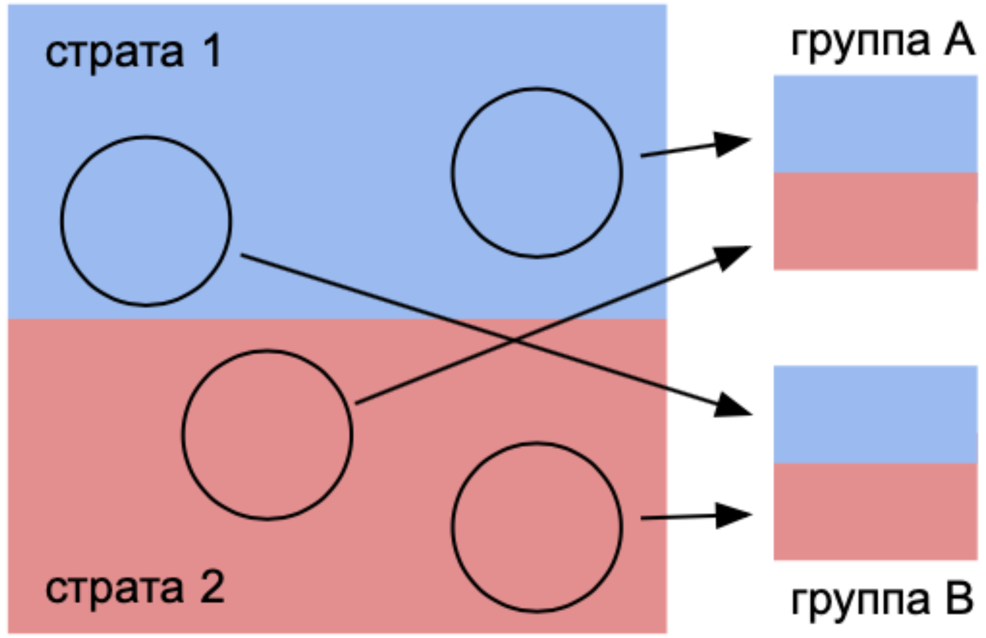
1. **Простое среднее.**

$$\overline{Y} = \frac{1}{n}\sum_{k=1}^{K}\sum_{j=1}^{n_k} Y_{kj}$$

2. **Взвешенное среднее** (стратифицированное среднее).

$$\hat{Y}_{strat} = \sum_{k=1}^{K} w_k \overline{Y}_k, \quad \overline{Y}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} Y_{kj}$$

Стратификация

	Обычное среднее $\overline{Y} = \frac{1}{n}\sum_{k=1}^{K}\sum_{j=1}^{n_k} Y_{kj}$	**Стратифицированное среднее** $\hat{Y}_{strat} = \sum_{k=1}^{K} w_k \overline{Y}_k$
Случайное семплирование 	Классический подход Без стратификации	**Постстратификация**
Стратифицированное семплирование 	Не контролирует вероятность ошибки I рода	**Стратификация**

Свойства

1. Оценка страт. среднего равна оценке выборочного среднего

$$\hat{Y}_{strat} = \sum_{k=1}^{K} w_k \overline{Y}_k = \sum_{k=1}^{K} w_k \frac{1}{n_k} \sum_{j=1}^{n_k} Y_{kj} = \sum_{k=1}^{K} \frac{n_k}{n} \frac{1}{n_k} \sum_{j=1}^{n_k} Y_{kj} = \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{n_k} Y_{kj} = \overline{Y}$$

2. Оценка мат. ожидания стратифицированного среднего равна среднему по ГС.

Более того, она является несмещенной оценкой среднего

$$E_{strat}(\hat{Y}_{strat}) = \sum_{k=1}^{K} p_k E_{strat}(\bar{Y}_k) = \sum_{k=1}^{K} p_k \mu_k = \mu.$$

Y: метрика

$\mu = E(Y)$: среднее по метрике в ГС

p_k: доля k страты в ГС = вероятность попадания в страту

Свойства стратификации. Дисперсия

$$\mathbb{D}(\bar{Y}) = \frac{1}{n}\sum_{k=1}^{K} w_k \sigma_k^2 + \frac{1}{n}\sum_{k=1}^{K} w_k (\mu_k - \mu)^2$$

внутригрупповая дисперсия

межгрупповая дисперсия

$$\mathbb{D}(\bar{Y} strat) = \frac{1}{n}\sum_{k=1}^{K} w_k \sigma_k^2$$

$$\mathbb{D}\left(\hat{Y}_{\text{post-strat}}\right) = \frac{1}{n}\sum_{k=1}^{K} w_k \sigma_k^2 + \frac{1}{n^2}\sum_{k=1}^{K}(1 - w_k)\sigma_k^2 + O\left(\frac{1}{n^2}\right)$$

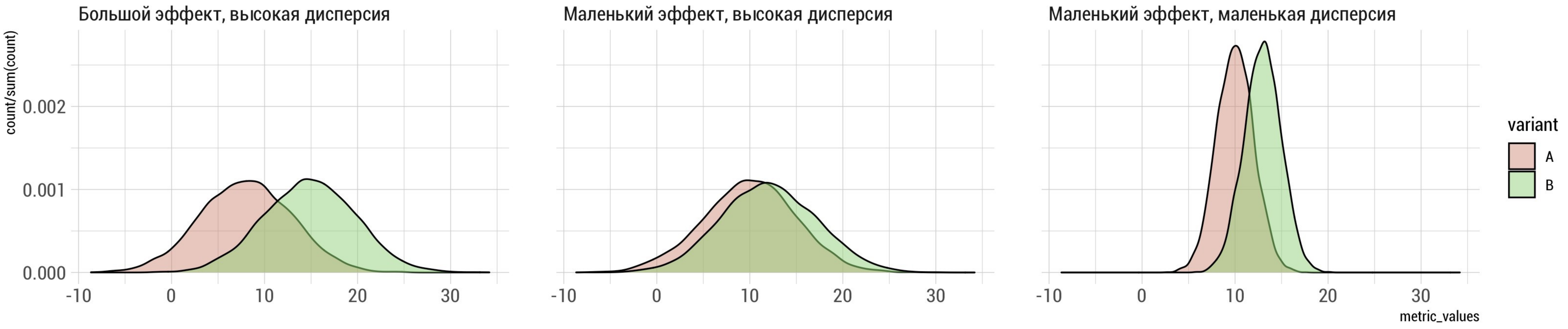
$$\mathbb{D}\left(\hat{Y}_{\text{strat}}\right) = \mathbb{D}\left(\hat{Y}_{\text{post-strat}}\right) + O\left(\frac{1}{n^2}\right) = \mathbb{D}(\bar{Y}) + O\left(\frac{1}{n}\right)$$

$$\mathbb{D}\left(\hat{Y}_{\text{strat}}\right) \leqslant \mathbb{D}\left(\hat{Y}_{\text{post-strat}}\right) \leqslant \mathbb{D}\left(\bar{Y}\right)$$

Дисперсия стратифицированного и пост-стратифицированного среднего меньше

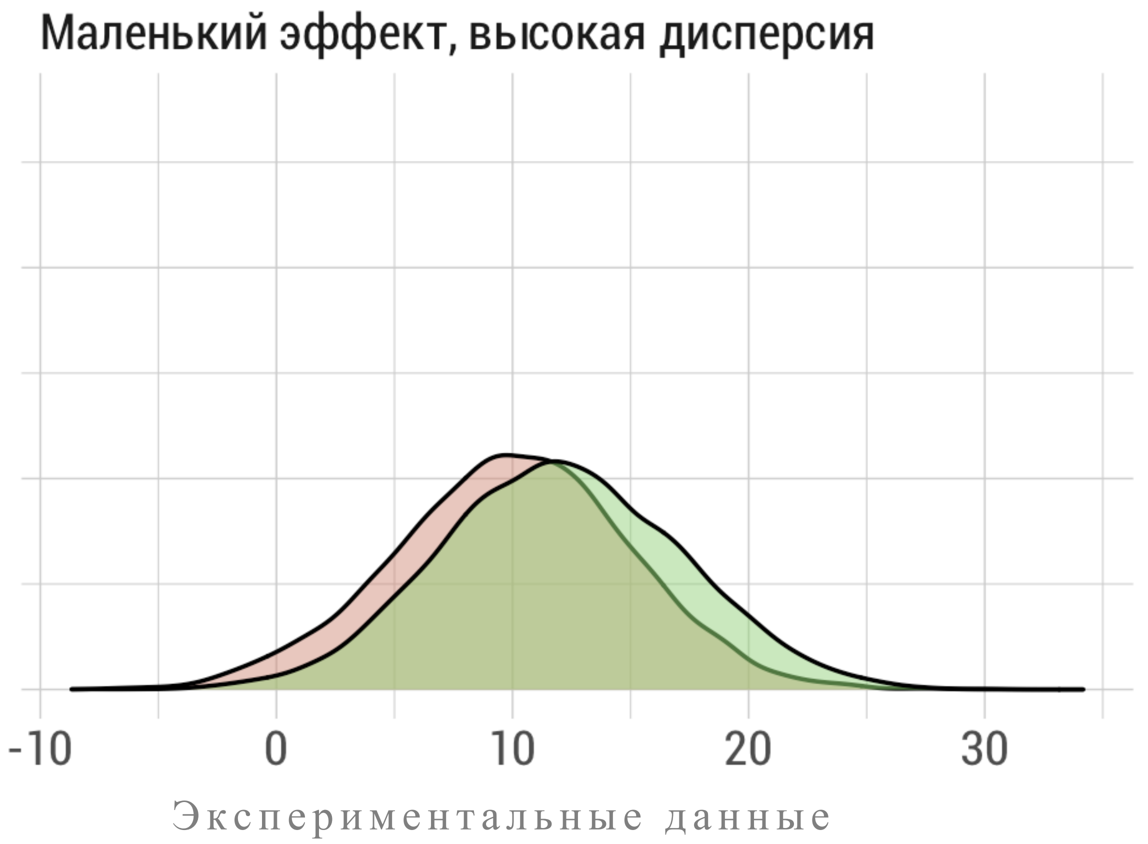
Сокращение дисперсии

Высокая дисперсия метрики мешает находить маленькие эффекты

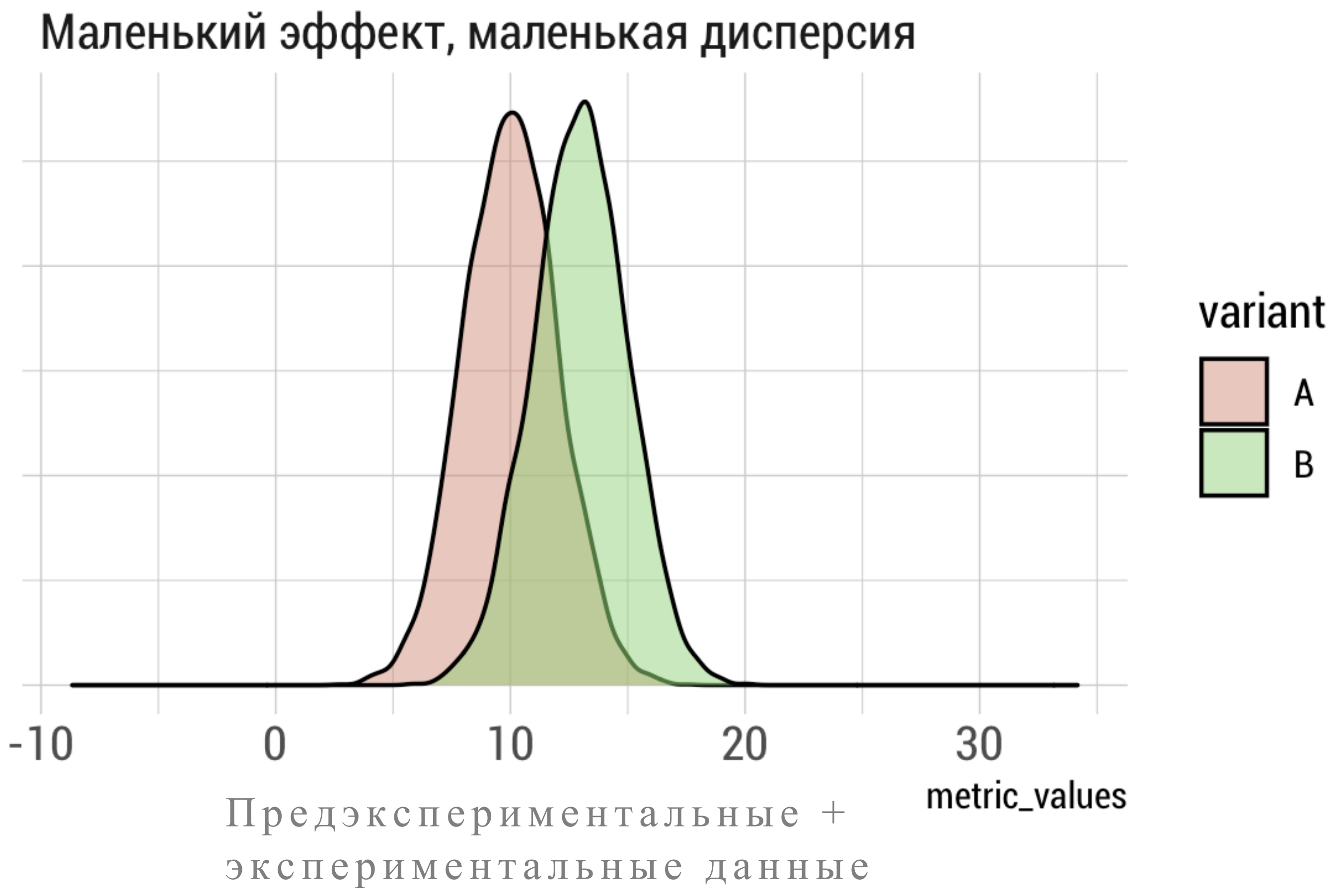


Сокращение дисперсии

Как сократить дисперсию ?
1) стратификация
2) использование ковариат



CUPED



CUPED

$$\bar{X}_{\text{cuped}} = \bar{X} - w\bar{Y} + wE[Y]$$

X - экспериментальная метрика

Y – метрика до эксперимента (ковариата)

$$w = \frac{\text{cov}(X, Y)}{V[Y]}$$

Мы смотрим не на изначальную бизнес метрику X, а измененную X_{cuped} метрику, которая чувствительнее

Ковариата (Y) не должна быть подвержена тритмент эффекту

1) Мы не сможем отделить отделить эффект тритмента от эффекта ковариаты на итоговую метрику
2) Если Y подвержена тритменту (растет или падает вместе с ним), то после вычета ковариаты, мы также уберем и тритмент эффект, связанный с ней, и получим смещенную оценку

CUPED

$$\bar{X}_{\text{cuped}} = \bar{X} - w\bar{Y} + wE[Y]$$

X - экспериментальная метрика

Y – метрика до эксперимента (ковариата)

$$w = \frac{\text{cov}(X, Y)}{V[Y]}$$

Дисперсия метрики X обусловлена дисперсией ковариаты Y и дисперсией неизвестных переменных. После применения CUPED, мы избавляемся от дисперсии Y и оставляем влияние других переменных

CUPED

При использовании метода CUPED учитывается поведение пользователя до начала эксперимента (ковариата Y) и во время эксперимента (фактическая метрика X)

После анализа зависимости между этими метриками дисперсия сокращается, что увеличивает чувствительность метрики

Таким образом, мы можем определить, как изменилось поведение пользователя во время эксперимента от среднего до экспериментального периода

Ограничения

- Неочевидный выбор ковариаты (чаще лучше всего метрика до эксперимента)
- Нужны исторические данные о пользователе (см. CUPAC, если их нет)
- Плохо подходит для бинарных величин (слабый эффект от понижения дисперсии, так низкая дисперсия в метриках конверсии итак низкая)