

Ковчег

Белов Василий

Получаем простую случайную выборку

01

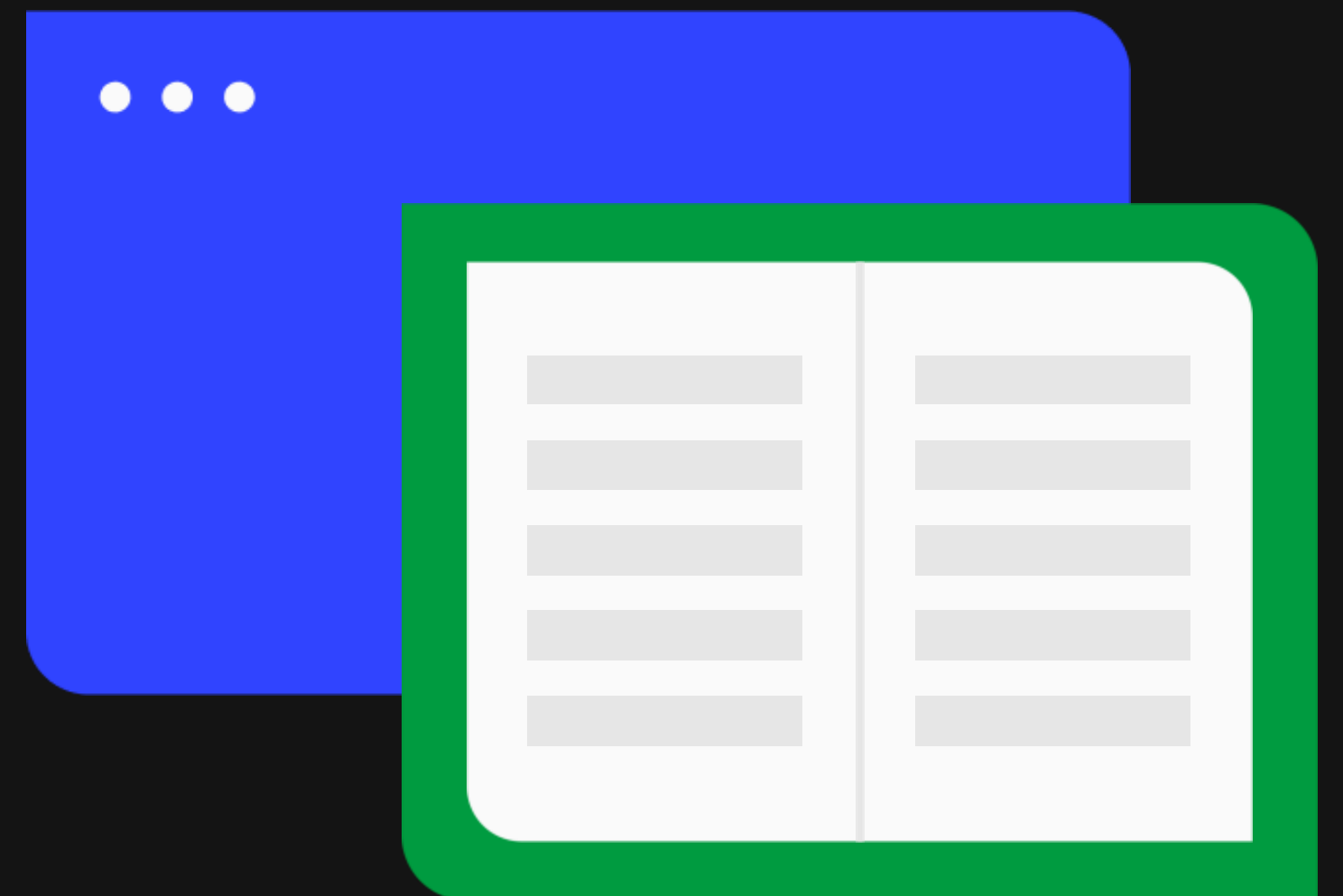
Определение генеральной совокупности: 80,000 клиентов, среди которых известна доля покупателей

02

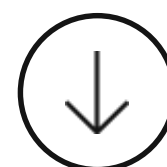
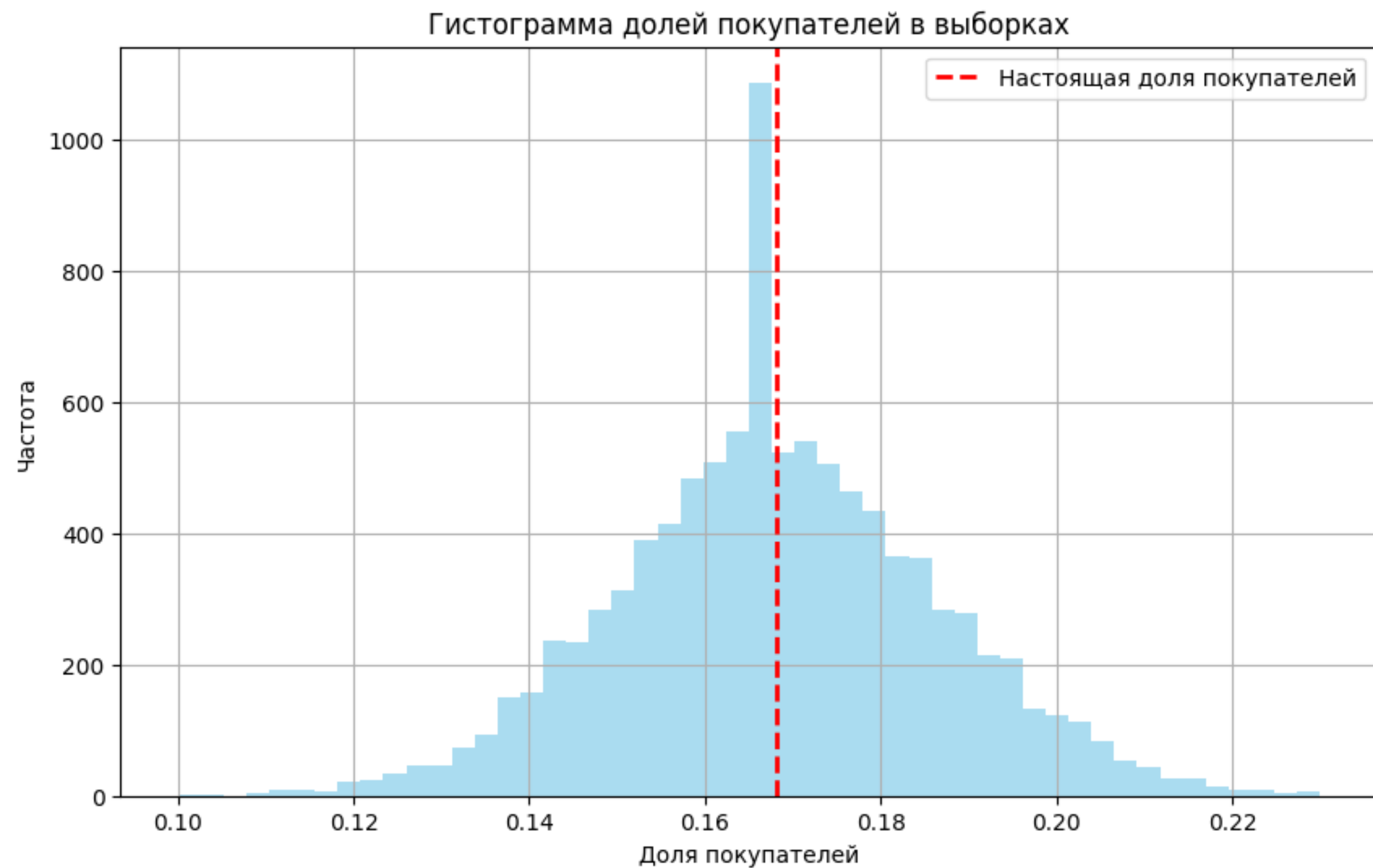
Формирование выборки:
Используется метод простой случайной выборки, где каждый клиент имеет равные шансы попасть в выборку.

03

Повторение эксперимента:
Провести $N = 10,000$ повторов, в каждом из которых формируется новая выборка из 400 клиентов и вычисляется доля покупателей.



Простая случайная выборка



Комментарии к результатам

Комментарии



Анализ результатов симуляции

Постоянное систематическое завышение или занижение

Центр распределения долей покупателей в выборках близок к истинной доле покупателей (красная вертикальная линия), что указывает на отсутствие систематического смещения.

Возможность большой ошибки

Да, метод простой случайной выборки может давать большую ошибку, особенно при небольшом объеме выборки. Вероятность большой ошибки определяется шириной распределения.

Частота большой ошибки

Чем шире распределение, тем больше вероятность получить выборку с большой ошибкой. Однако, если объем выборки достаточно большой, то вероятность большой ошибки будет низкой.

Получаем стратифицированную случайную выборку

01

Определение страт:

используем пол клиентов как критерий для деления на подгруппы (страты)

02

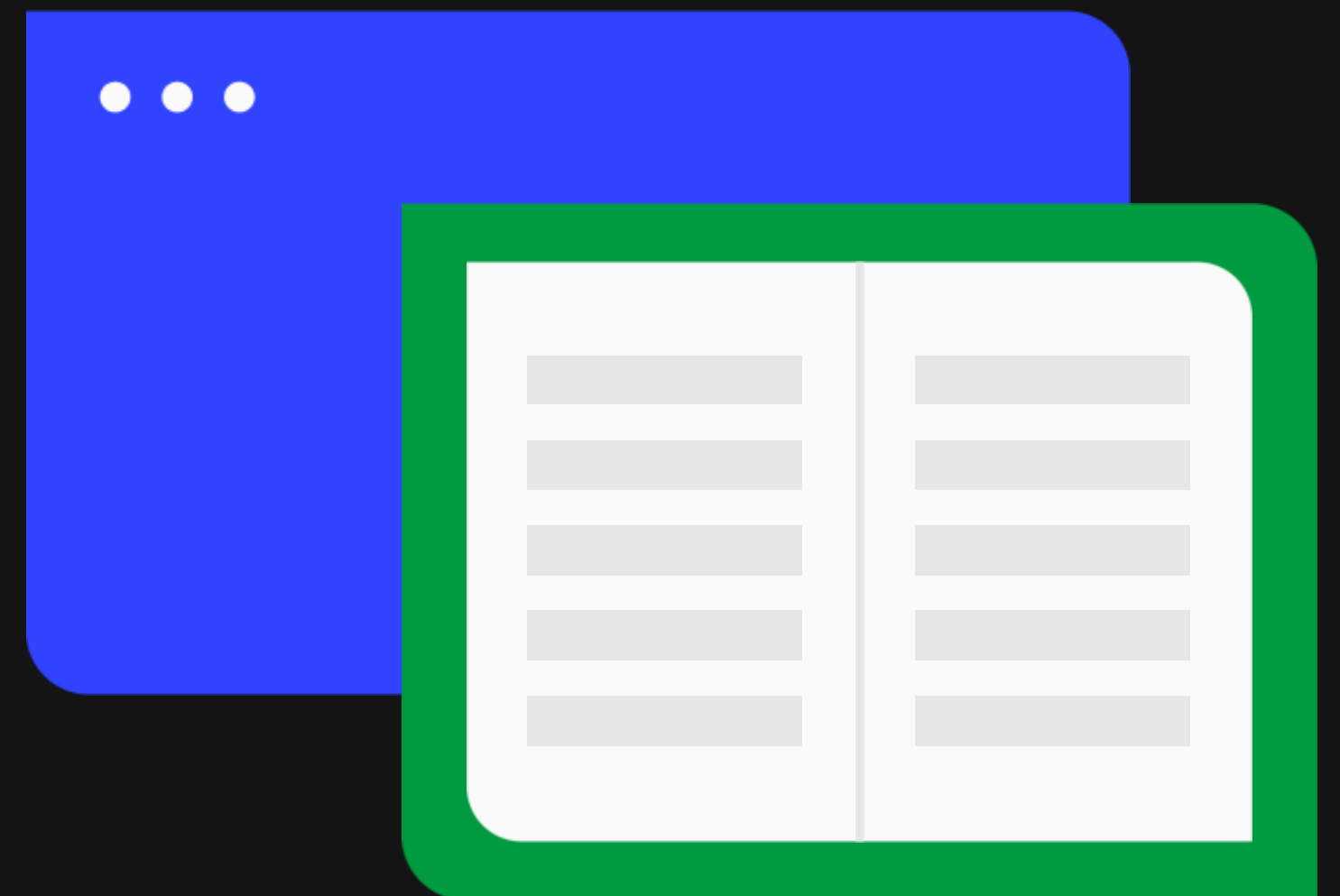
Формирование выборки из каждой страты:

случайным образом отбираем по 200 клиентов из каждой подгруппы

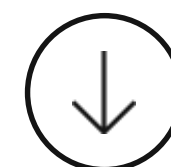
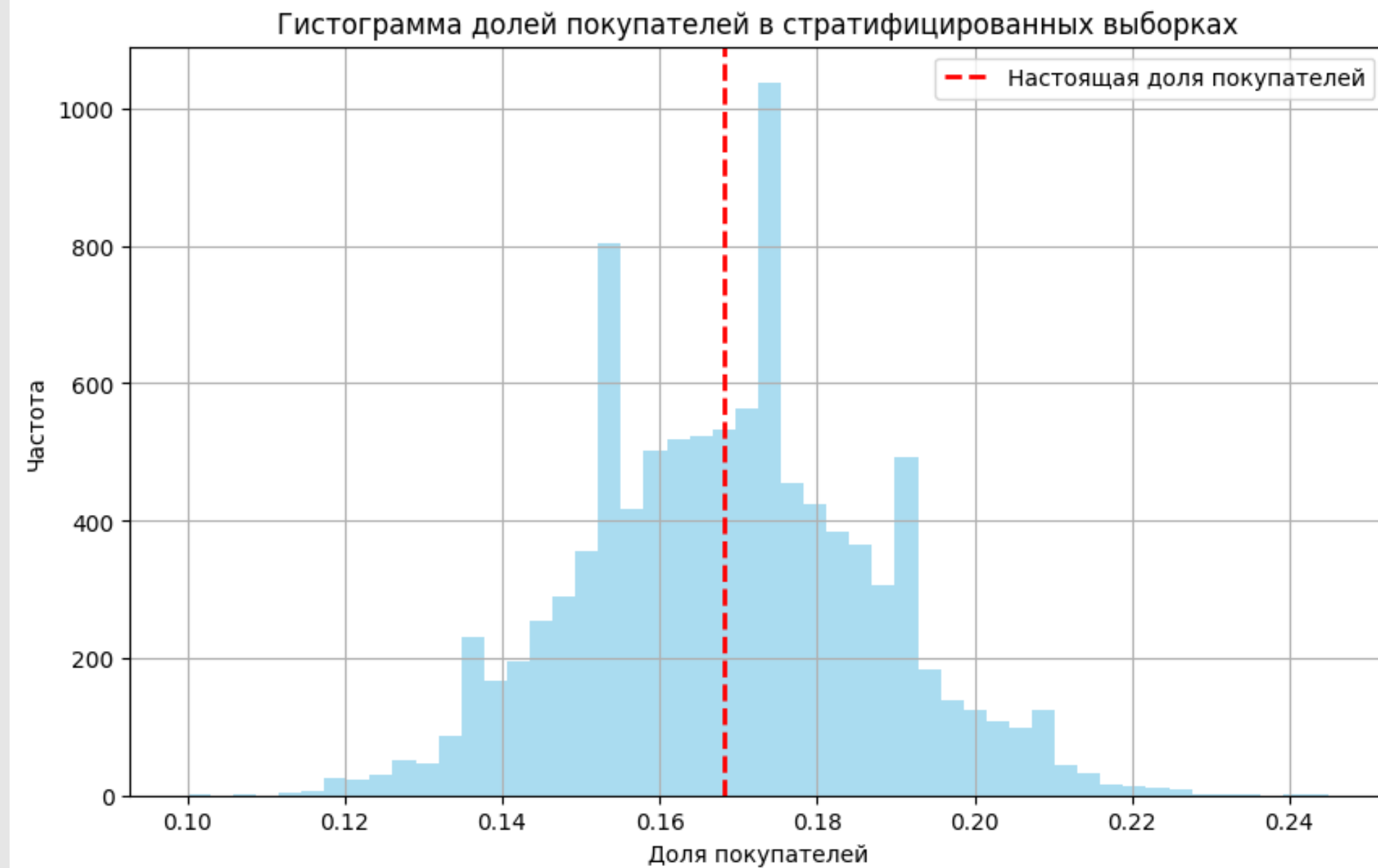
03

Повторение эксперимента:

Провести $N = 10,000$ повторов, в каждом из которых формируется новая выборка из 400 клиентов и вычисляется доля покупателей.



Стратифицированная случайная выборка



Комментарии к результатам

Комментарии



Анализ результатов симуляции

Постоянное систематическое завышение или занижение

Как и в случае с простой случайной выборкой, систематического смещения не наблюдается. Центр распределения выборочных долей покупателей совпадает с истинной долей.

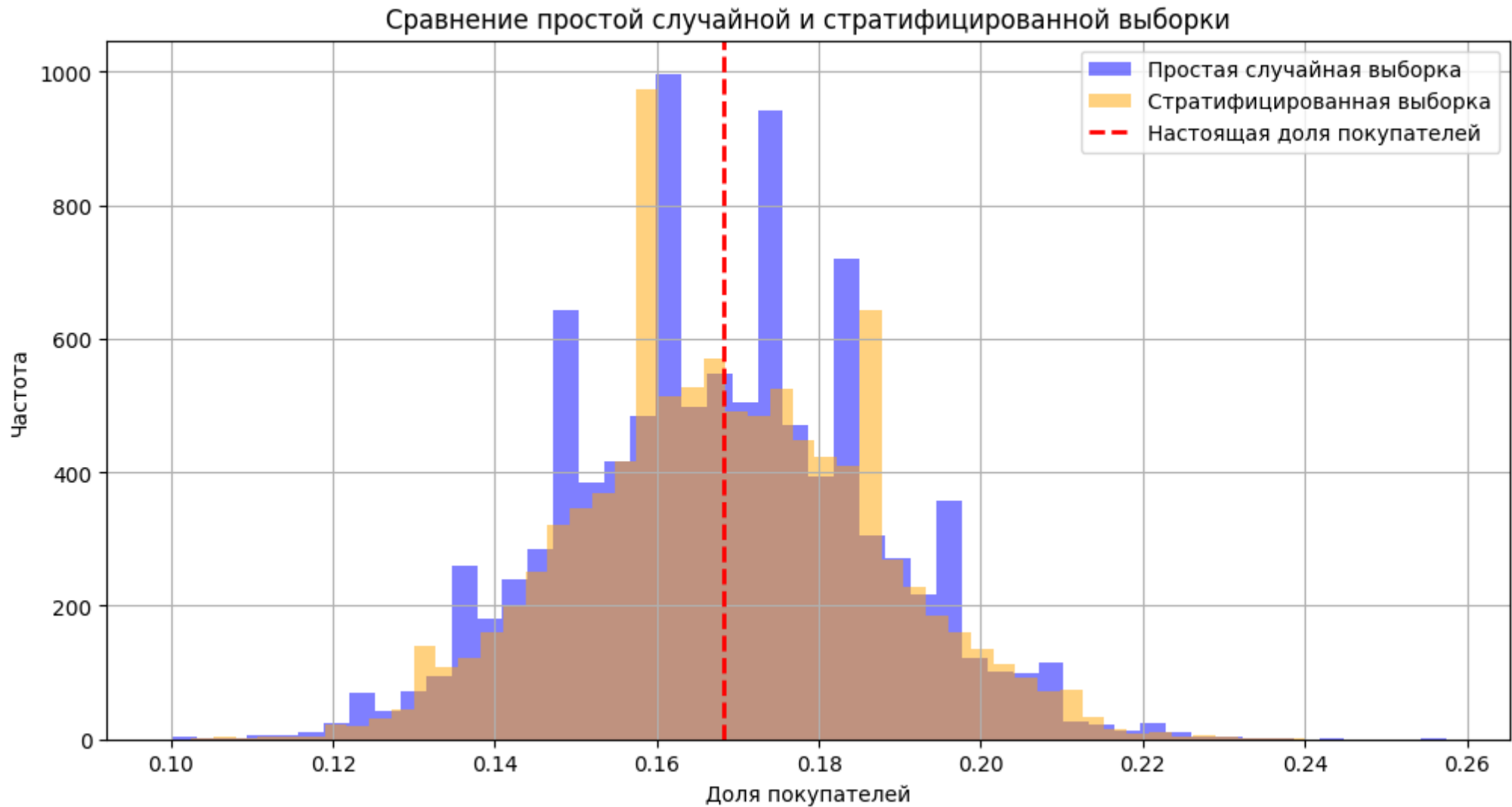
Возможность большой ошибки

Да, может. Однако, вероятность большой ошибки, судя по гистограмме, ниже, чем при использовании простой случайной выборки. Это связано с тем, что стратификация позволяет учесть неоднородность генеральной совокупности и уменьшить дисперсию оценок.

Частота большой ошибки

Чем шире распределение, тем больше вероятность получить выборку с большой ошибкой. Однако, если объем выборки достаточно большой, то вероятность большой ошибки будет низкой.

Случайная и стратифицированная случайная выборка на одной гистограмме



Комментарии к результатам



Комментарии



Анализ результатов симуляции

Точность методов

Центральная тенденция: Оба метода в среднем дают оценки доли покупателей, близкие к истинному значению (отмечено красной вертикальной линией). Это говорит об отсутствии систематической ошибки в обоих методах.

Дисперсия: Распределение долей покупателей в стратифицированной выборке имеет меньшую дисперсию, то есть данные более сгруппированы вокруг среднего значения. Это свидетельствует о большей точности метода стратифицированной выборки.

Интуитивное объяснение

Стратифицированный метод точнее за счет:

Деления генеральной совокупности на страты с низкой изменчивостью (например, по полу или возрасту), это уменьшает разброс внутри каждой группы.

Представленности всех групп:

Стратификация гарантирует, что все важные подгруппы присутствуют в выборке, что делает результаты более репрезентативными.

Уменьшения общей дисперсии: собственную оценку каждой страты позволяет снизить общую ошибку выборочного обследования.

Почему нельзя понять, какой из двух подходов лучше по одной выборке

Случайность: Одна выборка может не отражать истинное распределение в генеральной совокупности из-за случайных колебаний.

Необходимость статистической значимости: Для адекватного сравнения требуется множество повторов для оценки стабильности и надежности результатов.

Разные условия: Каждый метод может работать лучше или хуже в зависимости от структуры генеральной совокупности;

Зачем делать много повторов обоих методов?

01

Большое количество повторов необходимо для оценки стабильности результатов, выявления систематических ошибок и обеспечения статистической значимости. Повторные выборки позволяют получить множество оценок, что помогает выявить средние значения и разброс, а также определить, есть ли у метода тенденция к завышению или занижению результатов. Один раз проведенная выборка может быть аномальной или не репрезентативной, поэтому множество симуляций создает более полное представление о том, как метод работает в различных условиях и с разными подгруппами, что критически важно для надежности выводов.

