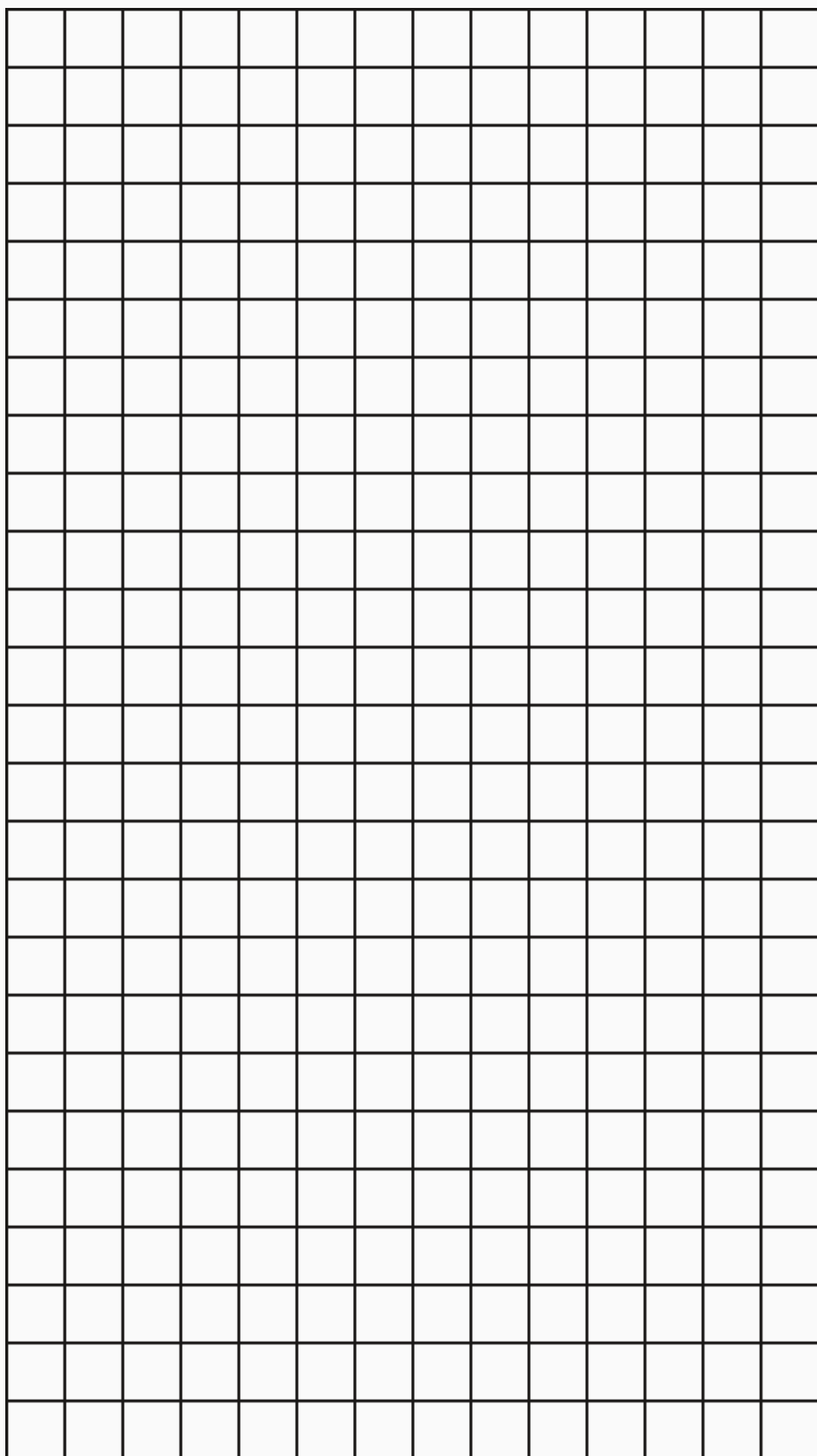
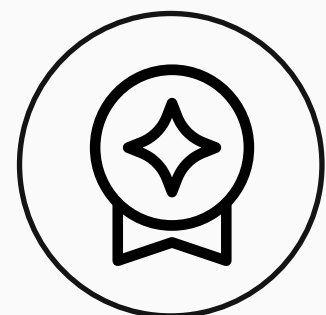


Кейс Ковчег

Введение в статистику

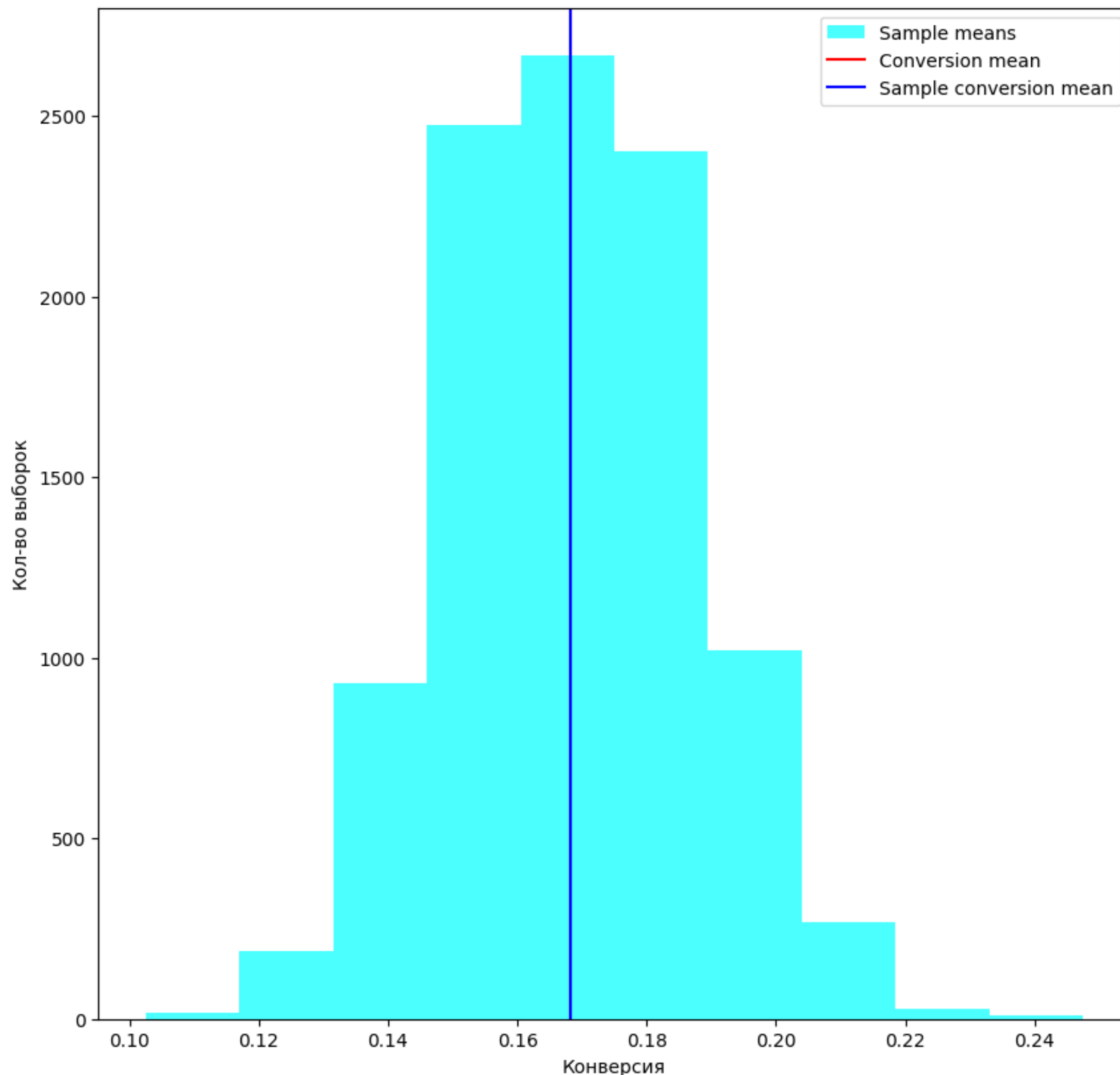


Способ получения простой случайной выборки



Для получения простой случайной выборки мы будем рассылать рекламные сообщения случайным 400 клиентам

Распределение среднего 10 000 простых случайных выборок
в 400 человек



Результаты простой случайной выборки



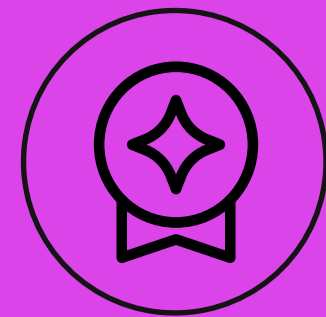
Систематическое завышение
или занижение не замечено



Simple random sample всегда
будет выдавать погрешность,
но в представленном виде
погрешность будет очень
небольшой, так как

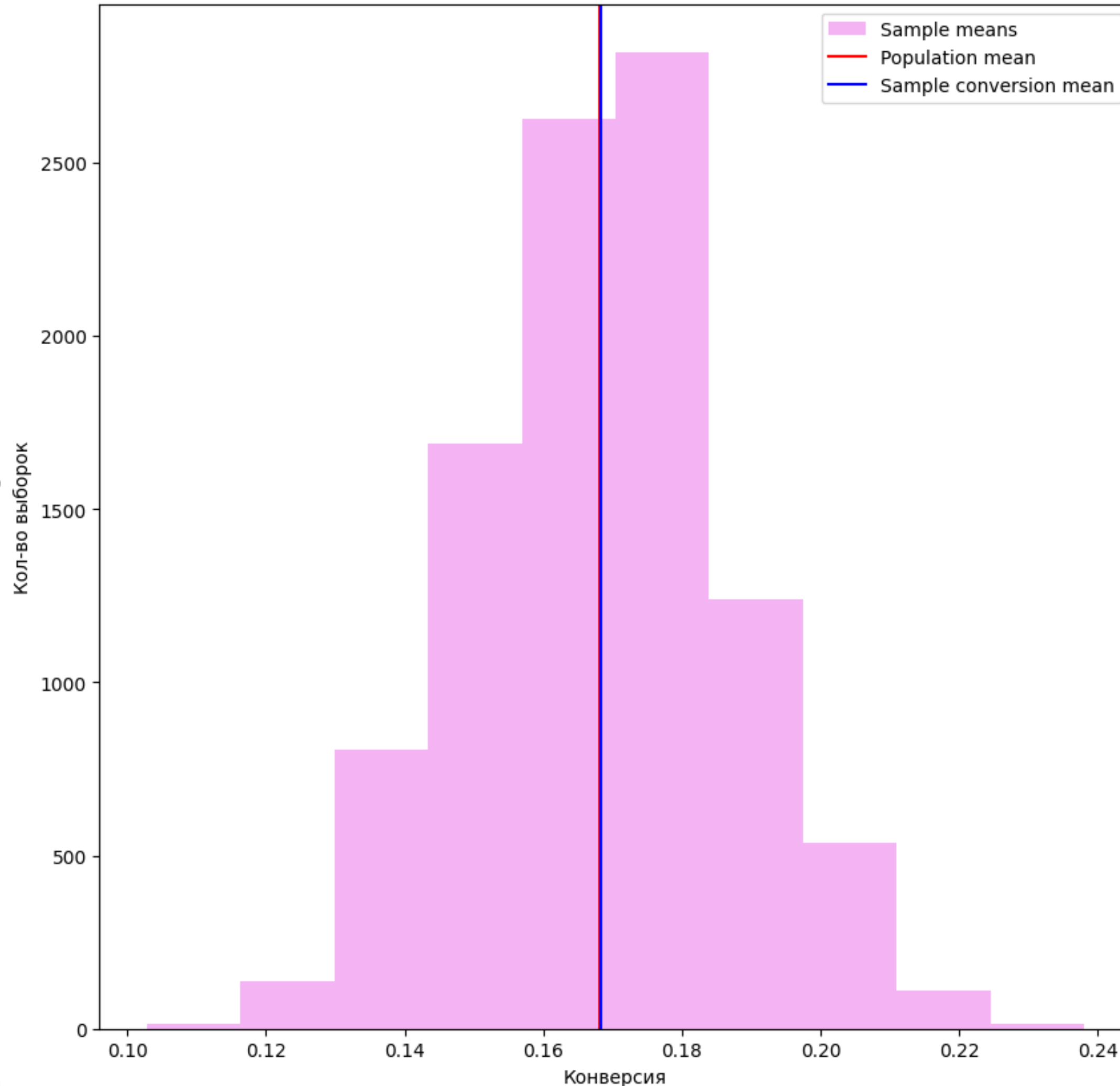
- Каждая выборка содержит
достаточно большое
количество клиентов(400)
- Проводится большое
число выборок(10 000)

Способ получения стратифицированной случайной выборки



Для получения стратифицированной случайной выборки необходимо разбить генеральную совокупность на страты и сохранять доли этих страт в каждой выборке идентичные их долям генеральной совокупности

Распределение среднего 10 000 стратифицированных случайных выборок
в 400 человек по половозрастным группам



Результаты стратифицированной случайной выборки

- Я разбил выборку на половозрастные группы для стратифицированного семплинга
- Систематическое завышение или занижение не замечено
- Стратифицированная случайная выборка выдает погрешность, но она будет меньше чем в простой, так как стратифицированная дополнительно сохраняет отношение между стратами, идентичное с ген совокупностью, поэтому значение среднего в ней приближено сильнее к среднему в ген совокупности

Сравнение обычной и стратифицированной выборок

- Можно заметить что стратифицированная случайная выборка имеет меньшее стандартное отклонение(0.01859) чем простая(0.01853), она сильнее “прижата” к среднему ген совокупности.
- Так же можно заметить, что различие в графиках плотности не велико, на основе этого можно предположить: страты имеют примерно равные доли в ген совокупности.
- В общем случае стратифицированная выборка не хуже простой
- Нельзя понять какая выборка лучше по одной итерации, тк есть большая погрешность, потому что выборки рандомные, и их точность увеличивается при большом количестве итераций благодаря большому количеству наблюдений

