

КОНТРОЛИРУЕМЫЕ ЭКСПЕРИМЕНТЫ

Люнченко Станислав

ОСНОВНЫЕ ТЕРМИНЫ КОНТРОЛИРУЕМЫХ ЭКСПЕРИМЕНТОВ



Рандомизатор – принцип распределения вариантов между объектами эксперимента (например, пользователям или страницам). Должен быть случайным.

Общий критерий оценки - количественный показатель цели эксперимента.

Параметр - контролируемая экспериментальная переменная, которая, как ожидается, влияет на общий критерий оценки

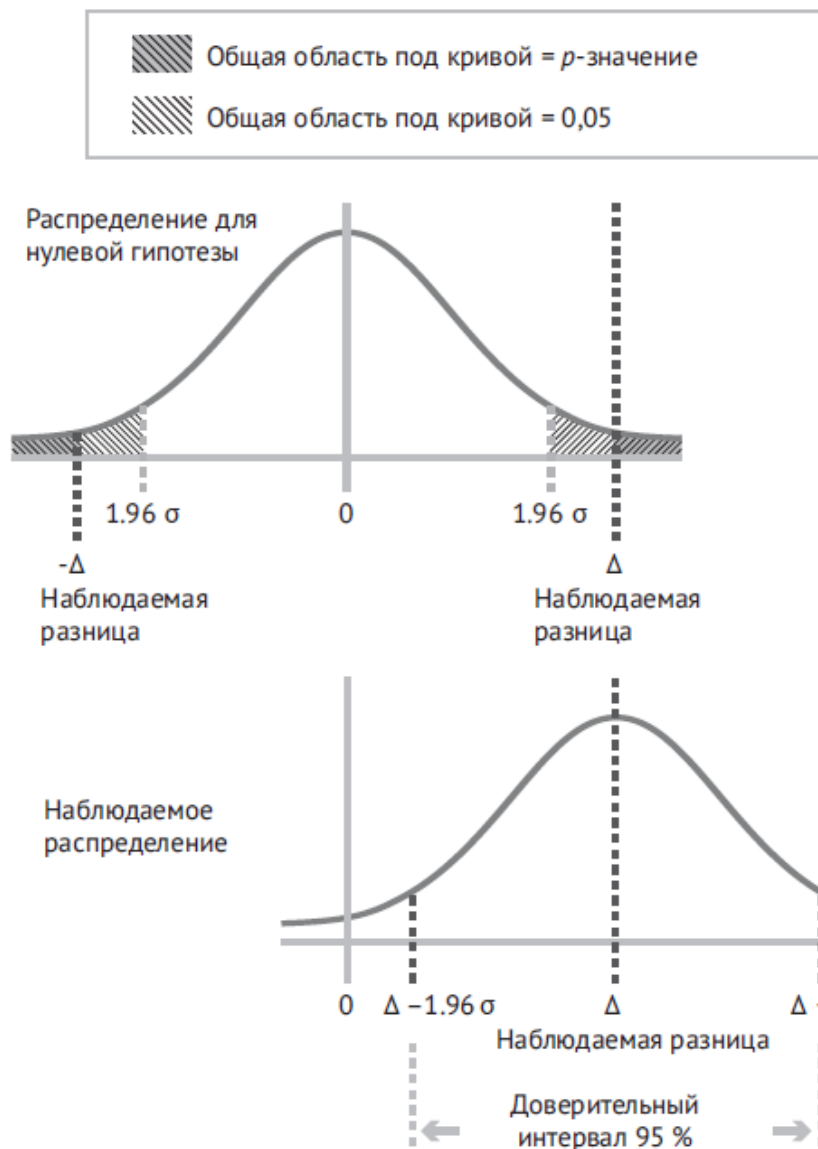
Вариант – тестируемое значение параметра.

Статистическая мощность - это вероятность обнаружения значимого различия между вариантами, когда оно действительно есть.

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ

Использование
p-value для оценки
статистической
значимости

Использование
95% доверительного
интервала



H₀ – нулевая гипотеза, средние значения критерия совпадают для обоих вариантов ($\Delta=0$).

H₁ – Альтернативная гипотеза, значение критерия для варианта В выше (лучше) чем для варианта А

P-value – вероятность получения наблюдаемой или более экстремальной разницы, если предположить, что нулевая гипотеза верна.

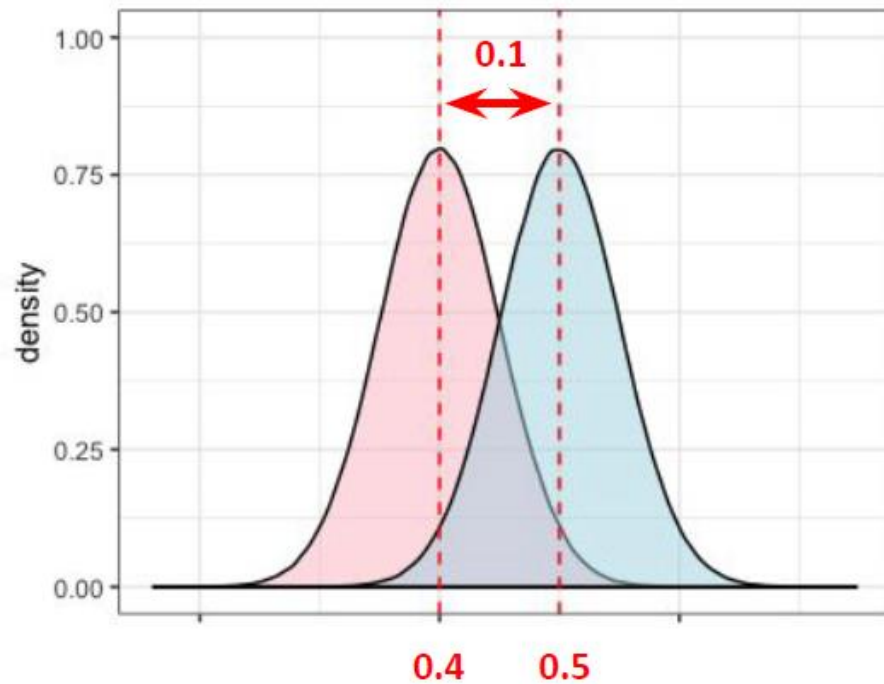
P-value – это не вероятность того что H₀ верна

Общепринятый научный стандарт – использовать значение **p-value** меньше **0.05**

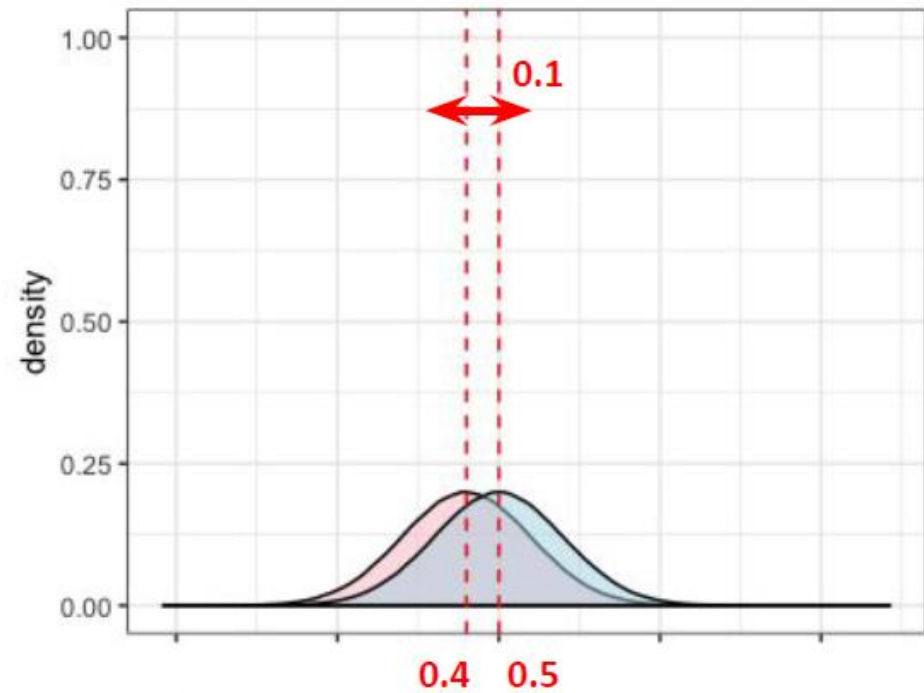
ПРИМЕР ВИЗУАЛИЗАЦИИ СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ

Если просто посчитать разницу средних значений,
насколько мы уверены в достоверности полученного результата?

Если мы еще раз проведем тест, то какова вероятность того, что мы сможем повторить результат?

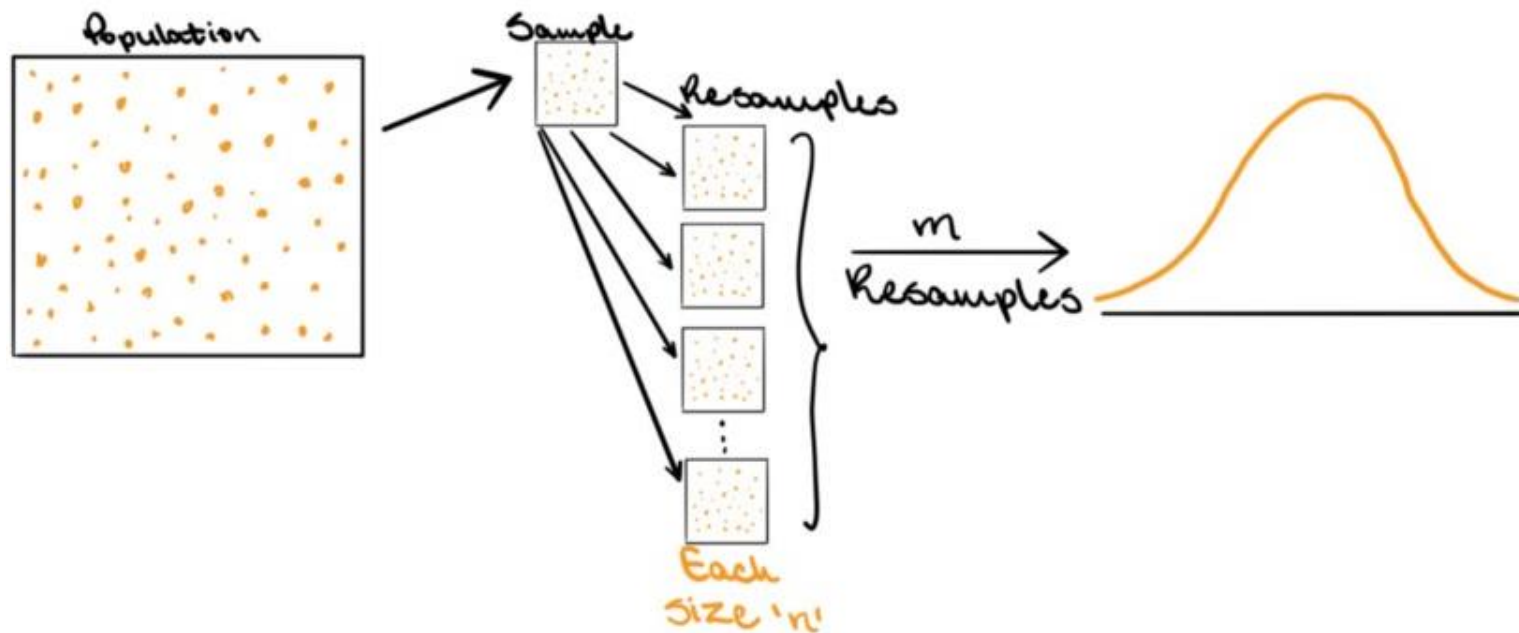
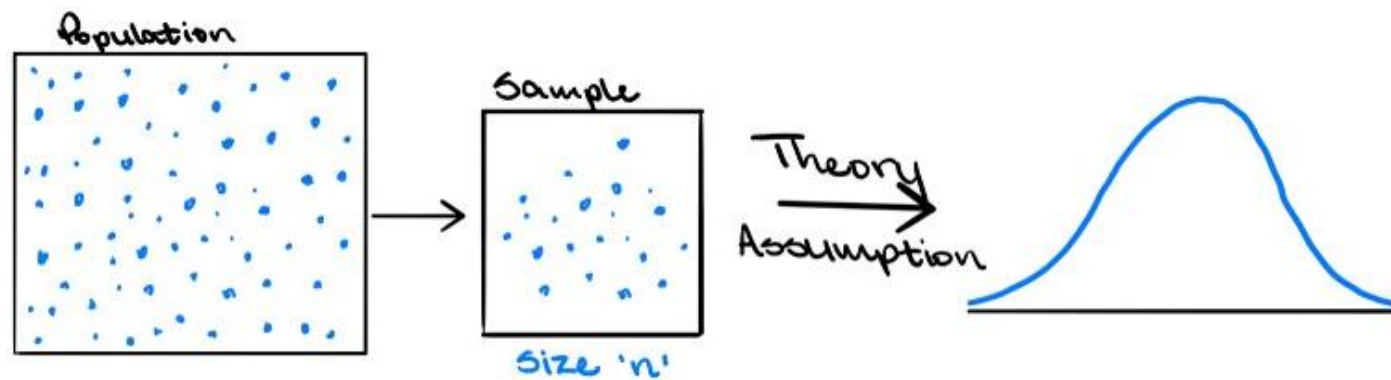


Маленький разброс значений →
разница в 0.1 значительна



Большой разброс значений → разница
в 0.1 незначительна

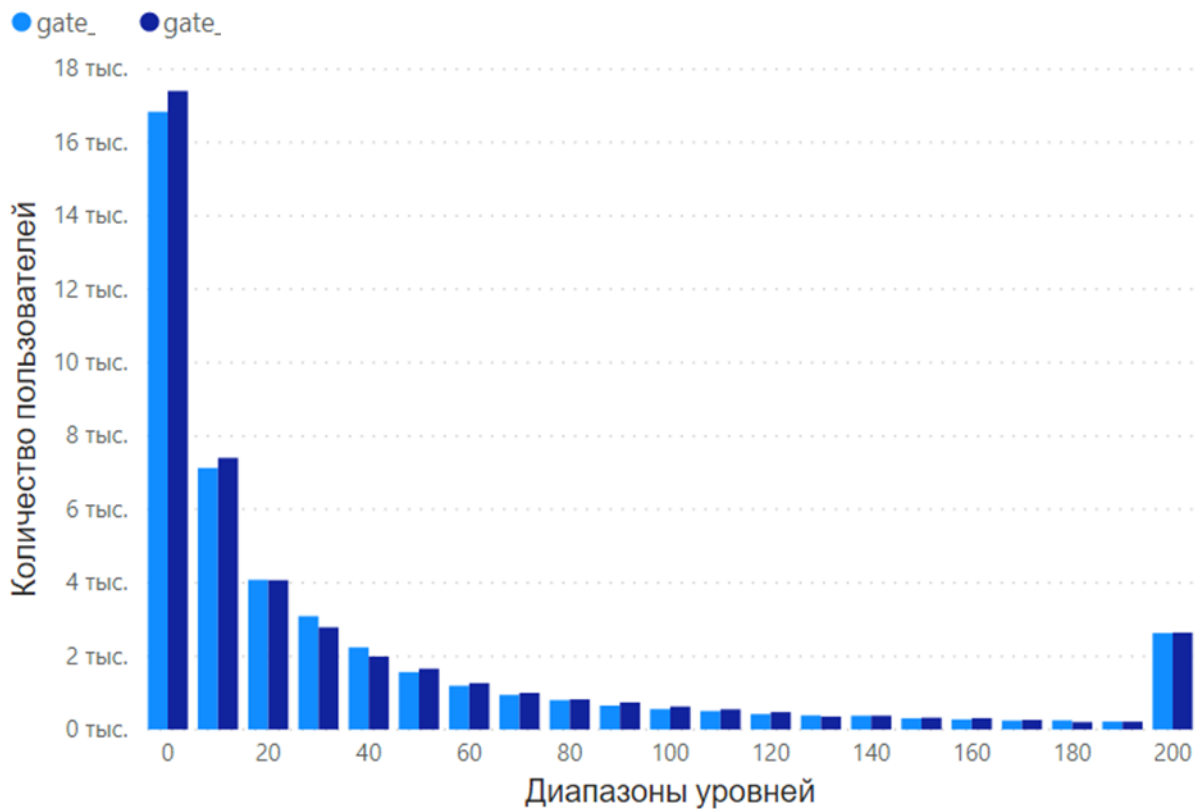
МЕТОД BOOTSTRAP



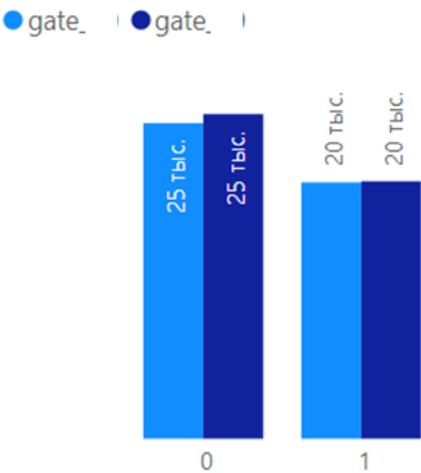
ЗАДАНИЕ 1 ПРОВЕСТИ ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ

Версия	Количество пользователей	Доля по кол-ву	Средний r1	Средний r2	Среднее gamerounds	СКО gamerounds
gate_	44700	50%	0,45	0,19	52	257
gate_	45489	50%	0,44	0,18	51	103
Всего	90189	100%	0,45	0,19	52	195

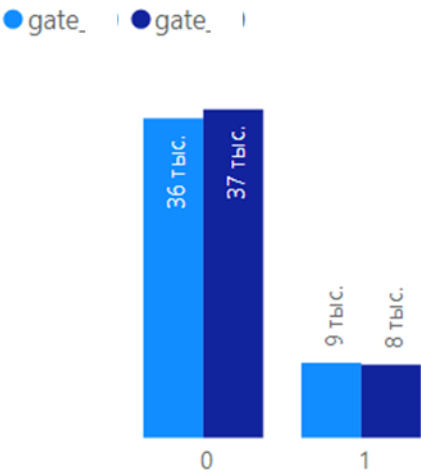
Распределение пользователей по пройденному уровню



Распределение retention_1

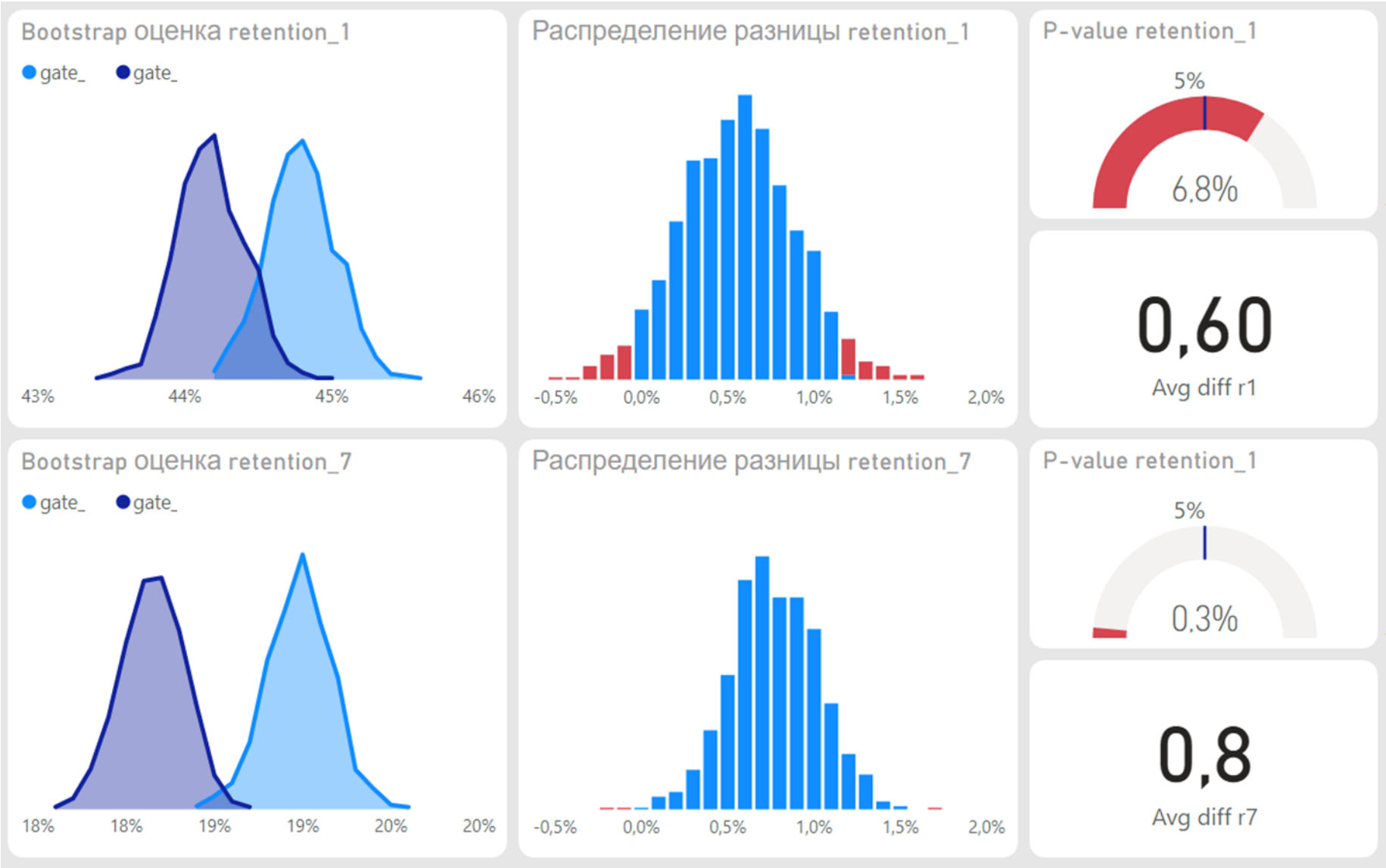


Распределение retention_7



По данным о результатах проведенного А/В тестирования построить представленные визуализации, сделать выводы о корректности рандомизации объектов

ЗАДАНИЕ 2 ПРОВЕСТИ АНАЛИЗ РЕЗУЛЬТАТОВ А/В-ТЕСТА



С использованием метода bootstrap определить разницу общих критериев оценки (retention_1 и retention_7), определить статистическую значимость полученных результатов.

Построить представленные визуализации, сделать выводы о целесообразности изменения значения анализируемого параметра с gate_30 на gate_40