Signed test (Знаковый тест)

Парные наблюдения
Смотрим, есть ли тенденция, что в паре первый элемент больше второго.

$$H_0 : median_1 = median_2$$ **=>** $$H_0 : p(X > Y) = 0.5$$

$$H_1 : median_1 < median_2$$

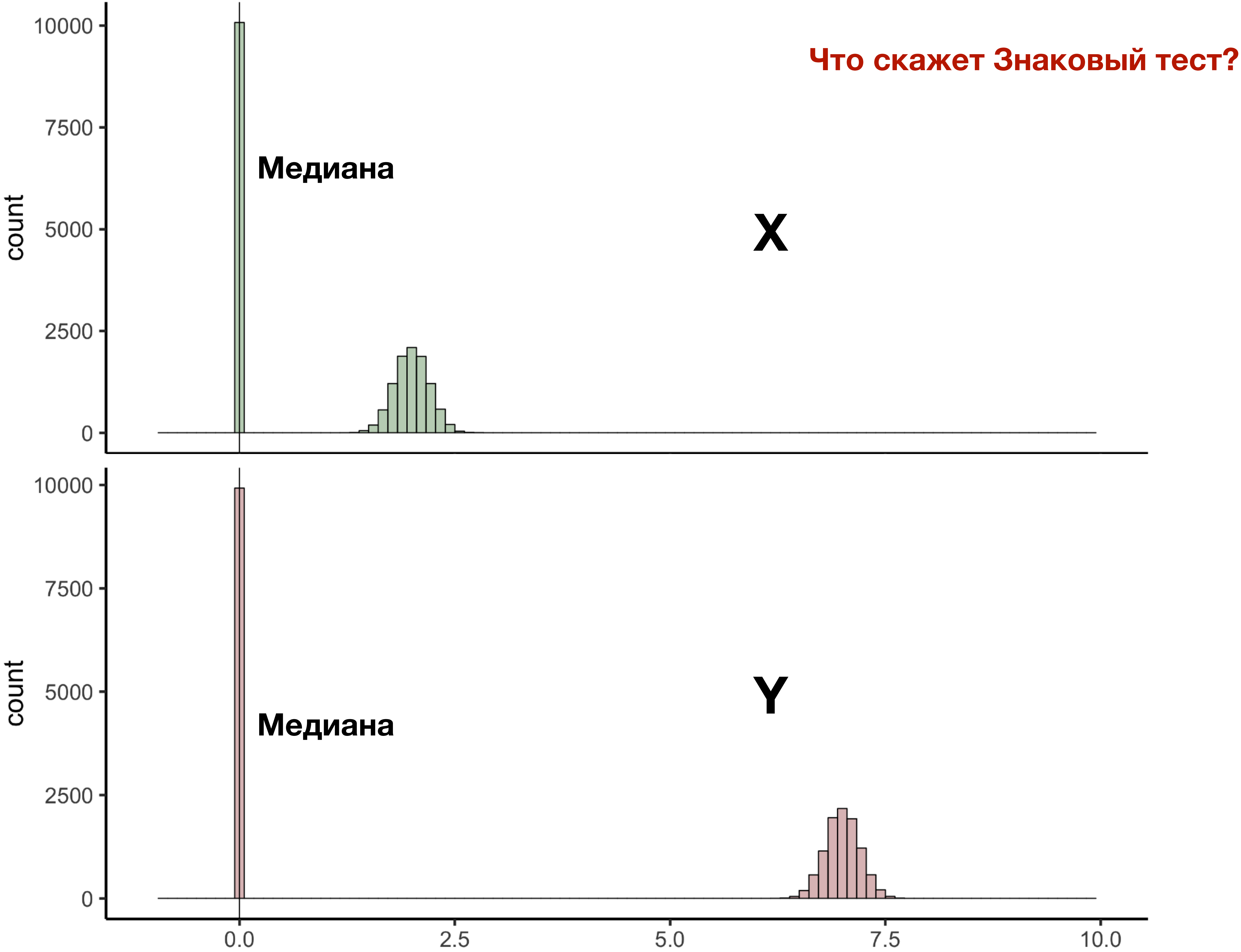
$$H_1 : median_1 > median_2$$

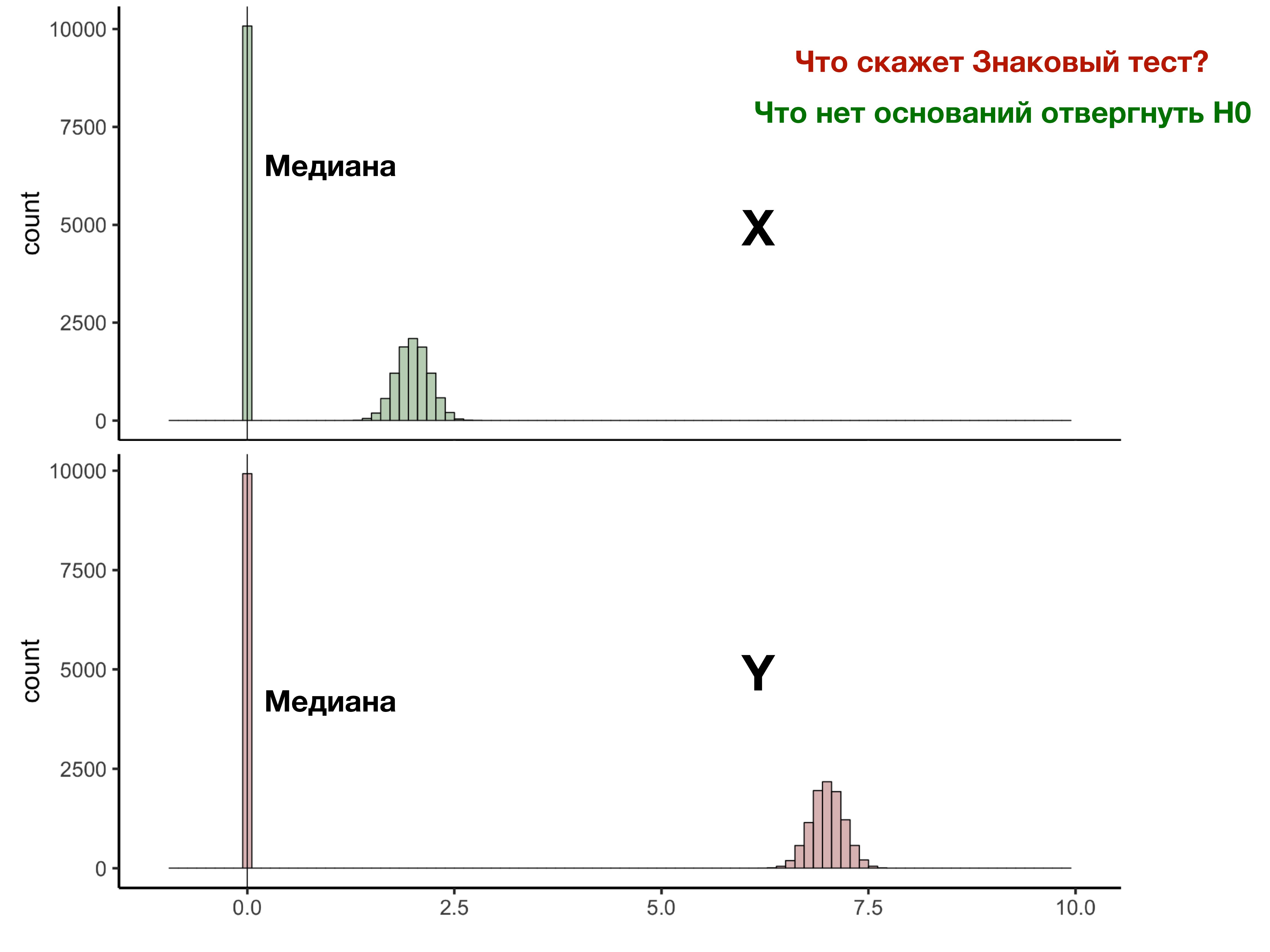
$$H_1 : median_1 \neq median_2$$

$$\hat{p} = \frac{\sum_i I(x_i < y_i)}{N}$$

Профессор в бизнес-школе хочет сравнить цены на новые учебники в университетском магазине и соответствующие цены в городе. Профессор случайно выбирает 12 названий учебников для курсов бизнес-школы и сравнивает их цены (в долларах) в университетском магазине и в случайно выбранном магазине в городе. Результаты приведены в таблице

Книга	Цена на кампусе	Цена в городе
1	55.00	50.95
2	47.50	45.75
3	50.50	50.95
4	38.95	38.50
5	58.70	56.25
6	49.90	45.95
7	39.95	40.25
8	41.50	39.95
9	42.25	43.00
10	44.95	42.25
11	45.95	44.00
12	56.95	55.60


Что скажет Знаковый тест?
10000
7500
5000
2500
0
count
Медиана
X
10000
7500
5000
2500
0
count
Медиана
Y
0.0
2.5
5.0
7.5
10.0


Что скажет Знаковый тест?
Что нет оснований отвергнуть Н0
Медиана
X
Медиана
Y
count
count
10000
7500
5000
2500
0
0.0
2.5
5.0
7.5
10.0

Wilcoxon signed-rank test (критерий Уилкоксона)

Парные наблюдения
Смотрим, распределены ли разницы между наблюдениями в паре симметрично около 0

$$H_0 : simmetric\ distribution\ around\ 0 (shift = 0)$$

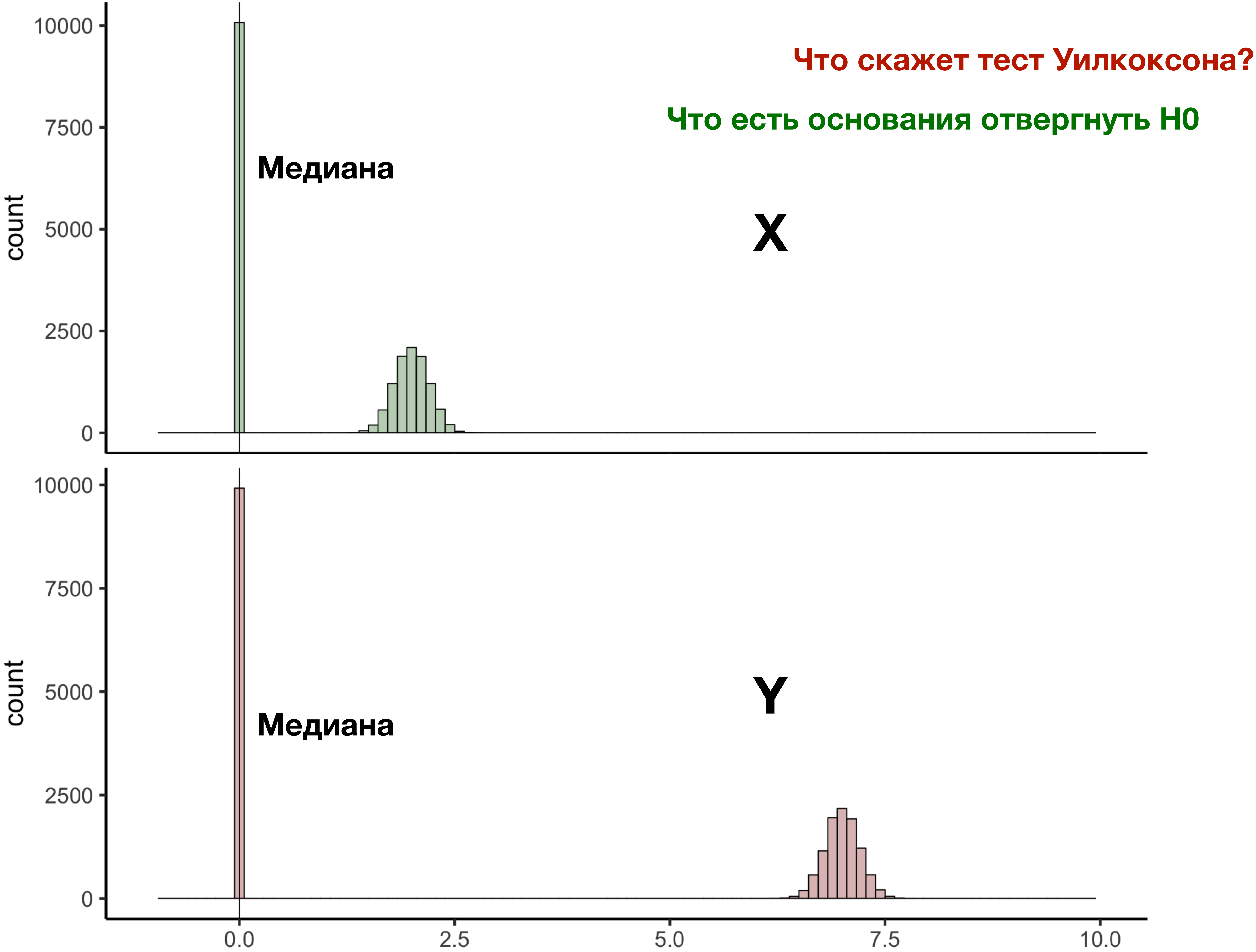
$$H_1 : shift < 0 \qquad H_1 : no \qquad H_1 : shift > 0$$

The magnitude of paired rank differences are either not symmetrically distributed, or are not distributed about zero, or both.

$$d_i = |x_i - y_i|$$

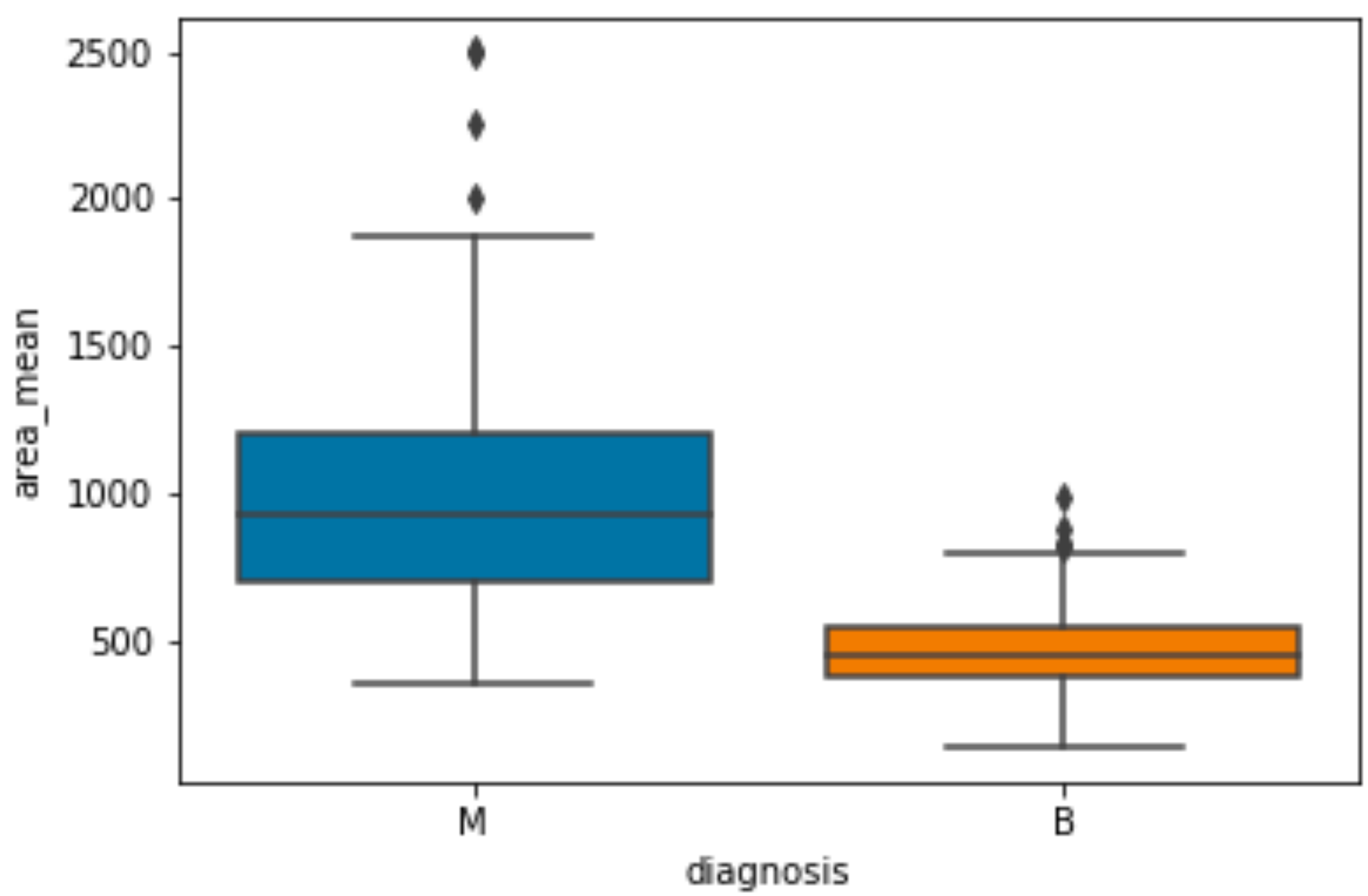
$$W = sgn(x_i - y_i) \cdot rank(d_i)$$

$$W \sim N(\mu = 0, \sigma = \sqrt{\frac{n(n+1)(2n+1)}{6}})$$


Что скажет тест Уилкоксона?
Что есть основания отвергнуть H0
Медиана
X
Медиана
Y
count
count
10000
7500
5000
2500
0
0.0
2.5
5.0
7.5
10.0

Что делать с этими тестами, чтобы не попадать на странные случаи?

Что делать с этими тестами, чтобы не попадать на странные случаи?



Всегда стройте boxplot прежде чем тестировать какую-то гипотезу

Профессор в бизнес-школе хочет сравнить цены на новые учебники в университетском магазине и соответствующие цены в городе. Профессор случайно выбирает 12 названий учебников для курсов бизнес-школы и сравнивает их цены (в долларах) в университетском магазине и в случайно выбранном магазине в городе. Результаты приведены в таблице

Книга	Цена на кампусе	Цена в городе
1	55.00	50.95
2	47.50	45.75
3	50.50	50.95
4	38.95	38.50
5	58.70	56.25
6	49.90	45.95
7	39.95	40.25
8	41.50	39.95
9	42.25	43.00
10	44.95	42.25
11	45.95	44.00
12	56.95	55.60

Mann-Whitney U-test
(критерий Манна-Уитни)

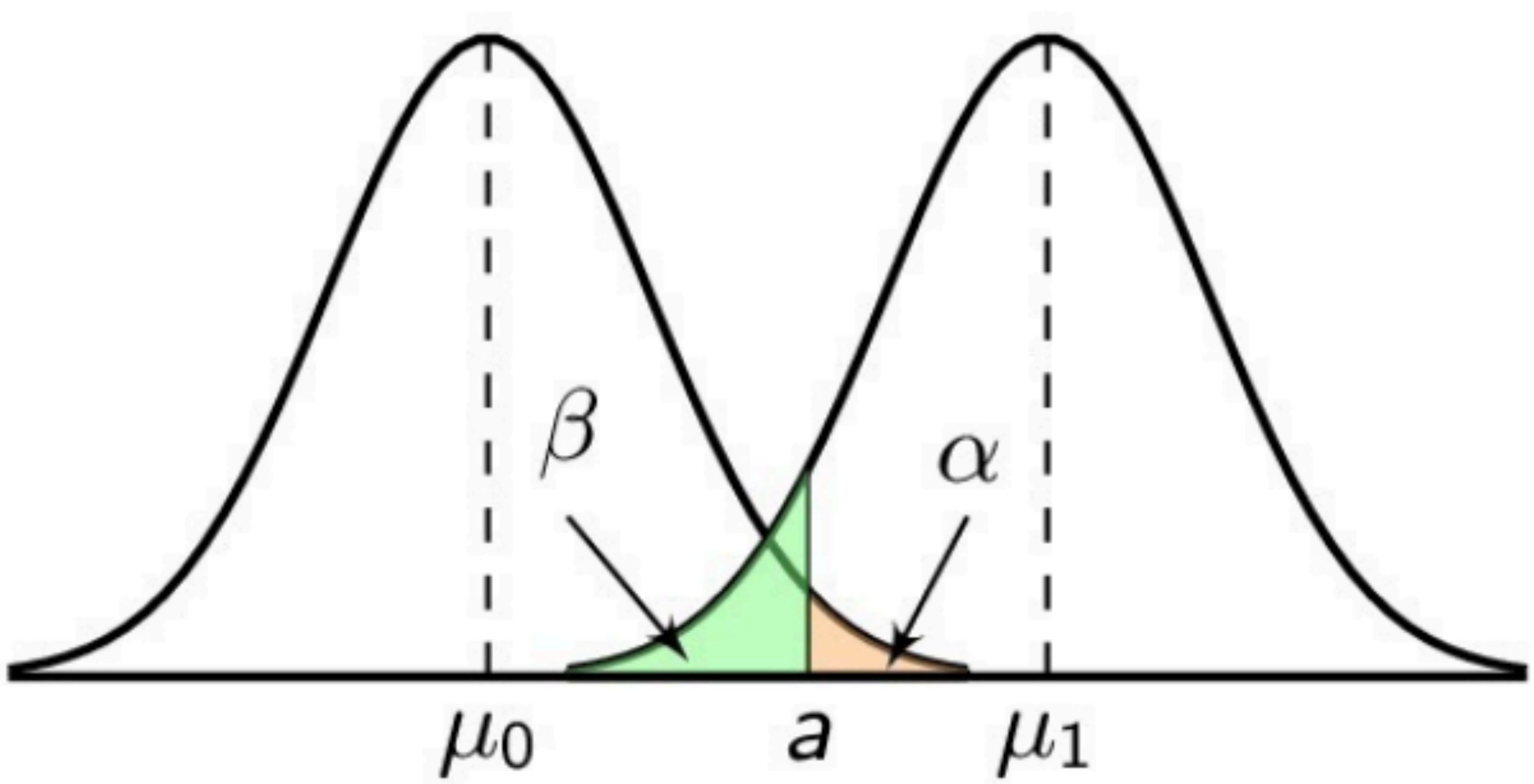
Две выборки
Нулевая и альтернативная гипотеза такая же, как в Вилкоксоне

Студентов обучают по двум методикам, A и B, причем занятия проходят в отдельных классах. Преподаватель интересуется различиями между двумя методиками и сравнивает оценки студентов. Есть достаточно оснований считать, что распределение оценок далеко от нормального.

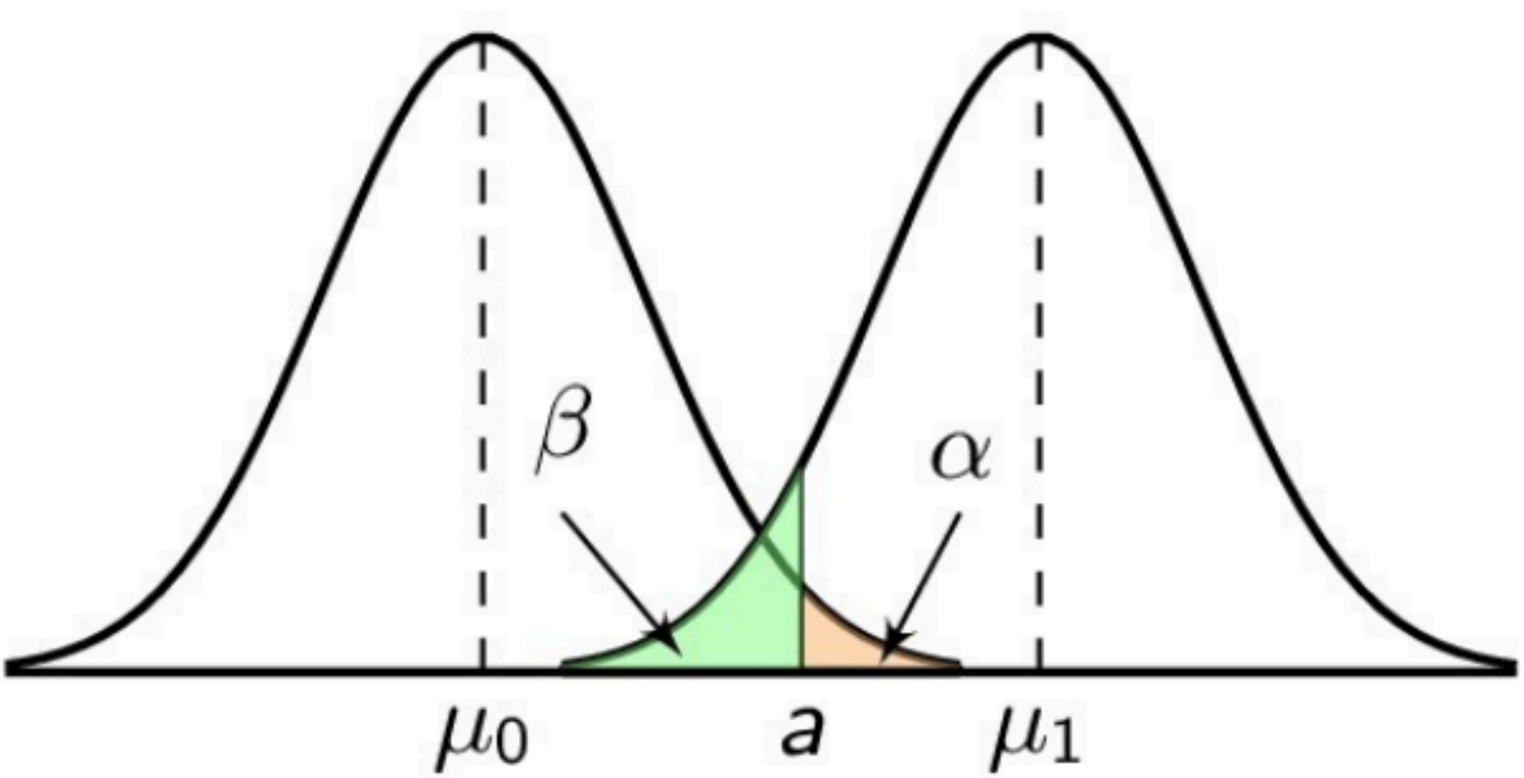
Методика А	48	39	37	36	59	69	60	45	38			
Методика Б	63	64	79	30	31	42	56	74	67	78	66	25

Чему равно значение тестовой статистики?

Подбор размера выборки для контроля alpha и beta



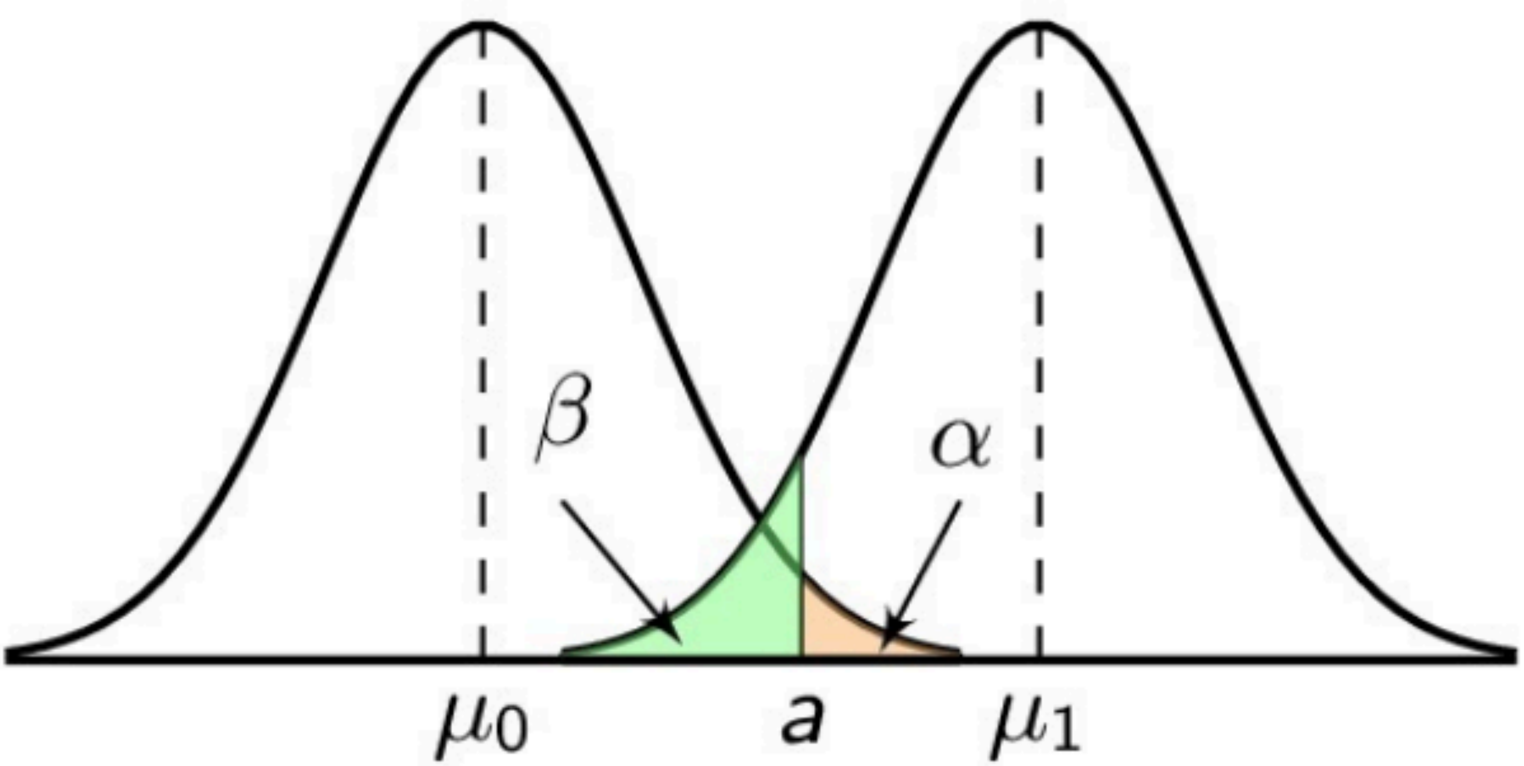
Подбор размера выборки для контроля alpha и beta



α **Вероятность ошибки первого рода**

β **Вероятность ошибки второго рода**

Подбор размера выборки для контроля alpha и beta



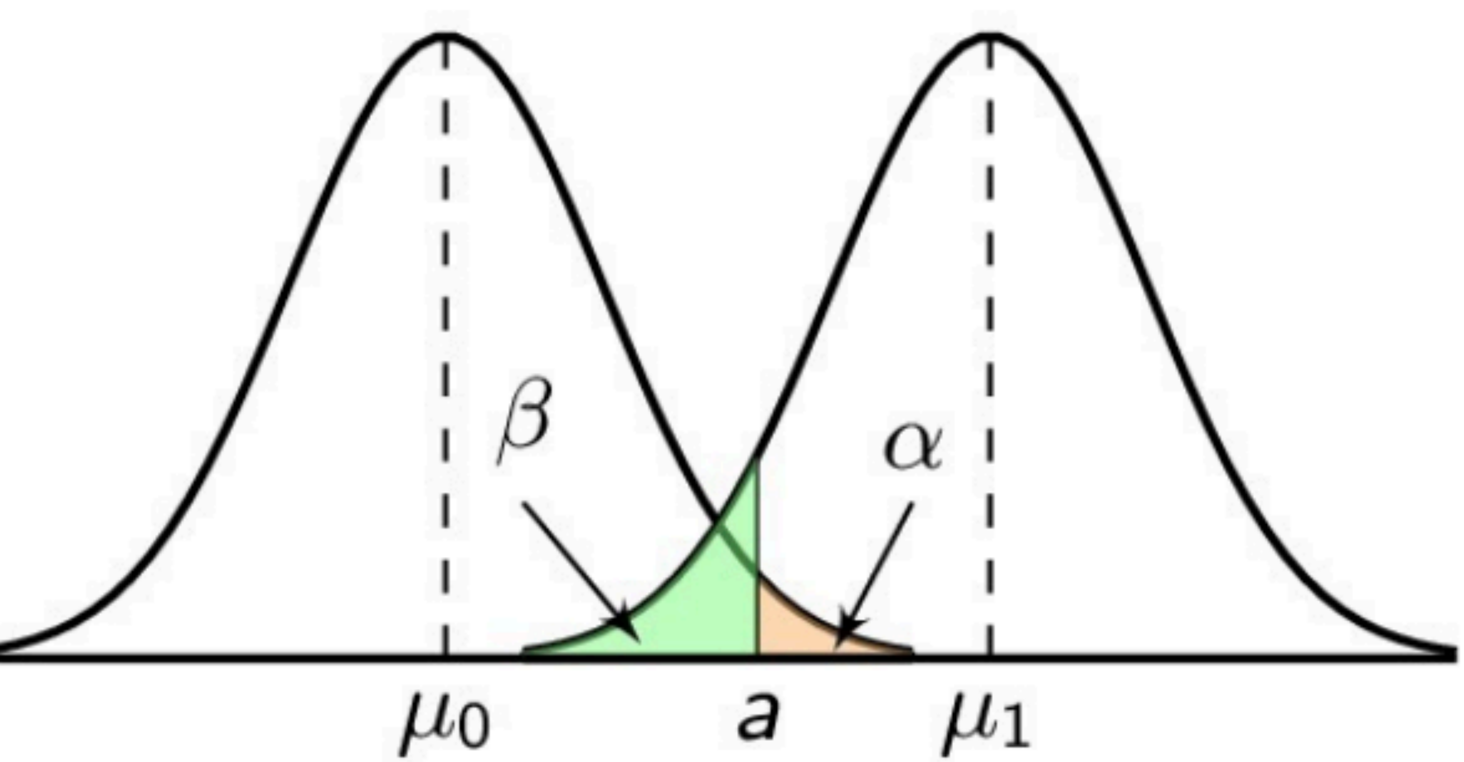
α **Вероятность ошибки первого рода**

β **Вероятность ошибки второго рода**

$$\alpha = P(\overline{X} > a \mid H_0)$$

$$\beta = P(\overline{X} < a \mid H_1)$$

Подбор размера выборки для контроля alpha и beta



Для нормального распределения

$$\mu_0 < \mu_1$$

$$\alpha = P(z > \frac{a - \mu_0}{SE})$$

$$\beta = P(z < \frac{a - \mu_1}{SE})$$

Один изобретатель утверждает, что его новая снегоуборочная машина может работать на одной унции бензина 300 минут без перерыва. Другой изобретатель сомневается в этом и планирует взять выборку из 50 машин. Он отвергнет утверждение первого изобретателя, если среднее время работы на одной унции бензина окажется меньше 296 минут. Если на самом деле новая машина в среднем работает без перерыва только 290 минут, а стандартное отклонение равно 20 минут, то чему равна сила теста?

Проводится исследование эффективности нового лекарства. Для этого используется две группы людей размера n. Лекарство считается возможным к использованию в случае, если оно уменьшает уровень холестерина хотя бы на 3 единицы. Предполагая стандартное отклонение равным 20 (и что лекарство не оказывает на него влияния), какой минимальный размер выборки необходим для того, чтобы проверить гипотезу на уровне значимости 0.05, сохраняя при этом мощность теста равной 0.85