

Ящик с усами (Box-plot)

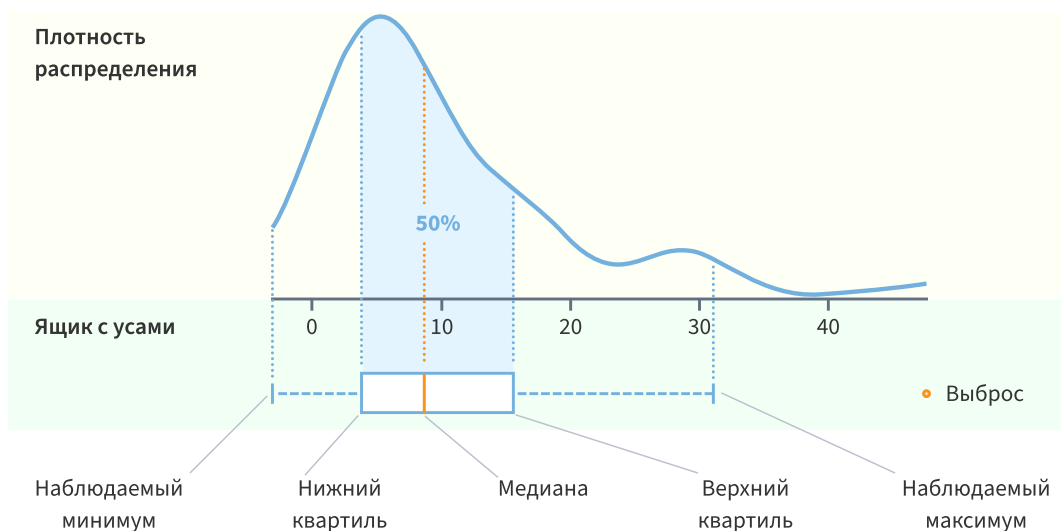
Синонимы: Ящик с бакенбардами, Диаграмма размаха, Ящиковая диаграмма, Box-and-whiskers diagram

Разделы: [Визуализация](#)

График «ящик с усами» позволяет очень компактно и наглядно представлять порядковые статистики одномерного закона распределения: квартили, медиану, наблюдаемые минимальное и максимальное значение выборки, а также отображать выбросы. При использовании «ящика с усами» не делается никаких предположения относительно закона распределения выборки, поэтому его можно рассматривать как инструмент непараметрической статистики.

График «ящик с усами» был предложен Джоном Тьюки (поэтому в литературе ее иногда называют еще диаграммой Тьюки) в 1969 году. Хотя «ящик с усами» считается менее информативным, чем гистограмма, но при этом является более компактным и простым в построении.

В основе графической структуры «ящика с усами» лежит прямоугольник (ящик) из боковых сторон которого отходят отрезки (усы). Если с помощью графика исследуется одна выборка, то «ящик» обычно изображается горизонтально. Если требуется визуализировать статистики для нескольких выборок с целью их сравнения, то «ящики» для каждого из них изображаются вертикально рядом друг с другом.



Изначально «ящик с усами» задумывался как способ представления так называемой **сводки пяти чисел** (five-number summary) — набора описательных статистик, характеризующих распределение исследуемой выборки. Сводка включает следующие элементы, отображаемые с помощью «ящика с усами»:

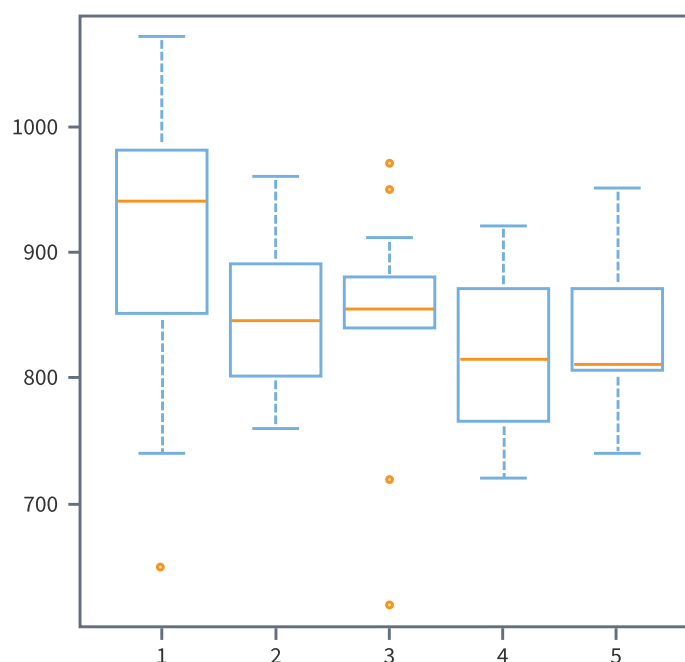
- минимальное наблюдаемое значение (0-й квартиль или 0-й процентиль) — минимальная точка данных выборки, отображается началом левого «уса»;
- максимальное наблюдаемое значение выборки (4-й квартиль или 100-й процентиль) — максимальная точка данных выборки, отображается концом правого «уса»;
- медиана (2-й квартиль или 50-процентиль) — отображается чертой, разделяющей «ящик» на две части;
- 1-й квартиль или 25-й процентиль — представляется левой стороной «ящика»;
- 3-й квартиль или 75-й процентиль — представляется правой стороной «ящика»;

Таким образом длина «ящика» представляет собой интерквартильный размах.

Длина «усов» на диаграмме характеризует разброс (вариацию) значений выборки. Расстояние между концами «усов», таким образом, представляет собой размах вариации. Если длина «усов» одинаковая, это говорит в пользу того, что распределение выборки симметричное. Дисбаланс между длинами «усов» говорит об асимметрии распределения.

Отметки, отображаемые на диаграмме кружками или звездочками, представляют собой выбросы данных.

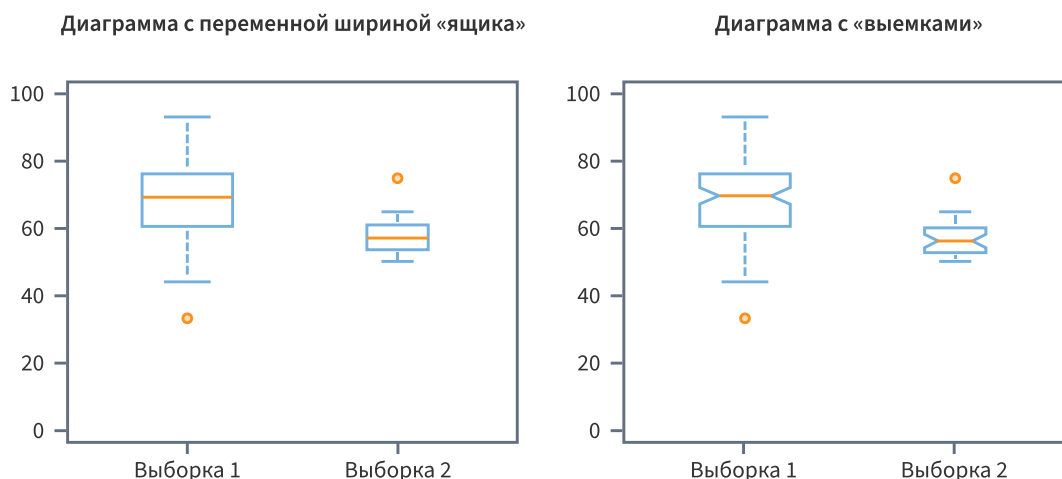
Часто диаграмму «ящик с усами» строят вертикально, представляя несколько «ящичков» рядом для разных выборок. Это очень удобно для сравнения статистических характеристик нескольких выборок.



Следует отметить, что каких-либо строгих правил или стандартов, регламентирующих построение диаграммы «ящик с усами», не существует. Например, в некоторых случаях могут не изображаться «усы» или выбросы, если представляемая ими информация о распределении не существенна для решаемой задачи.

Наиболее популярными модификациями классической диаграммы Тьюки являются:

- **диаграмма с переменной шириной «ящика»** — используется, когда несколько «ящичков с усами» отображаются вертикально рядом для сравнения распределений нескольких выборок. Если на классической диаграмме длина «ящичков» представляет интерквартильный размах каждой выборки, а ширина одинакова, то в диаграмме с переменной шириной, ширина «ящика» меняется в зависимости от объема выборки и обычно определяется как корень квадратный от числа ее элементов;
- **диаграмма с «выемками»** — применяется «выемка» или сужение рамки вокруг медианы с целью дать приблизительное представление о значимости разницы медиан; если выемки двух прямоугольников не перекрываются, это свидетельствует о статистически значимой разнице между медианами двух выборок. Ширина выемок пропорциональна интерквартильному размаху выборки и обратно пропорциональна квадратному корню из ее размера.



Кроме этого на диаграмме могут отображаться по мере необходимости и другие статистические характеристики выборки, такие как среднее значение.