

Квартили в описательной статистике¹

Г.И.Фалин, д.ф.м.н., проф.
кафедра теории вероятностей
механико-математический факультет
МГУ им М.В.Ломоносова (Москва)

А.И.Фалин, к.ф.м.н., доцент
кафедра общей математики
факультет ВМиК
МГУ им М.В.Ломоносова (Москва)

В соответствии с Федеральным государственным образовательным стандартом основного общего образования (ФГОС ООО) «предметные результаты изучения предметной области «Математика и информатика» должны отражать ... 8) овладение простейшими способами представления и анализа статистических данных; ... развитие умений ... описывать и анализировать массивы числовых данных с помощью подходящих статистических характеристик...». Как ФГОС ООО, так и Примерная основная образовательная программа образовательного учреждения (основная школа) не содержат конкретных деталей указаний о содержании статистического образования в школе. В учебных пособиях, используемых при преподавании статистики, рассматриваются лишь простейшие статистические характеристики числового набора (среднее, медиана, мода, дисперсия, размах). Однако, одна из основных характеристик положения числового набора, медиана, является частным случаем более интересных характеристик – квартилей. Эти величины позволяют лучше охарактеризовать не только положение, но и разброс чисел набора.

1. Определение квартилей

В статье [1] мы рассказали о различных мерах положения числового набора (среднем значении, медиане, моде) и провели их сравнительный анализ. В связи с этим мы упомянули, что более точно охарактеризовать положение и разброс чисел набора можно с помощью так называемых квартилей. В этой статье мы подробно расскажем о квартилях и связанных с ними понятиях.

¹ Г.И.Фалин, А.И.Фалин. Квартили в описательной статистике. *Математика*, 2011, №15, стр.8-14.

Начнём с того, что кратко повторим определение медианы числового набора. Неформально медиана числового набора $x = (x_1, \dots, x_n)$, упорядоченного по возрастанию, определяется как такое число, слева и справа от которого лежит одно и то же количество чисел набора. Точное определение длиннее:

- упорядочим по возрастанию рассматриваемый набор чисел; полученный набор (x_1, x_2, \dots, x_n) называется вариационным рядом;
- если этот набор состоит из нечётного количества чисел $n = 2k - 1$, то его медиана μ_x – это число x_k с номером k ;
- если же этот набор состоит из чётного количества чисел $n = 2k$, то его медиана μ_x – это число $\frac{x_k + x_{k+1}}{2}$, лежащее посередине отрезка $[x_k; x_{k+1}]$.

Если считать, что среднее арифметическое чисел x_k и x_{k+1} (т.е. k -го и $(k+1)$ -го членов набора) – это число $x_{k+\frac{1}{2}}$ с «номером» $\left(k + \frac{1}{2}\right)$, то оба случая можно объединить в один, сказав, что медиана – это $\frac{n+1}{2}$ -е в порядке возрастания число основного набора.

Значение медианы возрастает, когда её рассматривают как вторую квартиль Q_2 и в дополнение к ней вычисляют первую квартиль Q_1 (её обычно называют нижней) и третью квартиль Q_3 (её обычно называют верхней). Неформально говоря, квартили Q_1, Q_2, Q_3 делят исходный упорядоченный набор на 4 (примерно) равные части. Иначе говоря, нижняя квартиль – это медиана первой половины исходного набора, а верхняя квартиль – медиана второй половины исходного набора. Уточнить это неформальное определение нижней и верхней квартилей можно несколькими способами, которые обычно приводят к разным результатам (хотя и не очень сильно отличающимся) – общепринятого определения квартилей в описательной статистике нет (подробнее по этому поводу см., например, [2,3]). Скажем, в британском школьном учебнике [4], подготовленном экзаменационным центром OCR (Oxford, Cambridge and Royal Society of Arts), принято следующее определение.

Определение 1. *Чтобы найти квартили числового набора,*

1. *Нужно упорядочить числа исходного набора по возрастанию. Если некоторые числа набора повторяются, то они стоят одной группой, т.е. учитываются в наборе нужное количество раз.*
2. *Если набор содержит чётное количество чисел, то нужно разделить эту упорядоченную версию исходного набора на две равные (по числу элементов) половины. Медиана первой половины – это нижняя квартиль, а медиана второй половины – верхняя.*

3. Если набор содержит нечётное количество чисел, то нужно найти медиану и вычеркнуть её из набора (так что останется чётное количество чисел). После этого нужно оставшиеся числа разделить на две равные (по числу элементов) половины. Медиана первой половины – это нижняя квартиль, а медиана второй половины – верхняя.

Рассмотрим, например, набор 11, 11, 15, 15, 15, 15, 18, 20, 20, 160. Он состоит из 10 чисел, которые уже упорядочены по возрастанию. Первая половина – это набор 11, 11, 15, 15, 15. Он состоит из 5 чисел. Поэтому его медиана – третье по счёту число, т.е. 15 – это и будет нижняя квартиль Q_1 . Вторая половина – это набор 15, 18, 20, 20, 160. Он также состоит из 5 чисел. Поэтому его медиана – третье по счёту число, т.е. 20 – это и будет верхняя квартиль Q_3 . Вторая квартиль Q_2 – это медиана исходного набора, т.е. среднее арифметическое пятого и шестого чисел исходного набора:

$$Q_2 \equiv \mu_x = \frac{15+15}{2} = 15.$$
 Итак, для рассмотренного набора $Q_1 = 15$, $Q_2 = 15$, $Q_3 = 20$.

Возьмём теперь набор, содержащий нечётное количество чисел, например, 2, 7, 6, 2, 11, 8, 9, 4, 3 ($n=9$). После упорядочивания по величине мы получим набор: 2, 2, 3, 4, 6, 7, 8, 9, 11. Медиана этого набора (она будет второй квартилью Q_2) – это пятое по порядку число, т.е. 6. Вычёркивая медиану, мы получим набор из 8 чисел: 2, 2, 3, 4, 7, 8, 9, 11. Первая половина – это набор 2, 2, 3, 4. Он состоит из 4 чисел. Поэтому его медиана – среднее арифметическое второго и третьего чисел, т.е. 2,5 – это и будет нижняя квартиль Q_1 . Вторая половина – это набор 7, 8, 9, 11. Он также состоит из 4 чисел. Поэтому его медиана – среднее арифметическое второго и третьего чисел, т.е. 8,5 – это и будет верхняя квартиль Q_3 . Итак, для рассмотренного набора $Q_1 = 2,5$, $Q_2 = 6$, $Q_3 = 8,5$.

Именно это определение мы будем использовать дальше в нашей статье.

Обратим внимание на следующее обстоятельство. В рассмотренном только что примере ниже нижней квартили $Q_1 = 2,5$ находится два числа из рассматриваемого набора: $x_1 = 2$ и $x_2 = 2$. Эти два числа составляют $\frac{2}{9} \approx 22\%$ от общего количества чисел набора. Выше верхней квартили $Q_3 = 8,5$ находится тоже два числа из рассматриваемого набора: $x_8 = 9$ и $x_9 = 11$. Эти два числа составляют $\frac{2}{9} \approx 22\%$ от общего количества чисел набора. «Центральная» часть набора, которая состоит из чисел, лежащих между нижней и верхней квартилями, содержит 5 чисел, что составляет $\frac{5}{9} \approx 56\%$ от

общего количества чисел набора. Поэтому фразу «квартили делят набор на 4 равные части» нельзя понимать буквально.

Ещё один распространённый способ определения квартилей связан с определением медианы как $\frac{n+1}{2}$ -го в порядке возрастания числа основного набора. В соответствии с этим определением (назовём его **Определение 2**), i -я квартиль Q_i – это $\left(i \cdot \frac{n-1}{4} + 1\right)$ -е в порядке возрастания число основного набора. Иначе говоря, нижняя квартиль Q_1 – это $\frac{n+3}{4}$ -е в порядке возрастания число основного набора, медиана Q_2 – это (как и следовало ожидать) $\frac{n+1}{2}$ -е в порядке возрастания число основного набора, а верхняя квартиль Q_3 – это $\frac{3n+1}{4}$ -е в порядке возрастания число основного набора: $Q_1 = x_{\frac{n+3}{4}}$, $Q_2 = x_{\frac{n+1}{2}}$, $Q_3 = x_{\frac{3n+1}{4}}$. При этом, по определению, если «номер» $\frac{n+3}{4}$, $\frac{n+1}{2}$ или $\frac{3n+1}{4}$ имеет вид

- $k + \frac{1}{2}$, где k – некоторое натуральное число, то $x_{k+\frac{1}{2}} = \frac{1}{2}x_k + \frac{1}{2}x_{k+1}$;
- $k + \frac{1}{4}$, где k – некоторое натуральное число, то $x_{k+\frac{1}{4}} = \frac{3}{4}x_k + \frac{1}{4}x_{k+1}$;
- $k + \frac{3}{4}$, где k – некоторое натуральное число, то $x_{k+\frac{3}{4}} = \frac{1}{4}x_k + \frac{3}{4}x_{k+1}$.

Этот способ определения значений для дробных «номеров», когда значение x_{k+t} , где число k – натуральное, а число t лежит между 0 и 1, определяется по формуле $x_{k+t} = (1-t)x_k + tx_{k+1}$, называется линейной интерполяцией. Отметим, что в рассматриваемой ситуации число k называется целой частью числа $k+t$, а число t – дробной.

Применим это новое определение квартилей к рассмотренным ранее наборам.

Набор 11, 11, 15, 15, 15, 15, 18, 20, 20, 160 состоит из $n=10$ чисел.

Нижняя квартиль Q_1 равна

$$Q_1 = x_{\frac{13}{4}} = x_{3+\frac{1}{4}} = \frac{3}{4}x_3 + \frac{1}{4}x_4 = \frac{3}{4} \cdot 15 + \frac{1}{4} \cdot 15 = 15.$$

Напомним, что в соответствии с первым определением мы получили то же значение нижней квартили.

Верхняя квартиль Q_3 равна

$$Q_3 = x_{\frac{31}{4}} = x_{7+\frac{3}{4}} = \frac{1}{4}x_7 + \frac{3}{4}x_8 = \frac{1}{4} \cdot 18 + \frac{3}{4} \cdot 20 = 19,5.$$

Хотя первое определение дало для верхней квартили значение 20, отличие не очень большое.

Набор 2, 2, 3, 4, 6, 7, 8, 9, 11 состоит из $n=9$ чисел. Нижняя квартиль Q_1 равна $Q_1 = x_{\frac{12}{4}} = x_3 = 3$ (определение 1 дало для нижней квартили значение 2,5). Верхняя квартиль Q_3 равна $Q_3 = x_{\frac{28}{4}} = x_7 = 8$ (определение 1 дало для верхней квартили значение 8,5). Как и в первом примере, оба определения дают для квартилей близкие значения.

Термин «квартиль» (англ.: quartile) происходит от латинского слова *quartus* (в средневековой латыни: *quartilis*) – «четвёртый». От этой латинской основы произошли и слова: «квартал» – четвёртая часть года (три месяца), «квартет» – музыкальный ансамбль из четырёх исполнителей. В статистике термин «квартиль» появился в 1879 г. и обычно связывается с именем английского учёного Ф.Гальтона (F.Galton, 1822–1911).

2. Квартили в зарубежных школьных программах статистики

В странах с давней традицией преподавания статистики в школах понятие квартили и связанные с ним понятия (интерквартильный размах, квартильная диаграмма, выбросы и т.д.) включены в школьные программы изучения статистики. Например, английская (в других регионах Великобритании ситуация аналогична) школьная программа по статистике (см. [5], стр.35, 40, или [6], стр.13, 14) требует, чтобы учащиеся могли

- вычислять квартили и интерквартильный размах;
- строить, интерпретировать и использовать квартильные диаграммы;
- использовать квартили для выявления аномальных значений (выбросов).

Соответствующий материал излагается в британских школьных учебниках по статистике; например, учебник [7] содержит разделы:

- «Размах и квартили» (стр. 95),
- «Интерквартильный и интерперцентильный размах» (стр. 96-98),
- «Квартильная диаграмма» (стр.105-106),
- «Измерение асимметрии с использованием квартильного коэффициента» (стр.107),

а учебник [4] – разделы:

- «Интерквартильный размах» (раздел 3.3, стр.45-49),
- «Квартильные диаграммы» (раздел 3.4, стр.50-51),

- «Выбросы» (раздел 3.5, стр.52-53).

Задачи, которые предлагаются на экзаменах по статистике на получение аттестата о среднем образовании (General Certificate of Secondary Education – GCSE; это некоторый аналог отечественной ГИА) часто включают задания по указанным выше темам (см. , например, [8], задача 1В, или [9], экзамен 1Н, задача 8).

В зарубежных методических изданиях по статистике часто публикуются статьи (см., например, [10-12]), посвящённые преподаванию в школах понятия квартилей и связанных с ними более сложных понятий. В них обсуждается как методика преподавания, так и возникающие при этом проблемы с пониманием школьниками этих относительно сложных тем описательной статистики. В некоторых публикациях, скажем, в [13], на основе опыта преподавания в разных странах высказывается определённый скептицизм по поводу возможности и необходимости включения квартилей в школьный курс статистики.

3. Применение компьютеров для вычисления квартилей

Второе определение квартилей реализовано в электронных таблицах Microsoft Office Excel, где для подсчёта квартилей можно использовать стандартную функцию КВАРТИЛЬ.

Вернемся к ранее рассмотренному набору чисел 11, 11, 15, 15, 15, 15, 18, 20, 20, 160. Введём эти данные в ячейки с адресами A1, A2, ..., A10.

Чтобы найти первую квартиль введём в ячейку A11 формулу =КВАРТИЛЬ(A1:A10,1) (в зависимости от версии Excel для разделения параметров в функциях используется или запятая, или точка с запятой.) Значение функции (т.е. первой квартили) равно 15.

Чтобы найти вторую квартиль (медиану), введём в ячейку A12 формулу =КВАРТИЛЬ(A1:A10,2). Значение второй квартили на этом наборе также равно 15.

Для нахождения третьей квартили введём в ячейку A13 формулу =КВАРТИЛЬ(A1:A10,3). Её значение равно 19,5.

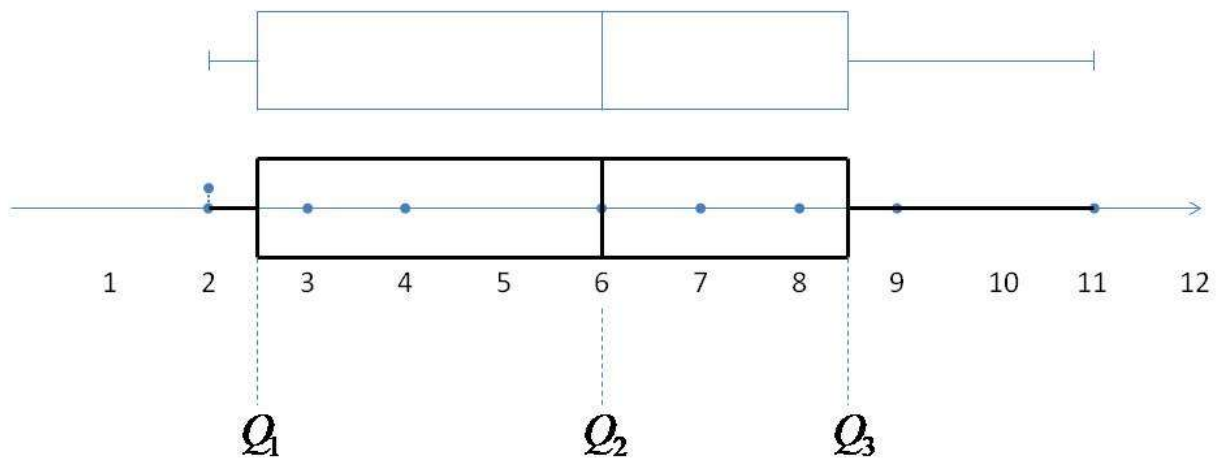
Отметим, что для функции КВАРТИЛЬ совершенно неважно, упорядочены или нет числа набора. Кроме того, эта функция позволяет найти наибольшее и наименьшее числа набора: формула КВАРТИЛЬ(A1:A10,0) даст наименьшее число набора, а формула КВАРТИЛЬ(A1:A10,4) – наибольшее.

4. Квартильная диаграмма

Обычно квартили изображают графически с помощью *квартильной диаграммы* (английский термин «box and whisker plot», «box and whisker diagram»; буквально – «ящик с усами»).

Чтобы нарисовать квартильную диаграмму, на числовой прямой нужно отметить квартили и нарисовать два смежных прямоугольника («ящика») одинаковой высоты или, что то же самое, нарисовать прямоугольник на отрезке $[Q_1; Q_3]$ и разделить его на два прямоугольника вертикальным отрезком, проходящим через медиану (см. рис.1). Высоту прямоугольника можно взять любой, хотя есть варианты квартильных диаграмм, когда высота несёт определённую смысловую нагрузку [14].

Рис.1



Кроме того, на числовой оси нужно отметить экстремальные значения исходного числового набора (т.е. наибольшее и наименьшее числа) и соединить их горизонтальными отрезками («усами») с серединами соответствующих вертикальных отрезков (проведённых через нижнюю и верхнюю квартили); концы этих отрезков обычно отмечают небольшими вертикальными отрезками.

Ситуация, изображённая на рис.1, соответствует набору 2, 7, 6, 2, 11, 8, 9, 4, 3, рассмотренному выше. Как было показано, в соответствии с Определением 1 (которое мы используем как основное), для этого набора $Q_1 = 2,5$, $Q_2 = 6$, $Q_3 = 8,5$.

Для большей наглядности мы отметили и числа исходного набора. Обратим внимание на число 2. Оно появилось в рассматриваемом наборе два

раза. Мы отразили это обстоятельство двумя точками – одна из них стоит на числовой оси, а вторая чуть выше её.

Часто получившуюся конфигурацию приподнимают над осью абсцисс (мы сделали это на рис.1) – это удобно для сравнения разных наборов, а иногда размещают вертикально.

Таким образом, квартильная диаграмма в наглядной форме показывает положение чисел набора на числовой оси:

- медиана указывает на «среднее» значение набора;
- ширина «ящичков» показывает разброс (примерно) 50% центральных, наиболее характерных значений;
- длина «усов» показывает насколько сильно выброшены в сторону (примерно) 25% наименьших и (примерно) 25% наибольших чисел набора;
- сравнивая между собой ширину левого и правого «ящичков», а также длины левого и правого «усов», можно судить о том, насколько несимметричным является рассматриваемый набор.

Графическое описание набора данных с помощью квартильной диаграммы было предложено в 1970 г. известным американским статистиком, профессором университета Принстона, Джоном Тьюки (John Tukey, 1915–2000). Этот метод стал общепринятым после опубликования в 1977 г. его книги [15].

5. Интерквартильный размах

С помощью верхней и нижней квартилей определяют важную меру рассеивания набора чисел – *интерквартильный размах* (соответствующий английский термин *interquartile range* переводят и как *межквартильный размах*). По определению, интерквартильный размах (его обычно обозначают *IQR*) – это разность $Q_3 - Q_1$:

$$IQR = Q_3 - Q_1.$$

Интерквартильный размах показывает, насколько разбросаны 50% «центральных» значений рассматриваемого набора чисел. Это понятие было введено в 1882 г. Ф.Гальтоном.

6. Выбросы

Выбросы – это числа, которые сильно отличаются от остальных чисел набора и в ситуации, которую описывает рассматриваемый набор чисел, являются необычными. Рассмотрим, например, следующий набор из 10 чисел: 63, 68, 62, 59, 64, 62, 67, 65, 94, 64 (это могут быть оценки ЕГЭ по математике для группы школьников). На рис. 2 эти числа изображены на числовой оси. Видно, что 9 оценок из 10 стоят плотной группой, а одна

оценка, 94, стоит далеко справа. Если бы мы сравнивали эту группу школьников с другой, в которой математика преподаётся по другой методике (например, чтобы понять, какая из двух методик лучше), то эту оценку было бы неразумно принимать в расчёт, т.к. столь высокая оценка, видимо, связана с математической одарённостью ученика и мало зависит от методики преподавания математики в школе. В рассматриваемой ситуации число 94 и будет выбросом.

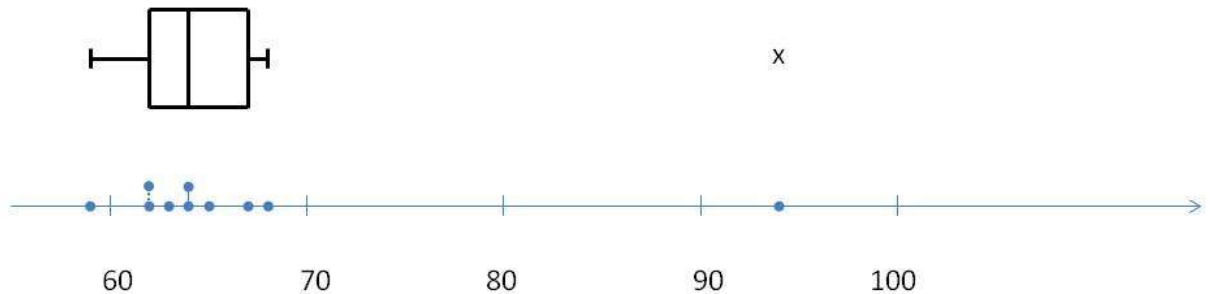
Однозначного ответа на вопрос, насколько далеко должно лежать число от основной массы значений набора, чтобы его можно было считать выбросом, нет. Джон Тьюки [15] предложил следующий подход, который является общепринятым в описательной статистике:

- числа набора, лежащие между нижней и верхней квартилями (т.е. в «ящике» квартильной диаграммы), являются наиболее характерными;
- числа, которые отклоняются от нижней и верхней квартилей не больше, чем на полтора интерквартильных размаха, т.е. удовлетворяют неравенству $f_* \leq x < Q_1$, где $f_* = Q_1 - 1,5 \cdot IQR$, или неравенству $Q_3 < x \leq f^*$, где $f^* = Q_3 + 1,5 \cdot IQR$, не столь характерны, но должны учитываться как значимые, т.к. их отклонения от типичных значений не очень большие;
- те числа, которые отклоняются от нижней и верхней квартилей больше, чем на полтора интерквартильных размаха, т.е. удовлетворяют неравенству $x < f_*$ или неравенству $x > f^*$, следует считать выбросами.

Чтобы наглядно отразить выбросы, квартильную диаграмму немного модифицируют, именно, если есть выбросы, то рисуют «усы» не до минимального и максимального чисел набора, а лишь до чисел набора, наиболее удалённых от соответствующих квартилей, но не дальше, чем на $1,5 \cdot IQR$. Таким образом, «усы» не переходят через барьеры f_* и f^* . Числа набора, которые не попадают на эту модифицированную квартильную диаграмму (т.е. выбросы), отмечают особо. Часто для этого используют не точки, а маленькие крестики.

Для рассматриваемого нами иллюстративного набора такая модифицированная диаграмма изображена на рис.2 (в нашем случае медиана равна 64, нижняя квартиль равна 62, верхняя квартиль равна 67, интерквартильный размах равен 5, имеется только один выброс, равный 94).

Рис.2



Если в анализируемом наборе есть выбросы, необходимо разобраться, почему они появились. Иногда выбросы являются следствием ошибок при сборе данных. В этом случае их нужно исправить или, если это невозможно, исключить выбросы из набора. Но чаще выбросы – это верные значения. В этом случае нужно особенно тщательно проанализировать причину их появления. Обычно это позволяет получить важные выводы о реальной ситуации, которую описывает анализируемый набор данных. В частности, необходимо понять, могут ли подобные экстремальные значения появиться в других подобных ситуациях.

Имея в виду проведённые выше рассуждения, уместно сделать несколько замечаний по поводу интерквартильного размаха:

- Важное достоинство интерквартильного размаха заключается в том, что его значение не зависит от выбросов.
- С другой стороны, оставшиеся 50% чисел набора игнорируются – это, конечно, недостаток этой меры рассеивания.
- Поэтому применять интерквартильный размах имеет смысл в тех случаях, когда выбросы не являются типичными значениями, характерными для ситуации, описываемой набором, который мы изучаем, и их следует игнорировать. Вообще, важно понимать, что математика – это только аппарат для исследования реальной ситуации. Если не учитывать важные свойства изучаемого объекта,

то формальные математические вычисления могут привести нас не к самым разумным выводам.

7. Асимметрия набора

Как мы уже отмечали, с помощью квартилей можно понять, насколько несимметричным является рассматриваемый набор.

Если медиана находится точно посередине между нижней и верхней квартилью (т.е. левый и правый «ящики» на квартильной диаграмме имеют одинаковые основания), то из (примерно) 50% «центральных» значений рассматриваемого набора чисел примерно половина лежит в левом «ящике», а вторая половина – в равном ему по размеру правом. Иначе говоря, эти «центральные» значения в целом расположены вокруг медианы симметрично.

Если нижняя квартиль находится от медианы дальше, чем верхняя (т.е. левый «ящик» на квартильной диаграмме больше правого), то левая половина «центральных» значений рассматриваемого набора чисел разбросана больше, чем правая. Иначе говоря, эти «центральные» значения в целом расположены вокруг медианы несимметрично, со скосом в левую сторону. Такой набор называют отрицательно асимметричным.

Если же верхняя квартиль находится от медианы дальше, чем нижняя (т.е. правый «ящик» на квартильной диаграмме больше левого), то правая половина «центральных» значений рассматриваемого набора чисел разбросана больше, чем левая. Иначе говоря, эти «центральные» значения в целом расположены вокруг медианы несимметрично, со скосом в правую сторону. Такой набор называют положительно асимметричным.

Асимметрию набора можно описать и количественно. В статистике для этого введено несколько разных величин (примерно так же, как положение набора можно описывать средним значением, медианой и модой). Например,

можно использовать отношение $\lambda = \frac{Q_2 - Q_1}{Q_3 - Q_2}$ (при $Q_3 \neq Q_2$). Говорят, что

точка Q_2 делит отрезок $[Q_1; Q_3]$ в отношении λ . Число λ показывает, во сколько раз левый «ящик» квартильной диаграммы шире правого:

- если $\lambda=1$, то эти ящики имеют одинаковую ширину и распределение является симметричным в смысле данного выше определения;
- если $\lambda > 1$, то «левый» ящик шире правого и распределение является отрицательно асимметричным в смысле данного выше определения;
- если $\lambda < 1$, то «правый» ящик шире левого и распределение является положительно асимметричным в смысле данного выше определения.

Единственный недостаток этого, в общем то естественного определения, заключается в том, что для симметричного распределения, т.е. с нулевой асимметрией, коэффициент равен 1, для положительно асимметричного

распределения коэффициент меньше 1, для отрицательно асимметричного распределения коэффициент больше 1. Хотелось бы ввести такую меру, чтобы для симметричного распределения, т.е. с нулевой асимметрией, она была бы равна 0, для положительно асимметричного распределения была бы больше 0 (т.е. положительна), для отрицательно асимметричного распределения была бы меньше 0 (т.е. отрицательна). Имея это в виду, в качестве меры асимметрии рассматривают число $k = \frac{1-\lambda}{1+\lambda}$. Его обычно

называют *квартильный коэффициент* или *асимметрия Боули* (A.L.Bowley, 1869–1957 – английский статистик и экономист). Через квартили квартильный коэффициент выражается следующим образом:

$$k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \equiv \frac{\text{верхняя квартиль} + \text{нижняя квартиль} - 2 \times \text{медиана}}{\text{верхняя квартиль} - \text{нижняя квартиль}}.$$

Квартильный коэффициент определён только для наборов, у которых верхняя квартиль не совпадает с нижней, т.е. «центральные» 50% значений набора разбросаны по отрезку $[Q_3; Q_1]$ ненулевой длины.

Возможные значения квартильного коэффициента лежат на отрезке $[-1; +1]$. Действительно, двойное неравенство $-1 \leq \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \leq 1$

равносильно двум неравенствам: $-Q_3 + Q_1 \leq Q_3 + Q_1 - 2Q_2$ и $Q_3 + Q_1 - 2Q_2 \leq Q_3 - Q_1$, которые после приведения подобных членов сводятся к двойному неравенству $Q_1 \leq Q_2 \leq Q_3$, которое, очевидно, истинно.

Если квартильный коэффициент равен 0, то это означает, что $Q_3 + Q_1 - 2Q_2 = 0$, т.е. $Q_3 - Q_2 = Q_2 - Q_1$. Но $Q_3 - Q_2$ – это ширина правого ящика, а $Q_2 - Q_1$ – левого. Поэтому равенство $Q_3 - Q_2 = Q_2 - Q_1$ означает симметрию набора в смысле определения, данного в начале этого раздела.

При изменении параметра λ от 0 до $+\infty$ функция $k(\lambda) = \frac{1-\lambda}{1+\lambda}$ монотонно убывает от 1 до -1 . Таким образом, значение квартильного коэффициента, близкое к -1 , равносильно тому, что значение λ велико, что, в свою очередь, означает, что левый «ящик» много шире правого, т.е. большую отрицательную асимметрию. Значение квартильного коэффициента, близкое к $+1$, равносильно тому, что значение λ очень мало, что, в свою очередь, означает, что правый «ящик» много шире левого, т.е. большую положительную асимметрию.

Особо подчеркнём, что при обсуждении симметрии или асимметрии числового набора мы принимали в расчёт только (примерно) 50% «центральных», наиболее типичных чисел, игнорируя числа, которые лежат ниже нижней квартили или выше верхней. В статистике есть и другие меры асимметрии, которые используют все числа набора. Наиболее простым из

них (он использует только основные меры положения и рассеивания) является коэффициент асимметрии Пирсона:

$$3 \frac{M_x - \mu_x}{s_x} = 3 \frac{\text{среднее значение} - \text{медиана}}{\text{стандартное отклонение}}.$$

Он равен нулю для наборов, у которых среднее значение совпадает с медианой, и только для них. Таким образом, этот подход считает характерным свойством симметричных наборов равенство среднего значения и медианы.

Литература

1. Г.Фалин, А.Фалин. О мерах положения числового набора. *Математика*, 2011 (принято к публикации).
2. J. Freund, B. Perles. A New Look at Quartiles of Ungrouped Data. *The American Statistician*, Vol. 41, No. 3 (Aug., 1987), pp. 200-203.
3. R.J. Hyndman, Y. Fan. Sample Quantiles in Statistical Packages. *The American Statistician*, Vol. 50, No. 4 (Nov., 1996), pp. 361-365.
4. S. Dobbs, J. Miller. *Statistics 1*. Cambridge University Press, 2009.
5. Specification. Edexcel GCSE in Statistics (2ST01). Edexcel Limited, 2008.
6. GCSE Specification. Statistics (for certification 2011 onwards). AQA, 2008.
7. GCSE Statistics. Complete Revision and Practice. Coordination Group Publication, 2010.
8. Edexcel GCSE Statistics. Paper 1H. Wednesday 18 June 2008.
9. Sample Assessment Materials. Edexcel GCSE in Statistics (2ST01). February 2010.
10. M. Pfannkuch. Comparing box plot distributions: a teacher's reasoning. *Statistics Education Research Journal*, 2006, 5(2), pp.27-45.
11. U. Kortenkamp, K. Rolka. Using technology in the teaching and learning of box plots. *Proceedings of CERME 6*, January 28th-February 1st 2009, Lyon France. INRP 2010, pp.1070-1080.
12. E. Langford. Quartiles in Elementary Statistics. *Journal of Statistics Education*, 2006, Volume 14, Number 3.
13. A.Bakker, R.Biehler, C.Konold. Should Young Students Learn About Box Plots? *Curricular Development in Statistics Education*, Sweden, 2004, pp.163-173.
14. R. McGill, J. W. Tukey, W.A. Larsen. Variations of Box Plots. *The American Statistician*, Vol. 32, No. 1. (Feb., 1978), pp. 12-16.
15. Дж.Тьюки. Анализ результатов наблюдений. Разведочный анализ. М.: Мир, 1981 (оригинальное издание: J.W.Tukey. *Exploratory Data Analysis*. Reading MA: Addison-Wesley Publishing Co., 1977).