

# Использование линейных преобразований при вычислении статистических характеристик наборов данных<sup>1</sup>

**Г.И.Фалин**, д.ф.м.н., проф.  
кафедра теории вероятностей  
механико-математический факультет  
МГУ им М.В.Ломоносова (Москва)  
<http://mech.math.msu.su/~falin>

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

Часто можно подобрать линейное преобразование набора данных, приводящее к набору более «простых» чисел (для которого легче вычислить как среднее значение, так и дисперсию). Мы приводим с доказательствами общие формулы, связывающие основные статистические характеристики исходного и преобразованного наборов, и на примерах демонстрируем пользу этих формул для решения задач.

**Ключевые слова:** описательная статистика, среднее значение, медиана, дисперсия, размах, линейное преобразование

## Введение

В описательной статистике основными характеристиками набора числовых данных  $x = (x_1, x_2, \dots, x_n)$  являются:

- *среднее значение*  $M_x = \frac{x_1 + x_2 + \dots + x_n}{n}$ ,
- *дисперсия*  $D_x = \frac{(x_1 - M_x)^2 + (x_2 - M_x)^2 + \dots + (x_n - M_x)^2}{n}$ .

Вместо дисперсии часто используют *среднее квадратичное отклонение* (его также называют *стандартным отклонением*)  $\sigma_x = \sqrt{D_x}$ .

Наряду со средним и дисперсией в качестве мер положения и разброса используют характеристики, полученные на основе вариационного ряда. Напомним, что *вариационный ряд* – это набор  $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ , который получится, если числа исходного набора  $x$  расположить в порядке

---

<sup>1</sup> Г.И.Фалин, А.И.Фалин. Использование линейных преобразований при вычислении статистических характеристик наборов данных. *Математика в школе*, 2012, №4, 66-73.

возрастания (таким образом,  $x_{(k)}$  – это  $k$ -е в порядке возрастания число исходного набора). С помощью вариационного ряда определяются:

- *медиана*

$$\mu_x = \begin{cases} x_{(k)} & , \text{ если } n = 2k - 1 \text{ – нечётное число,} \\ \frac{x_{(k)} + x_{(k+1)}}{2} & , \text{ если } n = 2k \text{ – чётное число,} \end{cases}$$

- *размах*  $R_x = x_{(n)} - x_{(1)} \equiv \max(x_1, \dots, x_n) - \min(x_1, \dots, x_n)$ .

В принципе вычисление этих величин является несложной арифметической задачей. Однако, если числа  $x_1, x_2, \dots, x_n$  – «некрасивые», то для наборов  $x$  *большой длины* вычисление таких характеристик как среднее и дисперсия даже с помощью калькулятора потребует относительно много усилий. Применение для обработки информации электронных таблиц Microsoft Office Excel не решает все проблемы (ошибки при вводе данных всё ещё возможны). В статистике известен приём, который позволяет упростить вычисления в случае, когда данные имеют линейную структуру. В школьной учебной литературе этот вопрос затронут лишь вскользь и без доказательств (см., например, [1], глава III, раздел 16). Цель нашей заметки – рассмотреть его подробнее. При этом мы стремились решить ещё одну задачу – показать, что описательная статистика – это математическая наука с определениями, теоремами, строгими доказательствами, излагаемыми на обычном алгебраическом языке формул. Этим объясняется некоторый формализм (который на первый взгляд может показаться неоправданным), особенно в разделе, посвящённом медиане и размаху.

### Линейные преобразования наборов данных

Пусть  $x = (x_1, x_2, \dots, x_n)$  – набор из  $n$  чисел. Предположим, что все числа этого набора могут быть получены из некоторого нового, вспомогательного, набора  $u = (u_1, u_2, \dots, u_n)$  применением одного и того же линейного преобразования  $x = au + b$ , т.е. верны следующие числовые равенства:

$$x_1 = au_1 + b, x_2 = au_2 + b, \dots, x_n = au_n + b. \quad (1)$$

Эту систему из  $n$  равенств можно записать и короче:  $x_i = au_i + b$ , где  $i = 1, 2, \dots, n$  (т.е. индекс  $i$  принимает последовательно все значения от 1 до  $n$ ).

Параметр  $a$  называется *масштабным коэффициентом* (или *масштабным параметром*; используют также и термин «*нормирующий параметр*»), а параметр  $b$  – *параметром сдвига* (или, короче, *сдвигом*; используют также и термин «*центрирующий параметр*»).

Например, если дан основной набор  $x_1 = 105, x_2 = 110, x_3 = 95$ , то числа этого набора можно записать так:  $x_1 = 100 + 5 = 100 + 5 \cdot 1$ ,  $x_2 = 100 + 10 = 100 + 5 \cdot 2$ ,  $x_3 = 100 - 5 = 100 + 5 \cdot (-1)$ . Иначе говоря,  $x_i = 100 + 5u_i$ , где  $u_1 = 1, u_2 = 2, u_3 = -1$ .

Чтобы в будущем исключить случаи, которые не представляют интереса для приложений, а лишь усложняют формулировки утверждений, мы будем рассматривать только линейные преобразования с ненулевым коэффициентом  $a$ .

Линейные преобразования наборов данных обладают важными свойствами:

**Свойство 1** (симметричность) Если набор  $(x_1, x_2, \dots, x_n)$  получен из набора  $(u_1, u_2, \dots, u_n)$  линейным преобразованием  $x = au + b$ , то набор  $(u_1, u_2, \dots, u_n)$  может быть получен из набора  $(x_1, x_2, \dots, x_n)$  линейным преобразованием

$$u = \frac{1}{a} \cdot x + \left( -\frac{b}{a} \right).$$

Обратим внимание на то, что коэффициент  $\frac{1}{a}$  не равен нулю.

**Свойство 2** (транзитивность) Если набор  $(x_1, x_2, \dots, x_n)$  получен из набора  $(u_1, u_2, \dots, u_n)$  линейным преобразованием  $x = a'u + b'$ , а набор  $(u_1, u_2, \dots, u_n)$ , в свою очередь, получен из набора  $(v_1, v_2, \dots, v_n)$  линейным преобразованием  $u = a''v + b''$ , то набор  $(x_1, x_2, \dots, x_n)$  может быть получен из набора  $(v_1, v_2, \dots, v_n)$  линейным преобразованием  $x = a'a''v + (a'b'' + b')$ .

Обратим внимание на то, что коэффициент  $a'a''$  не равен нулю (так как отличны от нуля коэффициенты  $a'$  и  $a''$ ).

**Свойство 3** (рефлексивность) Любой набор  $(x_1, x_2, \dots, x_n)$  можно получить из самого себя линейным преобразованием  $ax + b$ , с коэффициентами  $a = 1$ ,  $b = 0$ .

В высшей алгебре связь между парами объектов (любой природы), обладающую свойствами симметричности, транзитивности и рефлексивности называют отношением эквивалентности. Эквивалентные объекты в некотором смысле можно считать неотличимыми друг от друга (или различными представлениями одного и того же объекта). В нашем случае, наборы, отличающиеся друг от друга линейным преобразованием, в сущности задают один и тот же набор данных, но измеренных в разных единицах измерения.

Для приложений к описательной статистике важны свойства, связывающие средние значения (медианы, дисперсии, стандартные отклонения, размахи) исходного и преобразованного наборов данных.

### Среднее значение и линейные преобразования наборов данных

**Теорема 1.** Пусть  $M_x$  – среднее значение набора  $x = (x_1, x_2, \dots, x_n)$ , а  $M_u$  – среднее значение набора  $u = (u_1, u_2, \dots, u_n)$ , связанного с набором  $x$  соотношениями (1). Тогда

$$M_x = aM_u + b. \quad (2)$$

Иначе говоря, среднее значение  $M_x$  основного набора  $x$  получается из среднего значения  $M_u$  вспомогательного набора  $u$  с помощью того же линейного преобразования  $x = au + b$ , которое преобразует вспомогательный набор в основной.

**Доказательство.** Следующая цепочка формул связывает величины  $M_x$  и  $M_u$  между собой:

$$\begin{aligned} M_x &= \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{(au_1 + b) + (au_2 + b) + \dots + (au_n + b)}{n} \\ &= \frac{a(u_1 + u_2 + \dots + u_n) + nb}{n} = a \frac{u_1 + u_2 + \dots + u_n}{n} + b = aM_u + b. \end{aligned}$$

**Следствие 1.** Пусть  $(x_1, x_2, \dots, x_n)$  – некоторый набор чисел, а  $M_x$  – его среднее значение. Рассмотрим набор  $(u_1, u_2, \dots, u_n)$ , полученный из набора  $(x_1, x_2, \dots, x_n)$  применением следующего линейного преобразования:

$$u_i = x_i - M_x \Leftrightarrow x_i = M_x + u_i. \quad (3)$$

Величины  $u_i$  показывают отклонения величин  $x_i$  от их среднего значения. Поскольку среднее значение набора чисел в некотором смысле является его «центром», преобразование (3) называют *центрированием* исходного набора.

Операция центрирования соответствует следующим значениям нормирующего и центрирующего параметров  $a$  и  $b$ :  $a=1$ ,  $b=M_x$ . Поэтому для центрированного набора общее соотношение (2) примет вид:  $M_x = M_u + M_x$  или, что то же самое,  $M_u = 0$ .

Таким образом, среднее значение (а вместе с ним и обычная сумма) отклонений величин  $x_i$  от их среднего значения обязательно равна нулю.

Последнее свойство без доказательства отмечается в школьных учебниках (см., например, [1], гл. III, раздел 13, или [2], гл. 5, раздел 5.3).

Часто можно найти такое линейное преобразование исходного набора, что новый набор выглядит проще. Тогда найти среднее значение для нового набора легче, чем для исходного. Полученное общее соотношение (2) между средними значениями исходного и нового наборов позволяет найти среднее значение исходного набора меньшими вычислительными усилиями.

Рассмотрим примеры применения этого свойства (для простоты изложения мы ограничиваемся наборами небольшой длины).

**Пример 1.** Пусть дан набор  $x = (2007; 2011; 2003; 2006; 1998)$ , так что

$$x_1 = 2007, x_2 = 2011, x_3 = 2003, x_4 = 2006, x_5 = 1998.$$

Введём новый набор  $u = (u_1, u_2, u_3, u_4, u_5)$  по формуле:  $u_i = x_i - 2000$ . Иначе говоря, представим числа основного набора в виде  $x_i = 2000 + u_i \equiv au_i + b$ , где

$a=1$ ,  $b=2000$ . Тогда  $u_1=7, u_2=11, u_3=3, u_4=6, u_5=-2$ . Среднее значение этого набора найти гораздо легче:

$$M_u = \frac{7+11+3+6+(-2)}{5} = \frac{25}{5} = 5.$$

Поскольку  $x_i = u_i + 2000$ , аналогичное соотношение связывает средние значения:  $M_x = M_u + 2000$ . Поэтому для среднего значения исходного набора мы имеем:  $M_x = M_u + 2000 = 2005$ .

**Пример 2.** Пусть дан набор

$$x_1 = 2,91, x_2 = 3,07, x_3 = 3,02, x_4 = 3,05, x_5 = 3,01, x_6 = 3,03.$$

Введём набор  $(u_1, u_2, u_3, u_4, u_5, u_6)$  по формуле:  $u_i = 100(x_i - 3)$ . Иначе говоря, представим числа основного набора в виде  $x_i = 3 + 0,01u_i \equiv au_i + b$ , где  $a = 0,01, b = 3$ . Тогда  $u_1 = -9, u_2 = 7, u_3 = 2, u_4 = 5, u_5 = 1, u_6 = 3$ . Среднее значение этого набора найти гораздо легче:

$$M_u = \frac{-9+7+2+5+1+3}{6} = \frac{9}{6} = 1,5.$$

Поскольку  $x_i = 0,01u_i + 3$ , аналогичное соотношение связывает средние значения:  $M_x = 0,01M_u + 3$ . Поэтому для среднего значения исходного набора мы имеем:  $M_x = 0,01 \cdot 1,5 + 3 = 3,015$ .

### Дисперсия и линейные преобразования наборов данных

**Теорема 2.** Пусть  $D_x$  – дисперсия набора  $x = (x_1, x_2, \dots, x_n)$ , а  $D_u$  – дисперсия набора  $u = (u_1, u_2, \dots, u_n)$ , связанного с набором  $x$  соотношениями (1). Тогда

$$D_x = a^2 D_u. \quad (4)$$

**Доказательство.** Следующая цепочка формул связывает величины  $D_x$  и  $D_u$  между собой:

$$\begin{aligned} D_x &= \frac{(x_1 - M_x)^2 + (x_2 - M_x)^2 + \dots + (x_n - M_x)^2}{n} \\ &= \frac{(au_1 + b - aM_u - b)^2 + (au_2 + b - aM_u - b)^2 + \dots + (au_n + b - aM_u - b)^2}{n} \\ &= \frac{a^2(u_1 - M_u)^2 + a^2(u_2 - M_u)^2 + \dots + a^2(u_n - M_u)^2}{n} \\ &= a^2 \frac{(u_1 - M_u)^2 + (u_2 - M_u)^2 + \dots + (u_n - M_u)^2}{n} = a^2 D_u. \end{aligned}$$

**Следствие 2.** Из (4) для средних квадратичных отклонений мы имеем:

$$\sigma_x = |a| \sigma_u. \quad (5)$$

В частности, если масштабный множитель  $a$  равен 1 или  $-1$ , то  $D_x = D_u$ ,  $\sigma_x = \sigma_u$ .

**Следствие 3.** Пусть  $(x_1, x_2, \dots, x_n)$  – некоторый набор чисел,  $M_x$  – его среднее значение,  $D_x$  – дисперсия, а  $\sigma_x = \sqrt{D_x}$  – среднее квадратичное отклонение. Предположим, что дисперсия отлична от нуля и рассмотрим набор  $(u_1, u_2, \dots, u_n)$ , полученный из набора  $(x_1, x_2, \dots, x_n)$  применением следующего линейного преобразования:

$$u_i = \frac{x_i - M_x}{\sigma_x} \Leftrightarrow x_i = \sigma_x u_i + M_x. \quad (6)$$

Преобразование (6) соответствует следующим значениям нормирующего и центрирующего параметров  $a$  и  $b$ :  $a = \sigma_x$ ,  $b = M_x$ . Поэтому для набора  $u_1, u_2, \dots, u_n$  общие соотношения (2) и (4) примут вид:  $M_x = \sigma_x M_u + M_x$ ,  $D_x = \sigma_x^2 D_u$ . Поскольку  $\sigma_x^2 = D_x$ , эти соотношения равносильны тому, что  $M_u = 0$ ,  $D_u = 1$ . Набор с нулевым средним и единичной дисперсией называется стандартным. Соответственно, преобразование (6) называют *стандартизацией* исходного набора. Это преобразование играет важную роль в теории вероятностей.

**Пример 3.** Пусть дан набор

$$x_1 = 2007, x_2 = 2011, x_3 = 2003, x_4 = 2006, x_5 = 1998.$$

Как мы видели в Примере 1, числа этого набора можно представить в виде  $x_i = u_i + 2000$ , где  $u_1 = 7, u_2 = 11, u_3 = 3, u_4 = 6, u_5 = -2$ . Среднее значение нового набора мы нашли в Примере 1:  $M_u = 5$ . Теперь для его дисперсии мы имеем:

$$D_u = \frac{(7-5)^2 + (11-5)^2 + (3-5)^2 + (6-5)^2 + (-2-5)^2}{5} = \frac{94}{5} = 18,8.$$

Соответственно,  $\sigma_u = \sqrt{18,8} \approx 4,336$ . Поскольку масштабный множитель  $a = 1$ , для исходного набора мы имеем:  $D_x = D_u = 18,8$ ,  $\sigma_x = \sigma_u \approx 4,336$ .

**Пример 4.** Пусть дан набор

$$x_1 = 2,91, x_2 = 3,07, x_3 = 3,02, x_4 = 3,05, x_5 = 3,01, x_6 = 3,03.$$

Как мы видели в Примере 2 числа этого набора можно представить в виде  $x_i = 3 + 0,01u_i$ , где  $u_1 = -9, u_2 = 7, u_3 = 2, u_4 = 5, u_5 = 1, u_6 = 3$ . Среднее значение нового набора мы нашли в Примере 2:  $M_u = 1,5$ . Теперь для его дисперсии мы имеем:

$$D_u = \frac{(-9-1,5)^2 + (7-1,5)^2 + (2-1,5)^2 + (5-1,5)^2 + (1-1,5)^2 + (3-1,5)^2}{6} = \frac{155,50}{6} \approx 25,92.$$

Отсюда  $\sigma_u \approx 5,091$ , так что  $\sigma_x = 0,01\sigma_u \approx 0,0591$ , а  $D_x = 0,01^2 D_u \approx 0,002592$ .

### Медиана и размах и линейные преобразования наборов данных

**Теорема 3.** Пусть  $\mu_x$  и  $R_x$  – соответственно медиана и размах набора  $x = (x_1, x_2, \dots, x_n)$ , а  $\mu_u$  и  $R_u$  – соответственно медиана и размах набора  $u = (u_1, u_2, \dots, u_n)$ , связанного с набором  $x$  соотношениями (1). Тогда

$$\mu_x = a\mu_u + b \quad (7)$$

$$R_x = |a| \cdot R_u \quad (8)$$

**Доказательство.** Упорядочим числа основного набора  $x$  по возрастанию, т.е. образуем вариационный ряд  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Вычитая из всех чисел этого ряда одно и то же число  $b$  (параметр сдвига), мы получим цепочку неравенств

$$x_{(1)} - b \leq x_{(2)} - b \leq \dots \leq x_{(n)} - b. \quad (9)$$

Дальнейшие рассуждения будут зависеть от знака масштабного коэффициента  $a$ .

1 случай. Если  $a > 0$ , то деление всех чисел цепочки неравенств (9) на  $a$  даст:

$$\frac{x_{(1)} - b}{a} \leq \frac{x_{(2)} - b}{a} \leq \dots \leq \frac{x_{(n)} - b}{a}.$$

Поэтому, если  $u_{(k)}^* = \frac{x_{(k)} - b}{a}$  – член набора  $u = (u_1, u_2, \dots, u_n)$ , соответствующий  $k$ -му члену вариационного ряда  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , то верны неравенства:

$$u_{(1)}^* \leq u_{(2)}^* \leq \dots \leq u_{(n)}^*,$$

т.е. на самом деле число  $u_{(k)}^*$  является  $k$ -м членом вариационного ряда  $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(n)}$ , соответствующего набору  $u$ :  $u_{(k)}^* = u_{(k)}$ . Отсюда следует, что для набора  $u$  размах равен

$$R_u \equiv u_{(n)} - u_{(1)} = u_{(n)}^* - u_{(1)}^* \equiv \frac{x_{(n)} - b}{a} - \frac{x_{(1)} - b}{a} = \frac{1}{a} (x_{(n)} - x_{(1)}) = \frac{1}{a} R_x,$$

что (в рассматриваемом случае  $a > 0$ ) равносильно равенству (8).

Для того, чтобы связать между собой медианы  $\mu_x$  и  $\mu_u$  исходного и нового наборов, необходимо дополнительно отдельно рассмотреть случаи: (1)  $n$  – нечётное число и (2)  $n$  – чётное число.

(1) Если  $n = 2k - 1$  – нечётное число, то в соответствии с определением медианы мы имеем:  $\mu_u = u_{(k)} = u_{(k)}^* = \frac{x_{(k)} - b}{a} = \frac{\mu_x - b}{a}$ , что равносильно (7).

(2) Если  $n=2k$  – чётное число, то в соответствии с определением медианы мы имеем:

$$\begin{aligned}\mu_u &= \frac{u_{(k)} + u_{(k+1)}}{2} = \frac{u_{(k)}^* + u_{(k+1)}^*}{2} = \frac{\frac{x_{(k)} - b}{a} + \frac{x_{(k+1)} - b}{a}}{2} \\ &= \frac{x_{(k)} + x_{(k+1)} - 2b}{2a} = \frac{\frac{x_{(k)} + x_{(k+1)}}{2} - b}{a} = \frac{\mu_x - b}{a},\end{aligned}$$

и потому равенство (7) верно и в этом случае.

2 случай. Если  $a < 0$ , то почленное деление (9) на  $a$  изменит знаки неравенств на противоположные:

$$\frac{x_{(1)} - b}{a} \geq \frac{x_{(2)} - b}{a} \geq \dots \geq \frac{x_{(n)} - b}{a}.$$

Поэтому, если  $u_{(k)}^* = \frac{x_{(k)} - b}{a}$  – член набора  $\mathbf{u} = (u_1, u_2, \dots, u_n)$ , соответствующий  $k$ -му члену вариационного ряда  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , то верны неравенства:

$$u_{(1)}^* \geq u_{(2)}^* \geq \dots \geq u_{(n)}^* \Leftrightarrow u_{(n)}^* \leq u_{(n-1)}^* \leq \dots \leq u_{(1)}^*$$

т.е. на самом деле число  $u_{(k)}^*$  является  $(n - k + 1)$ -м членом вариационного ряда  $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(n)}$ , соответствующего набору  $\mathbf{u}$ :  $u_{(k)}^* = u_{(n-k+1)}$ . Если индекс  $n - k + 1$  обозначить какой-нибудь новой буквой, скажем,  $l$ , то  $k = n - l + 1$  и это соотношение примет вид:

$$u_{(l)} = u_{(n-l+1)}^* . \quad (10)$$

Отсюда следует, что в случае  $a < 0$  размах набора  $\mathbf{u}$  равен

$$R_u \equiv u_{(n)} - u_{(1)} = u_{(1)}^* - u_{(n)}^* \equiv \frac{x_{(1)} - b}{a} - \frac{x_{(n)} - b}{a} = -\frac{1}{a}(x_{(n)} - x_{(1)}) = \frac{1}{(-a)}R_x,$$

что (в рассматриваемом случае  $a < 0$ ) равносильно равенству (8).

Для того, чтобы связать между собой медианы  $\mu_x$  и  $\mu_u$  исходного и нового наборов, необходимо, как и ранее, дополнительно отдельно рассмотреть случаи: (1)  $n$  – нечётное число и (2)  $n$  – чётное число.

(1) Если  $n = 2k - 1$  – нечётное число, то, используя определение медианы и соотношение (10), мы имеем:  $\mu_u = u_{(k)} = u_{(n-k+1)}^* = \frac{x_{(k)} - b}{a} = \frac{\mu_x - b}{a}$ , так что равенство (7) верно и в этом случае.

(2) Если  $n=2k$  – чётное число, то, используя определение медианы и соотношение (10), мы имеем:

$$\begin{aligned}\mu_u &= \frac{u_{(k)} + u_{(k+1)}}{2} = \frac{u_{(n-k+1)}^* + u_{(n-(k+1)+1)}^*}{2} = \frac{\frac{x_{(k+1)} - b}{a} + \frac{x_{(k)} - b}{a}}{2} \\ &= \frac{\frac{x_{(k)} + x_{(k+1)} - 2b}{2}}{a} = \frac{\frac{x_{(k)} + x_{(k+1)}}{2} - b}{a} = \frac{\mu_x - b}{a},\end{aligned}$$

и потому равенство (7) верно и в этом случае.

**Замечание.** Соотношения (7) и (8) полностью аналогичны соотношениям (2) и (5). Это вполне соответствует тому, что медиана, как и среднее, является мерой положения чисел набора, а размах, как и стандартное отклонение, является мерой разброса.

**Пример 5.** Пусть дан набор

$$x_1 = 2,91, x_2 = 3,07, x_3 = 3,02, x_4 = 3,05, x_5 = 3,01, x_6 = 3,03.$$

Как мы видели в Примерах 2 и 4, линейное преобразование  $x = 3 + 0,01u$  с параметрами  $a = 0,01$ ,  $b = 3$  приводит к набору  $u = (-9; 7; 2; 5; 1; 3)$ .

Вариационный ряд для исходного набора выглядит так:

$$x_{(1)} = 2,91, x_{(2)} = 3,01, x_{(3)} = 3,02, x_{(4)} = 3,03, x_{(5)} = 3,05, x_{(6)} = 3,07.$$

Соответственно  $\min_{1 \leq i \leq 6} x_i = x_{(1)} = 2,91$ ,  $\max_{1 \leq i \leq 6} x_i = x_{(6)} = 3,07$  и потому размах

$$R_x = x_{(6)} - x_{(1)} = 3,07 - 2,91 = 0,16.$$

Вариационный ряд для набора  $u$  выглядит так:

$$u_{(1)} = -9, u_{(2)} = 1, u_{(3)} = 2, u_{(4)} = 3, u_{(5)} = 5, u_{(6)} = 7.$$

Соответственно  $\min_{1 \leq i \leq 6} u_i = u_{(1)} = -9$ ,  $\max_{1 \leq i \leq 6} u_i = u_{(6)} = 7$  и потому размах

$$R_u = u_{(6)} - u_{(1)} = 7 - (-9) = 16.$$

Как и следовало ожидать на основе формулы (8), величины  $R_x$  и  $R_u$  связаны соотношением:  $R_x = 0,01R_u$  (напомним, что в рассматриваемом примере масштабный параметр  $a$  равен 0,01).

Для медиан мы имеем: 
$$\mu_x = \frac{x_{(3)} + x_{(4)}}{2} = \frac{3,02 + 3,03}{2} = 3,025,$$

$$\mu_u = \frac{u_{(3)} + u_{(4)}}{2} = \frac{2 + 3}{2} = 2,5.$$
 Как и следовало ожидать на основе формулы (7),

величины  $\mu_x$  и  $\mu_u$  связаны соотношением:  $\mu_x = 0,01\mu_u + 3$  (напомним, что в рассматриваемом примере масштабный параметр  $a$  равен 0,01, а параметр сдвига  $b = 3$ ).

### Дополнительные задачи

**Задача 1.** В магазине продают семь видов сока в литровых упаковках по следующим ценам (в рублях):

$$x_1 = 28,90, x_2 = 34,90, x_3 = 29,90, x_4 = 32,90, x_5 = 39,90, x_6 = 24,90, x_7 = 38,90.$$

Определите среднюю цену литрового пакета сока, стандартное отклонение от этой цены, а также медиану и размах цен.

**Решение.** Все цены оканчиваются записью «,90» – это обычный приём розничной торговли, который позволяет создать впечатление, что товар дешевле, чем на самом деле. Например, для потребителя, в сущности, нет никакой разницы между 40 руб. и 39,90 руб. Однако психологически потребитель относит товар ценой 39,90 руб. в категорию 30 рублёвых товаров и может купить его, в то время как указание цены 40 рублей автоматически переводит товар в категорию 40 рублёвых товаров, что может быть неприемлемо для покупателя.

Имея в виду это соображение, введём округлённые цены:  $u_1 = x_1 + 0,10 = 29$ ,  $u_2 = x_2 + 0,10 = 35$ ,  $u_3 = x_3 + 0,10 = 30$ ,  $u_4 = x_4 + 0,10 = 33$ ,  $u_5 = x_5 + 0,10 = 40$ ,  $u_6 = x_6 + 0,10 = 25$ ,  $u_7 = x_7 + 0,10 = 39$ .

Значения этих округлённых цен колеблются вокруг числа 30. Имея в виду это соображение, введём ещё один набор  $v = (v_1, v_2, \dots, v_7)$  по формуле:  $u_i = 30 + v_i$ , так что  $v = (-1; 5; 0; 3; 10; -5; 9)$ . Значения набора  $x$  связаны со значениями набора  $v$  формулой:  $x_i = u_i - 0,1 = 30 + v_i - 0,1 = v_i + 29,90$ .

Среднее значение набора  $v$  легко подсчитать в уме:  $M_v = 3$ . Теперь для дисперсии набора  $v$  мы имеем (все вычисления можно проделать без калькулятора):  $D_v = \frac{178}{7} \approx 25,43$ . Соответственно,  $\sigma_v = \sqrt{D_v} \approx 5,04$ .

Чтобы подсчитать медиану и размах, образуем вариационный ряд вспомогательного набора  $v$  – это будет набор  $(-5; -1; 0; 3; 5; 9; 10)$ . Размах этого набора равен  $R_v = 10 - (-5) = 15$ . Далее, количество чисел в наборе  $v$  (как и в исходном наборе  $x$ ) равно 7, т.е. является нечётным числом. Поэтому медиана набора  $v$  – это четвёртый член соответствующего вариационного ряда, т.е.  $\mu_v = 3$ .

Теперь легко подсчитать статистические характеристики основного набора. Средние значения и медианы связаны тем же линейным соотношением  $x = v + 29,90$ , которое связывает числа наборов  $x$  и  $v$ :  $M_x = M_v + 29,90 = 32,90$ ,  $\mu_x = \mu_v + 29,90 = 32,90$ , а поскольку масштабный коэффициент при преобразовании набора  $v$  в набор  $x$  равен 1, их меры разброса совпадают:  $\sigma_x = \sigma_v \approx 5,04$ ,  $R_x = R_v = 15$ .

**Задача 2.** В витрине обувного магазина выставлены шесть пар ботинок по следующим ценам (в рублях):

$$x_1 = 3490, x_2 = 3990, x_3 = 5990, x_4 = 1990, x_5 = 4490, x_6 = 990.$$

Определите среднюю цену, стандартное отклонение от этой цены, а также медиану и размах цен.

**Решение.** Введём округлённые цены  $u_i = x_i + 10$ . В результате мы получим следующий набор чисел:  $\mathbf{u} = (3500, 4000, 6000, 2000, 4500, 1000)$ .

Все эти числа кратны 500. Имея в виду это соображение, введём новые переменные  $v_i$  по формуле:  $u_i = 500v_i$ . В результате мы получим следующий ряд чисел:  $\mathbf{v} = (7, 8, 12, 4, 9, 2)$ . Значения основного набора связаны со значениями этого набора соотношением  $x_i = u_i - 10 = 500v_i - 10$ .

Среднее значение набора  $\mathbf{v}$  легко подсчитать в уме:  $M_v = 7$ .

Теперь для дисперсии набора  $\mathbf{v}$  мы имеем (все вычисления можно проделать без калькулятора):  $D_v = \frac{64}{6} \approx 10,667$ , так что  $\sigma_v = \sqrt{D_v} \approx 3,266$ .

Чтобы подсчитать медиану и размах цен, образуем вариационный ряд вспомогательного набора  $\mathbf{v}$  – это будет набор (2;4;7;8;9;12). Размах этого набора равен  $R_v = 12 - 2 = 10$ . Далее, количество чисел как в наборе  $\mathbf{v}$  (как и в исходном наборе  $\mathbf{x}$ ) равно 6, т.е. является чётным числом. Поэтому медиана набора  $\mathbf{v}$  – это среднее арифметическое третьего и четвёртого членов соответствующего вариационного ряда, т.е.  $\mu_v = 7,5$ .

Теперь легко подсчитать статистические характеристики основного набора. Средние значения и медианы связаны тем же линейным соотношением  $x = 500v - 10$ , которое связывает числа наборов  $\mathbf{x}$  и  $\mathbf{v}$ :  $M_x = 500M_v - 10 = 3490$ ,  $\mu_x = 500\mu_v - 10 = 3740$ , а поскольку масштабный коэффициент при преобразовании набора  $\mathbf{v}$  в набор  $\mathbf{x}$  равен 500, их меры разброса есть:  $\sigma_x = 500\sigma_v \approx 1633$ ,  $R_x = 500R_v = 5000$ .

**Задача 3.** На метеорологической станции ежедневно в полдень фиксируется температура воздуха. Наблюдения в течение одной недели в январе дали следующие значения (в градусах Цельсия):  $-7, -9, -7, -5, -9, -10, -9$ . Определите среднюю температуру на этой неделе, стандартное отклонение от этой средней температуры, а также медиану и размах температур.

**Решение.** Пусть  $x_i$  – температура в  $i$ -й день недели,  $i=1, \dots, 7$ , так что, например,  $x_1 = -7, x_7 = -9$ . Чтобы не работать с отрицательными числами, введём абсолютные значения температур  $u_i$  по формуле:  $u_i = -x_i$ . В результате мы получим следующий ряд чисел:  $\mathbf{u} = (7; 9; 7; 5; 9; 10; 9)$ .

На глаз можно оценить, что эти числа колеблются вокруг значения 7 или 8. Имея в виду это соображение, введём новые переменные  $v_i$  по формуле:  $v_i = u_i - 8$  (с равным успехом можно было бы использовать формулу  $v_i = u_i - 7$ ; мы рекомендуем читателю самостоятельно проделать расчёты, используя это преобразование, и сравнить окончательные числовые значения статистических характеристик основного набора с полученными ниже). В

результате мы получим следующий ряд чисел:  $\mathbf{v} = (-1; 1; -1; -3; 1; 2; 1)$ . Значения основного набора связаны со значениями этого вспомогательного набора соотношением  $x_i = -v_i - 8$ .

Среднее значение набора  $\mathbf{v}$  легко подсчитать в уме:  $M_v = 0$ .

Теперь для дисперсии набора  $\mathbf{v}$  мы имеем (все вычисления можно проделать без калькулятора):  $D_v = \frac{18}{7} \approx 2,57$ , так что  $\sigma_v = \sqrt{D_v} \approx 1,60$ .

Чтобы подсчитать медиану и размах температур, образуем вариационный ряд вспомогательного набора  $\mathbf{v}$  – это будет набор  $(-3; -1; -1; 1; 1; 1; 2)$ . Размах этого набора равен  $R_v = 2 - (-3) = 5$ . Далее, количество чисел как в наборе  $\mathbf{v}$  (как и в исходном наборе  $\mathbf{x}$ ) равно 7, т.е. является нечётным числом. Поэтому медиана набора  $\mathbf{v}$  – это четвёртый член соответствующего вариационного ряда, т.е.  $\mu_v = 1$ .

Теперь легко подсчитать статистические характеристики основного набора. Средние значения и медианы связаны тем же линейным соотношением  $x = -v - 8$ , которое связывает числа наборов  $\mathbf{x}$  и  $\mathbf{v}$ :  $M_x = -M_v - 8 = -8$ ,  $\mu_x = -\mu_v - 8 = -9$ , а поскольку значение масштабного параметра линейного преобразования, связывающего наборы  $\mathbf{v}$  и  $\mathbf{x}$  равно  $(-1)$ , их меры разброса совпадают:  $\sigma_x = \sigma_v \approx 1,60$ ,  $R_x = R_v = 5$ .

### Заключение

Изложенная небольшая теория и её простые применения к решению задач позволяют лучше понять математические основы описательной статистики. Этот материал вполне может излагаться в 9 классе, в классах с углублённым изучением математики.

### Литература

1. Тюрин Ю.Н. и др. Теория вероятностей и статистика. М.: МЦНМО: АО «Московские учебники», 2004. – 256 с.
2. Дорофеев Г.В., Суворова С.Б., Бунимович Е.А. и др. Математика: алгебра. Функции. Анализ данных. Учебник для 9 кл. общеобразовательных учреждений. М.: Просвещение, 2005.—287 с.