

Вычисление средних хронологических значений – незаслуженно забытый способ статистической обработки¹

И.Г. Зенкевич

ФГАОУ ВО «Санкт-Петербургский государственный университет», Институт химии,
Университетский просп., 26, Санкт-Петербург 198504, Российская Федерация

Адрес для переписки: Зенкевич Игорь Георгиевич, E-mail: izenkevich@yandex.ru

Поступила в редакцию 21 января 2023 г., после доработки 31 марта 2023 г.

В реальной аналитической практике при выполнении так называемых параллельных определений существуют объективные причины (ограниченные время измерений, материальные ресурсы и другие факторы), которые не позволяют обеспечить достаточное число измерений, необходимых для их надежной статистической обработки. Особенно ненадежными для малых выборок данных представляются стандартные отклонения (чаще всего завышены). Однако если изменить характер их статистической обработки и вместо средних арифметических значений вычислять средние хронологические значения и соответствующие им стандартные отклонения, то этот прием позволяет получать более надежные результаты. Это требует ранжирования исходных данных не в последовательности их измерения, а по возрастанию.

Суть вычисления средних хронологических в том, что минимальное и максимальное значения исходной выборки заменяют одним числом, а именно – их средним арифметическим, которое далее усредняют вместе с остальными. В результате этого количество измерений формально уменьшается на единицу, но средние значения изменяются лишь незначительно. Значения же стандартных отклонений, вычисленные после такой модификации выборки, становятся более характеристичными и приближаются к значениям стандартных отклонений выборок данных большего объема.

Переход от средних арифметических к средним хронологическим приводит к существенно меньшим искажениям исходных выборок данных чем, например, исключение выбросов с использованием известного критерия 3 σ .

Ключевые слова: Статистическая обработка одномерных совокупностей данных, малые выборки, средние хронологические значения, стандартные отклонения, интерпретация.

¹ Данной работой редакция открывает своего рода дискуссионную трибуну: с общего согласия далее последуют комментарии рецензента д.ф.-м. н. А.Л. Померанцева и ответ на них автора.

For citation: *Analitika i kontrol'* [Analytics and Control], 2023, vol. 27, no. 1, pp. 51-58

DOI:10.15826/analitika.2023.27.1.005

Calculation of average chronological values – an undeservedly neglected method of statistical data processing

Igor G. Zenkevich

St. Petersburg State University, Institute of Chemistry,
26, Universitetskii Av., St. Petersburg, 198504, Russian Federation

Corresponding author: Igor G. Zenkevich, E-mail: izenkevich@yandex.ru

Submitted 21 January 2023, received in revised form 31 March 2023

In real analytical practice involving parallel determinations there are certain objective reasons (limited measurement time, available resources, etc.) that prevent performing sufficient number of measurements required for rigorous statistical data processing. Such parameters as standard deviations are particularly unreliable for small data sets (they are usually too high in comparison with the results obtained for more representative data sets). This problem can be minimized by changing the character of data processing, i.e. calculating so-called chronological averages instead of average arithmetical values. This, however, requires ranking the data not in the order as they were measured, but in the ascending order.

The essence of calculating chronological averages is that the minimum and the maximum values of the initial data set are replaced with a single number, namely their arithmetical average, which is then averaged with other data. As a result, the total number of data values decreases by one but the changes of the averages would be negligible. At the same time, the standard deviations of the modified data sets become more characteristic and approaching the standard deviations of extended data sets.

Replacing the arithmetical means by chronological average values leads to sufficiently smaller distortions of the initial data sets than, for example, in the case of excluding outliers by using the well-known “3s” criterion.

Keywords: Statistical processing of single-dimensional data sets, small data sets, average chronological values, standard deviations, interpretation.

ВВЕДЕНИЕ

Статистическая обработка результатов измерений – необходимая стадия представления аналитических данных. Разделы «Метрология химического анализа» (в различных формулировках) представлены во многих руководствах по аналитической химии, например, [1, т. 3, гл. 1; 2, гл. 6; 3] и др. Первая стадия обработки N результатов так называемых параллельных определений ($x_i, 1 \leq i \leq N$) обычно состоит в вычислении средних арифметических значений ($\langle x \rangle$), а в качестве меры их неопределенности – соответствующих средних квадратических (стандартных) отклонений (s_x):

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{i=N} x_i, \quad (1)$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{i=N} (\langle x \rangle - x_i)^2}{N-1}}$$

Кроме средних арифметических используют другие средние, в том числе средние квадратичные, реже – средние геометрические и средние гармонические. Здесь же можно упомянуть моду (наиболее часто повторяющееся число в выборке данных) и медиану (число, которое делит упорядоченную по возрастанию выборку на две равные части) [4]. Разновидностями перечисленных средних значений являются соответствующие взвешенные средние, учитывающие неодинаковые статистические веса исходных данных (например, кратность измерений), что умножает разнообразие вариантов. Для непрерывных величин $0 \leq a \leq A$ математический смысл среднего значения – высота прямоугольника, площадь которого равна площади под кривой $y = f(a)$ или, иначе, среднее значение интеграла:

$$\langle a \rangle = \frac{\int_0^A a f(a) da}{\int_0^A f(a) da} \quad (2)$$

В качестве примера экзотических средних значений можно привести вычисление результатов количественного анализа методом двойного внутреннего стандарта как среднего геометрического относительно его составляющих [5, 6]. Средние гармонические значения времен удерживания характеризуют режим газохроматографического анализа с программированием температуры [7]. В базе данных [8] для представления статистически обработанных газохроматографических индексов

удерживания использованы не средние арифметические значения и стандартные отклонения, а медиана и некоторая характеристика их отклонений от медианы (половина размаха).

Принято считать, что статистическая обработка одномерных массивов данных дает приемлемые значения средних значений и их стандартных отклонений при количестве измерений (по разным оценкам) не менее 10–30. Однако на практике это условие соблюдается далеко не всегда. Главной причиной оказывается трудоемкость определений и необходимые для этого затраты времени. Так, например, в классическом С,Н,N-элементном анализе (все операции проводят вручную) [9] принято ограничиваться всего двумя параллельными определениями. Если их результаты заметно различаются, проводят третье контрольное определение. То же относится к определению азота методом Дюма, который используют до настоящего времени [10], несмотря на то, что время одного определения достигает 40–70 мин (в зависимости от навески). В хроматографических количественных определениях ограничения не столь существенны, но, в зависимости от продолжительности цикла анализа, число дозирования параллельных проб редко превышает 3–5.

Такие послышки объясняют наличие актуальных проблем так называемых малых выборок данных. К их числу относится невозможность установить характер распределения данных (например, проверить его нормальность), а также сложности выявления или компенсации выбросов [11, 12]. Исключение или коррекция выбросов обычно приводят к существенным искажениям исходных массивов, так что правомерность использования таких приемов остается дискуссионной. Тем не менее, они рекомендованы даже в случае трех измерений [13]. На практике это означает, что если одно из трех измерений существенно отличается от двух других, то его целесообразно отбросить, а два оставшихся – усреднить.

При несколько большем числе измерений (3–5) последствия разброса результатов столь же неблагоприятны. Следует заметить, что в аналитической практике случаи аномально большого разброса результатов измерений, так и приводящие к ним обстоятельства, известны. Например, это относится к интенсивностям сигналов в искровой масс-спектрометрии из-за больших вариаций кинетических энергий образующихся ионов [14, 15],

однако здесь их разброс обычно компенсируют увеличением числа измерений. Другим примером могут служить результаты измерения площадей газохроматографических пиков химиками - аналитиками, не имеющими достаточной практики дозирования проб вручную. Здесь существенное увеличение числа измерений затруднительно, так как это приводит к большим затратам времени. Причиной разброса является непостоянство времени нахождения шприцев в нагретом испарителе хроматографа. Это особенность неоднократно была подтверждена студентами бакалавриата Института химии СПбГУ в ходе выполнения ими практических работ по газовой хроматографии в рамках курса лекций «Хроматографические методы разделения и очистки органических соединений» [16 – 18]. В указанных примерах нет никаких ограничений на вычисление обычных средних арифметических значений, но их стандартные отклонения оказываются аномально завышенными и, вследствие этого, неинформативными.

В подобных случаях желательно как-то модифицировать способ оценки отклонений экспериментальных данных от средних значений. Для этого целесообразно обратить внимание на их разновидность, крайне скудно обсуждаемую в справочной литературе и в Интернете. Изначально она была предложена для представления варьируемых во времени величин a , которые необходимо усреднить за время наблюдений. Следовательно, последовательность значений a_i в выборке данных детерминирована. Такие посылки определяют название этого варианта – **средние хронологические**

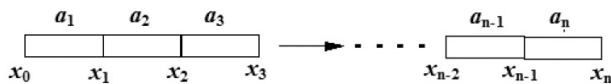


Рис. 1. Схема, иллюстрирующая суть средних хронологических значений. Необходимо вычислить среднее значение непрерывно изменяющейся во времени величины a в диапазоне $0 \leq a \leq A$. Для каждого из интервалов $a_1, a_2, a_3, \dots, a_{n-1}, a_n$ известны значения этой величины в начале и конце каждого интервала, т.е. для интервала a_1 – в точках x_0 и x_1 , для интервала a_2 – в точках x_1 и x_2 и т.д. Каждое значение a_i оценивают как среднее арифметическое значение этой величины на границах интервала, $a_i = (x_{i-1} + x_i) / 2$.

Fig. 1. The scheme illustrating the essence of average chronological values. It is necessary to calculate the average value of a continuously changeable in time variable a within the range of $0 \leq a \leq A$. The values of this variable are known at the beginning and at the end of each interval $a_1, a_2, a_3, \dots, a_{n-1}, a_n$, i.e. for the interval a_1 we know the variable's values at x_0 and x_1 points, for interval a_2 – the values at x_1 and x_2 , etc. Every value a_i is estimated as the average arithmetical value of the variable a at the first and the last points of the interval, $a_i = (x_{i-1} + x_i) / 2$.

величины, относящиеся к **временным последовательностям** (англоязычные термины **time series** или **dynamic series**).

Смысл средних хронологических величин можно проиллюстрировать рисунком 1. Весь диапазон вариаций непрерывной величины a разделим на n интервалов, $a_1, a_2, a_3, \dots, a_{n-1}, a_n$, границами которых являются значения $x_0, x_1, x_2, x_3, \dots, x_{n-1}, x_n$ (число таких границ равно $n + 1$):

Каждый из интервалов a_i ($1 \leq i \leq n$) можно характеризовать полусуммами значений a на его границах: $a_1 = (x_0 + x_1)/2, a_2 = (x_1 + x_2)/2, \dots, a_n = (x_{n-1} + x_n)/2$. Общее среднее значение $\langle a \rangle$ можно представить как среднюю величину a_i по всем интервалам: $\langle a \rangle = \sum_{i=1}^n a_i / n$ с пределами суммирования $1 \leq i \leq n$. Дальнейшие несложные преобразования последнего соотношения приводят к неожиданным выводам:

$$\langle a \rangle = \frac{\sum_{i=1}^n a_i}{n} = \frac{x_0}{2n} + \frac{\sum_{i=1}^{n-1} x_i}{n} + \frac{x_n}{2n} = \frac{1}{n} \left[\frac{(x_0 + x_n)}{2} + \sum_{i=1}^{n-1} a_i \right] \quad (3)$$

Если для вычисления среднего значения $\langle a \rangle$ мы пользуемся границами интервалов $\{x_i\}$, то $(n - 1)$ величин в точках $(1 \leq i \leq n-1)$ мы учитываем непосредственно, но результаты первого (x_0) и последнего (x_n) измерений – со вдвое меньшими статистическими весами ($1/2$). Этот факт можно представить иначе, учитывая, что $x_0/2 + x_n/2 = (x_0 + x_n)/2$, а такая полусумма представляет собой среднее арифметическое значение первого и последнего чисел серии, которое далее можно усреднять как одно число вместе с оставшимися. Общее количество усредняемых значений x_i при этом уменьшается на единицу (вместо $n + 1$ становится равным n). Если выборка данных симметрична, то средние значения $\langle a \rangle$ при этом не изменяются, если же нет, то $\langle a \rangle_{\text{арифм}} \neq \langle a \rangle_{\text{хрон}}$. Различия таких средних зависят от асимметрии выборки.

До настоящего времени примеры применения средних хронологических величин остаются относительно немногочисленными и преимущественно ограничены именно разрешенными для них временными последовательностями. Средним хронологическим величинам посвящены лишь единичные специальные публикации [19]; некоторые их описания можно найти в Интернете, чаще всего без ссылок на конкретные источники. К таким последовательностям относят, например, ежедневные колебания курсов валют, ежегодную динамику производства продукции черной металлургии, ежемесячные вариации цен, результаты почасовых измерений температуры в метеорологии и т.п.

Однако особенность средних хронологических величин заключается в том, что формально-математических запретов на их применение для статистической обработки массивов данных, не относящихся к временным последовательностям,

не существует. Такая реинтерпретация известного, но редко применяемого способа статистической обработки данных, позволяет расширить области его применения и является задачей настоящего сообщения.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ (исходные числовые данные)

Статистическую обработку числовых данных проводили, пользуясь стандартными ПО, позволяющими вычислять средние арифметические значения и их стандартные отклонения, в том числе Excel (Microsoft Office 2010) или Origin (версии 4.1 и 8.1).

Условия анализа методом обращенно-фазовой ВЭЖХ. Определение времен удерживания алкилфенилкетонов в различных изократических режимах проводили на жидкостном хроматографе Shimadzu LC-20 Prominence с диодно-матричным детектором и колонкой Phenomenex C18 длиной 250 мм, внутренним диаметром 4.6 мм, размер частиц сорбента 5 мкм. Расход элюента 1.0 мл/мин, температура колонки 30 °С. Пробы дозировали с использованием автосамплера SIL-20A/AC, объем проб 20 мкл, кратность измерений равна пяти.

В качестве **примеров вычисления средних хронологических значений** использовали: 1) Результаты параллельных измерений площадей хроматографических пиков при обсуждении возможностей уменьшения случайной составляющей погрешности количественных газохроматографических определений модифицированным способом внешнего стандарта за счет использования сигнала растворителя. Определения проведены студентами бакалавриата Института химии СПбГУ в ходе выполнения ими практических работ по газовой хроматографии в рамках курса лекций «Хроматографические методы разделения и очистки органических соединений» [16 – 18]. В этих же работах приведены условия газохроматографического анализа.

2) Результаты определения относительных оптических плотностей $A_{\text{отн}} = A(\lambda_1)/A(\lambda_2)$ при длинах волн $\lambda_1 = 254$ нм и $\lambda_2 = 220$ нм аналитов в обращенно-фазовой высокоэффективной жидкостной хроматографии (ОФ ВЭЖХ с УФ-детектированием) получены А. Деруишем. Использованы препараты ацетофенона, пропиофенона и бутирофенона (Sigma-Aldrich Rus LLC, Россия), которые растворяли в 2-пропанол («хч», Криохром, Санкт-Петербург), и дополнительно в подвижной фазе. Для приготовления элюента для ВЭЖХ использовали метанол («хч», Криохром, Санкт-Петербург) и деионизованную воду (удельное сопротивление 18.2 МОм·см) с добавкой 0.1 % муравьиной кислоты (98 % «для анализа», RapReac, Испания). Кратность измерений равна восьми.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Особенности вычисления средних хронологических значений. Вычисление (при необходимости) различных средних значений – достаточно простая задача. Некоторую проблему может представлять необходимость их дополнения соответствующими стандартными отклонениями. Небезынтересно сопоставить результаты вычисления различных средних значений для некоторой условной выборки данных, например, $x_i = \{1, 11, 3, 9, 5, 7\}$, для которой получаем следующие оценки (указаны с одной значащей цифрой после запятой):

Вид средних значений	Значение $\langle x \rangle$
Среднее арифметическое	6.0 ± 3.7
Среднее геометрическое	4.7
Среднее гармоническое	3.2
Среднее квадратическое	6.9

Даже из такого простейшего примера следует несколько полезных заключений. Все средние значения кроме средних квадратических меньше обычных средних арифметических, равно как и их стандартные отклонения. Таким образом, известно условие, что в интервал $\langle x \rangle \pm 2s_x$ должны попадать приблизительно 95 % всех значений исходной выборки (а в выбранном нами примере – практически все) выполняется только для средних арифметических. С другой стороны, значение $s_x = 3.7$ явно завышено. Если дополнить приведенные оценки значениями медианы (6) и размаха (5), то можно записать 6.0 ± 2.5 , так как медиана (6.0) в данном случае равна среднему арифметическому.

Чтобы включить в число рассматриваемых характеристик средние хронологические значения, исходные данные следует ранжировать по возрастанию, $x_i^* = \{1, 3, 5, 7, 9, 11\}$, после чего получаем значение 6.0. Подобное преобразование объясняет возможность вычисления средних хронологических величин не только для временных рядов, но и для иных совокупностей экспериментальных данных. Важное свойство средних хронологических величин – они ближе всего к средним арифметическим, чем другие средние значения, так что по этому критерию искажение исходной выборки данных минимально. Но если так, то в чем преимущества средних хронологических значений?

Главные отличия различных средних заключаются в алгоритме вычисления соответствующих им стандартных отклонений, который для средних хронологических не совсем очевиден. В этом случае значения $x_{\text{мин}}$ и $x_{\text{макс}}$ учитывают со статистическими весами 1/2, поэтому можно предположить, что соответствующие им вклады в величины стандартных отклонений, $[(\langle x \rangle - x_i)^2 / (N - 1)]$ также следует учитывать с коэффициентами 1/2. В результате формула для вычисления стандартных отклонений средних хронологических усложняется:

$$s_x = \sqrt{\frac{\sum_{i=2}^{N-1} (\langle x \rangle - x_i)^2 + (\langle x \rangle - x_{\min})^2 + (\langle x \rangle - x_{\max})^2}{2(N-1)}} \quad (4)$$

Если вычислить соответствующие величины (формулы 3 и 4) для приведенного выше примера $x_i^* = \{1, 3, 5, 7, 9, 11\}$, то получаем 6.0 ± 3.0 . Полученная оценка стандартного отклонения (3.0) меньше, чем в случае средних арифметических (3.7), но в интервал $\pm 2s_x$ попадают все значения (x_i) исходной выборки (вариант А).

С другой стороны, как отмечено выше, можно записать $x_1/2 + x_n/2 = (x_1 + x_n)/2$, а такая полусумма представляет собой среднее арифметическое значение первого и последнего чисел серии, которое далее можно усреднять как одно число вместе с оставшимися. Тогда получаем, что модифицированная серия измерений будет содержать уже не N , а $(N - 1)$ значений, но формула (1) для вычисления стандартных отклонений не претерпевает никаких изменений. Для рассматриваемого примера получаем $\langle x \rangle = 6.0$ и $s_x = 2.2$, так что интервал $\pm 2s_x$ уже не перекрывает все значения (x_i). Фактически, в этом варианте (Б) из исходной выборки исключены максимальное и минимальное значения, которые заменены их средним, что представляет собой гораздо большие искажения исходного набора данных, чем вариант А. Ближайшим аналогом такого способа усреднения представляется фильтр Olympus из программы Chromeleon (фирма Dionex), который просто отбрасывает максимальную и минимальную точки исходной выборки и усредняет оставшиеся². Он позиционируется производителем как устойчивый

² Автор благодарит Ю.А. Каламбета, указавшего на эту аналогию.

к промахам способ обработки данных. Поскольку такой способ сильно искажает исходные массивы, то его, скорее, следует расценивать как иллюстрацию актуальности обработки результатов параллельных измерений, по тем или иным причинам отличающихся аномально большим разбросом.

В качестве примера вычисления средних хронологических значений и их стандартных отклонений при решении реальных аналитических задач можно рассмотреть результаты определения относительных оптических плотностей $A_{\text{отн}}$ аналитов в обращенно-фазовой высокоэффективной жидкостной хроматографии (ОФ ВЭЖХ) с УФ-детектированием при длинах волн λ_1 и λ_2 [20 – 22]:

$$A_{\text{отн}} = A(\lambda_1) / A(\lambda_2) \approx S(\lambda_1) / S(\lambda_2) \quad (5)$$

где $S(\lambda_1)$ и $S(\lambda_2)$ – площади пиков характеризуемого аналита, регистрируемых на длинах волн λ_1 и λ_2 .

Значения $A_{\text{отн}}$ являются важным дополнением к хроматографическим параметрам удерживания (временам удерживания, коэффициентам емкости или индексам удерживания) при идентификации аналитов в ОФ ВЭЖХ. Их измерение возможно либо в результате последовательных дозирования одних и тех же проб, либо (при использовании детекторов типа «диодная матрица») в результате реконструкции аналитических сигналов при разных длинах волн. Однако выполнение достаточного числа параллельных измерений в данном случае осложнено большими затратами времени и значительным расходом элюента. В то же время, разброс параллельных определений площадей пиков часто оказывается неожиданно большим. Основной причиной этого следует считать погрешности интегрирования площадей пиков. Совокупность

Таблица 1

Результаты статистической обработки наборов относительных оптических плотностей $A_{\text{отн}} = A(254) / A(220)$ алкил-фенилкетонов $C_6H_5COCH_{2n+1}$ с вычислением различных средних значений

Table 1

The results of statistical processing of relative optical density datasets $A_{\text{rel}} = A(254) / A(220)$ of alkyl phenyl ketones $C_6H_5COCH_{2n+1}$ with calculation of different averages

Компонент $A_{\text{отн}} = A(254) / A(220)$	Среднее арифметическое значение $\langle A_{\text{отн, арифм}} \rangle$	Стандартное отклонение среднего арифметического	Среднее хронологическое значение $\langle A_{\text{отн, хрон}} \rangle$	Стандартное отклонение среднего хронологического (вариант А)
Ацетофенон 2.96, 2.88, 2.85, 2.95, 2.28* , 2.89, 2.98 , 2.70	2.81	0.23 (8.2)**	2.84	0.18 (6.3)
Пропиофенон 2.87, 2.89, 2.84, 2.42, 2.53, 2.11 , 2.72, 2.91	2.66	0.29 (10.9)	2.68	0.23 (8.6)
Бутирофенон 3.02 , 3.00, 2.99, 2.66, 2.61, 2.25 , 2.73, 2.94	2.78	0.27 (9.7)	2.80	0.22 (8.0)

*) Жирным шрифтом здесь и далее выделены максимальное и минимальное значения; **) в скобках указаны относительные стандартные отклонения, %.

этих факторов приводит к тому, что объемы выборок данных, во-первых, ограничены, а, во-вторых, стандартные отклонения $A_{отн}$ непропорционально велики. Это создает предпосылки замены средних арифметических средними хронологическими с вычислением стандартных отклонений по варианту А.

В табл. 1 приведены экспериментально определенные значения $A_{отн} = A(254) / A(220)$ трех *n*-алкилфенилкетонов $C_6H_5COC_nH_{2n+1}$ ($1 \leq n \leq 3$), их средние арифметические и средние хронологические значения в сочетании с соответствующими стандартными отклонениями. Отличия средних хронологических и средних арифметических составляют всего от +0.7 % до +1.1 %. Относительные стандартные отклонения средних арифметических варьируют от 8 до 11%, тогда как для средних хронологических в варианте А – (6–9) % (в 1.2–1.4 раза меньше). Тем не менее, большая часть исходных значений $A_{отн}$ попадает в интервалы $\langle A_{отн} \rangle \pm 2s_A$ за исключением минимальных 2.28 для ацетофенона, 2.11 для пропиофенона и 2.25 для бутирофенона. Однако, поскольку критерий $3s$ [3, 23] в рассматриваемом случае сохраняет свое значение, то все эти значения не могут быть отнесены к выбросам, так как они попадают в интервалы $\langle A_{отн} \rangle \pm 3s_A$.

Таким образом, вычисление средних хронологических вместо средних арифметических позволяет получать более реальные оценки относительных оптических плотностей и подтверждает выводы работ [20, 21]: значения $A_{отн}$ характеризуют не конкретные органические соединения, а типы хромофоров

в молекулах и, следовательно, применимы для групповой идентификации аналитов.

Второй пример целесообразности вычисления средних хронологических значений также относится к малым выборкам данных, а именно – к результатам определения площадей хроматографических пиков студентами бакалавриата Института химии СПбГУ в ходе выполнения ими практической работы по газовой хроматографии «Определение концентрации кумола в растворе изопропилового спирта модифицированным способом внешнего стандарта» [18].

Результаты статистической обработки шести наборов площадей хроматографических пиков различных компонентов с вычислением средних арифметических и средних хронологических значений для сравнения приведены в табл. 2. Выбранные наборы содержат от всего от четырех до семи значений, что представляется типичной кратностью большинства хроматографических измерений. Разброс данных (стандартные отклонения площадей пиков) определяется качеством (степенью изношенности) используемых для дозирования проб шприцев и персональной повторяемостью самого дозирования (временем выдерживания шприцев в нагретом испарителе хроматографа). Для приведенных в табл. 2 примеров относительные стандартные отклонения средних арифметических составляют от 1.4 % до 5.4 %.

Средние хронологические значения оказываются как незначительно меньше (от -0.1 % до -0.3 %), так и несколько больше средних арифметических (от

Таблица 2

Результаты статистической обработки наборов площадей хроматографических пиков различных соединений с вычислением различных средних значений

Table 2

The results of statistical processing of datasets of chromatographic peak areas with calculation of different averages

Площади пиков, S_i , мВ·мс	Среднее арифметическое значение, $\langle S_i \rangle_{арифм}$	Стандартное отклонение среднего арифметического	Среднее хронологическое значение, $\langle S_i \rangle_{хрон}$	Стандартное отклонение среднего хронологического (вариант А)
128904, 131147* , 121550 , 125160, 126953	126743	3660 (2.9)**	126841	2755 (2.1)
102424, 96663 , 107165 , 105439, 104995	103337	4099 (4.0)	103693	2724 (2.6)
199941, 186049, 207540 , 189010, 182562	193020	10402 (5.4)	192513	8241 (4.3)
337697, 363844, 358032, 342685, 334992 , 374056 , 360080	353055	14727 (4.2)	352810	12364 (3.5)
240441, 238455 , 241551, 246373	241705	3365 (1.4)	241469	1734 (0.7)
4652590, 4661557, 4102091 , 4572959, 4710197	4539879	249635 (5.5)	4573312	183621 (4.0)

*) Жирным шрифтом здесь и далее выделены максимальное и минимальное значения; **) в скобках здесь и далее указаны относительные стандартные отклонения, %.

+0.1 % до +0.3 %), что, как отмечено выше, определяется неодинаковой асимметрией выборок данных. Наибольшие различия средних зафиксированы для максимального пика растворителя (последняя строка таблицы), площади которого превышают $4 \cdot 10^6$ мВ·мс, а именно +0.7%. Однако все такие различия статистически незначимы, так как они заметно меньше стандартных отклонений соответствующих площадей. Относительные стандартные отклонения средних арифметических значений составляют от 1.4 % до 5.5 %, а средних хронологических – от 0.7 % до 4.3 %, т.е. меньше в 1.2 – 2.0 раза. Тем не менее, во всех шести сериях измерений не выявлены значения, которые по критерию $3s$ могли бы быть классифицированы как выбросы. На основании этого можно предположить, что стандартные отклонения средних хронологических величин лучше характеризуют малые выборки площадей хроматографических пиков. Отсюда следует, что их использование фактически позволяет обеспечить большую точность определения площадей пиков, чем это следует из результатов обычной статистической обработки.

В заключение небезынтересно заметить, что искажения выборок данных за счет замены средних арифметических средними хронологическими оказываются несоизмеримо меньшими, чем, например, последствия исключения выбросов по известному критерию $3s$. Тем не менее, подобное исключение выбросов в соответствии с этим и другим критериям представляет собой часто рекомендуемый прием статистической обработки данных [3, 23].

ЗАКЛЮЧЕНИЕ

В реальной аналитической практике число измерений, необходимых для их надежной статистической обработки, по объективным причинам (ограниченное время измерений, материальные ресурсы и др.) часто не может быть достигнуто. Особенно ненадежными (обычно завышенными) для малых выборок оказываются стандартные отклонения. Однако если изменить характер статистической обработки данных и вместо средних арифметических значений вычислять средние хронологические и соответствующие им стандартные отклонения, то этот прием позволяет получать более надежные результаты. Это требует ранжирования исходных данных не в последовательности их измерения, а по возрастанию.

Суть вычисления средних хронологических в том, что минимальное и максимальное значения выборки заменяют одним числом, а именно – их средним арифметическим значением, которое далее усредняют вместе с остальными. В результате этого количество измерений формально уменьшается на единицу. Значения же стандартных отклонений, вычисленные после такой модификации выборки, становятся более характеристичными (приближаются к значениям стандартных отклонений для выборок данных большего объема).

БЛАГОДАРНОСТИ

Автор благодарит канд. физ.-мат. наук Ю.А. Каламбета (Генеральный директор ООО «Амперсэнд», Москва) за ценные замечания к первому варианту рукописи и полезное обсуждение проблемы. Относительные оптические плотности алкилфенилкетонов измерены А. Деруишем в Ресурсном центре «Методы анализа состава вещества» Научного парка Санкт-Петербургского государственного университета. Автор благодарит сотрудников Центра за содействие.

ACKNOWLEDGEMENTS

The author thanks PhD Yuri A. Kalambet, the Director-in-Chief of JSC “Ampersend”, Moscow for valuable comments to the first version of the manuscript of useful discussion of the problem. Relative optical densities of *n*-alkyl phenyl ketones were measured by Abdennour Derouiche at the Center for Chemical and Material Researches of Research Park of St. Petersburg State University. The author is grateful to the staff of Center for assistance.

ЛИТЕРАТУРА

1. Аналитическая химия: в 3 т.; под ред. Л.Н. Москвина. М.: Академия, 2008-2010.
2. Практическое руководство для лабораторий. Специальные методы / В.Р. Лесс [и др.]: пер с нем. СПб.: Профессия, 2011. 471 с.
3. Причард Э., Барвик В. Контроль качества в аналитической химии: пер. с англ. СПб.: Профессия, 2011. 319 с.
4. Пасхавер И.С. Средние величины в статистике. М.: Статистика, 1979. 279 с.
5. Вигдергауз М.С., Краузе И.М. Развитие методов количественной интерпретации хроматограмм сложных смесей // Ж. аналит. химии. 1986. Т. 41, № 11. С. 2064-2074.
6. Zenkevich I.G., Makarov E.D. Chromatographic quantitation at losses of analyte during sample preparation. Application of the modified method of double internal standard // J. Chromatogr. A. 2007. V. 1150. P. 117-123.
7. Blumberg L.M. Temperature-Programmed Gas Chromatography. Weinheim: Wiley-VCH Verlag GmbH, 2010. 340 p.
8. The NIST 17 Mass Spectral Library (NIST17/2017/EPA/NIH). Software/Data Version (NIST17); NIST Standard Reference Database, Number 69, June 2017. National Institute of Standards and Technology, Gaithersburg, MD 20899. [Электронный ресурс]: <http://webbook.nist.gov> (дата обращения: 01. 2023 г.).
9. Климова В.А. Основные микрометоды анализа органических соединений. М.: Химия, 1975. 224 с.
10. Cyclization of arylhydrazones of cross-conjugated enynones: synthesis of luminescent styryl-1H-pyrazoles and propenyl-1H-pyrazoles / R.N. Itakhunov [et al.] // Org. Biomol. Chem. 2022. V. 20. P. 8693-8713.
11. He J., Tian C.-X. A statistical smoothing measure to eliminate outliers in motion trajectory tracking // Hum. Movement Sci. 1998. V. 17, № 2. P. 189-200.
12. Salkind N.J. Encyclopedia of Measurement & Statistics. Thousand-Oaks: SAGE Publ. Inc., 2006. 1416 p.
13. Anscombe F.J., Barron B.A. Treatment of outliers in samples of size three // J. Res. NBS B Math. Sci. 1966. V. 708, № 2. P. 141-147.

14. Adams F., Thomas J.M. Recent advances in analytical spark source mass spectrometry // *Philos. T. Roy. Soc. A*. 1981. V. 305. P. 509-519.

15. Becker J.S., Dietze H.-J. Inorganic mass spectrometric methods for trace, ultratrace, isotope, and surface analysis // *Int. J. Mass Spectrom.* 2000. V. 197, № 1–3. P. 1-35.

16. Сравнительная характеристика возможностей количественного газохроматографического анализа обычным и модифицированным методами внешнего стандарта / И.Г. Зенкевич [и др.] // *Аналитика и контроль*. 2019. Т. 23. № 2. С. 223-228.

17. Сравнительная характеристика различных вариантов количественного хроматографического анализа методом двойной стандартной добавки / И.Г. Зенкевич [и др.] // *Аналитика и контроль*. 2021. Т. 25, № 2. С. 146-154.

18. Уменьшение случайной составляющей погрешности количественных газохроматографических определений способом внешнего стандарта за счет использования сигнала растворителя / И.Г. Зенкевич [и др.] // *Аналитика и контроль*. 2022. Т. 26, № 2. С. 141-149.

19. Anghelache K., Manole A. Dynamic / chronological (time) series // *Rom. Stat. Rev.* 2020. № 10. P. 78-87.

20. Зенкевич И.Г., Косман В.М. Относительное поглощение при разных длинах волн – дополнительный УФ-спектральный параметр для идентификации органических соединений в обращенно-фазовой ВЭЖХ // *Ж. аналит. химии*. 1996. Т. 51, № 8. С. 870-875.

21. Зенкевич И.Г., Косман В.М. Новые возможности идентификации органических соединений по УФ-спектрам с использованием относительных оптических плотностей // *Ж. прикл. химии*. 1997. Т. 70, № 11. С. 1861-1869.

22. Новые возможности высокоэффективной жидкостной хроматографии в фармакопейном анализе / Г.И. Барам [и др.] // *Бюл. Эксперим. Биол. Мед.* 2003. Т. 135, № 1. С. 75-79.

23. Чарыков А.К. Математическая обработка результатов химического анализа. Л.: Химия, 1984. 168 с.

REFERENCES

1. *Analiticheskaya khimiya: v 3 t.* [Analytical Chemistry: Vols 1–3. ed.: Moskvina L.N.] Moscow, Akademia Press, 2008-2010. (in Russian).

2. Less W.R., Eckhardt S., Kettner M., Schmitt F., Walter B. *Die handlungsorientierte Ausbildung für Laborberufe. Band 2. Waldqualifikationen.* Würzburg: Vogel Business Media GmbH & Co KG, 2010. 508 p.

3. Prichard E., Barwick V. *Quality Assurance in Analytical Chemistry.* New York: John Wiley & Sons, Ltd. 2007. 316 p.

4. Paskhaver I.S. *Srednie velichiny v statistike* [Average values in statistics]. Moscow: Statistika Press, 1979. 279 p. (in Russian).

5. Vigdergauz M.S., Krauze I.M. Development of the methods of quantitative interpretation of chromatograms of complex mixtures. *J. Analyt. Chem. (Rus.)*. 1986. vol. 41. no. 11. pp. 2064-2074.

6. Zenkevich I.G., Makarov E.D. Chromatographic quantitation at losses of analyte during sample preparation. Application of the modified method of double internal standard. *J. Chromatogr. A*, 2007, vol. 1150, pp. 117-123. doi: 10.1016/j.chroma.2006.08.083.

7. Blumberg L.M. *Temperature-Programmed Gas Chromatography.* Weinheim: Wiley-VCH Verlag GmbH, 2010. 340 p.

8. The NIST 17 Mass Spectral Library (NIST17/2017/EPA/NIH). Software/Data Version (NIST17); NIST Standard Reference Database, Number 69, June 2017. National Institute of Standards and Technology, Gaithersburg, MD 20899: Available at: <http://webbook.nist.gov> (Accessed January 01, 2023).

9. Klimova V.A. *Osnovnye mikrometody analiza organicheskikh soedinenij. [Principal micro-methods of organic compound analysis]*. 2nd Edn. Moscow: Khimia Press, 1975. 224 p. (in Russian).

10. Itakhunov R.N., Odin I.S., Gusev D.M., Grabovskiy S.A., Gordon K.V., Vologzhanina A.V., Sokov S.A., Sosnin I.M., Golovanov A.A. Cyclization of arylhydrazones of cross-conjugated enynones: synthesis of luminescent styryl-1H-pyrazoles and propenyl-1H-pyrazoles. *Org. Biomol. Chem.* 2022, vol. 20, pp. 8693-8713. doi: 10.1039/d2ob01427k.

11. He J., Tian C.-X. A statistical smoothing measure to eliminate outliers in motion trajectory tracking. *Human Movement Sci.* 1998, vol. 17, no. 2, pp. 189-200. doi: 10.1016/S0167-9457(97)00029-8.

12. Salkind N.J. *Encyclopedia of Measurement & Statistics.* Thousand-Oaks: SAGE Publ., Inc. 2006. 1416 p.

13. Anscombe F.J., Barron B.A. Treatment of outliers in samples of size three. *J. Res. NBS B Math. Sci.*, 1966, vol. 70B, no. 2, pp. 141-147.

14. Adams F., Thomas J.M. Recent advances in analytical spark source mass spectrometry. *Philos. T. Roy. Soc. A*, 1981, vol. 305, pp. 509-519. doi: 10.1098/rsta.1982.0048.

15. Becker J.S., Dietze H.-J. Inorganic mass spectrometric methods for trace, ultratrace, isotope, and surface analysis. *Int. J. Mass Spectrom.*, 2000, vol. 197, no. 1–3, pp. 1-35. doi: 10.1016/S1387-3806(99)00246-8.

16. Zenkevich I.G., Devleshova N.A., Krivolapova Yu.V., Moskvichev D.O., Rubicheva L.G., Tyufyakov N.Yu. [Comparative characterization of quantitative gas chromatographic analysis capabilities using basic and modified external standard methods]. *Analytika i Kontrol' [Analytics and Control]*, 2019, vol. 23, no. 2, pp. 223-228. (in Russian). doi: 10.15826/analitika.2019.23.2.007.

17. Zenkevich I.G., Barkhatova D.D., Belysheva M.N., Kaminskii N.A., Karchuganova E.M., Klaving A.V., Kovalenko A.A., Krivovicheva V.S., Kuz'min A.A., Mel'nik M.V., Paramonova P.S., Popov R.A., Potapenkov V.V., Rashevskii A.A., Sysoeva A.A., Fedorova I.I., Firsov A.A. [Comparative characterization of different kinds of chromatographic quantification using the double standard addition method]. *Analytika i Kontrol' [Analytics and Control]*, 2021, vol. 25, no. 2, pp. 146-154. (In Russian). doi: 10.15826/analitika.2021.25.2.010.

18. Zenkevich I.G., Byvsheva S.V., Gerasimov A.I., Gladnev S.V., Grigoriev M.V., Gubina N.V., Didenko E.A., Kazantzev A.S., Kalutskaia T.D., Katernuk E.V., Koblova A.A., Krutin D.V., Malkova K.P., Metliaeva S.A., Odegova V.S., Smirnova D.S., Spivakovskiy V.A., Terno P.V. [Decreasing the uncertainty of gas chromatographic quantification using the solvent's signal in the method of external standard]. *Analitika i Kontrol' [Analytics and Control]*, 2022, vol. 26, no. 2, pp. 141-149. (in Russian). doi: 10.15826/analitika.2022.26.2.005.

19. Anghelache K., Manole A. Dynamic / chronological (time) series. *Rom. Stat. Rev.*, 2020, no. 10, pp. 78-87.

20. Zenkevich I.G., Kosman V.M. [Relative absorbance at the different wavelengths – additional UV-spectral parameter for identification of organic compounds in reversed phase HPLC]. *J. Analyt. Chem. (Rus.)*, 1996, vol. 51, no. 8, pp. 870-875. (in Russian).

21. Zenkevich I.G., Kosman V.M. [Novel possibilities in identification of organic compounds using UV-spectra with relative optical densities]. *J. Applied Chem. (Rus.)*, 1997, vol. 70, no. 11, pp. 1861-1869. (in Russian).

22. Baram G.I., Reikhardt D.V., Goldberg E.D., Isotov B.N., Rodinko M.O., Khazanov V.A. [New possibilities of high performance liquid chromatography in pharmacopoeia analysis]. *Bull. Exp. Biol. Med.*, 2003, vol. 135, no. 1, pp. 75-79. (in Russian).

23. Charykov A.K. *Matematicheskaya obrabotka rezul'tatov khimicheskogo analiza [Mathematical Processing of Chemical Analysis Results]*. Leningrad: Khimia Press, 1984. 168 p. (in Russian).