

УДК 006.91

С.Г. Рабинович

Нью Джерси, США

СТАНДАРТНОЕ ОТКЛОНЕНИЕ ИЛИ ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ?

В качестве показателя точности многократных измерений одни используют стандартное отклонение, другие - доверительный интервал результата измерения. В статье приведен анализ этих показателей и показано, что доверительный интервал гораздо лучше выражает точность измерений, чем стандартное отклонение. Использование стандартного отклонения было оправдано, когда для косвенных измерений с зависимыми аргументами не было способа построить доверительный интервал. Теперь такой способ есть. Он известен и применяется уже почти столетия. Поэтому сделан вывод, что во всех многократных измерениях можно и целесообразно применять один и тот же показатель точности – доверительный интервал.

Ключевые слова: многократные измерения, стандартное отклонение, доверительный интервал.

Введение

Измерения могут быть однократными и многократными. Оценивание точности результата измерения актуально для тех и других. Однако эта задача решается для них по-разному. Многократные измерения, как правило, точнее однократных, и вклад случайной составляющей в их неопределенность хорошо изучен. Поэтому в данной статье мы ограничимся рассмотрением многократных измерений и показателями точности их случайной составляющей. В принципе показателем точности многократных измерений может быть как стандартное отклонение результата измерения, так и доверительный интервал, покрывающий истинное значение измеряемой величины. **Цель данной статьи** – сравнить эффективность этих показателей.

Основной материал

1. Общие соображения. Будем иметь в виду сначала прямые измерения, так как эти измерения являются базой всех измерений. Стандартное отклонение и доверительный интервал – понятия математической статистики. В метрологии они используются для выражения неточности многократных измерений и, в частности, для оценивания вклада случайных погрешностей в общую неточность многократных измерений. Стандартное отклонение результата измерения представляет собой параметр функции распределения результата измерения как случайной величины, какой и является оценка истинного значения измеряемой величины. Этот параметр отражает в неявном виде возможный разброс оценок измеряемой величины при повторных ее измерениях в тех же самых условиях, но не связан ни с какой вероятностной оценкой возможного разброса. Доверительный интервал – это интервал, покрывающий истинное значение измеряемой величины. Он в явном виде дает границы возможного разброса повторных оценок измеряемой величины, причем с известной вероятностью того, что истинное

значение измеряемой величины находится внутри этого интервала. Поэтому доверительный интервал интуитивно более понятен, чем стандартное отклонение. Кроме того, интервал как предел возможной погрешности измерения, применим и для однократных измерений. Стандартное отклонение для однократных измерений неприменимо.

Перейдем к более конкретному анализу. Основываясь на известных математических результатах [1], в [2] был получен показатель точности оценки дисперсии S^2 выборки в зависимости от числа наблюдений n :

$$\varepsilon = s[S^2]/\sigma^2 = \sqrt{2/(n-1)}.$$

Эту зависимость иллюстрируют приведенные в табл. 1 значения ε , вычисленные для произвольно взятого ряда числа наблюдений:

Таблица 1

Зависимость показателя надежности ε от числа измерений

n	3	5	7	10	15	20	30	40	50
ε	1,0	0,71	0,54	0,47	0,37	0,32	0,26	0,22	0,2

Аналогичным путем можно получить стандартное отклонение оценки стандартного отклонения среднего арифметического:

$$s[S(\bar{x})] = \sigma(\bar{x})/\sqrt{2(n-1)}.$$

Точность стандартного отклонения оценки стандартного отклонения среднего будем выражать в форме отношения этой оценки к истинному значению стандартного отклонения среднего выборки. Обозначим это соотношение символом φ :

$$\varphi = s[S(\bar{x})]/\sigma(\bar{x}) = 1/\sqrt{2(n-1)}.$$

Сравнивая выражения для φ и ε , получим

$$\varphi = \varepsilon/2.$$

Так что, не повторяя расчеты, нужные значения φ легко пересчитать из приведенного выше ряда для ε .

Например, показатель надежности ϕ для оценки стандартного отклонения среднего в выборке объемом 10 из совокупности с нормальным распределением составляет 23%, т.е. стандартное отклонение оценки стандартного отклонения среднего достигает почти четверти этой оценки. Следовательно, в повторной такой же по объему выборке оценка стандартного отклонения может значительно отличаться от оценки стандартного отклонения первой выборки. При такой неопределенности нет возможности судить о возможном разбросе результатов измерений.

Более наглядно предел возможного различия оценок стандартного распределения дает доверительный интервал, который можно построить на основе распределения χ^2 Пирсона. Техника построения такого интервала известна и приведена, например, в [3]. В иллюстративном примере со стандартным отклонением $\tilde{\sigma} = 1.2 \times 10^{-5}$ среднего в выборке объемом 10 при доверительной вероятности 0,9 границы интервала оказались равными $[0,88; 2,0] \cdot 10^{-5}$. Следовательно, в повторных выборках того же объема из того же распределения оценки стандартных отклонений среднего с довольно высокой вероятностью могут отличаться почти в два раза.

Мы рассмотрели пример с выборкой объемом 10. На практике выборки большего объема встречаются редко; чаще встречаются выборки меньшего объема, у которых возможный разброс оценок стандартного отклонения среднего еще больше. Следовательно, стандартное отклонение среднего является ненадежным показателем точности практических измерений. Странно, что на практике такой значительный разброс оценок стандартного отклонения среднего не принимается во внимание.

При научных исследованиях число наблюдений тоже редко когда превышает 10. Но иногда можно встретить измерения с числом наблюдений порядка 100 и больше. При таком большом числе наблюдений показатель ϕ становится приемлемым и надежность оценки стандартного отклонения можно считать достаточной. Но о возможном разбросе результатов измерений все равно судить трудно.

Перейдем к анализу доверительных интервалов в качестве показателя точности измерений. Доверительный интервал строится обычно на основе распределения Стьюдента

$$t = \frac{\bar{x} - A}{S(\bar{x})},$$

где A – истинное значение измеряемой величины, а $S(\bar{x})$ – стандартное отклонение среднего выборки.

Примечательно, что в формулу Стьюдента входят только интегральные оценки параметров выборки \bar{x} и $S(\bar{x})$. Поэтому это распределение мало чувствительно к форме распределения исходных данных и пригодно для любых симметричных выпук-

лых распределений. Доверительный интервал, получаемый по распределению Стьюдента, учитывает зависимость \bar{x} и $S(\bar{x})$ от числа наблюдений, т.к. эта зависимость выражается в числе степеней свободы выборки и отражается на ширине интервала. Поэтому доверительный интервал является однозначным и точным в вероятностном смысле показателем точности измерений. Последнее означает, что известна вероятность того, что построенный доверительный интервал содержит истинное значение измеряемой величины; эта вероятность выбирается экспериментатором заранее исходя из конкретных целей измерения. Чаще всего доверительная вероятность принимается равной 95%. В ответственных измерениях – 99%, а если требуется, то и еще более высокой.

Таким образом, в качестве показателя точности результата измерения доверительный интервал значительно информативнее стандартного отклонения.

2. Показатели точности косвенных измерений. Косвенные измерения могут быть с независимыми непосредственно измеряемыми величинами (для краткости называемые аргументами) и с зависимыми аргументами. В первом случае можно вычислить как стандартное отклонение результата измерения, так и доверительный интервал. В случае с зависимыми аргументами стандартное отклонение вычислить можно, но для этого в расчеты нужно ввести коэффициенты корреляции между аргументами. Доверительный же интервал построить нельзя, т.к. неизвестно, как найти число степеней свободы с учетом коэффициентов корреляции.

Многочисленные косвенные измерения с зависимыми аргументами очень часто используются при научных исследованиях. Поскольку в этом случае нельзя было построить доверительный интервал, то в качестве показателя точности измерения ученые привыкли использовать стандартное отклонение. К тому же другой возможности просто не было. Однако почти столетия назад был найден другой метод обработки экспериментальных данных для зависимых косвенных измерений – метод приведения [3]. Этот метод позволяет построить доверительный интервал для результата зависимых косвенных измерений и, как показано в [3], имеет много других преимуществ по сравнению с прежним методом. В частности, он точнее традиционного, так как использует всю информацию, содержащуюся в экспериментальных данных, и проще в вычислительном плане так как устраняет необходимость в коэффициентах корреляции.

Суть метода приведения можно пояснить следующим образом. Прежде всего, заметим, что при зависимых косвенных измерениях аргументы всегда измеряют группами так, чтобы каждая группа составляла одно событие. Например, чтобы найти коэффициент линейного расширения какого-то метал-

ла, измеряют длину стержня из данного металла и его температуру при разных температурах. При этом каждый раз длина стержня и его температура должны измеряться одновременно и соответствовать друг другу. В этом отражается тот факт, что в данном измерении длина и температура являются зависимыми аргументами. Их связь можно было учесть при вычислении стандартного отклонения результата с помощью коэффициента корреляции. Метод приведения использует те же самые группы наблюдений, полученные при согласованных измерениях аргументов, но использует их по-другому. Согласно этому методу каждая группа таких наблюдений подставляется в уравнение, связывающее измеряемую величину с аргументами, и дает одно значение измеряемой величины. Другая группа дает еще одно значение этой величины, а если мы имеем n групп, то в итоге получим n значений измеряемой величины. Этот ряд эквивалентен ряду наблюдений, получаемому при прямых измерениях, и к его обработке применимы все хорошо разработанные в статистике методы для прямых измерений. В частности, становится известным число степеней свободы и можно построить доверительный интервал, используя распределение Стьюдента.

Так что в настоящее время уже нет никакой необходимости использовать стандартное отклонение в качестве показателя точности измерений даже при косвенных измерениях; во всех случаях можно построить доверительный интервал.

Заключение

Оценка стандартного отклонения результата многократного измерения всегда необходима при обработке экспериментальных данных. В частности, эта оценка нужна для построения доверительного интервала. Однако в качестве показателя точности результата измерения эта оценка имеет принципиальный недостаток: она не дает никакой информации о ее надежности. Казалось бы, можно выйти из положения, взяв за оценку границу доверительного интервала, покрывающего истинное значение стандартного отклонения. Однако этот путь не нашел применения из-за того, что занижает точность измерения и усложняет расчеты. Другой недостаток стандартного отклонения состоит в том, что оно не дает представления о возможном разбросе результатов измерения.

Доверительный интервал имеет точно вычисляемые границы и дает ясное представление о возможном разбросе результатов измерения. Поэтому доверительный интервал является хорошим показателем точности измерения. В прошлом его нельзя было построить для зависимых косвенных измерений. Однако Метод приведения снял это ограничение, он позволяет построить доверительный интервал для зависимых косвенных измерений. Таким образом, для практических измерений во всех случаях можно и нужно пользоваться доверительным интервалом в качестве показателя точности измерений.

Особое положение с измерениями в научных целях. Здесь часто делается так много измерений, что стандартное отклонение становится достаточно надежным и поэтому его можно использовать в качестве показателя точности измерения. Но это только косвенный показатель точности и он не дает осязательного представления о возможном разбросе результатов измерения и поэтому уступает доверительному интервалу. Тем не менее, стандартное отклонение обычно используется научными работниками. Возможно, это вызвано тем, что в прошлом для зависимых измерений другого варианта просто не было. Но с появлением метода приведения это ограничение отпало. К тому же метод приведения точнее традиционного, а в вычислительном плане он проще традиционного. Поэтому во всех видах многократных измерений можно и целесообразно использовать один и тот же показатель точности - доверительный интервал.

Список литературы

1. Крамер Г. Математические методы статистики / Г. Крамер. – 2-е изд. – М.: Мир, 1975. – 648 с.
2. Захаров И.П. Сравнительный анализ методов обработки экспериментальных данных при косвенных некоррелированных измерениях / И.П. Захаров, С.Г. Рабинович // Системи обробки інформації: зб. наук. пр. – Х.: ХУПС, 2011. – Вип. 1 (91). – С. 33-37.
3. Rabinovich S.G. Evaluating measurement accuracy: a practical approach / S.G. Rabinovich. – New York: Springer, 2010. – 271 p.

Поступила в редколлегию 24.12.2011

Рецензент: д-р техн. наук, проф. И.П. Захаров, Харьковский национальный университет радиоэлектроники, Харьков, Украина.

СТАНДАРТНІ ВІДХИЛЕННЯ АБО ДОВІРЧИЙ ІНТЕРВАЛ!

С.Г. Рабінович

Як показник точності багаторазових вимірювань одні використовують стандартне відхилення, інші – довірчий інтервал результату вимірювання. У статті наведено аналіз цих показників і показано, що довірчий інтервал набагато краще висловлює точність вимірювань, ніж стандартне відхилення. Використання стандартного відхилення можна розглядати як пережиток минулого, коли для непрямих вимірювань з залежними аргументами не було способу побудувати довірчий інтервал. Тепер такий спосіб є. Він відомий і застосовується вже майже півстоліття. Тому зроблено висновок, що у всіх багаторазових вимірюваннях можна та доцільно застосовувати один і той же показник точності – довірчий інтервал.

Ключові слова: багаторазові вимірювання, стандартне відхилення, довірчий інтервал.

STANDARD DEVIATION OR CONFIDENCE INTERVAL?

S.G. Rabinovich

Two different indicators are commonly used to characterize accuracy of multiple measurements - standard deviation and confidence interval. This paper analyzes both indicators and shows that the confidence interval better expresses measurement accuracy. Usage of the standard deviation was unavoidable in the past, when there was no known technique to construct confidence intervals for indirect measurements with dependent arguments. Now this technique exists and has been used for almost half a century. Therefore, we conclude that presently, the confidence interval should be used as the only indicator of measurement accuracy in all cases of multiple measurements.

Keywords: *multiple measurements, standard deviation, confidence interval.*