ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 50779.60— 2017 (ИСО 13528:2015)

Статистические методы

ПРИМЕНЕНИЕ ПРИ ПРОВЕРКЕ КВАЛИФИКАЦИИ ПОСРЕДСТВОМ МЕЖЛАБОРАТОРНЫХ ИСПЫТАНИЙ

(ISO 13528:2015, Statistical methods for use in proficiency testing by interlaboratory comparison, MOD)

Издание официальное

Москва
Стандартинформ
2017

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Открытым акционерным обществом «Научно-исследовательский центр контроля и диагностики технических систем» (АО «НИЦ КД») на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 4, при участии Федерального государственного унитарного предприятия «Уральский научно-исследовательский институт метрологии» (ФГУП «УНИИМ»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 125 «Применение статистических методов»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 12 сентября 2017 г. № 1061-ст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО 13528:2015 «Статистические методы для применения при проверке квалификации посредством межлабораторных испытаний» (ISO 13528:2015 «Statistical methods for use in proficiency testing by interlaboratory comparison», MOD) путем внесения отклонений, объяснение которых приведено во введении к настоящему стандарту.

Международный стандарт разработан техническим комитетом ТС 56 «Dependability» международной электротехнической комиссии (IEC).

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2012 (пункт 3.5).

Сведения о соответствии ссылочных национальных и межгосударственных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте, приведены в дополнительном приложении ДА

5 ВЗАМЕН ГОСТ Р ИСО 13528—2010

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Стандартинформ, 2017

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения . . . 1
2 Нормативные ссылки . . . 1
3 Термины и определения . . . 2
4 Основные принципы . . . 4
5 Руководство по планированию эксперимента для программ проверки квалификации . . . 5
6 Предварительный анализ образцов и результатов при проверке квалификации . . . 9
7 Определение приписанного значения и его стандартной неопределенности . . . 13
8 Определение критериев оценки работы лабораторий . . . 18
9 Вычисления статистик функционирования лабораторий . . . 22
10 Графические методы описания индексов функционирования . . . 29
11 Планирование эксперимента и анализ программ проверки квалификации для качественных показателей (включая номинальные и порядковые свойства) . . . 34
Приложение А (обязательное) Обозначения . . . 39
Приложение В (обязательное) Однородность и стабильность образцов для проверки квалификации . . . 41
Приложение С (обязательное) Робастный анализ . . . 47
Приложение D (справочное) Дополнительное руководство к статистическим процедурам . . . 56
Приложение Е (справочное) Иллюстративные примеры . . . 60
Приложение ДА (справочное) Сведения о соответствии ссылочных национальных и межгосударственных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте . . . 80
Библиография . . . 81

Введение

0.1 Цели проверки квалификации

Проверка квалификации посредством межлабораторных сравнительных испытаний предназначена для определения способности участников (которыми могут быть лаборатории, контролирующие органы или физические лица) проводить испытания или измерения, а также для проверки их работы. Во введении к *ГОСТ ISO/IEC 17043—2013* приведены основные задачи проверки квалификации лабораторий, например:

- определение оценок характеристик функционирования лабораторий;
- выявление проблем в лабораториях;
- установление результативности и сопоставление методов испытаний или измерений;
- обеспечение дополнительного доверия заказчиков лаборатории;
- подтверждение заявленной неопределенности;
- обучение участвующих лабораторий.

Применяемые статистические и аналитические методы должны быть приемлемыми для указанных целей.

0.2 Обоснование для выбора критериев проверки квалификации

Для проверки квалификации используют различные стратегии. Несмотря на то что детали расчетов разные, в большинстве схем проверки квалификации отклонение результатов измерений участника от приписанного значения сравнивают с критерием для того, чтобы определить, является ли отклонение поводом для беспокойства. Поэтому стратегии, используемые для выбора приписанного значения и критерия оценки отклонений, очень значимы. В частности, важно определить, следует ли выбирать приписанное значение и критерий независимо от результатов измерений участников или их следует определять на основе результатов измерений, представленных участниками. В настоящем стандарте рассмотрены обе стратегии. В разделах 7 и 8 рассмотрены преимущества и недостатки выбора приписанных значений и критериев оценки отклонений независимо от результатов измерений участников. В целом выбор приписанных значений и критериев независимо от результатов измерений участников является предпочтительным. Это, в частности, касается критериев, используемых для оценки отклонения от приписанного значения, таких как стандартное отклонение оценки квалификации или допуск на погрешность измерений, для которых выбор, основанный на пригодности использования для конечного пользователя, особенно важен.

0.3 Настоящий стандарт и *ГОСТ ISO/IEC 17043*

Настоящий стандарт дополняет *ГОСТ ISO/IEC 17043* в части требований к статистическим методам валидации образцов для проверки квалификации, анализа результатов, общей статистической отчетности. В приложении В *ГОСТ ISO/IEC 17043—2013* кратко описаны общие статистические методы, используемые в программах проверки квалификации. Настоящий стандарт предназначен в качестве дополнения к *ГОСТ ISO/IEC 17043* и содержит рекомендации по применению статистических методов для проверки квалификации.

Определение проверки квалификации, приведенное в *ГОСТ ISO/IEC 17043*, совпадает с представленными в настоящем стандарте замечаниями, которые описывают различные типы проверки квалификации и набор планов эксперимента, которые могут быть использованы. Настоящий стандарт не охватывает все цели, планы эксперимента, матрицы и измеряемые величины. Методы, представленные в настоящем стандарте, предназначены для широкого применения, особенно для вновь создаваемых программ проверки квалификации. Предполагается, что статистические методы, используемые в конкретных программах проверки квалификации, будут совершенствоваться с изменением программ, планов эксперимента, критериев оценки, а графические методы — уточняться для учета потребностей целевой группы участников, органов аккредитации и регулирующих органов.

В настоящем стандарте приведено руководство по проверке квалификации химических аналитических лабораторий [1], а также дополнительно представлен набор процедур, позволяющих использовать валидированные методы измерений и работы с данными. Настоящий стандарт представляет статистические методы и содержит указания по их применению, расширенные, по мере необходимости, в соответствии с упомянутыми документами и *ГОСТ ISO/IEC 17043*. В *ГОСТ ISO/IEC 17043* приведены

требования относительно проверки квалификации для отдельных лиц и проверяющих органов; в приложении В — сведения, относящиеся к качеству результатов.

В настоящем стандарте приведены статистические методы, согласующиеся со статистическими методами других международных стандартов, особенно разработанных комитетом ИСО/ТС 69/ SC6, в частности с методами, представленными в серии стандартов *ГОСТ Р ИСО 5725*. Методы, приведенные в настоящем стандарте, согласованы с *ГОСТ Р 54500-3 / Руководство ИСО/МЭК 98-3* (GUM).

0.4 Статистическая экспертиза

В соответствии с требованиями *ГОСТ ISO/IEC 17043—2013* провайдер проверки квалификации должен быть компетентным в области проведения межлабораторных сличений, наделен полномочиями по осуществлению статистической экспертизы, привлечению персонала для выполнения статистического анализа. Однако ни в *ГОСТ ISO/IEC 17043*, ни в настоящем стандарте не установлены требования к необходимости проведения проверки. В некоторых случаях обозначено наличие специального образования у провайдера, но, как правило, требования к проверке могут быть выполнены и техническими специалистами, знакомыми с основными статистическими понятиями и имеющими опыт работы или обученными общим методам анализа данных, применяемым в программе проверки квалификации. Если физическое лицо привлекают к статистической обработке или анализу данных, то даже при наличии специального образования в области статистики очень важно, чтобы этот специалист имел опыт в области межлабораторных сравнительных испытаний. Обычное обучение статистическим методам не включает в себя методы, применяемые при межлабораторных сравнительных испытаниях, и описание причин возникновения ошибок измерений, которые происходят при проверке квалификации и часто остаются невыясненными. В настоящем стандарте не могут быть представлены весь спектр возникающих проблем и опыт, накопленный при работе, связанной с межлабораторными сравнительными испытаниями.

0.5 Программное обеспечение

Компьютерное программное обеспечение, которое необходимо для статистического анализа данных проверки квалификации, весьма разнообразно — от электронных таблиц для выполнения арифметических операций, используемых в случае известных опорных значений, до сложного программного обеспечения, применяемого для итерационных расчетов или других современных численных методов. Для большинства методов, перечисленных в настоящем стандарте, достаточно использования электронных таблиц с настройкой для работы с конкретными программами или для проведения анализа; некоторые методы требуют применения компьютерных приложений, которые находятся в свободном доступе. Во всех случаях пользователи должны проверять точность своих расчетов, особенно при использовании специальных программ. Тем не менее даже тогда, когда методы, приведенные в настоящем стандарте, являются надлежащими и правильно реализованными соответствующими компьютерными приложениями, их не следует применять без тщательной технической и статистической экспертизы, достаточной для выявления и изучения отклонений, которые могут произойти на любом этапе проверки квалификации.

0.6 Отклонения от ИСО 13528:2015

Из раздела 2 исключены стандарты, которые нецелесообразно применять в соответствии с требованиями национальной системы стандартизации.

ПОПРАВКИ, ВНЕСЕННЫЕ В НАЦИОНАЛЬНЫЕ СТАНДАРТЫ РОССИЙСКОЙ ФЕДЕРАЦИИ

03 СОЦИОЛОГИЯ. УСЛУГИ. ОРГАНИЗАЦИЯ ФИРМ И УПРАВЛЕНИЕ ИМИ. АДМИНИСТРАЦИЯ. ТРАНСПОРТ

ОКС 03.120.30

Поправка к ГОСТ Р 50779.60—2017 (ИСО 13528:2015) Статистические методы. Применение при проверке квалификации посредством межлабораторных испытаний

В каком месте	Напечатано	Должно быть
Пункт 7.6.2	(см. 4.4.1)	—
Подпункт 8.6.2.3, пример	***менее 0,4 мg/дм***	***менее 0,4 мg/дм^3***
Пункт 9.2.2, перечисление b)	9,6 для 9,7 для	9.6 для 9.7 для
Пункт 9.3.2	$\delta_E < D < \delta_E$	$-\delta_E < D < \delta_E$
Приложение Е.3, второй абзац	[см. *R* 3.1.1]	—

(ИУС № 2 2020 г.)

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Статистические методы

ПРИМЕНЕНИЕ ПРИ ПРОВЕРКЕ КВАЛИФИКАЦИИ ПОСРЕДСТВОМ МЕЖЛАБОРАТОРНЫХ ИСПЫТАНИЙ

Statistical methods. Use in proficiency testing by interlaboratory comparison

Дата введения — 2018—12—01

1 Область применения

В настоящем стандарте приведены статистические методы, используемые провайдерами проверки квалификации лабораторий для разработки программ проверки квалификации и анализа полученных данных при выполнении этих программ. В настоящем стандарте приведены рекомендации по интерпретации данных проверки квалификации, и он предназначен для использования участниками данных программ и органами по аккредитации.

Процедуры, приведенные в настоящем стандарте, предназначены для демонстрации того, что полученные результаты измерений лабораторий (проверяющих органов, физических лиц) подтверждают (или не подтверждают) качество их работы. Настоящий стандарт применим к проверке квалификации, результатами которой являются количественные или качественные результаты наблюдений, выполняемые на образцах.

Примечание — Процедуры, установленные в настоящем стандарте, могут быть применены к экспертным оценкам, когда мнения или заключения экспертов представлены в форме, допускающей объективное сопоставление с некоторой независимой эталонной величиной или согласованной статистикой: например, при классификации образцов для проверки квалификации на несколько категорий с помощью контроля или при определении во время контроля принадлежности (или нет) образцов для проверки квалификации одному источнику и объективного сопоставления результатов классификации. В этом случае могут быть применены положения настоящего стандарта, относящиеся к номинальным (качественным) свойствам.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие документы:

ГОСТ ISO/IEC 17043—2013 Оценка соответствия. Основные требования к проведению проверки квалификации

ГОСТ ISO Guide 35—2015 Стандартные образцы. Общие и статистические принципы сертификации (аттестации)

ГОСТ ИСО/МЭК 17025—2009 Общие требования к компетентности испытательных и калибровочных лабораторий

ГОСТ Р ИСО 5725-1—2002 Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Основные положения и определения

ГОСТ Р ИСО 5725-2—2002 Точность (правильность и прецизионность) методов и результатов измерений. Часть 2. Основной метод определения повторяемости и воспроизводимости стандартного метода измерений

ГОСТ Р ИСО 5725-5—2002 Точность (правильность и прецизионность) методов и результатов измерений. Часть 5. Альтернативные методы определения прецизионности стандартного метода измерений

ГОСТ Р ИСО 16269-4—2017 Статистические методы. Статистическое представление данных. Часть 4. Выявление и обработка выбросов

Издание официальное

ГОСТ Р 54500.3—2011/Руководство ИСО/МЭК 98-3:2008 Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения

ГОСТ Р ИСО 7870-2—2015 Статистические методы. Контрольные карты. Часть 2. Контрольные карты Шухарта

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1 **межлабораторные сравнительные испытания**[1)] (interlaboratory comparison): Организация, выполнение и оценка результатов измерений или испытаний одного и того же или нескольких подобных образцов двумя или более лабораториями в соответствии с заранее установленными условиями.

3.2 **проверка квалификации** (proficiency testing): Оценка характеристики функционирования участника по заранее установленным критериям посредством межлабораторных сравнительных испытаний.

Примечание — В настоящем стандарте термин «проверка квалификации» употребляется в широком понимании и включает следующее (перечень может быть дополнен):

- количественную программу, когда целью является определение количественной оценки одной или нескольких измеряемых величин каждого образца для проверки квалификации;
- качественную программу испытаний, когда целью является идентификация или описание одной или нескольких качественных характеристик образца для проверки квалификации;
- последовательную программу, когда один или несколько образцов для проверки квалификации последовательно распределяют для испытаний или измерений и возвращают провайдеру проверки квалификации через определенные интервалы;
- параллельную программу, когда образцы для проверки квалификации распределяют для выполнения одновременных испытаний или измерений в течение определенного периода времени;
- единоразовое задание, когда образцы проверки квалификации возможно предоставить только один раз;
- непрерывную программу, когда образцы для проверки квалификации предоставляют через регулярные интервалы времени;
- выборочный контроль, когда выборки отбирают для последующего анализа, а цели программы проверки квалификации включают оценку выполнения выборочного контроля;
- интерпретацию данных, когда предоставляют наборы данных и обрабатывают другую информацию для обеспечения их интерпретации (или получения иного результата).

3.3 **приписанное значение** (assigned value): Значение, приписываемое конкретному свойству образца для проверки квалификации.

3.4 **стандартное отклонение для оценки квалификации** (standard deviation for proficiency assessment): Мера рассеяния, используемая при оценке результатов проверки квалификации.

Примечание 1 — Стандартное отклонение может быть интерпретировано как стандартное отклонение результатов, полученных некоторой гипотетической совокупностью лабораторий, работающих в точном соответствии с установленными требованиями.

Примечание 2 — Стандартное отклонение для оценки квалификации применяют только к результатам, полученным по шкале отношений или шкале разности.

Примечание 3 — Не во всех программах проверки квалификации оценка квалификации основана на анализе рассеяния результатов.

[*ГОСТ ISO/IEC 17043—2013*]

[1)] Синонимами являются термины «межлабораторные сличительные испытания», «межлабораторные сличения».

3.5 **погрешность измерений** (measurement error): Разность между результатом и опорным значением величины.

3.6 **максимально допустимая погрешность** (maximum permissible error): Максимальное значение погрешности измерений по отношению к известному опорному значению величины, разрешенное процедурами спецификации или нормативными документами для данного метода (методики) измерений, средства измерений или измерительной системы.

3.7 ***z*-индекс** (z-score): Стандартизованная мера характеристики функционирования, вычисленная с использованием результата измерений участника, приписанного значения и стандартного отклонения для оценки квалификации.

Примечание — Общая вариация *z*-индекса (иногда обозначается *z'* и произносится *z*-штрих) формируется путем объединения неопределенности приписанного значения со стандартным отклонением для оценки квалификации до вычисления *z*-индекса.

3.8 **Дзета-индекс** (Zeta score): Стандартная мера характеристики функционирования, вычисленная с использованием результата измерений участника, приписанного значения и суммарной стандартной неопределенности результата и приписанного значения.

3.9 **доля допустимого предельного индекса** (proportion of allowed limit score): Стандартизованная мера характеристики функционирования, вычисленная с использованием результатов участника, приписанного значения и критерия для погрешности измерений при проверке квалификации.

Примечание — Для единичных результатов характеристика функционирования выражена в виде отклонения от приписанного значения (*D* или *D* %).

3.10 **сигнал к действиям** (action signal): Признак, указывающий на то, что по результатам проверки квалификации необходимо выполнить какие-либо действия.

Пример — Значение z-индекса более двух подразумевает необходимость проведения анализа возможных причин происходящего, а значение индекса равное трем — выполнения корректирующих действий.

3.11 **согласованная величина** (consensus value): Величина, полученная на основе набора результатов при межлабораторных сравнительных испытаниях.

Примечание — Термин «согласованная величина» обычно используют для описания оценок параметров положения и разброса, полученных по результатам участников раунда в межлабораторных сравнительных испытаниях, но может быть также использован для значений, полученных по результатам установленного подмножества таких результатов или, например, ряда экспертных лабораторий.

3.12 **выброс** (outlier): Наблюдение в совокупности, которое признано несовместимым с остальными членами совокупности.

Примечание 1 — Выброс может появиться за счет случайности из рассматриваемой совокупности, принадлежать другой совокупности, быть результатом некорректной записи или грубой ошибки.

Примечание 2 — Во многих программах проверки квалификации термин «выброс» используют для обозначения сигнала к действиям. Но это не всегда так. Сигнал к действиям может появиться и при отсутствии выбросов.

[*ГОСТ Р ИСО 5725-1:2002*, добавлено примечание]

3.13 **участник** (participant): Лаборатория, организация или физическое лицо, которые получают образцы для проверки квалификации и представляют результаты на рассмотрение провайдеру проверки квалификации.

3.14 **образец для проверки квалификации** (proficiency test item): Проба, продукт, искусственный объект (артефакт), стандартный образец, часть оборудования, эталон, набор данных или другая информация, используемые для проверки квалификации.

Примечание — В большинстве случаев термин «образец для проверки квалификации» соответствует определению термина «стандартный образец».

3.15 **провайдер проверки квалификации**[1)] (proficiency testing provider): Организация, которая несет ответственность за все задачи по разработке и выполнению программы проверки квалификации.

[1)] Синонимом является термин «провайдер межлабораторных сличительных испытаний».

3.16 **программа проверки квалификации** (proficiency testing scheme): Процедура проверки квалификации, разработанная и реализованная за один или несколько раундов в определенной области испытаний, измерений, калибровки или контроля.

Примечание — Программа проверки квалификации может охватывать конкретный вид испытаний, калибровки, контроля или ряд испытаний, калибровок или контрольных операций на образцах для проверки квалификации.

3.17 **стандартный образец; *RM*** (reference material, RM): Материал (вещество), достаточно однородный и стабильный по отношению к одному или нескольким определенным свойствам для того, чтобы использовать его в соответствии с назначением в измерительном процессе.

Примечание 1 — Стандартный образец — общее понятие.

Примечание 2 — Свойства могут быть охарактеризованы количественными или качественными величинами (например, идентичность веществ или категорий).

Примечание 3 — Использование может включать калибровку (градуировку) измерительной системы, определение пригодности методики измерений, приписывание значений свойств другим материалам и контроль качества.

3.18 **сертифицированный стандартный образец**[1]**; *CRM*** (certified reference material, CRM): Стандартный образец, одно или несколько определенных свойств которого установлены метрологически обоснованной процедурой, сопровождаемый паспортом, в котором приведено значение этого свойства, связанной с ним неопределенности и утверждение о метрологической прослеживаемости.

Примечание — Термин «значения» охватывает как количественные, так и качественные характеристики, такие как идентичность или последовательность. Неопределенность такой характеристики может быть выражена с помощью вероятности или уровня доверия.

4 Основные принципы

4.1 Основные требования к статистическим методам

4.1.1 Используемые статистические методы должны соответствовать целям и быть статистически обоснованными. Все статистические предположения, на которых основаны методы или планы проверки, должны быть указаны в программе проверки квалификации, и их применение должно быть обосновано.

Примечание — Статистически обоснованный метод имеет прочную теоретическую основу, известные свойства в ожидаемых условиях использования, и опирается на предположения или условия, при которых данные могут быть применены для целей обработки.

4.1.2 Статистические методы планирования эксперимента и анализа данных должны соответствовать целям программы проверки квалификации.

4.1.3 Провайдер проверки квалификации должен предоставить участникам описание используемых методов расчета, объяснение общей интерпретации результатов, а также обоснование всех ограничений, касающихся интерпретации результатов. Эта информация должна быть представлена в каждом отчете для каждого раунда программы проверки квалификации или в отдельных процедурах и должна быть доступной для участников.

4.1.4 Провайдер проверки квалификации должен гарантировать, что все программное обеспечение является адекватным.

4.2 Базовая модель

4.2.1 Если в программах проверки квалификации для заданного образца результатом проверки квалификации является единственный результат, представляющий собой количественную величину, то базовая модель имеет вид:

$$x_i = \mu + \varepsilon_i, \quad (1)$$

где x_i — результат проверки квалификации i-го участника;

μ — истинное значение измеряемой величины;

ε_i — погрешность измерений i-го участника, подчиняющаяся распределению в соответствии с моделью.

[1] Синонимом является термин «аттестованный стандартный образец».

Примечание 1 — Общие модели для ε включают: нормальное распределение $\varepsilon_i \sim N(0,\sigma^2)$ со средним, равным 0, и дисперсией, одинаковой или различной для каждой лаборатории; более общие модели включают «загрязненное (с выбросом)» нормальное распределение, представляющее собой сочетание нормального распределения с распределением, представляющим собой совокупность ошибочных результатов.

Примечание 2 — Основой модели сравнительной оценки с использованием *z*-индексов и σ_{pt} является предположение о том, что в «идеальной» генеральной совокупности квалифицируемых лабораторий межлабораторное стандартное отклонение должно быть σ_{pt} или меньше.

Примечание 3 — Данная модель отличается от базовой модели, установленной в стандартах серии *ГОСТ Р ИСО 5725*, так как она не использует термин «лабораторное смещение» B_i. Это происходит потому, что термины «лабораторное смещение» и «остаточная погрешность» не могут быть применены в случае единственного наблюдения. Однако если рассматривают результаты участников в нескольких раундах или при проверке используют несколько образцов, может быть полезно включение лабораторного смещения.

4.2.2 Для порядковых или качественных результатов могут быть использованы другие модели или вообще отсутствовать необходимые статистические модели.

4.3 Общие подходы к сравнительной оценке результатов

4.3.1 Существуют три различных общих подхода к оценке показателей в программе проверки квалификации. Применение того или иного подхода зависит от целей программы проверки квалификации. Ниже перечислены эти подходы, как то оценка показателя на основе сопоставления:

a) с независимо выведенным критерием;

b) показателями других участников;

c) заявленной неопределенностью результатов измерений.

4.3.2 Для определения приписанного значения и определения критерия сравнительной оценки общие подходы могут быть использованы по-разному. Например, если приписанное значение является робастным средним участника и сравнительную оценку выполняют на основе σ_{pt} или и δ_E, где δ_E — заранее определенная допустимая погрешность измерений и $\sigma_{pt} = \delta_E/3$. Аналогично, в некоторых случаях приписанное значение может быть опорным значением, а σ_{pt} — робастным стандартным отклонением результатов участника. В подходе c) 4.3.1, использующем неопределенность измерений, приписанное значение обычно соответствует опорному значению.

5 Руководство по планированию эксперимента для программ проверки квалификации

5.1 Введение

Проверка квалификации связана с оценкой показателей участников и по существу не связана со смещением или прецизионностью (хотя и они могут быть оценены в некоторых конкретных случаях). Показатели участников оценивают с помощью статистической сравнительной оценки их результатов на основе измерений или интерпретаций, которые они выполняют на образцах при проверке квалификации. Показатель часто выражают в виде индексов, которые дают возможность единой интерпретации результатов измерений, позволяя тем самым сопоставлять различные результаты измерений с единой базой. Индекс обычно определяют на основе сопоставления разности полученного результата участника, зафиксированного в отчете, и приписанного значения со стандартным отклонением или с оценкой неопределенности результатов измерений. Анализ индексов за несколько раундов программы проверки квалификации может дать информацию, свидетельствующую о наличии систематических ошибок (смещений) или низкой долговременной прецизионности в лабораториях.

В разделах 5—10 приведены рекомендации по разработке программ проверки квалификации и статистической обработке результатов для количественных величин, включая вычисление и интерпретацию различных индексов. В разделе 11 приведены рекомендации для качественных программ проверки квалификации, включая программы с порядковыми данными для количественных величин.

5.2 Основы статистического планирования эксперимента

5.2.1 В соответствии с *4.4.4.1 ГОСТ ISO/IEC 17043—2013* «следует разработать методы статистического планирования эксперимента, отвечающие целям программы проверки квалификации, учитывающие тип данных (качественные или количественные, включая порядковые и категоризированные

данные), статистические предположения, особенности ошибок и ожидаемое количество результатов». Следовательно, в программах проверки квалификации с различными целями и источниками ошибок следует применять различные планы эксперимента.

Ниже приведены планы эксперимента для общих целей. Но другие цели также возможны.

Пример 1 — Для программы проверки квалификации, в которой сопоставляют результаты участников с предварительно определенным опорным значением в пределах границ, которые устанавливают до начала раунда, для планирования эксперимента требуются метод получения внешне определенного опорного значения, методы установления границ и расчета индексов.

Пример 2 — Для программы проверки квалификации, в которой сопоставляют результаты участников с комбинированными результатами группы в одном и том же раунде и границами, которые устанавливают до начала раунда, для планирования эксперимента необходимо предусмотреть определение приписанного значения по комбинированным результатам, а также метод установления границ и метод индексов.

Пример 3 — Для программы проверки квалификации, в которой сопоставляют результаты участников с комбинированными результатами группы в одном и том же раунде и границами, определенными на основе изменчивости результатов участников, необходимо рассмотреть расчеты приписанного значения и соответствующей меры разброса, а также метод определения индекса.

Пример 4 — Для программы проверки квалификации, в которой сопоставляют результаты участников с приписанным значением, полученным на основе их собственной неопределенности, при планировании эксперимента, необходимо рассмотреть способ получения приписанного значения и его неопределенности и использование неопределенности измерений участников при получении индекса.

Пример 5 — Для программы проверки квалификации с целью сопоставления различных методов измерений при планировании эксперимента, необходимо рассмотреть соответствующие обобщенные статистики и процедуры их вычисления.

5.2.2 При проверке квалификации используют различные типы данных, включая количественные, номинальные (категоризированные) и порядковые. Для количественных переменных некоторые результаты могут быть определены по интервальной или относительной шкале либо шкале отношений. В некоторых случаях при измерении количественных величин могут быть получены только дискретные или прерывистые наборы данных (например, при последовательном разбавлении), однако во многих случаях эти данные могут быть обработаны с помощью методов, применяемых к непрерывным данным.

Примечание 1 — Интервальная шкала для количественных величин представляет собой шкалу, на которой можно определить интервалы (разности) значений, но нельзя определить отношения величин, например, шкала Цельсия. Для шкалы отношений значимыми являются как интервалы, так и отношения величин, например, шкала Кельвина или наиболее общие единицы измерения длины.

Примечание 2 — Для качественных величин применяют классификацию по категориям, при этом упорядочивание значений не имеет смысла, например, упорядочивание наименований видов бактерий. Для порядковой шкалы порядок расположения величин имеет значение, но разности между ними не могут быть определены, например, такие понятия, как «большая, средняя, маленькая величины», могут быть упорядочены, но разность между величинами не может быть определена, кроме как в виде промежуточных значений.

5.2.3 Кроме того, программы проверки квалификации могут быть использованы для других целей (см. *0.1 ГОСТ ISO/IEC 17043—2013*). Для всех установленных целей конкретной программы проверки квалификации должен быть определен подходящий план эксперимента.

5.3 Анализ статистического распределения результатов

5.3.1 В соответствии с *4.4.4.2 ГОСТ ISO/IEC 17043—2013* методы статистического анализа данных должны быть согласованы со статистическими предположениями относительно данных. Большая часть методов общего статистического анализа данных, используемых при проверке квалификации, предполагает, что набор результатов участников подчиняется приблизительно нормальному распределению или, по крайней мере, унимодальному и достаточно симметричному либо результаты могут быть преобразованы к таким данным. Общее дополнительное предположение состоит в том, что распределение результатов измерений представляет собой смесь, включающую результаты из совокупности ошибочных измерений, являющихся выбросами. Обычно интерпретация в баллах основана на предположении о нормальности распределения, но только для основного предполагаемого распределения для компетентных участников.

5.3.1.1 Как правило, нет необходимости в проверке того, что результаты принадлежат нормальному распределению, но очень важно проверить (хотя бы визуально) симметричность распределения. Если нельзя проверить симметричность плотности распределения, то провайдеру проверки квалификации следует использовать методы, которые устойчивы к асимметрии плотности распределения (см. приложение С).

5.3.1.2 Если ожидаемое распределение для программы проверки квалификации не является достаточно симметричным (с учетом загрязнения выбросами), провайдер программы квалификации должен выбрать методы анализа данных, которые учитывают ожидаемую асимметрию и устойчивы к выбросам, и методы расчета индексов, учитывающие особенности ожидаемого распределения результатов участников.

Такие методы могут включать:

- преобразование, обеспечивающее переход к приближенно симметричному распределению;
- методы оценки, устойчивые к асимметрии распределения;
- методы оценки, учитывающие предположения о распределении (например, метод максимального правдоподобия, подходящий для предполагаемого распределения, при необходимости — исключение выбросов).

Пример 1 — Результаты, полученные на основе разбавления растворов, например при расчетах в микробиологических исследованиях или в методах анализа иммунологического статуса, часто имеют логарифмически нормальное распределение, и первым этапом анализа данных может быть их логарифмическое преобразование.

Пример 2 — Результаты подсчета небольшого количества частиц могут быть распределены в соответствии с распределением Пуассона, и поэтому критерии оценки могут быть определены по таблицам вероятностей Пуассона, на основе расчета среднего для группы участников.

5.3.1.3 В некоторых случаях калибровки результаты участников могут иметь распределение, описанное в процедуре выполнения измерений (например, экспоненциальное, прямоугольное, треугольное или другое распределение); это распределение следует указывать во всех протоколах оценки.

5.3.2 В соответствии с *4.4.4.2 ГОСТ ISO/IEC 17043—2013* провайдер проверки квалификации должен обосновать все используемые статистические предположения. Эти доказательства могут быть основаны, например, на данных наблюдений, результатах предыдущих раундов программы проверки квалификации или технической литературе.

Примечание — Обоснование предположений о виде распределения является менее строгим, чем доказательство справедливости этого предположения.

5.4 Исследования для небольшого количества участников

5.4.1 План эксперимента для программы проверки квалификации должен устанавливать минимальное количество участников, необходимых для достижения поставленных целей, а также альтернативные подходы в том случае, если количество участников будет недостаточным [*см. 4.4.4.3, перечисление b) ГОСТ ISO/IEC 17043—2013*]. Статистические методы, которые подходят для большого количества участников, не следует применять при недостаточном количестве участников. Опасность состоит в том, что результаты, определяемые при небольшом количестве участников, могут быть недостаточно достоверными, а оценка участника может быть сопоставлена с группой, неподходящей для сравнения.

Примечание — В [2] приведены полезные рекомендации для программ проверки квалификации с небольшим количеством участников. Приписанное значение рекомендуется устанавливать на основе независимых достоверных измерений: например, на основе использования сертифицированных стандартных образцов, независимо от назначенного посредством калибровки национальным метрологическим институтом, или с помощью гравиметрической подготовки образцов. Кроме того, стандартное отклонение для оценки квалификации не может быть определено на основе наблюдаемого разброса результатов участников одного раунда программы проверки квалификации.

5.4.2 Минимальное количество участников, необходимое для выполнения различных статистических методов, зависит от различных ситуаций, в частности:

- используемых статистических методов, например, робастных методов или выбранной стратегии удаления выбросов;
- опыта участников конкретной программы проверки квалификации;

- опыта работы провайдера проверки квалификации с матрицами, измеряемыми величинами, методами или группой участников;
- определения приписанного значения и стандартного отклонения или того и другого.

Дальнейшие рекомендации по методам обработки результатов при небольшом количестве участников приведены в D.1 приложения D.

5.5 Рекомендации по выбору формы отчета

5.5.1 В соответствии с *4.6.1.2 ГОСТ ISO/IEC 17043—2013*, провайдеры проверки квалификации должны предоставить участникам подробные инструкции по выполнению измерений и оформлению отчета о результатах проверки квалификации так же, как для большинства повседневно выполняемых измерений, за исключением особых случаев.

Это требование в некоторых ситуациях может затруднить получение точной оценки прецизионности и правильности измерений участников или квалификации при выполнении процедуры измерений. Провайдер проверки квалификации должен установить форму отчета для программы проверки квалификации, но, по возможности, использовать единицы измерений, знакомые большинству участников, и выбрать форму отчета, минимизирующую ошибки при переписывании данных и другие ошибки. Может быть установлено автоматическое предупреждение об использовании ненадлежащих единиц измерения.

Примечание 1 — Целью некоторых программ проверки квалификации является оценка способности участника выполнять стандартный метод, использующий определенную единицу измерений или конкретное количество значащих цифр.

Примечание 2 — Количество ошибок переписывания при сборе данных можно существенно уменьшить или совсем устранить за счет использования электронных систем отчетности, которые позволяют участникам вводить свои данные самостоятельно.

5.5.2 Если программа проверки квалификации требует выполнения репликаций измерений на образцах, используемых при проверке квалификации, участник обязан внести в отчет все результаты репликаций измерений. Это может потребоваться в том случае, когда надо оценить прецизионность измерений участника на репликациях измерений известных образцов для проверки квалификации или когда для процедуры измерений необходимо выделить отчет о многократных наблюдениях. В таких ситуациях провайдер проверки квалификации также может попросить участника сообщить выборочное среднее своих результатов (или другую оценку параметра положения) и неопределенность измерений для анализа данных.

5.5.3 Если результаты в отчете записывают в виде «больше» или «меньше» ограничения (например, уровень калибровки или ограничение количественной величины) или если числовые результаты необходимы для оценки в виде индексов, провайдер должен определить, каким образом результаты должны быть обработаны.

5.5.3.1 Для цензурированных данных (см. Е.1 приложения Е) провайдер должен либо принять утвержденные процедуры обработки данных и определения индексов, либо требовать от участников результаты измерений вместо или в дополнение к обычно фиксируемым в отчете значениям.

Примечание 1 — Процедуры обработки индексов могут не учитывать такие данные.

Примечание 2 — Требования участников о фиксировании в отчете числовых значений вне диапазона обычно указываемых значений (например, ниже границы числовых значений) позволят использовать статистические методы, но могут приводить к определению индексов, которые не отражают уровень обслуживания участником потребителей.

5.5.3.2 Если при использовании согласованных статистик цензурированных значений достаточно много, может оказаться невозможно выполнить оценку, так как эти данные влияют на робастность метода. В тех случаях, когда количество цензурированных результатов достаточно велико и влияет на робастность метода, результаты должны быть оценены с использованием статистических методов, которые позволяют получить несмещенную оценку при наличии цензурированных данных [3], или результаты не должны быть оценены. В случае сомнений относительно выбранной процедуры провайдеру проверки квалификации необходимо рассчитать суммарные статистики и оценки показателей в соответствии с каждой альтернативной статистической процедурой, потенциально применимой в данной ситуации, и исследовать значимость различий полученных оценок.

5.5.3.3 Если ожидаются или наблюдаются цензурированные результаты, такие как «меньше заявленного значения», в план эксперимента программы проверки квалификации следует включать положения для определения индексов и/или производить другие действия с цензурированными данными, сообщаемыми участниками, а участники должны быть уведомлены об этих положениях.

Примечание — В Е.1 приложения Е приведен пример некоторых подходов к анализу цензурированных данных. В этом примере показаны робастные согласованные статистики с тремя различными подходами: с удалением цензурированных данных, с заменой цензурированных данных на максимальные значения (указываемые после знака «<») и с заменой цензурированных данных на половину максимальных значений.

5.5.4 Как правило, количество значащих цифр, указанных в отчете, определяют при разработке плана эксперимента программы проверки квалификации.

5.5.4.1 При определении количества значащих цифр, указываемых в отчете, необходимо обеспечить, чтобы ошибка округления была пренебрежимо малой по сравнению с ожидаемыми вариациями результатов участников.

Примечание — В некоторых ситуациях корректный отчет является частью определения квалификации участника, а количество значащих цифр и знаков после запятой может быть различным.

5.5.4.2 Если количество значащих цифр, указанное участником в отчете о результатах измерений в обычных условиях, затрудняет обработку данных провайдером (например, когда в соответствии с процедурой измерений необходимо фиксировать небольшое количество значащих цифр), провайдер проверки квалификации может сообщить участнику количество значащих цифр, которые требуется указать.

Пример — В процедуре измерений установлено, что результаты измерений можно указывать с точностью до 0,1 g, что приводит к появлению большого количества (> 50 %) одинаковых результатов измерений и таким образом к неправильному расчету робастных средних и стандартных отклонений. Поэтому провайдер проверки квалификации может потребовать от участников указывать в отчете результаты измерений с точностью до 2-го или 3-го знака после запятой для получения достаточно достоверных оценок параметров положения и разброса.

5.5.4.3 Если допускается, что различные участники фиксируют в отчете результаты измерений с указанием различного количества значащих цифр, провайдеру проверки квалификации следует принять это во внимание при определении согласованных статистик (таких как приписанное значение и стандартное отклонение для оценки квалификации).

6 Предварительный анализ образцов и результатов при проверке квалификации

6.1 Однородность и стабильность образцов при проверке квалификации

6.1.1 Провайдер проверки квалификации должен гарантировать, что партии образцов для проверки квалификации являются достаточно однородными, стабильными и пригодными для целей программы проверки квалификации. Провайдер должен оценить однородность и стабильность образцов с использованием критериев, позволяющих гарантировать, что неоднородность и нестабильность образцов не оказывают отрицательного влияния на оценку работы участника. При оценке однородности и стабильности следует использовать один или несколько из следующих подходов:

a) экспериментальные исследования, описанные в приложении В, или альтернативные экспериментальные методы, которые обеспечивают равную им или большую уверенность в однородности и стабильности образцов;

b) использование аналогичных образцов в предыдущих раундах программы проверки квалификации, подтверждающие приемлемость образцов для текущего раунда;

c) оценка данных участников в текущем раунде программы проверки квалификации для доказательства согласованности с предыдущими раундами, то есть для доказательства возможных изменений после подготовки отчета или в порядке производства и всех неожиданных отклонений, характерных для неоднородности или нестабильности.

Примечание 1 — Эти подходы могут быть применены с использованием соответствующих статистических методов и технического обоснования для каждого подхода. В течение срока действия программы проверки квалификации подход может быть применен многократно, накопленный опыт со временем снижает первоначальную потребность в экспериментальных исследованиях.

Примечание 2 — Эксперимент [см. перечисления a)—c)] является разумной основой до тех пор, пока выполняются следующие требования:

1 В процессе изготовления партий образцов для проверки квалификации не возникают изменения, которые могут повлиять на однородность образцов.

2 Материалы, используемые в производстве образцов для проверки квалификации, не изменяются таким образом, что могут повлиять на однородность образцов.

3 Не появляются нарушения однородности, выявленные с помощью проверки на однородность или по реакции участников.

4 Требования к однородности материала регулярно анализируют с учетом предполагаемого использования материала при проведении анализа, чтобы гарантировать, что однородность, достигнутая за счет производственного процесса, остается пригодной для целей проверки квалификации.

Пример — Если в предыдущих раундах программы проверки квалификации использованы проверенные образцы, которые были в наличии, и продемонстрировано, что их однородность и стабильность являются достаточными, и участники проверки не изменились в новом раунде, то, если межлабораторное стандартное отклонение в текущем раунде не превышает стандартного отклонения в предыдущих раундах, это является доказательством достаточной однородности и стабильности образцов в текущем раунде.

6.1.2 Для программ проверки квалификации при калибровке, когда один и тот же артефакт используют несколько участников, провайдер проверки квалификации должен гарантировать его стабильность в течение всего раунда или иметь процедуры идентификации и оценки его нестабильности в процессе выполнения раунда программы проверки квалификации. Необходимо также исследовать тенденции изменения отдельных образцов и измеряемых величин, такие как дрейф. По возможности для уверенности в стабильности следует учитывать влияние использования одинаковых артефактов из различных партий.

6.1.3 Все результаты измерений величин (или свойств) должны быть проверены на однородность и стабильность. Однако если показано, что некоторое подмножество свойств может обеспечить хорошую индикацию однородности и стабильности для всех наблюдаемых свойств в этом раунде, программы проверки квалификации оценки, описанные в 6.1.1, могут быть ограничены этим подмножеством свойств. Проверяемые измеряемые величины должны быть чувствительными к источникам неоднородности и нестабильности в процессе работы с образцами для проверки квалификации. Некоторые важные случаи:

a) если результатом измерений является доля, характерная особенность которой ее небольшое значение, эта величина может быть трудно управляемой и более чувствительной к изменениям однородности;

b) если испытуемый образец нагревается во время работы с ним, то в качестве измеряемой величины следует выбрать величину, чувствительную к неравномерному нагреву;

c) если измеряемая величина характеризует свойство, зависящее от отстаивания, выпадения осадка и других явлений, зависящих от времени подготовки образца, то это свойство следует проверять косвенным способом.

Пример — В программе проверки квалификации при определении содержания токсичных металлов в почве результат зависит от влажности образца почвы. Последовательная проверка содержания влаги в почве может быть признана достаточной для обеспечения адекватной стабильности содержания токсичных металлов.

Примечание — Пример проверки на однородность и стабильность с использованием статистических методов, рекомендованных в приложении В, приведен в Е.2 приложения Е.

6.2 Анализ различных методов измерений

6.2.1 Если все участники проверки должны сообщать значение одной и той же измеряемой величины, приписанное значение обычно должно быть одинаковым для всех участников. Но если участникам предоставлен выбор собственного метода измерений, то возможно, что единственное приписанное значение для каждого исследуемого вещества или свойства не будет подходить для всех участников. Однако существует вероятность того, что при использовании различных методов измерений результаты будут несопоставимыми. В этом случае провайдер проверки квалификации может использовать для каждого метода измерений свое приписанное значение.

Примеры

a) Медицинские исследования, в которых при установлении диагноза для одного и того же исследуемого материала различные методы измерений дают различные результаты и имеют различные диапазоны нормальных значений.

b) Оперативно определяемые измеряемые величины, такие как выщелачиваемые токсичные металлы в почвах, для которых доступны различные стандартные методы измерений и прямое сравнение величин не предполагается, по программе проверки квалификации устанавливают без ссылки на установленный метод испытаний.

6.2.2 При разработке программы проверки квалификации следует рассматривать необходимость использования различных приписанных значений (например, сделать заготовки для отчетов по конкретным методам), а также при анализе данных по каждому раунду.

6.3 Удаление грубых ошибок

6.3.1 В *B.2.5 ГОСТ ISO/IEC 17043—2013* и согласованом протоколе IUPAC рекомендуют удалять очевидные грубые ошибки из набора данных на ранней стадии анализа данных перед использованием любой робастной процедуры или любого теста по выявлению статистических выбросов. Как правило, эти результаты следует рассматривать отдельно (например, по согласованию с участником). Некоторые грубые ошибки можно исправить, но это должно быть сделано только в соответствии с утвержденной политикой и процедурами.

Примечание — Очевидные грубые ошибки, такие как результаты, указанные в отчете в неправильных единицах измерений, спутанные результаты различных образцов, возникают в большинстве раундов проверки квалификации, и эти результаты только ухудшают применение последующих статистических методов.

6.3.2 При наличии подозрений в том, что результат является грубой ошибкой, он должен быть сохранен в наборе данных и подвергнут последующей обработке в соответствии с 6.4—6.6.

6.4 Визуальный анализ данных

6.4.1 В качестве первого этапа анализа данных провайдер должен организовать визуальный анализ данных, проводимый человеком, имеющим адекватные технические и статистические знания и опыт. Этот анализ необходим для подтверждения ожидаемого распределения результатов, а также для выявления аномалий или непредвиденных источников изменчивости. Например, бимодальное распределение может быть свидетельством того, что данные представляют собой смесь результатов из различных совокупностей, полученных различными методами, или использования загрязненных проб, или применения плохо сформулированных инструкций. Эта ситуация должна быть разрешена до выполнения анализа или оценки.

Примечание 1 — Наиболее полезным и широко доступным способом определения вида распределения является гистограмма. Она позволяет определить унимодальность и симметричность распределения, а также выявить необычные выбросы (см. 10.2). Однако интервалы, используемые для составления гистограммы, и количество результатов в интервале влияют на качество гистограммы, при небольшом количестве точек гистограммы ее может быть трудно построить. Более полезным для выявления возможной бимодальности или отсутствия симметрии является график плотности (см. подраздел 10.3).

Примечание 2 — Также могут быть полезны и другие методы, например, график кумулятивной функции распределения или диаграмма «стебель с листьями». Некоторые графические методы анализа данных приведены в Е.3 и Е.4 приложения Е.

6.4.2 Если проведение визуального анализа всех наборов данных невозможно, должна быть предусмотрена процедура предупреждения появления неожиданной вариации в наборе данных, например с помощью сравнительного анализа неопределенности приписанного значения и ее сопоставления с критериями или с предыдущими раундами программы проверки квалификации.

6.5 Робастные статистические методы

6.5.1 Робастные статистические методы могут быть использованы для описания центральной части нормального распределения набора результатов без идентификации выбросов и исключения их из последующего анализа. На 1-м этапе большинство используемых робастных методов в качестве параметров применяют оценки медианы и размаха для центральных 50 % результатов — это показатели положения центра и разброса данных, аналогичные среднему и стандартному отклонению. В общем случае при проведении анализа надо отдавать предпочтение робастным методам, а не тем, для которых перед использованием требуется удаление выбросов.

Примечание — Стратегии, которые используют классические статистики, такие как стандартное отклонение, после удаления выбросов, как правило, приводят к заниженной оценке дисперсии для приблизительно нормальных данных; робастные статистики, как правило, дают несмещенные оценки дисперсии.

6.5.2 В качестве простых оценок применимы медиана, масштабированная медиана абсолютного отклонения *MAD* и нормированный межквартильный размах *nIQR*. Алгоритм *A* преобразует исходные данные с помощью процесса, называемого винсоризацией, для получения альтернативных оценок среднего и стандартного отклонения для данных, не подчиняющихся нормальному распределению, и является наиболее полезным, когда ожидаемая доля выбросов составляет менее 20 %. Методы Q_n и *Q* для оценки стандартного отклонения (см. приложение С) особенно полезны в той ситуации, когда в большинстве случаев (> 20 %) результаты являются несоответствующими или когда специалисты не могут достоверно проанализировать данные. Другие методы, описанные в приложении С, также достаточно эффективны, если ожидаемая доля экстремальных значений составляет более 20 % (см. приложение D).

Примечание — Медиана, среднее абсолютное отклонение, межквартильный размах и масштабированная медиана абсолютного отклонения имеют большую изменчивость, чем среднее и стандартное отклонения при применении к данным, подчиняющимся распределению близкому к нормальному. Более сложные робастные оценки дают более высокую эффективность для данных с распределением близким к нормальному, обеспечивая при этом большую устойчивость при использовании выбросов за счет использования медианы и межквартильного размаха.

6.5.3 Выбор статистических методов является обязанностью провайдера проверки квалификации. Робастные среднее и стандартное отклонения могут быть использованы для различных целей, оценка работы участников — лишь одна из них. Робастные средние и стандартные отклонения могут быть также использованы в качестве общих статистик для различных групп участников или для конкретных методов.

Примечание — Более подробно робастные процедуры описаны в приложении С. В Е.3 и Е.4 приложения Е приведены примеры, иллюстрирующие использование робастных статистических методов, приведенных в приложении С.

6.6 Методы работы с выбросами для индивидуальных результатов

6.6.1 Проверка на выбросы может быть использована при визуальном анализе аномальных значений или в случае исключения выброса для обеспечения устойчивости к экстремальным значениям при вычислении общих статистик. При использовании методов выявления выбросов следует подтвердить выполнение предположений, лежащих в их основе, чтобы обеспечить пригодность этих методов для целей программы проверки квалификации, в частности многие критерии используют предположение о нормальности распределения данных.

Примечание — В *ГОСТ Р ИСО 16269-4* и *ГОСТ Р ИСО 5725-2* приведено несколько процедур идентификации выбросов, применимых к межлабораторным данным.

6.6.2 Если робастные методы не применяют (см. 6.5.1), то допускается использование стратегии отклонения выбросов, которая основана на исключении из обработки выбросов, обнаруженных с помощью соответствующих критериев с высоким уровнем доверия, и дальнейшем применении простых статистик, таких как среднее и стандартное отклонения. При использовании стратегий исключения выбросов провайдер проверки квалификации должен:

a) документировать критерии и уровень доверия, применяемые для выявления исключаемого выброса;

b) устанавливать ограничения на долю исключенных данных при применении тестов выявления выбросов в случае их использования;

c) демонстрировать, что полученные оценки параметров положения и (при необходимости) масштаба обладают достаточными свойствами (включая эффективность и смещение) для целей программы проверки квалификации.

Примечание — В *ГОСТ Р ИСО 5725-2* приведены рекомендации по выбору уровня доверия, подходящего для отклонения выбросов в межлабораторных исследованиях по определению прецизионности методов испытаний. В частности, в соответствии с *ГОСТ Р ИСО 5725-2* рекомендуется исключать данные только в том случае, когда при выявлении выброса был использован уровень доверия 99 %, если нет другого основания отказаться от конкретного результата.

6.6.3 В тех случаях, когда исключение выброса является частью процедуры обработки данных и результат удаляют как выброс, работу участника тем не менее следует оценивать в соответствии с критериями, используемыми для всех участников программы проверки квалификации.

Примечание 1 — Выбросы среди зарегистрированных значений часто идентифицируют с помощью критерия Граббса для выявления выбросов в соответствии с *ГОСТ Р ИСО 5725-2*. Оценку в этой процедуре определяют с использованием стандартного отклонения всех участников, включая возможные выбросы. Поэтому данную процедуру необходимо применять, когда показатели работы участников соответствуют ожиданиям, основанным на результатах предыдущих раундов, а количество выбросов невелико (один или два выброса с каждой стороны среднего). Обычные таблицы для критерия Граббса предполагают одно применение для возможного выброса или два — для установленного параметра положения, но допустимо последовательное применение этого критерия. Если таблицы Граббса применяют последовательно, вероятности ошибки I рода для испытаний можно не применять.

Примечание 2 — Если результаты репликаций совпадают или если в раунде программы проверки квалификации использованы идентичные образцы, это является основанием для использования критерия Кохрена для повторяющихся выбросов в соответствии с *ГОСТ Р ИСО 5725-2*.

Примечание 3 — Выбросы также могут быть идентифицированы с помощью робастных или непараметрических методов: например, если вычислены робастное среднее и стандартное отклонение значения, отклоняющиеся от вычисленного среднего более чем на три робастных стандартных отклонения, могут быть идентифицированы как выбросы.

7 Определение приписанного значения и его стандартной неопределенности

7.1 Выбор метода определения приписанного значения

7.1.1 В 7.3—7.7 приведены пять методов определения приписанного значения. За выбор метода отвечает провайдер проверки квалификации.

Примечание — Подразделы 7.3—7.6 аналогичны подходам, используемым для определения характеристик свойств сертифицированных стандартных образцов, описанных в *ГОСТ ISO Guide 35*.

7.1.2 Альтернативные методы определения приписанного значения и его неопределенности могут быть использованы при условии, что они имеют строгое статистическое обоснование, а используемый метод описан в документированном плане для программы проверки квалификации, и это описание полностью известно участникам. Независимо от метода, используемого для определения приписанного значения, всегда целесообразно проверить обоснованность выбора приписанного значения для конкретного раунда программы проверки квалификации. Более подробное обсуждение этого положения приведено в 7.8.

7.1.3 Подходы к определению приписанных значений для качественных данных рассмотрены в 11.3.

7.1.4 Метод определения приписанного значения и соответствующей ему неопределенности должен быть указан в каждом отчете для участников или четко описан в протоколе программы проверки квалификации, доступном всем участникам.

7.2 Определение неопределенности приписанного значения

7.2.1 В руководстве по выражению неопределенности измерения *ГОСТ Р 54500.3* приведены рекомендации по оценке неопределенности. В *ГОСТ ISO Guide 35* представлены рекомендации по определению неопределенности приписанного значения для аттестованных значений показателей свойств, которые могут быть использованы при разработке многих программ проверки квалификации.

7.2.2 Общая модель приписанного значения и его неопределенности описывается уравнениями (2) и (3).

Модель приписанного значения может быть рассчитана следующим образом:

$$x_{pt} = x_{chap} + \delta_{hom} + \delta_{trans} + \delta_{stab}, \quad (2)$$

где x_{pt} — приписанное значение;

x_{chap} — показатель свойства, полученный при характеризации (определении приписанного значения);

δ_{hom} — погрешность, вызванная различиями образцов для проверки квалификации;

δ_{trans} — погрешность, вызванная нестабильностью условий транспортирования;

δ_{stab} — погрешность, вызванная нестабильностью процесса проведения проверки квалификации.

Соответствующая модель неопределенности приписанного значения имеет вид:

$$u(x_{pt}) = \sqrt{u_{char}^2 + u_{hom}^2 + u_{trans}^2 + u_{stab}^2}, \quad (3)$$

где $u(x_{pt})$ — стандартная неопределенность приписанного значения;
u_{char} — стандартная неопределенность, соответствующая характеризации;
u_{hom} — стандартная неопределенность, связанная с различиями между образцами;
u_{trans} — стандартная неопределенность, соответствующая нестабильности условий транспортирования образцов;
u_{stab} — стандартная неопределенность, соответствующая нестабильности процесса проведения проверки квалификации.

Примечание 1 — Ковариация источников неопределенности или даже незначительных источников неопределенности может привести к другой модели в конкретных случаях. В некоторых ситуациях все компоненты неопределенности могут быть равны нулю или ничтожно малы.

Примечание 2 — При расчете стандартного отклонения результатов участников σ_{pt} составляющие неопределенности, обусловленные неоднородностью, условиями транспортирования и нестабильностью процесса проверки, в значительной степени отражаются в изменчивости результатов участников. В этом случае достаточно учитывать неопределенность характеризации в соответствии с 7.3—7.7.

Примечание 3 — Обычно предполагают, что провайдер проверки квалификации гарантирует, что изменения, связанные с нестабильностью процесса или условиями транспортирования образцов, пренебрежимо малы по сравнению со стандартным отклонением для оценки квалификации; то есть гарантируют, что δ_{trans} и δ_{stab} пренебрежимо малы. Если это требование удовлетворено, u_{stab} и u_{trans} можно считать равными нулю.

7.2.3 Приписанное значение может иметь смещение, которое не учтено в выражении (2). Это следует по возможности учитывать при разработке программы проверки квалификации. Если проведена корректировка смещения приписанного значения, то неопределенность, соответствующая такой корректировке, должна быть включена в оценку неопределенности приписанного значения.

7.3 Состав образцов

7.3.1 Образец для проверки квалификации может быть подготовлен путем смешивания материалов с различными известными уровнями содержания вещества в установленных пропорциях или добавления установленного количества вещества к основному материалу.

7.3.1.1 Приписанное значение x_{pt} получают путем расчета на основе содержания веществ в используемых материалах. Такой подход особенно важен, когда отдельные образцы для проверки квалификации подготовлены описанным способом и пропорции используемых материалов известны.

7.3.1.2 Следует предпринять необходимые действия для того, чтобы обеспечить следующее:

a) основной материал должен быть практически свободен от дополнительных компонентов, или их доли должны быть точно известны;

b) компоненты должны быть смешаны до однородности (если это необходимо);

c) все существенные источники ошибок должны быть идентифицированы (например, не всем известно, что стекло поглощает соединения ртути, так что концентрация соединения ртути в водном растворе может измениться, если раствор находится в стеклянном контейнере);

d) не существует неблагоприятного взаимодействия между компонентами и основным веществом;

e) свойства образцов для квалификационных испытаний, содержащих дополнительный материал, должны быть аналогичны свойствам образцов, обычно используемых заказчиком. Например, чистые материалы, добавляемые к природной основе, часто извлечь легче из такой же субстанции природного происхождения. Если существует возможность возникновения такой ситуации, провайдер должен обеспечить пригодность образцов для используемых методов проверки квалификации.

7.3.1.3 Если в образцах для проверки квалификации добавки более слабо связаны, чем в обычно проверяемых образцах, или находятся в другой форме, предпочтительнее использовать другой подход для подготовки образцов.

7.3.1.4 Определение приписанного значения по составу материала образцов является одним из случаев общего подхода к характеризации сертифицированных стандартных образцов, описанного в *ISO Guide 35*, когда единственная лаборатория определяет приписанное значение с использованием основного метода измерения. Существуют и другие способы определения приписанного значения единственной лабораторией с помощью основного метода (см. 7.5)

7.3.2 Если приписанное значение рассчитывают на основе состава материала образцов для проверки квалификации, оценки стандартной неопределенности u_{char} определяют с помощью сочетания неопределенностей, используемых в соответствующей модели. Например, при проверке квалификации лабораторий, выполняющих химические измерения, основными являются неопределенности, связанные с гравиметрическими измерениями, измерениями объема и чистотой всех материалов, используемых при составлении образца. Стандартную неопределенность приписанного значения $u(x_{pt})$ рассчитывают по уравнению (3).

7.4 Сертифицированный стандартный образец

7.4.1 Если образец для проверки квалификации является сертифицированным стандартным образцом *CRM*, его сертифицированное значение x_{CRM} используется в качестве приписанного значения x_{pt}.

Ограничения такого подхода вызваны следующим:

- обеспечить каждого участника сертифицированным стандартным образцом потребует больших средств;
- обеспечить долговременную стабильность образца часто очень сложно, что может приводить к нецелесообразности замены образца;
- сертифицированные стандартные образцы могут быть известны участникам, и поэтому важно скрыть их истинные характеристики.

7.4.2 При использовании *CRM* в качестве образца для проверки квалификации стандартную неопределенность приписанного значения выводят из информации о неопределенности, указанной в сертификате. Сертификат должен включать в себя информацию о компонентах из уравнения (3) и информацию о предполагаемом использовании образца для целей программы проверки квалификации.

7.5 Результаты одной лаборатории

7.5.1 Приписанное значение может быть определено единственной лабораторией с использованием эталонного метода, например, основного метода. Используемый эталонный метод должен быть полностью описан и понятен с утверждением полной неопределенности и документированной метрологической прослеживаемостью, подходящей для целей программы проверки квалификации. Эталонный метод должен быть совместим со всеми методами измерений, используемыми участниками.

7.5.1.1 Приписанное значение рассчитывают как среднее арифметическое, определяемое для двух или большего количества образцов или условий измерений и при достаточном количестве репликаций измерений.

7.5.1.2 Неопределенность характеризации представляет собой соответствующую оценку неопределенности для эталонного метода и условий исследования в соответствии с планом эксперимента.

7.5.2 Приписанное значение x_{pt} образца для проверки квалификации может быть получено с помощью единственной лаборатории и посредством использования подходящего метода измерений по данным калибровки на соответствие опорному значению, соответствующему сертифицированному стандартному образцу.

Такой подход предполагает, что *CRM* совместим со всеми методами измерений, используемыми участниками.

7.5.2.1 Это определение требует проведения серии испытаний, проводимых в одной лаборатории, на образцах для проверки квалификации и *CRM* с использованием одного и того же метода и в условиях повторяемости.

Если x_{CRM} — приписанное значение для *CRM*;

x_{pt} — приписанное значение образца для проверки квалификации;

d_i — разность между средним арифметическим результатов для образца проверки квалификации и *CRM* для *i*-го образца;

$\bar{d}$ — среднее арифметическое разностей d_i, тогда

$$x_{pt} = x_{CRM} + \bar{d}. \quad (4)$$

Примечание — x_{CRM} и $\bar{d}$ независимы, за исключением того редкого случая, когда *CRM* получен экспертной лабораторией.

7.5.2.2 Стандартную неопределенность характеризации выводят на основе неопределенности измерений, используемых для определения приписанного значения. Такой подход позволяет установить

приписанное значение таким образом, чтобы была обеспечена метрологическая прослеживаемость с сертифицированным значением *CRM* и стандартной неопределенностью, полученной в соответствии с уравнением (5):

$$u_{char} = \sqrt{u_{CRM}^2 + u_d^2}. \quad (5)$$

Пример, приведенный в Е.5 приложения Е, показывает, как требуемая неопределенность может быть вычислена в простом случае, когда приписанное значение образца для проверки квалификации устанавливают прямым сопоставлением с единственным *CRM*.

7.5.3 Если опорное значение присваивают до начала раунда последовательной программы проверки квалификации, а затем опорное значение проверяют, используя ту же измерительную систему, то разность полученных значений должна быть меньше, чем удвоенная неопределенность этой разности (то есть результаты должны быть метрологически совместимыми). В таких случаях провайдер проверки квалификации может принять решение об использовании в качестве приписанного значения среднего арифметического результатов измерений с соответствующей неопределенностью. Если результаты не являются метрологически совместимыми, провайдер должен исследовать причину различий и принять соответствующие меры, включая использование альтернативных методов определения приписанного значения и его неопределенности или прекращения раунда.

7.6 Согласованное значение по данным экспертных лабораторий

7.6.1 Приписанное значение может быть определено с применением межлабораторных сравнительных испытаний с привлечением экспертных лабораторий в соответствии с *ISO Guide 35* для использования межлабораторного сравнения для характеризации *CRM*. Сначала подготавливают образцы для проверки квалификации и распределяют их между участниками. Некоторые из этих образцов отбирают случайным образом, и затем их анализирует группа экспертов с использованием протокола, в котором устанавливают количество образцов и репликаций и все другие соответствующие условия. Каждая эталонная лаборатория должна обеспечить стандартную неопределенность своих результатов.

7.6.2 Если экспертные лаборатории сообщают единственный результат и в соответствии с протоколом не требуется обеспечить достаточную информацию о неопределенности результатов или при наличии свидетельств либо предположении о недостаточной достоверности полученной неопределенности, согласованная величина должна быть получена в соответствии с методами, приведенными в 7.7, применяемыми к набору результатов экспертных лабораторий. Если каждая из экспертных лабораторий получает более одного результата (например, включая репликации), провайдер в программе проверки квалификации должен установить альтернативный метод определения приписанного значения, его неопределенности, и допускать наличие выбросов или других отклонений от ожидаемого распределения результатов, который должен быть статистически обоснованным (см. 4.4.1).

7.6.3 Если экспертные лаборатории вместе с результатами измерений указывают их неопределенности, оценка приписанного значения путем согласования результатов является сложной проблемой, и для ее решения существует много подходов, включая, например, применение взвешенного среднего арифметического, невзвешенного среднего арифметического, процедур, допускающих большой разброс данных, и процедур, допускающих наличие выбросов или ошибочных результатов и оценок неопределенности [4]. Провайдер проверки квалификации должен соответствующим образом установить процедуру оценки, которая должна:

a) включать проверку достоверности указанных оценок неопределенности, например, проверку соответствия в полной мере указанной неопределенности наблюдаемому разбросу результатов;

b) использовать взвешенные процедуры с весами, соответствующими степени достоверности и размеру указанной неопределенности, например, равные веса в случае близких неопределенностей, либо низкой или неизвестной достоверности (см. 7.6.2);

c) допускать возможность того, что указанные неопределенности могут не в полной мере соответствовать наблюдаемому (полному) разбросу, например, путем включения дополнительного члена;

d) допускать возможность появления неожиданных удаленных значений в указанных результатах или неопределенностях;

e) иметь обоснованную теоретическую основу;

f) иметь продемонстрированные свойства (например, на данных испытаний или на моделях), необходимые для целей программы проверки квалификации.

7.7 Приписанное значение на основе согласованного значения результатов участников

7.7.1 При таком подходе приписанным значением x_{pt} для образца, используемого в раунде программы проверки квалификации, является оценка параметра положения (например, робастное среднее, медиана или среднее арифметическое), сформированная на основе результатов, представленных участниками раунда, и рассчитанная с использованием процедуры в соответствии с планом эксперимента, приведенным в приложении С. Методы, описанные в 6.2—6.6, следует использовать до объединения результатов для обеспечения достаточной согласованности результатов.

7.7.1.1 В некоторых случаях провайдер проверки квалификации использует только группу участников, которые определены как надежные по некоторым критериям, например, на основе статуса аккредитации или предварительной работы. Методы настоящего подраздела применимы к таким ситуациям, в том числе и к определению размера группы.

7.7.1.2 Вместо методов, приведенных в приложении С, могут быть использованы и другие методы при условии, что они имеют статистическое обоснование и указаны в отчете.

7.7.1.3 Преимущества такого подхода заключаются в том:

a) что для получения приписанного значения никаких дополнительных измерений не требуется;

b) подход может быть особенно полезен для стандартизованной, оперативно определяемой измеряемой величины, так как часто нет более надежного способа получения равноценных результатов.

7.7.1.4 Недостатками данного подхода являются следующие:

a) между участниками могут отсутствовать необходимые соглашения;

b) согласованное значение может включать неизвестное смещение, вызванное общим использованием ошибочного метода, и это смещение не будет учтено в стандартной неопределенности приписанного значения;

c) согласованное значение может иметь смещение вследствие применения методов, дающих смещение при определении приписанного значения;

d) трудно определить метрологическую прослеживаемость согласованного значения. Поскольку результат всегда связан с результатами отдельных лабораторий, четкое установление прослеживаемости за их пределами может быть обеспечено только тогда, когда провайдер проверки квалификации имеет полную информацию об используемых калибровочных стандартах и контролирует условия других соответствующих методов, применяемых всеми участниками, участвующими в определении согласованного значения.

7.7.2 Стандартная неопределенность приписанного значения зависит от используемой процедуры. Если требуется самый общий подход, провайдер проверки квалификации должен рассмотреть вопрос об использовании бутстреп-метода для оценки стандартной погрешности приписанного значения. Подробное описание бутстреп-метода приведено в [5], [6].

Примечание — Пример использования бутстреп-метода приведен в Е.6 приложения Е.

7.7.3 Если приписанное значение определяют на основе робастного среднего, вычисляемого с использованием процедур, приведенных в С.2, С.3 или С.5 приложения С, оценку стандартной неопределенности приписанного значения и (x_{pt}) можно определить по формуле

$$u(x_{pt}) = 1{,}25\frac{s^*}{\sqrt{p}}, \qquad (6)$$

где s^* — робастное стандартное отклонение результатов. (Здесь «результатом» является среднее арифметическое всех результатов измерений, выделенных на образце для проверки квалификации.)

Примечание 1 — В данной модели, когда приписанное значение и робастное стандартное отклонение определяют по результатам участников, предполагается, что неопределенность приписанного значения включает влияние неопределенностей, соответствующих неоднородности, условиям транспортирования и нестабильности.

Примечание 2 — Значение коэффициента 1,25 выбрано на основе стандартного отклонения медианы или эффективности медианы как оценки среднего в большом наборе результатов, описываемых нормальным распределением. Следует понимать, что эффективность более сложных робастных методов может быть намного больше, чем методов, использующих медиану с поправочным коэффициентом менее 1,25. Тем не менее рекомендуется применять этот коэффициент, поскольку результаты проверки квалификации, как правило, не подчиняются строго нормальному распределению и содержат неизвестные доли результатов из различных распределений («загрязненные результаты»). Коэффициент 1,25 обеспечивает получение гарантированной (завышенной) оценки с учетом возможного загрязнения. Провайдер проверки квалификации может обосновать использование меньшего значения коэффициента или другого уравнения в зависимости от опыта и используемой робастной процедуры.

Примечание 3 — Пример получения робастного среднего по результатам участников приведен в Е.3 приложения Е.

7.8 Сопоставление приписанного значения с независимым опорным значением

7.8.1 Если для установления приписанного значения используют методы, приведенные в 7.7, при наличии независимой оценки x_{ref}, например, из сведений о подготовке образцов или об опорном значении, согласованное значение следует сопоставлять с x_{ref}.

Если для установления приписанного значения используют методы, приведенные в 7.3—7.6, робастное среднее арифметическое x^*, полученное по результатам раунда, следует сопоставлять с приписанным значением после каждого раунда программы проверки квалификации.

Стандартную неопределенность разности $x_{diff} = (x_{ref} - x_{pt})$ или $x_{diff} = (x^* - x_{pt})$ определяют в следующем виде:

$$u_{diff} = \sqrt{u^2(x_{ref}) + u^2(x_{pt})}, \quad (7)$$

где $u(x_{ref})$ — неопределенность опорного значения для сопоставления;

$u(x_{pt})$ — неопределенность приписанного значения.

Примечание — Пример сопоставления опорного значения с согласованным значением включен в Е.7 приложения Е.

7.8.2 Если разность больше удвоенной соответствующей ей неопределенности, следует исследовать возможные причины этого явления. Возможными причинами могут быть:

- наличие смещения соответствующего метода измерений;
- общего смещения результатов участников;
- ошибки в понимании ограничений метода измерений при использовании метода, приведенного в 7.3;
- смещения в результатах экспертных лабораторий при использовании подходов, приведенных в 7.5 или 7.6;
- опорное значение и приписанное значение не связаны с одним и тем же метрологическим эталоном.

7.8.3 В зависимости от выявленной причины провайдер проверки квалификации должен решить, следует ли оценивать результаты раунда или нет, а также (для непрерывных программ проверки квалификации) нужно ли изменить план эксперимента для последующих программ проверки квалификации. Если разность достаточно велика и может повлиять на оценку функционирования лаборатории или можно предположить наличие существенного смещения у методов измерений, используемых участниками, разность следует указать в отчете. В таких случаях эту разность следует учитывать при разработке будущих программ проверки квалификации.

8 Определение критериев оценки работы лабораторий

8.1 Подходы к определению критериев оценки

8.1.1 Основным подходом во всех случаях является сопоставление результатов измерений, выполненных на образце x_i, с приписанным значением x_{pt}. Для оценки работы разность сравнивают с допустимой погрешностью измерений. Это сравнение обычно выполняют для стандартизованной статистики (например, z, z', ζ, E_n) в соответствии с 9.4—9.7. Также эту разность можно сравнивать с определенным критерием (D или D % сравнивают с δ_E) в соответствии с 9.3. Альтернативным подходом является сопоставление разности с неопределенностью результатов, заявленной участниками, объединенной с неопределенностью приписанного значения E_n и ζ.

8.1.2 Если обязательные требования или соответствие цели заданы в виде стандартного отклонения, то оно может быть использовано непосредственно как σ_{pt}. Если обязательные требования или цели заданы в виде максимально допустимой погрешности измерений, то для получения σ_{pt} критерий может быть поделен на значение границы зоны действия. Установленная максимально допустимая погрешность может быть использована непосредственно в качестве δ_E для использования с D или D %. Преимущества такого подхода для непрерывных программ состоят в следующем:

а) оценки функционирования лабораторий имеют последовательную интерпретацию с точки зрения соответствия конкретной цели в каждом раунде;

b) оценки функционирования зависят от ожидаемой вариации при оценке разброса по фиксированным результатам.

Пример — Если критерием оценки работы является максимально допустимая погрешность и 3,0 — граница зоны действия при оценке с z-индексом, то для определения σ_{pt} значение установленного критерия делят на 3,0.

8.1.3 Если критерий оценки работы основан на согласованных статистиках из текущего или предыдущего раундов программы проверки квалификации, то предпочтительной статистикой является робастная оценка стандартного отклонения результатов участников. При использовании такого подхода, как правило, удобнее всего использовать при оценке работы z-индекс и установить в качестве стандартного отклонения для оценки квалификации σ_{pt} выборочное стандартное отклонение.

8.2 Чувствительность экспертов

8.2.1 Значение максимально допустимой погрешности или стандартного отклонения для оценки квалификации может быть установлено как значение, которое соответствует такому уровню работы, который, по мнению органа по аккредитации или технических экспертов провайдера, является приемлемым для участников.

8.2.2 Установленная максимальная допустимая погрешность может быть преобразована в стандартное отклонение для оценки квалификации путем деления ее предельного значения на число кратное σ_{pt}, которое используют для определения границ зоны действия (или неприемлемого результата). Аналогично установленное значение σ_{pt} может быть преобразовано в δ_E.

8.3 Оценка на основе данных предыдущих раундов проверки квалификации

8.3.1 Стандартное отклонение для оценки квалификации σ_{pt} и максимально допустимая погрешность δ_E могут быть определены на основе данных предыдущих раундов проверки квалификации для одной и той же измеряемой величины со сопоставимыми значениями показателей, если участники используют согласованные процедуры измерений. Это полезный подход, если нет согласия между экспертами относительно соответствия установленным целям. Преимущества данного подхода:

- оценки основаны на разумных ожиданиях качества работы участников;
- критерии оценки не изменяют во всех раундах из-за случайных вариаций или изменений в составе участников;
- критерии оценки не меняют для различных провайдеров, если для испытаний или калибровки утверждены два или более провайдеров.

8.3.2 Анализ предыдущих раундов программы проверки квалификации должен включать рассмотрение показателей компетентных участников и не затронутых влиянием новых участников или не затронутых случайными изменениями, например, сокращением группы или других факторов, присущих конкретному раунду. Заключения могут быть сделаны субъективно, на основе изучения данных предыдущих раундов, объективно — в соответствии среднему или регрессионной модели, установленной для измеряемой величины. Линия регрессии может быть прямой или кривой [7]. Стандартные отклонения и относительные стандартные отклонения следует рассматривать с учетом значений, более подходящих области значений измеряемых величин. Соответствующая максимально допустимая погрешность может быть рассчитана таким же способом.

8.3.3 Если критерии для оценки показателя функционирования основаны на согласованных статистиках предыдущих раундов программы проверки квалификации, следует использовать робастную оценку стандартного отклонения.

Примечание 1 — Алгоритм *S* (см. С.4 приложения С) обеспечивает определение робастного общего стандартного отклонения и применим в том случае, когда все рассматриваемые предыдущие раунды программы проверки квалификации имеют одинаковое среднее стандартное отклонение или (если для оценки использованы относительные отклонения) то же относительное стандартное отклонение.

Примечание 2 — Пример вывода значения по данным предыдущих раундов программы проверки квалификации приведен в Е.8 приложения Е.

8.4 Использование общей модели

8.4.1 Значение стандартного отклонения для оценки квалификации может быть получено из общей модели воспроизводимости метода измерений. Преимуществами этого метода являются объективность и согласованность измеряемых величин, а также его эмпирическая основа. В зависимости от используемой модели такой подход можно рассматривать как частный случай соответствия целевому критерию.

8.4.2 Все средние стандартные отклонения, выбранные с помощью общей модели, должны быть обоснованы. Если результаты очень большой или очень маленькой доли участников попали в зону сигнала к действиям и сигнала предупреждения, провайдер проверки квалификации должен гарантировать, что это согласуется с целью программы проверки квалификации.

8.4.3 Конкретная оценка, учитывающая особенности проблемы измерений, как правило, предпочтительнее подхода на основе общей модели. Следовательно, перед использованием общей модели следует изучить возможность применения подходов, приведенных в 8.2, 8.3 и 8.5.

Пример — Кривая Хорвица.

Общая модель для применения при проведении химических исследований описана Хорвицем [8] и модифицирована Томпсоном [7]. Такой подход дает общую модель воспроизводимости аналитических методов, которая может быть использована для вывода выражения для стандартного отклонения воспроизводимости:

$$\sigma_R = \begin{cases} 0{,}22c, & \text{если } c < 1{,}2 \cdot 10^{-7}, \\ 0{,}02c^{0{,}8495}, & \text{если } 1{,}2 \cdot 10^{-7} \le c \le 0{,}138, \\ 0{,}01c^{0{,}5}, & \text{если } c > 0{,}138, \end{cases} \quad (8)$$

где c — массовая доля химических соединений, которые необходимо определить, $0 \le c \le 1$.

Примечание 1 — Модель Хорвица является эмпирической, основанной на наблюдениях в течение длительного периода времени в совместных испытаниях с большим количеством параметров. Значения σ_R являются средними верхними границами межлабораторной изменчивости в совместных испытаниях в тех случаях, когда совместные испытания не имеют значимых проблем. Следовательно, значение σ_R не может быть походящим критерием при определении квалификации в программе проверки квалификации.

Примечание 2 — Пример определения значения на основе модифицированной модели Хорвица приведен в Е.9 приложения Е.

8.5 Использование стандартного отклонения повторяемости и воспроизводимости на основе данных предварительных совместных исследований прецизионности метода измерений

8.5.1 Если метод измерений, используемый в программе проверки квалификации, стандартизован, и информация о стандартном отклонении повторяемости σ_r и стандартном отклонении воспроизводимости σ_R метода доступна, стандартное отклонение для оценки квалификации σ_{pt} может быть вычислено с использованием следующего выражения:

$$\sigma_{pt} = \sqrt{\sigma_R^2 - \sigma_r^2(1 - 1/m)}, \quad (9)$$

где m — количество репликаций измерений, которые каждый участник должен выполнить в раунде программы проверки квалификации.

Примечание — Данное выражение получено из основной модели случайных воздействий в соответствии с *ГОСТ Р ИСО 5725-2*.

8.5.2 Если стандартные отклонения повторяемости и воспроизводимости зависят от среднего арифметического результатов испытаний, функциональные зависимости должны быть получены с помощью методов, приведенных в *ГОСТ Р ИСО 5725-2*. Эти зависимости должны быть использованы для расчета значений стандартных отклонений повторяемости и воспроизводимости соответствующего приписанного значения, используемого в программе проверки квалификации.

8.5.3 Для валидации вышеуказанных методов должно быть проведено совместное исследование в соответствии с требованиями *ГОСТ Р ИСО 5725-2* или аналогичного документа.

Примечание — Пример представлен в Е.10 приложения Е.

8.6 Использование данных, полученных в том же раунде программы проверки квалификации

8.6.1 В данном подходе стандартное отклонение для оценки квалификации σ_{pt} вычисляют по результатам участников, полученных в одном и том же раунде программы проверки квалификации.

При использовании такого подхода, как правило, удобнее всего использовать индексы, характеризующие работу лаборатории, такие как *z*-индексы. Как правило, для расчета σ_{pt} используют робастную оценку стандартного отклонения результатов, представленных всеми участниками, которая получена с использованием методов, перечисленных в приложении С. В общем случае оценки с *D* или *D* % и использующие δ_E не подходят в таких ситуациях, однако P_A может быть применима в качестве стандартизованного индекса для сопоставления со всеми измеряемыми величинами (см. 9.3.6).

8.6.2 Использование результатов участников может приводить к определению неподходящих критериев оценки функционирования лаборатории. Провайдер проверки квалификации должен обеспечивать, чтобы σ_{pt}, используемое для оценки работы лаборатории, соответствовало целям проверки квалификации.

8.6.2.1 Провайдер проверки квалификации должен установить ограничение на наименьшее значение σ_{pt} для случая, когда робастное стандартное отклонение очень мало. Это ограничение должно быть выбрано таким образом, чтобы, когда погрешность измерений является удовлетворительной в подавляющем количестве случаев, *z*-индекс имел значение менее 3,0.

***Пример** — В программе проверки квалификации лаборатория определяет плотность ткани, измеряемой величиной является количество нитей на 1 см ткани. Робастное стандартное отклонение в некоторых раундах может быть небольшим (менее 1 см ткани), а погрешность менее четырех нитей на сантиметр считают незначительной. Провайдер определяет, что робастное стандартное отклонение используют как σ_{pt}, в этом случае σ_{pt} менее 1,3 нити на 1 см, поэтому σ_{pt} = 1,3.*

8.6.2.2 Провайдер проверки квалификации должен установить ограничение на наибольшее используемое значение σ_{pt} или на результаты измерений, которые следует оценивать как приемлемые (нет сигнала), в том случае, когда значение робастного стандартного отклонения очень большое. Это ограничение должно быть выбрано таким образом, чтобы результаты, которые не соответствуют целям, воспринимались как сигнал к действиям.

8.6.2.3 Иногда провайдер проверки квалификации может устанавливать верхнюю и нижнюю границы на результаты, которые могут быть оценены как приемлемые (нет сигнала предупреждения или сигнала к действиям), когда симметричные интервалы включают результаты, не соответствующие установленным целям.

***Пример** — При определении качества не питьевой (технической) воды в схеме проверки квалификации лабораторий установлено, что результаты должны находиться в пределах ± 3σ_{pt} от робастного среднего результатов участников. Однако поскольку в некоторых случаях размах приемлемых результатов может включать 0 мг/дм^3, все результаты менее 10 % установленного значения должны вызывать сигнал к действиям (являются неприемлемыми). Образец для проверки квалификации составляет 4,0 мг/дм^3 регламентированного вещества. Робастное среднее участников составляет 3,2 мг/дм^3 и σ_{pt} = 1,1 мг/дм^3. Следовательно, возможно представление участником результата 0,0 мг/дм^3, который находится в пределах ± 3σ_{pt}, но все результаты менее 0,4 мг/дм будут признаны как неприемлемые.*

8.6.3 Основными преимуществами такого подхода являются простота и успешное применение во многих ситуациях. Иногда он может быть единственно возможным.

8.6.4 Существует несколько недостатков данного подхода:

a) значение σ_{pt} может существенно меняться при переходе от раунда к раунду программы проверки квалификации, что затрудняет участникам использование значения *z*-индекса для выявления трендов, которые сохраняются в течение нескольких раундов;

b) стандартные отклонения могут быть недостоверными, когда количество участников программы проверки квалификации мало или когда объединяют результаты различных методов измерений. Например, если *p* = 20, стандартное отклонение для данных из нормального распределения может изменяться на ± 30 % от его истинного значения от одного раунда программы проверки квалификации к другому;

c) использование мер рассеяния, полученных на основе данных, может приводить к приблизительно постоянной пропорции приемлемых индексов. Очевидно, что в этом случае плохое функционирование лаборатории может быть не обнаружено с помощью индексов, а в целом хорошие показатели могут стать причиной того, что хороший участник получит плохую оценку;

d) отсутствие полезных интерпретаций относительно соответствия конечному использованию результатов.

Примечание — Примеры использования данных участников приведены в комплексном примере (см. Е.3 приложения Е).

8.7 Мониторинг межлабораторных соглашений

8.7.1 Для проверки работы участников, а также оценки преимуществ программы проверки квалификации для участников, провайдеру проверки квалификации следует применять процедуру мониторинга межлабораторных соглашений, чтобы отслеживать изменения в работе лабораторий и обеспечивать обоснованность применения статистических процедур.

8.7.2 Результаты, полученные в каждом раунде программы проверки квалификации, необходимо использовать для расчета оценок стандартных отклонений воспроизводимости метода измерений (и повторяемости, при необходимости), используя робастные методы, описанные в приложении С. Эти оценки необходимо наносить на график последовательно или в виде временных рядов вместе со значениями стандартного отклонения повторяемости и воспроизводимости, полученными при исследовании прецизионности в соответствии с *ГОСТ Р ИСО 5725-2* (при наличии) и/или σ_{pt}, если использованы методы, приведенные в 8.2—8.4.

8.7.3 Полученные графики должны быть исследованы провайдером проверки квалификации. Если графики показывают, что прецизионность значений, полученных в конкретном раунде проверки квалификации, в два или более раз больше значений, ожидаемых на основе ранее полученных данных или из опыта, то провайдер проверки квалификации должен исследовать причину того, что согласование в данном раунде хуже, чем в предыдущем. Аналогично тенденция к улучшению или ухудшению значения прецизионности становится основанием для анализа наиболее вероятных причин этого явления.

9 Вычисления статистик функционирования лабораторий

9.1 Общие положения при определении качества работы лабораторий

9.1.1 Статистики, используемые при определении качества работы лабораторий, должны соответствовать цели(ям) программы проверки квалификации.

Примечание — Статистики функционирования более полезны, если участники и заинтересованные стороны понимают их структуру и вывод.

9.1.2 Индексы функционирования должны быть легко анализируемыми для всех уровней измеряемой величины и различных раундов программы проверки квалификации.

9.1.3 Результаты участников должны быть определены и проанализированы относительно выполнения предположений, используемых при разработке программы проверки квалификации, чтобы обеспечить значимость статистики функционирования. Например, проверить, не существует ли каких-либо доказательств ухудшения образца для проверки квалификации, или не смешаны ли группы участников, или нет ли серьезных нарушений статистических предположений о свойствах данных.

9.1.4 В общем случае нецелесообразно использовать методы оценки, которые намеренно устанавливают фиксированную долю результатов для получения сигнала к действиям.

9.2 Ограничения неопределенности приписанного значения

9.2.1 Если стандартная неопределенность $u(x_{pt})$ приписанного значения велика по сравнению с критерием оценки функционирования лаборатории, то существует риск того, что некоторые участники получат сигнал к действиям и сигнал предупреждения из-за неточностей в определении приписанного значения, а не по причине плохой работы участника. Поэтому стандартная неопределенность приписанного значения должна быть определена заранее и сообщена всем участникам (см. *4.4.5 и 4.8.2 ГОСТ ISO/IEC 17043—2013*).

Если соблюдены приведенные ниже критерии, то неопределенность приписанного значения можно считать пренебрежимо малой и ее можно не учитывать при интерпретации результатов раунда проверки квалификации

$$u(x_{pt}) < 0{,}3\,\sigma_{pt} \quad \text{или} \quad u(x_{pt}) < 0{,}1\,\sigma_E. \tag{10}$$

Примечание — $0{,}3\sigma_{pt} = 0{,}1\,\delta_E$, если $|z| \geq 3{,}0$ — зона сигнала к действиям.

9.2.2 Если этот критерий не выполняется, то провайдер проверки квалификации должен выполнить перечисленные ниже действия, гарантируя при этом, что любые предпринятые действия попрежнему соответствуют согласованной политике в области оценки функционирования лаборатории в программе проверки квалификации:

a) следует выбрать метод определения приписанного значения таким образом, чтобы неопределенность приписанного значения удовлетворяла неравенствам (10);

b) необходимо использовать неопределенность приписанного значения при анализе результатов программы проверки квалификации (см. 9.5 для *z*-индексов, 9,6 для ζ-индексов или 9,7 для E_n-индексов);

c) если приписанное значение получено по результатам участников и из-за различий между подгруппами участников неопределенность приписанного значения слишком велика, следует использовать значения неопределенности для каждой подгруппы участников (например, сгруппировать участников по видам используемых методов измерений).

Примечание — В гармонизированном протоколе IUPAC [1] описана конкретная процедура обнаружения бимодальности, основанная на проверке графика плотности ядра с установленной «шириной полосы пропусканий»;

d) должны информировать участников о том, что неопределенность приписанного значения не является незначительной и это может повлиять на оценку их работы.

Если неприменимо ни одно из перечислений a)—d), то участники должны быть проинформированы, что приписанное значение и индексы функционирования не могут быть определены достоверно.

Примечание — Методы, представленные в данном разделе, приведены в Е.3 и Е.4 приложения Е.

9.3 Оценка отклонения (погрешности измерений)

9.3.1 Пусть x_i представляет результат измерений характеристики свойства образца для проверки квалификации (или среднее арифметическое репликаций измерений), указанных в отчете *i*-го участника в одном из раундов программы проверки квалификации. Тогда простая мера качества работы участника может быть вычислена как разность между результатом x_i и приписанным значением x_{pt}

$$D_i = x_i - x_{pt}. \quad (11)$$

D_i можно интерпретировать как погрешность измерений результатов в той степени, в которой приписанное значение можно рассматривать как обычное или опорное значение количественной величины.

Разность D_i может быть выражена в тех же единицах, что и приписанное значение, или в процентах, вычисленных по формуле

$$D_i\,\% = 100\,(x_i - x_{pt})/x_{pt}\,\%. \quad (12)$$

9.3.2 Разность D или D %, как правило, сопоставляют с критерием δ_E, основанным на соответствии цели или данных предыдущих раундов программы проверки квалификации; критерий, обозначаемый здесь как δ_E, представляет собой припуск к погрешности измерений. Если $\delta_E < D < \delta_E$, то работу лаборатории можно считать приемлемой (сигнал отсутствует). (Тот же критерий применяют для D % в зависимости от выражения для δ_E.)

9.3.3 Значение δ_E тесно связано с σ_{pt}, используемым для расчета z-индексов (9.4), если σ_{pt} определено на основе соответствия цели или ожидания от предыдущих раундов. Соотношение определяется критерием оценки для z-индексов. Например, если $z \geq 3$ определяют как границу сигнала к действиям, то $\delta_E = 3\,\sigma_{pt}$ или эквивалентно $\sigma_{pt} = \delta_E/3$. Различные выражения для δ_E являются обычными при проверке квалификации в области медицины или в функциональных требованиях к методам измерений и продукции.

9.3.4 Преимущества D как статистики и δ_E как критерия функционирования состоят в том, что у участников существует интуитивное понимание этих статистик, так как они непосредственно связаны с погрешностью измерений и являются общими в качестве критериев соответствия цели. Преимущество статистики D % состоит в том, что она является интуитивно понятной, стандартизирована для уровня измеряемой величины и связана с общими причинами погрешности (например, некорректная калибровка или смещение при разбавлении растворов).

9.3.5 Недостатками этой статистики может быть то, что она не является общепринятой при проверке квалификации во многих странах и областях измерений, а также и то, что D не стандартизована, и это не дает возможности простого просмотра отчетов для определения сигналов к действиям в программах проверки квалификации с несколькими аналитами, или в том случае, где критерии соответствия цели могут изменяться в зависимости от уровня измеряемой величины.

Примечание — Использование D и D % обычно предполагает симметричность функции распределения результатов участников в том смысле, что областью приемлемых значений является диапазон — $\delta_E < D < \delta_E$.

9.3.6 Для сравнения различных уровней измеряемой величины в том случае, когда критерий соответствия цели может изменяться, или для объединения раундов или измеряемых величин D и D % могут быть преобразованы в стандартизованный индекс функционирования, который показывает разности с критерием для измеряемых величин. Для этого следует вычислить «процент допустимого отклонения» P_A для каждого результата:

$$P_{Ai} = (D_i/\delta_E) \cdot 100\ \%. \qquad (13)$$

Следовательно, $P_A \geq 100$ % или $P_A \leq -100$ % указывает на сигнал к действиям (неприемлемости функционирования).

Примечание 1 — Можно сопоставлять значения P_A для различных уровней измеряемых величин и различных раундов программы проверки квалификации или отображать их на графиках. Эти оценки функционирования аналогичны по использованию и интерпретации z-индексам, которые имеют общий критерий оценки, такой как $z \leq -3$ или $z \geq 3$ для сигналов к действиям.

Примечание 2 — Часто используют вариации этой статистики, особенно в медицине, где проверку квалификации проводят более часто с использованием большого количества аналитов.

Примечание 3 — Иногда целесообразно использовать абсолютное значение P_A для того, чтобы отразить приемлемость (или неприемлемость) результатов относительно приписанного значения.

9.4 z-индексы

9.4.1 z-индекс для результата проверки квалификации x_i вычисляют по следующей формуле

$$z_i = \frac{(x_i - x_{pt})}{\sigma_{pt}}, \qquad (14)$$

где x_{pt} — приписанное значение;

σ_{pt} — стандартное отклонение для оценки квалификации.

9.4.2 Общепринятая интерпретация z-индекса состоит в следующем (см. *B.4.1.1 приложения B ГОСТ ISO/IEC 17043—2013*):

- результат считают приемлемым, если $|z| \leq 2{,}0$;
- результат находится в зоне предупреждения (сигнал предупреждения), если $2{,}0 < |z| < 3{,}0$;
- результат считают неприемлемым (сигнал к действиям), если $|z| \geq 3{,}0$.

Участникам необходимо проверить процедуры измерений при появлении предупреждающего сигнала, так как он служит признаком появившейся или повторяющейся проблемы.

Примечание 1 — В некоторых случаях в качестве границы зоны сигнала к действиям провайдеры используют z-индекс, равный 2.

Примечание 2 — Выбор в качестве критерия σ_{pt} допускает вышеуказанную интерпретацию. Этот критерий широко используют для оценки квалификации, и такая интерпретация очень похожа на границы контрольных карт.

Примечание 3 — Обоснование для использования границ 2,0 и 3,0 для z-индексов состоит в следующем. Правильно выполненные измерения предполагают получение результатов, которые могут быть описаны (после преобразования при необходимости) распределением со средним x_{pt} и стандартным отклонением σ_{pt}. Тогда z-индексы тоже подчиняются нормальному распределению со средним 0 и стандартным отклонением 1. В этом случае в среднем только около 0,3 % результатов выйдут за пределы интервала $-3{,}0 \leq z \leq 3{,}0$ и только около 5 % результатов выйдут за пределы интервала $-2{,}0 \leq z \leq 2{,}0$. Так как вероятность выхода за границы ± 3,0 очень мала, то вероятность случайного появления сигнала к действиям очень низка, и такой выход означает появление реальных аномалий.

Примечание 4 — Предположения, на которых основана эта интерпретация, относятся только к гипотетическому распределению квалифицируемых лабораторий, но не к распределению наблюдаемых результатов. Не требуется предположений о самих наблюдаемых результатах.

Примечание 5 — Если истинная межлабораторная изменчивость менее σ_{pt}, то вероятность ошибок сокращается.

Примечание 6 — Если стандартное отклонение для оценки квалификации устанавливают в соответствии с 8.2 или 8.4, оно может существенно отличаться от (робастного) стандартного отклонения результатов и доли результатов, выходящих за границы ± 2,0 и 3,0, также могут существенно отличаться от 5 и 0,3 % соответственно.

9.4.3 Провайдер проверки квалификации определяет правила округления для полученных z-индексов на основе количества значащих цифр для результатов приписанного значения и стандартного отклонения для проверки квалификации. Эта информация должна быть доступна всем участникам.

Примечание — Обычно редко используют более двух знаков после запятой.

9.4.4 Если в качестве стандартного отклонения результатов участников используют σ_{pt} и количество участников проверки слишком велико, провайдер проверки квалификации может проверить нормальность распределения, используя реальные результаты или z-индексы. С другой стороны, если количество участников слишком мало, сигнал к действиям может вообще не появиться. В этом случае более полезными являются графические методы, которые дают возможность проследить за индексами функционирования в течение нескольких раундов, что может обеспечить более полезные данные о квалификации участников, чем результаты отдельных раундов.

9.5 z′-индексы

9.5.1 Если появляются опасения по поводу неопределенности приписанного значения $u(x_{pt})$, например если $u(x_{pt}) > 0{,}3\sigma_{pt}$, то эта неопределенность может быть учтена путем добавления ее в знаменатель при расчете z-индекса, который в этом случае называют z′-индексом и вычисляют (в обозначениях 9.4) следующим образом:

$$z'_i = \frac{x_i - x_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}}. \qquad (15)$$

Примечание — Если x_{pt} и/или σ_{pt} вычисляют по результатам участников, то индексы функционирования коррелируют с результатами отдельных участников, потому что эти результаты влияют на робастные среднее и стандартное отклонения. Корреляция для отдельного участника зависит от весового коэффициента, назначенного этому участнику в объединенной статистике. По этой причине индекс функционирования, включающий неопределенность приписанного значения без учета корреляции, дает заниженный индекс и заниженные результаты по сравнению с ситуацией, когда ковариацию учитывают. Например, если $u(x_{pt}) = 0{,}3\ \sigma_{pt}$, то снижение оценки z′-индекса составляет около 10. Таким образом, формулу (15) следует использовать, если x_{pt} и/или σ_{pt} определяют по результатам участников.

9.5.2 Индексы D и D % также могут быть модифицированы для учета неопределенности приписанного значения с помощью следующей формулы, обеспечивающей получение δ'_E:

$$\delta'_E = \sqrt{\delta_E^2 + U^2(x_{pt})}, \qquad (16)$$

где $U(x_{pt})$ — расширенная неопределенность приписанного значения x_{pt}, вычисленная с коэффициентом охвата $k = 2$.

9.5.3 z′-индекс можно интерпретировать тем же способом, что и z-индекс (см. 9.4), и использовать те же критические значения 2,0 и 3,0 в зависимости от программы проверки квалификации. Аналогично D и D % следует затем сравнивать с δ'_E (см. 9.3).

9.5.4 Сравнение формул для расчета индексов z и z′ в соответствии с 9.4 и 9.5 для раунда программы проверки квалификации показывает, что z′-индекс всегда меньше z-индекса. Отношение z′-индекса к z-индексу составляет

$$\frac{\sigma_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}}.$$

В случае введения ограничений на неопределенность приписанного значения в соответствии с 9.2.1 это значение находится в интервале:

$$0{,}96 < \frac{\sigma_{pt}}{\sqrt{\sigma_{pt}^2 + u^2(x_{pt})}} < 1{,}00.$$

В этом случае z′-индекс почти совпадает с z-индексом, и можно сделать вывод, что при оценке функционирования неопределенностью приписанного значения можно пренебречь.

Если требования 9.2.1 на неопределенность приписанного значения не выполнены, то разность значений z′-индекс и z-индекс может быть столь существенной, что когда z-индекс достигает значений

2 или 3, что соответствует «сигналу предупреждения» и «сигналу к действиям», z'-индекс не достигает таких критических значений и, следовательно, не дает никаких сигналов.

В общем случае для ситуаций, когда приписанное значение и/или σ_{pt} не определяют по результатам участников, использование z'-индекса может быть предпочтительнее, поскольку когда критерий в соответствии с 9.2.1 выполнен, разность между z'-индексом и z-индексом пренебрежимо мала.

9.6 Дзета-индекс

9.6.1 Применение дзета-индекса полезно, если целью программы проверки квалификации является оценка способности участника получать результаты, близкие к приписанному значению в пределах указанной участником неопределенности.

С учетом обозначений, использованных в 9.4, ζ-индекс вычисляют по следующей формуле

$$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u^2(x_i) + u^2(x_{pt})}}, \qquad (17)$$

где $u(x_i)$ — собственная оценка лабораторией стандартной неопределенности ее результата x_i;

$u(x_{pt})$ — стандартная неопределенность приписанного значения x_{pt}.

Примечание 1 — Если приписанное значение x_{pt} рассчитывают на основе согласованного значения результатов участников, то x_{pt} коррелирует с результатами участников. Корреляция для отдельного участника зависит от весового коэффициента этого участника в приписанном значении и в меньшей степени от неопределенности приписанного значения. По этой причине оценки функционирования, включая неопределенность приписанного значения, но не включая припуск на корреляцию, являются заниженными оценками, то есть оценками при отсутствии корреляции. Занижение оценки несущественное, если неопределенность приписанного значения мала; при использовании робастных методов для большинства внешних участников опасность получения неблагоприятных значений индекса функционирования очень мала. Следовательно, формула (17) может быть использована с согласованными статистиками без учета корреляции.

Примечание 2 — Дзета-индекс отличается от E_n-индекса (см. 9.7) использованием стандартных неопределенностей $u(x_i)$ и $u(x_{pt})$, а не расширенных неопределенностей $U(x_i)$ и $U(x_{pt})$. Значение ζ-индекса выше 2 или ниже – 2 может быть вызвано систематическим смещением метода или плохой оценкой неопределенности результатов измерений участником. Следовательно, ζ-индекс обеспечивает строгую оценку полного результата, предоставленного участником.

9.6.2 Использование дзета-индексов позволяет осуществлять прямую оценку способности лаборатории обеспечить корректные результаты, то есть результаты, согласованные с x_{pt} в пределах неопределенности их результатов измерений. Дзета-индексы могут быть интерпретированы с использованием тех же значений 2,0 и 3,0, как и z-индексы, или с умножением на коэффициент охвата, используемым при оценке расширенной неопределенности. Однако неблагоприятный дзета-индекс может указывать либо на большое отклонение x_i от x_{pt}, либо на недооценку неопределенности участником, или на то и другое одновременно.

Примечание — Для провайдера полезно получить дополнительную информацию о достоверности зафиксированных в отчете неопределенностей. Полезные указания по такой оценке приведены в 9.8.

9.6.3 Дзета-индексы могут быть использованы в сочетании z-индексами как дополнительное средство для улучшения работы участников. Если участник получает z-индекс, многократно превышающий критическое значение 3,0, необходимо проанализировать методику исследований шаг за шагом и получить для нее оценку неопределенности. Оценка неопределенности позволит идентифицировать те шаги в процессе измерений, в которых появляется большая неопределенность, и участники могут увидеть, где необходимо затратить усилия, чтобы добиться улучшений. Если дзета-индексы участника также повторно превышают критическое значение 3,0, это означает, что оценка неопределенности участника не включает в себя все существенные источники неопределенности (то есть пропущено что-то важное). И наоборот, если участник повторно получает дзета-индекс более или равный трем, а дзета-индекс — менее двух, это указывает на то, что участник точно оценивает неопределенность своих результатов, но его результаты не соответствуют ожидаемому уровню работы для программы проверки квалификации. Это может быть в случае, например, если участник использует метод скрининга в процессе измерений, а другие участники применяют количественные методы. Если участник считает, что неопределенность его результатов обоснована, то никаких действий проводить не требуется.

Примечание — Если дзета-индекс использован самостоятельно, он может быть интерпретирован только для проверки соответствия неопределенности результатов участника конкретному наблюдаемому отклонению и не может быть использован как признак соответствия результатов конкретного участника целям испытаний.

Определение соответствия целям может быть сделано отдельно (например, участником или органом по аккредитации) с помощью проверки отклонений разности ($x - x_{pt}$) или объединенных стандартных неопределенностей по сравнению с целевой неопределенностью.

9.7 E_n-индексы

9.7.1 E_n-индексы могут быть использованы в том случае, если целью программы проверки квалификации является оценка способности участников получать результаты, близкие к приписанному значению в пределах заявленной ими расширенной неопределенности. Эта статистика является обычной для проверки квалификации при калибровке, но может быть использована и для других видов проверки квалификации. Эту статистику вычисляют по формуле

$$(E_n)_i = \frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}}, \qquad (18)$$

где x_{pt} — приписанное значение, определенное в эталонной лаборатории;
$U(x_i)$ — расширенная неопределенность результатов участника x_i;
$U(x_{pt})$ — расширенная неопределенность приписанного значения x_{pt}.

Примечание — Непосредственное объединение расширенных неопределенностей не соответствует требованиям [6] и не эквивалентно расчету объединенной расширенной неопределенности, за исключением случая, когда коэффициент охвата и число эффективных степеней свободы одинаковы для $U(x_i)$ и $U(x_{pt})$.

9.7.2 E_n-индекс следует интерпретировать с осторожностью, поскольку он является отношением двух отдельных (независимых) показателей функционирования. Числитель представляет собой отклонение результата от приписанного значения (рассмотрение интерпретации этого показателя приведено в 9.3). Если участник правильно определил $U(x_i)$ и провайдер проверки квалификации правильно определил $U(x_{pt})$, знаменатель представляет собой объединенную расширенную неопределенность, которая не должна быть больше отклонений числителя. Таким образом, значения $E_n \geq 1{,}0$ или $E_n \leq -1{,}0$ могут указывать на необходимость анализа оценок неопределенности или на необходимость коррекции выполнения измерений; аналогично $-1{,}0 < E_n < 1{,}0$ следует рассматривать как признак успешной работы только в том случае, если неопределенности валидированы и отклонение ($x_i - x_{pt}$) меньше, чем необходимо заказчику участника.

Примечание — Несмотря на то что интерпретация E_n-индексов может быть достаточно сложной, это не препятствует их использованию. Включение информации о неопределенности в интерпретацию результатов проверки квалификации может играть важную роль в улучшении понимания участниками неопределенности измерений и ее оценки.

9.8 Оценки неопределенности результатов участников

9.8.1 Применение *ГОСТ ИСО/МЭК 17025* обеспечивает более глубокое понимание неопределенности измерений. Использование оценок неопределенности при оценке функционирования лабораторий распространено в программах проверки квалификации в различных областях калибровки (так же как E_n-индексов, но при проверке квалификации испытательных лабораторий не применялось). ζ-индексы, описанные в 9.6, и E_n-индексы, описанные в 9.7, представляют собой варианты оценки результатов по отношению к заявленной неопределенности.

9.8.2 Некоторые провайдеры проверки квалификации признали, что указание лабораториями в отчете неопределенности результатов при проверке квалификации является полезным. Это может быть полезно даже тогда, когда неопределенности не используют в расчетах. Существуют несколько целей сбора такой информации:

a) органы по аккредитации могут гарантировать, что участники укажут неопределенности, которые соответствуют их области аккредитации;

b) участники могут проанализировать свою неопределенность по отношению к другим участникам и получить возможность определить, учтены ли в их оценке неопределенности все возможные составляющие неопределенности и не завышена ли неопределенность некоторых составляющих;

c) проверка квалификации может быть использована для подтверждения заявленной неопределенности, и это легче сделать, когда неопределенность указана вместе с результатом.

Примечание — Пример анализа данных, когда неопределенности указаны, приведен в Е.3 приложения Е.

9.8.3 Если x_{pt} определено в соответствии с процедурами, установленными в 7.3—7.6, и $u(x_{pt})$ соответствует критерию 9.2.1, то маловероятно, что результат участника будет иметь меньшую стандартную неопределенность, тогда $u(x_{pt})$ может быть использована в качестве нижнего предела при скрининге, то есть как u_{min}. Если приписанное значение определяют по результатам участников (см. 7.7), то провайдер проверки квалификации должен определить границы реального скрининга для u_{min}.

Примечание — Если $u(x_{pt})$ включает изменчивость вследствие неоднородности или нестабильности, $u(x_i)$ участников может быть меньше u_{min}.

9.8.4 Маловероятно также, что все участники указали стандартную неопределенность, более чем в 1,5 раза превышающую робастное стандартное отклонение участников 1,5 s^*, это значение может быть использовано как реальная верхняя граница для скрининга зафиксированных неопределенностей, называемая u_{max}.

Примечание — Коэффициент 1,5 является верхним пределом изменчивости стандартных отклонений, который можно ожидать для согласованного стандартного отклонения десяти или более результатов на основе квадратного корня процентили *F*-распределения. Провайдер проверки квалификации может использовать и другое значение этого коэффициента.

9.8.5 Если для определения отклоняющих неопределенностей используют u_{min} или u_{max} или другие критерии, провайдер проверки квалификации должен объяснить участникам, что указанная ими неопределенность $u(x_i)$ должна быть валидирована, даже если она менее u_{min} или более u_{max}. Если это происходит, участники и все заинтересованные стороны должны проверить результат или оценку неопределенности. Аналогично, если указанная неопределенность больше u_{min} и меньше u_{max}, это не обеспечивает ее валидацию. Это всего лишь индикатор.

9.8.6 Провайдеры проверки квалификации могут также обратить внимание на необычно высокие или низкие значения неопределенности на основе:

- установленных квантилей, соответствующих указанным неопределенностям (например, ниже 5-й процентили и выше 95-й процентили для указанных стандартных или расширенных неопределенностей);
- границ, построенных на предполагаемом распределении, со шкалой, созданной на разбросе указанных неопределенностей;
- требуемой неопределенности результатов измерений.

Примечание — Поскольку маловероятно, что неопределенности подчиняются нормальному распределению, будет необходимо выполнение преобразования при использовании границ, полученных на основе приближенно нормального или нормального распределения, например, границы на диаграмме «ящик с усами» основаны на межквартильном размахе и имеют вероятностную интерпретацию только тогда, когда распределение является приближенно нормальным.

9.9 Комбинированные индексы функционирования

9.9.1 Обычно в пределах одного раунда программы проверки квалификации должны быть получены результаты для нескольких образцов или нескольких измеряемых величин. В этом случае результаты для каждого образца проверки квалификации и каждой измеряемой величины необходимо интерпретировать в соответствии с 9.3—9.7; то есть результаты для каждого образца и каждой измеряемой величины следует рассматривать отдельно.

9.9.2 Существуют случаи, когда в программу проверки квалификации включены два или более образцов со специально разработанными уровнями для измерения других аспектов работы лаборатории, например, для исследования повторяемости, систематической погрешности или линейности модели. Например, два аналогичных образца могут быть использованы в программе проверки квалификации с применением графика Юдена (см. 10.5). В такой ситуации провайдер проверки квалификации должен предоставить участникам полное описание плана эксперимента и используемых процедур.

9.9.3 Рекомендуется использовать графические методы, описанные в 10, если результаты получены по нескольким исследуемым образцам или нескольким измеряемым величинам, при условии, что они тесно связаны между собой и/или получены одним тем же методом. Значения показателей объединяют

подобного рода способами, которые не скрывают информацию о высоких значениях отдельных показателей. Таким образом, может быть получена дополнительная информация о работе лабораторий, например, как корреляция между результатами для различных измеряемых величин, не очевидная по данным таблиц для отдельных показателей.

9.9.4 В схемах проверки квалификации, в которых использовано большое количество измеряемых величин для оценки качества работы, могут быть применены подсчет или соотношение количества сигналов к действиям и предупреждения.

9.9.5 Комбинированные индексы функционирования либо поощрительные или штрафные индексы следует использовать с осторожностью, так как может быть трудно описать лежащие в основе статистические предположения. В то же время комбинированные индексы функционирования в случае использования нескольких образцов и единственной измеряемой величины могут иметь ожидаемые распределения и быть полезны для выявления постоянного смещения, усреднение или суммирование индексов для нескольких измеряемых величин и одних и тех же или разных образцов может замаскировать смещение результатов для измеряемой величины. Таким образом, метод расчета, интерпретации и ограничений всех комбинированных или штрафных индексов должен быть понятен участникам.

10 Графические методы описания индексов функционирования

10.1 Применение графических методов

Для подготовки графиков в соответствии с 10.2 и 10.3 провайдер обычно использует индексы, полученные в каждом раунде программы проверки квалификации. Использование индексов функционирования $P_Д$, z, z', ξ и E_n в таких графиках дает преимущество использования стандартизованных осей, что позволяет упростить их представление и интерпретацию. Графики должны быть доступны участникам, предоставляя каждому возможность видеть расположение собственных результатов по отношению к результатам других участников. При этом могут быть использованы буквенные и числовые коды так, чтобы каждый участник имел возможность идентифицировать свои результаты, но не мог идентифицировать результаты других участников. Графики могут быть использованы провайдером или органом по аккредитации для того, чтобы делать выводы об общей эффективности программы проверки квалификации и определения необходимости анализа критерия оценки квалификации.

10.2 Гистограммы результатов или индексов функционирования

10.2.1 Гистограмма представляет собой общий статистический прием, полезный при проведении анализа результатов проверки квалификации с двух точек зрения. Во-первых, график полезен на предварительном этапе анализа для проверки обоснованности статистических предположений или при наличии отклонений, которые невозможно предвидеть, таких как бимодальное распределение, значительная доля выбросов или необычная асимметрия.

Использование гистограмм в отчетах участников полезно для программ проверки квалификации с небольшим или средним количеством участников (менее 100), что дает участникам возможность сопоставить свою работу с результатами других участников: например, с помощью выделения блока данных в вертикальных столбцах, представляющих результаты участников, или (при небольшой группе участников) используя индивидуальные характеристики для каждого участника.

10.2.2 При построении гистограмм могут быть использованы результаты участников или индексы функционирования. Использование результатов участников имеет преимущество, состоящее в том, что они непосредственно связаны с представленными данными и могут быть оценены без дальнейших вычислений и преобразований индекса функционирования в погрешность измерений. Преимущество гистограмм, основанных на индексах функционирования, состоит в том, что они связаны с оценкой функционирования: их можно сравнивать по измеряемым величинам и раундам программы проверки квалификации.

Размах и размеры интервалов, используемых для диаграммы, должны быть определены для каждого набора данных на основе изменчивости и количества результатов. Это можно сделать и на основе данных предыдущих проверок квалификации, но в большинстве случаев после первичного анализа разбиение на интервалы необходимо корректировать. Если при построении гистограммы используют индексы функционирования, полезно применить шкалу на основе стандартного отклонения для оценки квалификации и выделить точки сигналов предупреждения и сигналов к действиям.

10.2.3 Масштаб и интервалы на графике следует выбирать таким образом, чтобы бимодальность (если она присутствует) могла быть обнаружена, не создавая ситуаций появления ложных сигналов тревоги, связанных с разрешающей способностью средств измерений или небольшим количеством результатов.

Примечание 1 — Форма гистограммы зависит от выбранной ширины интервалов и положения их границ (при постоянной ширине интервалов эти положения зависят от начальной точки). Если ширина интервалов слишком мала, то гистограмма будет включать много невысоких столбиков (использование слишком больших интервалов), не обеспечивая существенного различия столбиков по высоте вблизи среднего. Изменение ширины интервалов сразу приводит к изменению их высоты, особенно там, где набор данных мал и/или имеется некоторая группировка данных.

Примечание 2 — Примеры гистограмм приведены в Е.3 приложения Е.

10.3 График ядерной плотности

10.3.1 График ядерной плотности представляет сглаженную кривую, описывающую общую форму плотности распределения набора данных. Для определения ядерной плотности каждую точку данных заменяют заданным распределением (как правило, нормальным) с центром в этой точке и стандартным отклонением σ_k; σ_k обычно называют «шириной полосы». Все распределения объединяют и результирующее распределение масштабируют таким образом, чтобы площадь под кривой плотности была равна 1, это позволяет получить оценку плотности, которая представляет собой гладкую кривую.

10.3.2 Для подготовки графика ядерной плотности необходимо выполнить следующие этапы. Предположим, что набор данных X состоит из p значений x_1, x_2, ..., x_p, используемых для построения графика. Обычно это результаты участников, но могут быть индексы функционирования, полученные на основе этих результатов.

i) Выбирают подходящую «ширину полосы» σ_k.

Для этого используют два подходящих способа:

a) для общего контроля устанавливают $\sigma_k = 0{,}9\ s^*/p^{0{,}2}$, где s^* — робастное стандартное отклонение величин x_1, ..., x_p, вычисленное с использованием процедур, приведенных в С.2 или С.3 приложения С;

b) при проверке набора данных с большими модами, которые имеют важное значение для сопоставления с критериями оценки работы, устанавливают $\sigma_k = 0{,}75\sigma_{pt}$ при использовании z-индекса или ζ-индекса или $\sigma_k = 0{,}25\delta_E$ при использовании D и D %.

Примечание 1 — Вариант a) предложен Сильверманом [9], который рекомендует выбирать s^* на основе нормализованного межквартильного интервала n/QR. Другие правила выбора «ширины столбца» (длины интервала для диаграммы), которые обеспечивают аналогичные результаты, предложенные Скоттом [10], сведены к замене множителя 0,9 на 1,06. В [10] описан близкий к оптимальному, но гораздо более сложный метод выбора ширины столбца. На практике визуальные различия незначительны, и выбор зависит от наличия необходимого программного обеспечения.

Примечание 2 — Вариант b) предложен в руководстве IUPAC [1].

ii) Устанавливают на графике значения $q_{\min}$ и $q_{\max}$ так, что $q_{\min} \le \min(x_1, ..., x_p) - 3\sigma_k$, а $q_{\max} \ge \max(x_1, ..., x_p) + 3\sigma_k$.

iii) Выбирают количество точек n_k для построения кривой, обычно $n_k = 200$ достаточно, если нет выбросов в пределах заданного диапазона $(q_{\min}, q_{\max})$.

iv) Вычисляют положение на графике значений q_i от q_1 до q_{n_k}

$$q_i = q_m(i-1)\frac{(q_{n_k} - q_1)}{n_k - 1}. \qquad (19)$$

v) Вычисляют n_k значений h_i от h_1 до h_{n_k}

$$h_i = \frac{1}{p}\sum_{j=1}^{p}\varphi\left(\frac{x_j - q_i}{\sigma_k}\right) \quad i = 1, ..., n_k, \qquad (20)$$

где $\varphi(.)$ — плотность стандартного нормального распределения.

vi) На график наносят пары значений h_i и q_i.

Примечание 1 — Иногда полезно указать на графике отдельные точки данных. Обычно эти точки наносят под кривой плотности в виде вертикальных черточек («коврика»), но можно наносить их прямо на кривую плотности.

Примечание 2 — График плотности лучше всего формировать с помощью программного обеспечения. Приведенные этапы вычислений могут быть выполнены посредством электронных таблиц, если объемы данных невелики. Имеющееся в собственности и в свободном доступе программное обеспечение часто включает построение графиков плотности на основе материалов одинаковой ширины по умолчанию. В более современном программном обеспечении для построения графиков плотности могут быть использованы приведенный алгоритм или вычисления, основанные на методах свертки.

Примечание 3 — Примеры графиков ядерной плотности приведены в Е.3, Е.4 и Е.6 приложения Е.

10.3.3 Форма кривой указывает на распределение, которому принадлежат данные. Моды выглядят в виде пиков, выбросы тоже имеют вид пиков, но удаленных от основного массива данных.

Примечание 1 — График плотности чувствителен к выбранному значению σ_k. Если это значение слишком мало, график покажет много небольших возвышений, что может сделать неясным положение истинной моды.

Примечание 2 — Как и в случае гистограмм при составлении графика плотности лучше использовать массивы среднего или большого объема, так как наборы данных небольшого объема (десять и менее) могут включать небольшие выбросы или очевидные моды, в частности, если в качестве основы для выбора ширины столбца использовано робастное стандартное отклонение.

10.4 Штриховые графики для стандартизованных индексов функционирования

10.4.1 Штриховые графики подходят для представления индексов функционирования большого количества аналогичных характеристик на одном графике. Эти графики применяют в тех ситуациях, когда существуют общие свойства индексов участника: например, если участник имеет несколько высоких значений z-индекса, указывающих на недостаточное качество работы, то есть участник имеет положительное смещение.

10.4.2 Для подготовки штрихового графика собирают значения стандартизованных индексов функционирования и отражают их на графике, как показано на рисунке Е.10 приложения Е, для каждого участника индексы объединяют в одну группу. Другие стандартизованные индексы функционирования, такие как D % и P_A, могут быть нанесены на график для тех же целей.

10.4.3 Если в раунде программы проверки квалификации выполняют репликации измерений, результаты могут быть использованы для расчета и составления графика меры прецизионности, например k-статистики в соответствии с *ГОСТ Р ИСО 5725-2* или с масштабированной мерой стандартного отклонения робастного среднего, так как это определено в алгоритме S (см. С.4 приложения С).

Примечание — Пример штрихового графика для z-индексов приведен в Е.11 приложения Е.

10.5 График Юдена

10.5.1 Если в раунде проверки квалификации были проверены два аналогичных образца, то график Юдена обеспечивает очень информативный графический метод исследования результатов. Этот график может быть полезен для демонстрации коррелированности (или независимости) результатов на различных образцах исследования причин появления сигналов к действиям.

10.5.2 На график наносят результаты участника или z-индексы, полученные на одном образце проверки квалификации, вместе с результатами, или z-индексы, полученные на другом образце. Для облегчения интерпретации на график наносят вертикальную и горизонтальную линии, которые делят плоскость на четыре квадранта. Линии проходят через приписанные значения или медианы для двух распределений результатов, или через 0 при использовании z-индексов.

Примечание — Для правильной интерпретации графика Юдена важно, чтобы два используемых в проверке образца имели аналогичные (или идентичные) уровни измеряемой величины; это объясняется тем, что природа систематической погрешности измерений одна и та же в интервале измерений. Графики Юдена могут быть полезны для различных уровней измеряемой величины при наличии последовательной систематической погрешности, но могут ввести в заблуждение, если ошибка калибровки не является последовательно положительной или отрицательной по всей области уровней измеряемой величины.

10.5.3 После построения графика Юдена проводят его анализ. Для этого проверяют график:

а) на наличие точек, которые отличаются от остальных данных. Если участник некорректно использует метод испытаний, то его результаты имеют систематическую погрешность, и точки находятся далеко от остальных, в нижнем левом или верхнем правом квадранте. Точки, отстоящие далеко от остальных и находящиеся в верхнем левом и нижнем правом квадрантах, представляют участников, у которых повторяемость выше, чем у большинства участников; методы измерений показывают

различную чувствительность к компонентам образцов или то (иногда), что участники случайно перепутали образцы;

b) предмет выявления признаков взаимосвязи между результатами образцов для проверки квалификации (например, все точки находятся приблизительно вокруг наклонной линии). Если признаки взаимосвязи существуют, это означает, что у участников существуют смещения, на которые одинаково воздействуют образцы. Если же никакой зависимости между результатами не наблюдается (все точки расположены внутри окружности, обычно с более высокой плотностью в центре), то погрешности измерений для двух образцов в значительной степени независимы. Если визуального анализа недостаточно, зависимость можно проверить с помощью статистики ранговой корреляции;

c) участников точек, расположенных по диагонали или другим упорядоченным способом, для закрытых групп. Наличие таких точек указывает на различия в методах.

Примечание 1 — Если в исследованиях все участники используют один и тот же метод или графики результатов получены по единственному методу, а результаты лежат вдоль прямой линии, это означает, что метод измерений не был установлен должным образом. Исследование метода испытаний может позволить улучшить общую воспроизводимость метода.

Примечание 2 — Пример графика Юдена приведен в Е.12 приложения Е.

10.6 Графики стандартных отклонений повторяемости

10.6.1 Если участниками раунда программы проверки квалификации выполнены репликации измерений, результаты могут быть использованы для построения графика идентификации всех участников, у которых выборочное среднее и стандартное отклонения существенно отличаются от остальных.

10.6.2 На графике изображают внутрилабораторное стандартное отклонение s_i для каждого участника в зависимости от соответствующего выборочного среднего x_i участника. В качестве альтернативы вместо стандартного отклонения может быть использован размах репликаций.

Пусть

x^* — робастное среднее значений x_1, x_2, ..., x_p, вычисленное в соответствии с алгоритмом А;

w^* — робастное объединенное значение величин s_1, s_2, ..., s_p, вычисленное в соответствии с алгоритмом S.

Предположим, что данные подчиняются нормальному распределению. Для нулевой гипотезы (результаты лабораторий относятся к одной генеральной совокупности, или не существует различий между лабораториями по внутрилабораторным средним или стандартным отклонениям) статистика

$$\left(\sqrt{m}\,\frac{x_i - x^*}{w^*}\right)^2 + \left(\sqrt{2(m-1)}\,\ln\left(\frac{s_i}{w^*}\right)\right)^2. \qquad (21)$$

имеет распределение, близкое к χ^2-распределению с двумя степенями свободы. Следовательно, критическую область с уровнем значимости приблизительно 1 % можно изобразить на графике, откладывая

$$s = w^* \exp\left\{\pm\frac{1}{\sqrt{2(m-1)}}\sqrt{\chi^2_{2;0,99} - \left(\sqrt{m}\,\frac{x - x^*}{w^*}\right)^2}\right\} \qquad (22)$$

по оси стандартного отклонения в точках на оси среднего x из интервала

$$\text{от } x = x^* - w^*\sqrt{\frac{\chi^2_{2;0,99}}{m}} \quad \text{до } x = x^* + w^*\sqrt{\frac{\chi^2_{2;0,99}}{m}}. \qquad (23)$$

Примечание — Данная процедура основана на методике, предложенной Нуландом в [11]. Для распределения стандартного отклонения в методе Нуланда использовано обычное нормальное распределение, что может образовать критическую область, содержащую отрицательные стандартные отклонения. Приведенный метод использует приближение для распределения стандартного отклонения, которое исключает эту возможность, но критическая область уже не является эллипсом, как в оригинале. Кроме того, для центральной точки использованы робастные значения вместо простых выборочных средних, как в оригинале.

10.6.3 График может выявить участников, в результатах которых смещение в условиях повторяемости слишком велико. Если количество репликаций достаточно велико, этот метод может также

идентифицировать участников, у которых повторяемость очень мала. Однако поскольку количество репликаций, как правило, невелико, интерпретация подобных случаев затруднена.

Примечание — Пример графика стандартных отклонений повторяемости приведен в Е.13 приложения Е.

10.7 Разделенные пробы

10.7.1 Разделенные пробы используют при необходимости для выполнения детального сравнения двух участников или в том случае, когда проверка квалификации невозможна и требуется внешняя верификация. Получают пробы нескольких материалов, представляющих широкую амплитуду исследуемых свойств. Каждую пробу разбивают на две части, и каждая лаборатория выполняет несколько (не менее двух) репликаций измерений на каждой пробе.

В том случае, если количество участников более двух, одну из лабораторий следует рассматривать как экспертную, а другие лаборатории сопоставляют свои результаты с ее результатами описанным ниже методом.

Примечание 1 — Этот метод исследования является общим, иногда его называют по-другому — «метод парных сравнений» или «метод двусторонних сравнений».

Примечание 2 — План эксперимента разделенных проб не следует путать с планом эксперимента с разделенным уровнем, описанным в стандартах серии *ГОСТ Р ИСО 5725*, когда участникам поставляют два образца, немного отличающихся по уровням.

10.7.2 Данные эксперимента с разделенной пробой следует использовать для построения графиков, отображающих различия между репликациями измерений, для каждого из двух участников и средними результатами участников для каждой пробы. Двумерные графики с использованием полного диапазона концентраций могут иметь такую шкалу, которая затруднит идентификацию важных различий между участниками, в этом случае более полезными могут быть графики различий или процентов различий между результатами двух участников. Дальнейший анализ зависит от сделанных выводов.

10.8 Графические методы объединения индексов после нескольких раундов программы проверки квалификации

10.8.1 При необходимости объединения стандартизованных индексов функционирования по результатам нескольких раундов программы проверки квалификации провайдер проверки квалификации может рассмотреть подготовку графиков в соответствии с 10.8.2 или 10.8.3. Использование графиков, в которых индексы функционирования для нескольких раундов программы проверки квалификации объединены, может допускать наличие трендов и других особенностей результатов, которые необходимо идентифицировать и которые не могут быть выявлены при исследовании индексов в каждом раунде отдельно.

Примечание — При использовании «текущих индексов» или «накопленных индексов», в которых объединены индексы функционирования, полученные участником в нескольких раундах программы проверки квалификации, индексы следует отобразить на графике. Участник может иметь несоответствие, которое обнаруживается при работе с образцом для проверки квалификации в одном раунде, но не обнаруживается в других раундах; «бегущий индекс» может скрыть это несоответствие. Тем не менее в некоторых случаях (например, при частом повторении раундов) «сглаживание» случайных выбросов индексов может быть полезным для более наглядной демонстрации основных показателей.

10.8.2 Контрольная карта Шухарта является эффективным методом идентификации проблем, вызывающих большие отклонения значений *z*-индекса. Рекомендации по построению карт Шухарта и правила построения границ действия приведены в *ГОСТ Р ИСО 7870-2.*

10.8.2.1 Для подготовки карты Шухарта стандартизованные индексы, такие как *z*-индексы или P_A-индексы, для участника наносят на карту в виде отдельных точек вместе с границами зон предупреждения и действия в соответствии с программой проверки квалификации. Если в каждом раунде измеряют несколько характеристик, индексы для различных характеристик могут быть представлены на одном графике, но точки для различных характеристик должны быть нанесены с использованием различных символов и/или различных цветов. Если в один раунд проверки включено несколько образцов проверки квалификации, индексы функционирования могут быть нанесены в виде нескольких точек в каждый момент времени. В этом случае на график можно добавить линию, соединяющую средние индексы в каждый момент времени.

10.8.2.2 Обычно правило интерпретации контрольной карты Шухарта состоит в том, что наблюдаемое значение считают выходящим за установленные границы зоны сигнала к действиям, если:

a) единственная точка выходит за пределы зоны сигнала к действиям (± 3,0 для z-индексов или 100 % для P_A);

b) две из трех последовательных точек лежат вне зоны сигнала предупреждения (± 2,0 для z-индексов или 70 % для P_A);

c) шесть последовательных результатов либо все положительны, либо все отрицательны.

10.8.2.3 Если контрольная карта Шухарта показывает, что наблюдаемая характеристика выходит за установленные границы, участник должен исследовать возможные причины этого явления.

Примечание — Стандартное отклонение для оценки квалификации σ_{pt} не обязательно является стандартным отклонением разностей ($x_i - x_{pt}$), так что уровни вероятностей, которые обычно соответствуют границам зон предупреждения и действия карт Шухарта, не могут быть применены.

10.8.3 Если уровень характеристики свойства изменяется от одного раунда программы проверки квалификации к другому, графики стандартизованных индексов функционирования, таких как z-индексы или P_A-индексы, в зависимости от приписанного значения будут это отражать, если смещение результатов участника изменяется вместе с уровнем. Если более чем один образец включается в один и тот же раунд, индексы функционирования могут быть представлены на графике независимо.

Примечание 1 — Может быть полезным отмечать на карте результаты текущего раунда другим символом или цветом, чтобы их отличать от предыдущих раундов.

Примечание 2 — Пример такой карты с использованием P_A-индекса приведен в Е.14 приложения Е. Такую карту легко использовать и для z-индекса. Только изменения надо наносить по вертикальной шкале.

11 Планирование эксперимента и анализ программ проверки квалификации для качественных показателей (включая номинальные и порядковые свойства)

11.1 Вид качественных данных

Довольно часто при проверке квалификации используют свойства, которые идентифицируют по качественной шкале. Среди видов качественных данных различают программы проверки квалификации, в которых:

- результаты фиксируют по категориальной шкале (иногда называемой «номинальной шкалой»), где характеристика свойства не имеет числового значения (например, тип вещества или организма);
- фиксируют наличие или отсутствие свойства, определяемого по субъективным критериям или значению сигнала при выполнении процедуры измерений. Эту ситуацию можно рассматривать как частный случай категориальной или порядковой шкалы с двумя значениями (также называемой «дихотомической» или «бинарной»);
- результаты фиксируют по порядковой шкале, эти результаты могут быть упорядочены, но для них арифметические соотношения неприменимы. Например, порядковая шкала может иметь значения «высокий», «средний», «низкий».

Такие программы проверки квалификации требуют особого внимания на этапах планирования эксперимента, выбора приписанного значения и оценки показателей функционирования (индексов), так как:

- приписанные значения очень часто основаны на мнении экспертов;
- статистическая обработка, предназначенная для непрерывных значений и количественных данных, не применима к качественным данным. Например, при определении результатов по порядковой шкале не имеет смысла применять средние и стандартные отклонения, даже если результаты могут быть ранжированы.

В следующих подразделах приведены рекомендации по планированию, выбору приписанного значения и оценке функционирования для схем проверки квалификации с использованием качественных характеристик.

Примечание — Рекомендации для порядковых данных не применяют к результатам измерений, которые основаны на количественной шкале с дискретными показателями (см. 5.2.2).

11.2 Статистическое планирование эксперимента

11.2.1 Для программ проверки квалификации, в которых мнение экспертов используют при определении приписанного значения или оценке отчетов участников, необходимо собрать комиссию из квалифицированных экспертов и дать ей время для обсуждения и выработки согласованного мнения. Там, где есть необходимость полагаться на мнение отдельных экспертов при выборе индексов или назначении величин, провайдер проверки квалификации должен дополнительно обеспечить оценку и проверку согласованности мнений различных экспертов.

Пример — В программе проверки квалификации в клинике, где для диагностики используют микроскоп, для оценки предметных стекол, предоставляемых участникам, используют экспертное заключение, которое обеспечивает соответствующий клинический диагноз для образцов проверки квалификации. Провайдер проверки квалификации может выбрать и раздать членам экспертной комиссии образцы вслепую (без указания участника) для обеспечения согласованности диагноза или проводить периодические совещания для получения согласованной оценки всех членов экспертной комиссии.

11.2.2 Для программ проверки квалификации, в которых участники фиксируют простые, однозначные, категоризированные или порядковые результаты, провайдеру проверки квалификации следует рассмотреть возможность:

- обеспечения двух или более образцов проверки квалификации в раунде;
- запроса результатов репликаций измерений для каждого образца проверки квалификации в соответствии с количеством установленных заранее репликаций измерений.

Любая из этих стратегий позволяет подсчитывать результат для каждого участника, который может быть использован либо в анализе данных, либо для расчета индексов. Использование двух или более образцов обеспечивает дополнительную информацию об особенностях ошибок, а также позволяет определить более сложные индексы при оценке квалификации.

Пример 1 — В программе проверки квалификации фиксируют наличие или отсутствие загрязняющего вещества, предоставленные образцы содержат некоторый диапазон уровней загрязняющего вещества на каждом уровне его содержания как функцию уровня содержания загрязняющего вещества. Это может быть использовано, например, для предоставления информации участникам о возможности обнаружения выбранным методом испытаний загрязняющего вещества или для получения средней вероятности обнаружения и последующего определения индексов функционирования, которые в свою очередь могут быть распределены среди участников на основе оценок вероятностей конкретных моделей отклика.

Пример 2 — Проверка квалификации для судебно-медицинских исследований часто требует сопоставления образцов на предмет того, получены они из одного и того же источника или из различных источников (например, отпечатки пальцев, ДНК, гильзы от пули, следы и т. д.). В большинстве случаев возможен ответ «не определено». Программа проверки квалификации может включать в себя несколько образцов из различных источников, и участников просят для каждой пары образцов установить, принадлежат ли они одному и тому же источнику, различным источникам или их источник не может быть определен. Это позволяет дать объективные оценки в виде числа (или %) правильных или неправильных заключений или количества правильных решений о соответствии или отклонении. Затем могут быть установлены критерии их функционирования по степени пригодности использования или сложности задачи.

11.2.3 Однородность должна быть подтверждена анализом соответствующей выборки из образцов, каждый из которых должен продемонстрировать ожидаемое свойство. Для некоторых качественных показателей, например наличие или отсутствие чего-либо, может быть возможна проверка однородности с помощью измерений количественных показателей, таких как микробиологический подсчет объектов или спектр поглощения выше заданного порога. В таких ситуациях могут быть целесообразны испытания на однородность или демонстрация всех результатов выше или ниже заданного значения.

11.3 Приписанное значение для качественных показателей

11.3.1 Приписанные значения для образцов могут быть установлены на основе:

a) экспертной оценки;

b) использования стандартных образцов в качестве образцов для проверки квалификации;

c) сведений о происхождении или подготовке образца(ов);

d) использования моды или медианы результатов участника (медиана подходит только для порядковых значений).

Для получения достоверных результатов может быть использован также любой другой способ выбора приписанного значения. Ниже рассмотрена каждая из перечисленных выше стратегий.

Примечание — Эти способы обычно не подходят для обеспечения количественной информации о неопределенности приписанного значения в программах проверки квалификации, использующих качественные показатели. Тем не менее в соответствии с 11.3.2—11.3.5 необходима такая базовая информация о достоверности приписанного значения, чтобы участники могли взвешенно оценить, может ли полученный плохой результат быть связан с ошибкой при установлении приписанного значения.

11.3.2 Значения, присвоенные на основании заключения экспертов, обычно должны базироваться на согласованном мнении всех квалифицированных экспертов. Любое значимое расхождение между членами комиссии должно быть записано в отчете по результатам раунда. Если комиссия не может достичь соглашения в отношении конкретного образца, провайдер может рассмотреть альтернативный способ присвоения приписанного значения из перечисленных в 11.3.1. Если этот способ не подходит, образец не должен быть использован для оценки работы участников.

Примечание — В некоторых случаях приписанное значение может определить единственный эксперт.

11.3.3 Если в качестве образца для проверки квалификации участникам предоставляют стандартный образец в качестве приписанного значения для данного раунда, следует использовать соответствующее опорное значение или сертифицированное значение. Информация, представленная вместе со стандартным образцом, относящаяся к обоснованию правильности установления приписанного значения, должна быть доступна для участников следующих раундов.

Примечание — Ограничения данного подхода приведены в 7.4.1.

11.3.4 Если образцы получены из известного источника, приписанное значение может быть определено на основании информации об источнике получения материала. Провайдер должен сохранять записи о происхождении, транспортировании и обработке используемых материалов. В связи с этим следует соблюдать осторожность для подтверждения загрязнения образцов, которое может привести к неверным результатам участников. Сведения об источнике и/или детали подготовки образцов должны быть доступны участникам после завершения раунда или по запросу, или в качестве части отчета о раунде по проверке квалификации.

Пример — Образцы вина, представленные участникам программы проверки квалификации для проверки его подлинности, могут быть закуплены непосредственно у изготовителя в указанном регионе или через коммерческого поставщика, способного обеспечить свидетельства подлинности.

11.3.4.1 По возможности рекомендуется проводить подтверждающие испытания или измерения, особенно если использование образца может привести к его загрязнению. Например, образец, идентифицированный как экземпляр одного вида микроорганизмов, растений или животных, как правило, должен быть проверен на наличие отклика для других соответствующих видов. Такие испытания должны быть по возможности максимально чувствительными, чтобы гарантировать либо отсутствие загрязняющих видов, либо количественное определение уровня загрязнения.

11.3.4.2 Провайдер должен предоставить информацию о каких-либо обнаруженных загрязнениях или сомнениях о происхождении образца, которые могут поставить под угрозу его использование.

Примечание — Более подробное описание таких образцов выходит за рамки области применения настоящего стандарта.

11.3.5 В качестве приписанного значения для результатов по категориальной или порядковой шкале может быть использована мода (наиболее частое наблюдение), для результатов на порядковой шкале в качестве приписанного значения также может быть использована медиана. При использовании этих статистик в отчет о раунде проверки квалификации следует включать указание доли результатов, использованных при выборе приписанного значения. Нецелесообразно рассчитывать средние значения или стандартные отклонения для результатов проверки квалификации для качественных показателей, в том числе для порядковых значений, поскольку для этих величин неприменимы арифметические операции.

11.3.6 Если приписанные значения определяют на основе измеримых величин (например, наличие или отсутствие), то приписанное значение обычно может быть определено достаточно точно, то есть с низкой неопределенностью. Статистические расчеты при определении неопределенности могут быть применены для уровней измеряемой величины как «неопределенный» или «сомнительный».

11.4 Оценка функционирования и определения индексов для качественных показателей

11.4.1 Оценка работы участников в программе проверки квалификации, использующей качественные показатели, частично зависит от характера требуемого отчета. В некоторых программах проверки квалификации, где требуется значимое количество оценок участников и выводы должны быть тщательно проанализированы и точно сформулированы, отчеты участников могут быть переданы экспертам для общей оценки и направлены для общего учета. С другой стороны, для оценки работы участника иногда достаточно выяснить, насколько точно его результаты совпадают с приписанным значением для соответствующего образца проверки квалификации. Ниже приведены рекомендации относительно оценки функционирования и определения индексов для целого ряда обстоятельств.

11.4.2 Для экспертной оценки отчетов участников требуется один или несколько экспертов для анализа каждого отчета участника для каждого образца и назначения оценки или индекса функционирования. В такой программе проверки квалификации провайдер должен гарантировать:

- что конкретный участник эксперту неизвестен, в частности отчет не должен включать информацию, по которой эксперт может идентифицировать участника;
- анализ, маркировка и оценка функционирования соответствуют ранее установленным критериям и по возможности соответствующим целям:
- положения 11.3.2 в отношении согласованности между экспертами выполнены;
- при необходимости для участника предусмотрена возможность обжалования мнения конкретного эксперта и/или проведения повторного анализа заключений вблизи важных границ функционирования.

11.4.3 При наличии единственного зафиксированного результата качественного показателя на основе приписанного значения могут быть использованы две системы определения индекса:

i) каждый результат считают приемлемым (или успешным), если он точно совпадает с приписанным значением, и неприемлемым, имеющим неблагоприятный индекс функционирования, в противном случае.

Пример — В программе проверки квалификации при определении наличия или отсутствия загрязняющего вещества правильному результату присваивают единицу, а неверному — нуль;

ii) результаты, точно совпадающие с приписанным значением, считают приемлемыми, и им присваивают соответствующий индекс, результатам, не точно совпадающим с приписанным значением, присваивают индекс, который зависит от особенностей несоответствия. Такая система должна присваивать более низкие индексы более хорошей работе для согласования такой системы с другими типами индексов (например, z-индексы, P_A-индексы, ξ и E_n).

Пример 1 — В программе проверки квалификации при оценке клинической патологии провайдер присваивает оценку 0 для абсолютно точной идентификации микробиологического вида, оценку 1 для неверного результата с правильным лечением (например, при идентификации другого микробиологического вида с правильным и аналогичным лечением) и оценку 3 для неверной идентификации, приводящей к неправильному лечению пациента. Эта система оценок, как правило, требует экспертной оценки характера несоответствий, по возможности полученной до присвоения оценки.

Пример 2 — В программе проверки квалификации, в которой возможны шесть ранжированных ответов, соответствующему присваивают индекс 0, и индекс увеличивается на 2 для каждого следующего значения из возможных шести (так, ближайший результат к приписанному значению имеет индекс 2).

Участникам должны быть предоставлены их индивидуальные индексы функционирования для каждого образца. При репликациях наблюдений могут быть представлены суммарные индексы для каждого результата.

11.4.4 Если для каждого образца зафиксировано несколько результатов репликаций или если каждому участнику предоставлено несколько образцов, провайдер может рассчитать и использовать комбинированные индексы функционирования или суммарный индекс. Комбинированные или суммарные индексы функционирования могут быть вычислены, например, в виде:

- суммы всех индексов для всех образцов;
- суммы для каждого назначенного уровня функционирования;
- доли правильных результатов;
- метрического расстояния, определенного на основе разностей результатов и приписанных значений.

Пример — В качестве метрического расстояния иногда используют такую статистику для качественных данных, как коэффициент Гауэра [12]. Она позволяет объединить количественные и качественные данные на основе объединения аналогичных индексов. Для категоризированных или бинарных данных индекс равен 1 в случае точного соответствия категории и 0 в противоположном случае; для порядковых данных индекс равен 1 минус разность рангов, деленная на количество имеющихся рангов; для интервальной шкалы или шкалы отношений индекс равен 1 минус абсолютная величина разности, деленная на наблюдаемый размах всех значений. Эти индексы находятся в интервале от 0 до 1, их суммируют и используют сумму, деленную на количество используемых переменных. Также может быть применен вариант с использованием весовых коэффициентов.

Комбинированные индексы функционирования можно рассматривать как общую оценку функционирования лаборатории. Например, работу можно считать приемлемой при наличии установленной доли (как правило, высокой) правильных результатов, если это соответствует целям программы проверки квалификации.

11.4.5 Для предоставления участникам информации об их работе могут быть использованы графические методы или представлены общие данные в отчете по раунду.

Примечание — Пример анализа порядковых данных приведен в Е.15 приложения Е.

Приложение А
(обязательное)

Обозначения

d — разность измеряемой величины для образца, используемого в испытаниях при проверке квалификации и приписанного значения для *CRM*;

$\bar{d}$ — выборочное среднее разности измеренных величин и приписанного значения для *CRM*;

D — разность результата участника и приписанного значения $x - x_{pt}$;

D % — отношение разности результата участника и приписанного значения к приписанному значению, %;

δ_E — критерий максимально допустимой погрешности для разностей;

δ_{hom} — погрешность, связанная с различиями между образцами для проверки квалификации;

δ_{stab} — погрешность, связанная с нестабильностью в процессе испытаний;

δ_{trans} — погрешность, связанная с нестабильностью условий транспортирования;

E_n — индекс стандартизованной погрешности, который включает неопределенности результатов участников и приписанного значения;

g — количество образцов, используемых при проверке однородности;

m — количество повторных измерений, выполняемых на образце;

p — количество участников раунда программы проверки квалификации;

P_A — доля допустимой погрешности D/δ_E (может быть выражена в процентах);

s_r — оценка стандартного отклонения повторяемости;

s_R — оценка стандартного отклонения воспроизводимости;

s_s — оценка стандартного отклонения изменчивости между образцами;

s^* — робастная оценка стандартного отклонения участника;

$S_{\bar{x}}$ — стандартное отклонение выборочных средних арифметических;

s_w — выборочное или лабораторное стандартное отклонение;

σ_k — стандартное отклонение ширины полосы, используемое для построения графиков ядерной плотности распределения;

σ_L — межлабораторное (между участниками) стандартное отклонение;

σ_{pt} — стандартное отклонение для оценки квалификации;

σ_r — стандартное отклонение повторяемости;

σ_R — стандартное отклонение воспроизводимости;

u_{hom} — стандартная неопределенность, соответствующая различиям между образцами, используемыми при проверке квалификации;

u_{stab} — стандартная неопределенность, соответствующая нестабильности условий проведения проверки квалификации;

u_{trans} — стандартная неопределенность, соответствующая нестабильности условий транспортирования;

$u(x_i)$ — стандартная неопределенность результата i-го участника;

$u(x_{pt})$ — стандартная неопределенность приписанного значения;

$u(x_{ref})$ — стандартная неопределенность опорного значения;

$U(x_i)$ — расширенная неопределенность зафиксированных результатов i-го участника;

$U(x_{pt})$ — расширенная неопределенность приписанного значения;

$U(x_{ref})$ — расширенная неопределенность опорного значения;

w_t — размах результатов измерений на различных экземплярах одного образца;

w^* — робастная оценка повторяемости участника;

x — результат измерений (в общем случае);

x_{char} — значение характеристики, полученное в соответствии с определением приписанного значения;

x_{CRM} — приписанное значение для показателя *CRM*;

x_i — результат измерений *i*-го участника;

x_{pt} — приписанное значение;

x_{ref} — опорное значение для установленной цели;

x^* — робастная оценка среднего участника;

$\overline{x}$ — среднее арифметическое набора результатов;

z — индекс, используемый для оценки квалификации;

z' — модифицированный z-индекс, включающий неопределенность приписанного значения;

ζ — дзета-индекс, модифицированный z-индекс, включающий неопределенность результата участника и приписанного значения.

Приложение В
(обязательное)

Однородность и стабильность образцов для проверки квалификации

В.1 Общая процедура проверки однородности

В.1.1 Для проверки однородности при подготовке образцов следует выполнять приведенную ниже процедуру.

Выбирают свойство (или свойства) или измеряемую(ые) величину(ны) для проверки однородности.

Для проверки однородности выбирают лабораторию и используемый метод измерений. Метод измерений должен иметь достаточно малое стандартное отклонение повторяемости s_r, чтобы любая значительная неоднородность могла быть обнаружена. Отношение стандартного отклонения повторяемости метода измерений к стандартному отклонению для оценки квалификации должно составлять менее 0,5 в соответствии с рекомендациями гармонизированного протокола IUPAC (или 1/6 δ_E). Так как это не всегда возможно, то провайдер должен использовать большее количество репликаций измерений.

Подготавливают и упаковывают достаточное количество образцов для проверки квалификации для раунда программы проверки квалификации и для проверки их однородности.

Из готовых упакованных образцов случайным образом отбирают g образцов, $g \geq 10$. Количество образцов для проверки однородности может быть уменьшено при наличии соответствующих данных предыдущих проверок однородности для образцов, подготовленных с применением тех же самых процедур.

Подготавливают $m \geq 2$ исследуемых порций для каждого образца, используя приемы, соответствующие исследуемому образцу, чтобы минимизировать различия между исследуемыми порциями.

Случайным образом отбирают $g \cdot m$ исследуемых порций и определяют результаты измерений на каждой порции, получая серию результатов измерений в условиях повторяемости.

Вычисляют общее выборочное среднее $\bar{x}$, стандартное отклонение s_w и стандартное отклонение s_s в соответствии с В.3.

В.1.2 Если нет возможности выполнения репликаций измерений, например в случае разрушающих испытаний, в качестве s_s можно использовать стандартное отклонение результатов. В такой ситуации важно иметь метод, обладающий достаточно низким стандартным отклонением повторяемости s_r.

В.2 Критерий проверки однородности

В.2.1 Следует использовать следующие три проверки, чтобы обеспечить валидацию данных испытаний на однородность:

a) проверяют результаты измерений для каждой исследуемой порции для выявления тренда или дрейфа; если выявлен тренд, необходимо предпринять соответствующие корректирующие действия по отношению к методу измерений или соблюдать осторожность в интерпретации результатов;

b) проверяют результаты для выборочных средних образцов для проверки квалификации в порядке получения, если существует явный тренд, который может привести к тому, что образец превысит критерий, установленный в В.2.2, или возникнут обстоятельства, препятствующие использованию образца, то (i) устанавливают индивидуальные значения для каждого образца или (ii) исключают поврежденные образцы и повторно проверяют оставшиеся образцы на однородность; или (iii) поступают в соответствии с положениями В.2.4, если тренд охватывает все образцы;

c) сопоставляют разности результатов репликаций (или размах при выполнении более двух репликаций) и при необходимости используют критерий Кохрена для проверки наличия статистически значимых различий между результатами репликаций измерений (см. *ГОСТ Р ИСО 5725-2*). Если разность результатов репликаций велика, для всех пар следует найти техническое обоснование этого явления и при необходимости исключить отличающуюся группу из анализа или удалить одну точку, если $m > 2$ и высокая дисперсия обусловлена единственным выбросом.

Примечание — Если $m > 2$ и исключено одно наблюдение, то при дальнейших расчетах необходимо учесть дисбаланс между s_w и s_s.

В.2.2 Сопоставляют стандартное отклонение s_s со стандартным отклонением для оценки квалификации σ_{pt}. Образцы для проверки квалификации можно считать однородными если:

$$s_s \leq 0{,}3\ \sigma_{pt}. \qquad (B.1)$$

Примечание 1 — Коэффициент 0,3 означает, что при выполнении критерия выборочного стандартного отклонения в изменчивость оценка функционирования не превышает 10 %, поэтому эти оценки вряд ли влияют на оценку.

Примечание 2 — Аналогично s_s можно сравнить с δ_E:

$$s_s \leq 0{,}1\ \delta_E. \qquad (B.2)$$

B.2.3 Для учета фактической выборочной погрешности и повторяемости при проверке однородности может быть полезно расширить критерий. В этих случаях необходимо выполнить следующие действия:

a) вычислить $\sigma^2_{allow} = (0{,}3\ \sigma_{pt})^2$;

b) $c = F_1\sigma^2_{allow} + F_2 s^2_w$,

где s_w — выборочное стандартное отклонение, вычисляемое в соответствии с B.3. Значения F_1 и F_2 определяют по таблице B.1 для выбранного количества образцов для проверки квалификации и с каждым образцом, исследуемым два раза [13].

Таблица B.1 — Коэффициенты F_1 и F_2 для использования при проверке однородности

gm	20	19	18	17	16	15	14	13	12	11	10	9	8	7
F1	1,59	1,60	1,62	1,64	1,67	1,69	1,72	1,75	1,79	1,83	1,88	1,94	2,01	2,10
F2	0,57	0,59	0,62	0,64	0,68	0,71	0,75	0,80	0,86	0,93	1,01	1,11	1,25	1,43

Если $m > 2$, F_2 в перечислении b) B.2.3 и таблице B.1 необходимо заменить на $F_{2m} = (F_{g-1,\ g(m-1),\ 0{,}95-1})/m$, где $(F_{g-1,\ g(m-1),\ 0{,}95-1})$ — расширенное значение, соответствующее вероятности 0,05 для случайной величины, подчиняющейся F-распределению с $(g-1)$ и $g(m-1)$ степенями свободы.

Примечание — Значения F_1 и F_2 в таблице B.1 получены из стандартных статистических таблиц следующим образом:

$F_1 = \chi^2_{0{,}95(g-1)}$, где $\chi^2_{0{,}95(g-1)}$ — значение, которое случайная величина, подчиняющаяся χ^2-распределению с $(g-1)$ степенями свободы, превышает с вероятностью 0,05. $F_2 = (F_{0{,}95\ (g-1;g)} - 1)/2$, где $F_{0{,}95(g-1;g)}$ — значение, которое случайная величина, подчиняющаяся F-распределению с $(g-1)$ и g степенями свободы, превышает с вероятностью 0,05;

c) если $s_s > \sqrt{c}$, это является свидетельством того, что партия образцов для испытаний недостаточно однородная.

B.2.4 Если значение σ_{pt} неизвестно заранее, например, когда σ_{pt} является робастным стандартным отклонением результатов участников, провайдеру проверки квалификации следует выбрать другие критерии для определения наличия достаточной однородности. Соответствующие процедуры могут включать в себя:

a) проверку наличия статистически значимых различий между образцами, например, с помощью критерия Фишера и дисперсионного анализа для $\alpha = 0{,}05$;

b) использование информации из предыдущих раундов проверки квалификации для оценки σ_{pt};

c) использование данных эксперимента на прецизионность (таких как стандартное отклонение воспроизводимости по *ГОСТ Р ИСО 5725-2*);

d) принятие риска наличия неоднородных образцов и проверку критерия после вычисления согласованного значения σ_{pt}.

B.2.5 Если критерий достаточной однородности не выполнен, провайдер должен принять решение о выполнении одного из следующих действий:

a) включение стандартного отклонения образца в стандартное отклонение для оценки квалификации путем вычисления σ'_{pt} в соответствии с (B.3). Важно, чтобы эта информация была доведена до участников

$$\sigma'_{pt} = \sqrt{\sigma^2_{pt} + s^2_s}; \qquad (B.3)$$

b) включение s_s в неопределенность приписанного значения и использование z' или δ_E' для получения оценки функционирования участника (см. 9.5);

c) если σ_{pt} — робастное стандартное отклонение результатов участников, то неоднородность между образцами включают в σ_{pt}. Таким образом критерий приемлемой однородности становится более слабым и должен быть использован с осторожностью.

Если ни одно из действий, приведенных в перечислении a)—c), не может быть применено, следует отказаться от применения образца и повторить подготовку после устранения причины неоднородности.

B.3 Формулы для проверки однородности

Оценки стандартных отклонений s_w и s_s могут быть вычислены с использованием дисперсионного анализа, как показано ниже. Метод применяют для выбранного количества образцов g с репликациями m измерений каждого.

Данные проверки однородности представлены значениями переменной $x_{t,k}$
где t — соответствует образцу (t = 1, 2, ..., g);
k — соответствует исследуемой порции (k = 1, 2, ..., m).

Вычисляют выборочное среднее и выборочную дисперсию для образца:

$$\bar{x}_t = \frac{1}{m}\sum_{k=1}^{m} x_k,$$
$$s_t^2 = \frac{1}{m}\sum_{k=1}^{m}(x_k - \bar{x}_t)^2, \quad \text{(B.4)}$$

а также оценку дисперсии между исследуемыми порциями:

$$w_t^2 = \frac{1}{(m-1)}\sum_{k=1}^{m}(x_k - \bar{x}_t)^2. \quad \text{(B.5)}$$

Вычисляют общее выборочное среднее:

$$\bar{\bar{x}} = \frac{1}{g}\sum_{t=1}^{g}\bar{x}_t, \quad \text{(B.6)}$$

оценку дисперсии выборочных средних:

$$s_x^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x}_t - \bar{\bar{x}})^2, \quad \text{(B.7)}$$

а также оценку выборочной дисперсии:

$$s_w^2 = \frac{1}{g}\sum_{t=1}^{g}s_t^2. \quad \text{(B.8)}$$

Определяют оценку объединенной дисперсии s_s и s_w:

$$s_{s,w}^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x}_t - \bar{\bar{x}})^2 + \left(1 - \frac{1}{m}\right)s_w^2 = s_s^2 + s_w^2. \quad \text{(B.9)}$$

Определяют оценку:

$$s_s^2 = s_{s,w}^2 - s_w^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}(\bar{x}_t - \bar{\bar{x}})^2 - \frac{1}{m}s_w^2. \quad \text{(B.10)}$$

Примечание — В том случае если $s_s^2 < 0$, то лучше использовать $s_s = 0$.

В общем случае, если m = 2, следует использовать приведенные ниже вычисления.

Определяют выборочные средние по порциям:

$$\bar{x}_t = (x_{t,1} + x_{t,2})/2 \quad \text{(B.11)}$$

и размах между исследуемыми порциями как

$$w_t = \left|x_{t,1} - x_{t,2}\right|. \quad \text{(B.12)}$$

Вычисляют общее выборочное среднее:

$$\bar{\bar{x}} = \frac{1}{g}\sum_{t=1}^{g}\bar{x}_t. \quad \text{(B.13)}$$

Вычисляют оценку стандартного отклонения оценок $\bar{x}_t$

$$s_x = \sqrt{\sum_{t=1}^{g}(\bar{x}_t - \bar{\bar{x}})^2/(g-1)} \quad \text{(B.14)}$$

и выборочное стандартное отклонение по всем образцам:

$$s_w = \sqrt{\sum_{t=1}^{g}w_t^2/(2g)}. \quad \text{(B.15)}$$

Суммирование в формулах (B.13), (B.14) и (B.15) ведется по образцам (t = 1, 2, ..., g).

Вычисляют оценку стандартного отклонения между образцами:

$$s_s = \max\left(0, \sqrt{s_x^2 - (s_w^2/2)}\right). \quad \text{(B.16)}$$

Примечание 1 — Если s_x меньше s_w, то разность $\left(s_x^2 - \left(s_w^2/2\right)\right)S_s^2$ будет отрицательной. Это возможно при высокой однородности. В этом случае принимают $s_S = 0$.

Примечание 2 — Вместо размахов можно использовать оценки стандартных отклонений исследуемых порций

$$s_t = w_t/\sqrt{2}.$$

Примечание 3 — Пример приведен в Е.2 приложения Е.

В.4 Процедуры проверки стабильности

В.4.1 Общие положения

Ниже приведены общие рекомендации по проверке выполнения требований к стабильности, установленные в 6.1. Положения, установленные в 6.1.3 в отношении исследуемых свойств, применимы к любой экспериментальной проверке на стабильность в течение раунда проверки квалификации и стабильности в процессе транспортирования.

В.4.1.1 Если предыдущие экспериментальные исследования, опыт или знания позволяют быть уверенными в том, что нестабильность маловероятна, экспериментальную проверку стабильности можно проводить только в случае появления значительных изменений в процессе раунда проверки квалификации и после раунда. Если такой уверенности нет, исследования влияния транспортирования на стабильность и стабильности типового раунда проверки квалификации могут иметь форму планируемых исследований, проводимых до распространения образцов проверки квалификации, либо для каждого раунда, либо на ранних стадиях планирования и определения возможности согласованных условий транспортирования и хранения. Провайдеры проверки квалификации могут также проводить проверку на наличие нестабильности путем анализа приведенных в отчете результатов измерений на наличие тренда с результатами измерений.

В.4.1.2 При проверке стабильности могут быть выполнены следующие действия:

- все показатели, используемые в программе проверки квалификации, должны быть проверены на стабильность (или их стабильность должна быть подтверждена другим способом). Это может быть сделано на основе опыта работы и технических заключений на основе знаний матрицы (или артефакта) и измеряемой величины;
- при проверке следует использовать не менее двух образцов, если изменчивость между образцами является достаточно большой; большее количество образцов или репликаций следует использовать при наличии сомнений относительно повторяемости (например, если s_w или $s_r > 0{,}5\ \sigma_{pt}$).

Примечание — В *ISO Guide 35* приведены стратегии минимизации влияния на исследования стабильности долгосрочных изменений процесса измерений, в том числе изохронные исследования или использование стабильных стандартных образцов.

В.4.2 Процедура проверки стабильности при проведении раунда проверки квалификации

В.4.2.1 Удобным способом проверки стабильности при проведении раунда проверки квалификации являются испытания небольшого количества образцов для проверки квалификации после окончания раунда и сравнение их с образцами до проведения раунда, чтобы гарантировать, что никаких изменений образцов за время проведения раунда не произошло. Проверка может также включать проверку влияния на образцы условий транспортирования с помощью сохранения образцов для исследования влияния продолжительности условий транспортирования. Для исследования влияния только условий транспортирования сравнивают образцы, поставленные для раунда испытаний, с образцами, сохраняемыми в контролируемых условиях.

Примечание 1 — Провайдер может использовать результаты проверки однородности до проведения раунда проверки квалификации вместо выбора и измерения конкретного набора образцов.

Примечание 2 — Данный способ в равной степени относится к программам проверки квалификации при испытаниях и калибровке.

В.4.2.2 Если провайдер включает в процедуру оценки стабильности поставленные образцы в соответствии с В.4.2.1, то влияние условий транспортирования таким образом будет учтено при оценке стабильности. Если влияние условий транспортирования проверяют отдельно, то следует использовать процедуру, установленную в В.6.

В.4.2.3 Процедура по проверке базовой стабильности, использующая измерения до и после раунда проверки квалификации, включает следующие действия:

a) случайным образом отбирают $2g$ образцов для раунда, где $g \geq 2$;

b) выбирают единственную лабораторию, использующую единственный метод измерений с хорошей промежуточной прецизионностью;

c) выполняют измерения на g образцах до плановой даты представления образцов участникам. Репликации измерений должны быть выполнены случайным образом;

d) оставшиеся *g* образцов сохраняют в условиях, аналогичных предполагаемым условиям хранения у участников;

e) в максимально сжатые сроки после даты возвращения результатов участников выполняют измерения на оставшихся *g* образцах в той же лаборатории, используя тот же метод измерений и то же количество репликаций измерений как в перечислении a), все репликации измерений проводят случайным образом;

f) вычисляют средние арифметические $\bar{y}_1$ и $\bar{y}_2$ результатов для двух групп (до и после раунда) соответственно.

B.4.2.4 В процедуре, установленной в B.4.2.3, могут быть сделаны следующие изменения:

a) 1-я группа из *g* образцов может быть исключена, если имеются другие результаты измерений набора образцов, выполненные в той же лаборатории и с тем же методом измерений, например, можно использовать данные проверки однородности;

b) для обеспечения большей уверенности в стабильности могут быть использованы условия, обеспечивающие ускоренные изменения образцов;

c) 2-я группа образцов может быть дополнительно помещена в условия, ожидаемые при транспортировании, для проверки их воздействия на образцы;

d) могут быть использованы все другие планы и условия, которые вместе с выбранными критериями проверки стабильности обеспечивают большую или равную стабильность.

B.5 Критерий проверки стабильности

B.5.1 Сравнивают общее среднее арифметическое результатов измерений, полученных до проверки однородности, с общим средним арифметическим результатов, полученных при проверке стабильности. Образцы можно считать стабильными, если:

$$\left|\bar{y}_1 - \bar{y}_2\right| \le 0{,}3\sigma_{pt} \quad \text{или} \quad \left|\bar{y}_1 - \bar{y}_2\right| \le 0{,}1\delta_E. \tag{B.17}$$

B.5.2 Если существует возможность, что промежуточная прецизионность метода измерений (или неопределенность результатов измерений образца) вносит свой вклад в несоответствие критерию, то следует выполнить одно из следующих действий:

a) использовать исследование изохронной стабильности (см. *ISO Guide 35*);

b) увеличить неопределенность приписанного значения для учета возможной нестабильности;

c) расширить критерий приемки путем добавления неопределенности разности к σ_{pt} в соответствии со следующей формулой

$$\left|\bar{y}_1 - \bar{y}_2\right| \le 0{,}3\sigma_{pt} + 2\sqrt{u^2(\bar{y}_1) + u^2(\bar{y}_2)}. \tag{B.18}$$

Примечание — Коэффициент 2 в формуле (B.18) является коэффициентом охвата расширенной неопределенности разности, обеспечивающим уровень доверия около 95 %, при вычислении комбинированной неопределенности предполагается, что $\bar{y}_1$ и $\bar{y}_2$ являются независимыми.

B.5.3 Если неравенства, приведенные в формулах (B.17) и (B.18), не выполнены, необходимо осуществить следующие действия:

- количественно оценить влияние нестабильности и учесть его при определении оценок (например, используя *z′*-индексы);
- проверить процедуры подготовки и хранения образцов для выявления возможности их улучшений;
- не определять оценки функционирования участников.

B.5.4 Критерии, приведенные в B.5.1 или B.5.2, могут быть заменены подходящим статистическим критерием, позволяющим установить различия двух наборов данных, при условии, что этот критерий учитывает повторяемость измерений и обеспечивает идентификацию стабильности по крайней мере не менее, чем в формуле (B.18).

Примечание — Обычно *t*-критерий, выявляющий существенные различия наборов данных с уровнем достоверности 95 %, использующий средние для каждого образца, обеспечивает такое же или более высокое выявление нестабильности, чем в формуле (B.18), при условии, что количество используемых образцов более или равно 3.

B.6 Стабильность в условиях транспортирования

B.6.1 Провайдер должен проверить воздействие условий транспортирования на образцы по крайней мере на ранних стадиях программы проверки квалификации. При такой проверке необходимо (при возможности) сопоставить образцы, сохраняемые в помещении провайдера, с образцами, поставляемыми и возвращаемыми участниками. Также могут быть проведены исследования на основе выдерживания образцов в разумных пределах в условиях транспортирования.

B.6.2 При определении оценки функционирования участника необходимо исследовать воздействие на образцы всех известных условий транспортирования. Значимое увеличение неопределенности, связанной с транспортированием образцов, следует включить в неопределенность приписанного значения.

В.6.3 В том случае, если проверка стабильности условий транспортирования включает сравнение результатов для двух групп образцов, одну группу подвергают воздействию условий транспортирования, а другую — нет, критерием достаточной стабильности при транспортировании является критерий, установленный в В.5.1 или В.5.2.

Примечание 1 — Если приписанное значение и стандартное отклонение оценки квалификации определяют по результатам участников (например, с помощью робастных методов), то среднее арифметическое и стандартное отклонения для оценок квалификации отражают любое смещение и увеличение изменчивости (соответственно), вызванные условиями транспортирования образцов.

Примечание 2 — Пример проверки стабильности приведен в Е.2 приложения Е.

Приложение С
(обязательное)

Робастный анализ

С.1 Введение

Межлабораторные сравнительные испытания представляют собой особый анализ данных. В то время как большинство межлабораторных сравнительных испытаний представляют данные, подчиняющиеся унимодальному и приблизительно симметричному распределению в задачах проверки квалификации, большая часть наборов данных включает часть результатов, неожиданно далеко отстоящих от основного набора данных. Причины появления таких данных могут быть различными: например, появление новых, менее опытных участников проверки, появление новых и, возможно, менее точных методов измерений, непонимание некоторыми участниками инструкции или неправильная обработка образцов. Такие отличающиеся результаты (выбросы) могут быть весьма изменчивы, в этом случае применение традиционных статистических методов, в том числе вычисление среднего арифметического и стандартного отклонений, может дать недостоверные результаты.

Провайдерам рекомендуется (см. 6.5.1) использовать статистические методы, устойчивые к выбросам. Большинство таких методов предложено в книгах по математической статистике, и многие из них успешно использованы в задачах проверки квалификации. Обычно робастные методы обеспечивают дополнительную устойчивость при обработке данных из асимметричных распределений с выбросами.

В данном приложении описано несколько простых в применении методов, используемых в задачах проверки квалификации и имеющих различные возможности в отношении определения устойчивости оценок при наличии данных из загрязненных совокупностей (например, эффективности и пороговой точки). Методы представлены в порядке возрастания сложности (первый — самый простой, последний — самый сложный), и в порядке убывания эффективности, поэтому наиболее сложные оценки требуют доработки для повышения их эффективности.

Примечание 1 — В приложении D приведена дополнительная информация об эффективности, пороговых точках и чувствительности к небольшим модам — трем важным показателям различных робастных методов определения оценки функционирования.

Примечание 2 — Робастность является свойством алгоритма определения оценки, а не свойством полученных оценок, поэтому не совсем корректно называть средние значения и стандартные отклонения, рассчитанные с помощью такого алгоритма, робастными. Однако, чтобы избежать использования чрезмерно громоздких терминов, в настоящем стандарте применены термины «робастное среднее» и «робастное стандартное отклонение». Следует учитывать, что это означает оценки среднего или стандартного отклонения, полученные в соответствии с робастным алгоритмом.

С.2 Простые устойчивые к выбросам оценки для среднего и стандартного отклонений совокупности

С.2.1 Медиана

Медиана является наиболее простой, высоко устойчивой к выбросам оценкой среднего для симметричного распределения. Обозначим медиану $med(x)$. Для определения $med(x)$ по совокупности из p данных необходимо:

i) расположить p данных в порядке неубывания:

$x_{\{1\}}, x_{\{2\}}, \ldots, x_{\{p\}}$;

ii) вычислить

$$med(x) = \begin{cases} x_{\{(p+1)/2\}}, & \text{если } p \text{ — нечетное,} \\ \dfrac{\left[x_{\{p/2\}} + x_{\{1+p/2\}}\right]}{2}, & \text{если } p \text{ — четное.} \end{cases} \quad (C.1)$$

С.2.2 Абсолютное отклонение от медианы *MADe*

Абсолютное отклонение от медианы $MADe(x)$ обеспечивает определение оценки стандартного отклонения генеральной совокупности для данных из нормального распределения и является высоко устойчивым при наличии выбросов. Для определения $MADe(x)$ вычисляют:

i) абсолютные значения разностей d_i (i = 1, ..., p)

$$d_i = \left|x_i - med(x)\right|; \quad (C.2)$$

ii) $MADe(x)$

$$MADe(x) = 1{,}483\ med(d). \quad (C.3)$$

Если у половины или большего количества участников результаты совпадают, то $MADe(x) = 0$, и следует использовать оценку *nIQR* в соответствии с С.2.3, стандартное отклонение, полученное после исключения выбросов, или процедуру, описанную в С.5.2.

С.2.3 Нормированный межквартильный размах *nIQR*

Данный метод определения робастной оценки стандартного отклонения аналогичен методу определения $MADe(x)$. Эту оценку получить немного проще, поэтому ее часто используют в программах проверки квалификации. Данную оценку определяют как разность 75-го процентиля (или 3-го квартиля) и 25-го процентиля (или 1-го квартиля) результатов участника. Данную статистику называют нормированным межквартильным размахом *nIQR* и вычисляют по формуле

$$nIQR(x) = 0{,}7413(Q_3(x) - Q_1(x)), \quad (C.4)$$

где $Q_1(x)$ — 25-й процентиль выборки x_i ($i = 1, 2, ..., p$);

$Q_3(x)$ — 75-й процентиль выборки x_i ($i = 1, 2, ..., p$).

Если 75-й и 25-й процентили совпадают, то $nIQR = 0$ [как и $MADe(x)$], а для вычисления робастного стандартного отклонения следует использовать альтернативную процедуру, такую как арифметическое стандартное отклонение (после исключения выбросов), или процедуру, описанную в С.5.2.

Примечание 1 — Для расчета *nIQR* требуется сортировка данных только один раз в отличие от вычисления *MADe*, но *nIQR* имеет пороговую точку в 25 % (см. приложение D), в то время как у *MADe* пороговая точка 50 %. Поэтому *MADe* устойчива при значительно более высокой доле содержания выбросов, чем *nIQR*.

Примечание 2 — При $p < 30$ обе оценки обладают заметным отрицательным смещением, неблагоприятно влияющим на оценки участников при проверке квалификации.

Примечание 3 — Различные пакеты статистических программ используют различные алгоритмы расчета квартилей и, следовательно, могут давать оценки *nIQR* с некоторыми различиями.

Примечание 4 — Пример использования робастных оценок приведен в Е.3 приложения Е.

С.3 Алгоритм *A*

С.3.1 Алгоритм *A* с итеративной шкалой

Данный алгоритм дает робастные оценки среднего и стандартного отклонения на основе используемых данных.

Для выполнения алгоритма *A* p данные располагают в порядке неубывания

$$x_{\{1\}}, x_{\{2\}}, ..., x_{\{p\}}.$$

Полученные по этим данным робастное среднее и робастное стандартное отклонения обозначают x^* и s^*.

Вычисляют начальные значения для x^* и s^* по формулам:

$$x^* = med(x_i) \ (i = 1, 2, ..., p), \quad (C.5)$$

$$s^* = 1{,}483\ med\left|x_i - x^*\right|, \ (i = 1, 2, ..., p). \quad (C.6)$$

Примечание 1 — Алгоритмы *A* и *S*, приведенные в настоящем приложении, соответствуют *ГОСТ Р ИСО 5725-5* с добавлением критерия остановки: при совпадении до 3-го знака после запятой среднего и стандартного отклонения вычисления прекращают.

Примечание 2 — В некоторых случаях более половины результатов x_i будут идентичны (например, количество нитей в образцах ткани или количество электролитов в образцах сыворотки крови). В этом случае начальное значение $s^* = 0$ и робастная процедура будут некорректными. Если начальное значение $s^* = 0$, допустимо заменить выборочное стандартное отклонение после проверки всех очевидных выбросов, которые могут сделать стандартное отклонение неоправданно большим. Такую замену проводят только для начального значения s^* и после этого итеративный алгоритм применяют в соответствии с описанием.

Вычисляют новые значения x^* и s^*. Для этого вычисляют

$$\delta = 1{,}5\ s^*. \quad (C.7)$$

Для каждого x_i ($i = 1, 2, ..., p$) вычисляют

$$x_i^* = \begin{cases} x^* - \delta, & \text{если } x_i < x^* - \delta, \\ x^* + \delta, & \text{если } x_i > x^* + \delta, \\ x_i & \text{во всех остальных случаях.} \end{cases} \quad (C.8)$$

Вычисляют новые значения x^* и s^*

$$x^* = \sum_{i=1}^{p} x_i^* / p, \quad (C.9)$$

$$s^* = 1{,}134\sqrt{\sum_{i=1}^{p}(x_i^* - x^*)^2/(p-1)}, \quad (C.10)$$

где суммирование производят по i.

Робастные оценки x^* и s^* получают на основе итеративных, то есть повторных вычислений x^* и s^* в соответствии с (С.7)—(С.10) до тех пор, пока процесс не начнет сходиться, то есть разности предыдущих и последующих значений x^* и s^* не станут пренебрежимо малы. Обычно итеративные вычисления прекращают при совпадении в предыдущих и последующих значениях трех знаков после запятой.

Альтернативные критерии сходимости могут быть определены в соответствии с требованиями к плану эксперимента и к отчету по результатам проверки квалификации.

Примечание — Примеры использования алгоритма *A* приведены в Е.3 и Е.4 приложения Е.

С.3.2 Варианты алгоритма *A*

Итеративный алгоритм *A*, приведенный в С.3.1, имеет скромную разбивку (примерно 25 % для больших наборов данных [14]) и начальную точку для s^*, предложенную в С.3.1, для наборов данных, где $MADe(x) = 0$ может серьезно ухудшить устойчивость при наличии нескольких выбросов в наборе данных. Если в наборе данных ожидаемая доля выбросов составляет более 20 % или если начальное значение s^* подвержено неблагоприятному влиянию экстремальных выбросов, то следует рассмотреть следующие варианты:

i) замена *MADe* на $med\left(\left|x_i - \bar{x}\right|\right)$ при $MADe = 0$ либо использование альтернативной оценки в соответствии с С.5.1 или арифметического стандартного отклонения (после исключения выбросов);

ii) если при оценке робастное стандартное отклонение не используют, следует применять *MADe* [исправленное в соответствии с *i*)], и не изменяют s^* во время итерации. Если при оценке используют робастное стандартное отклонение, заменяют s^* в соответствии с С.5 оценкой Q и не изменяют s^* во время итерации.

Примечание — Вариант, приведенный в перечислении ii), улучшает пороговую точку алгоритма *A* до 50 % [14], что позволяет применять алгоритм при наличии высокой доли выбросов.

С.4 Алгоритм *S*

Данный алгоритм применяют к стандартным отклонениям (или размахам), которые вычисляют, если участники представляют результаты m репликаций измерений измеряемой величины образца или в исследовании используют m идентичных образцов. Алгоритм позволяет получить робастное объединенное значение стандартных отклонений или размахов.

Имеющиеся p стандартных отклонений или размахов располагают в порядке неубывания

$$w_{\{1\}}, w_{\{2\}}, \ldots, w_{\{p\}}.$$

Обозначим робастное объединенное значение w^*, а v — число степеней свободы, соответствующее каждому w_i. (Если w_i — размах, то $v = 1$. Если w_i — стандартное отклонение для m результатов испытаний, то $v = m - 1$.) Значения ξ и η определяют в соответствии с алгоритмом, приведенным в таблице С.1.

Вычисляют начальное значение w^*:

$$w^* = med(w_i), \ (i = 1, 2, \ldots, p). \quad (C.11)$$

Примечание — Если более половины w_i имеют значения, равные нулю, то начальное значение w^* равно нулю, а робастный метод является некорректным. Если начальное значение w^* равно нулю, то после устранения выбросов, которые могут повлиять на выборочное среднее, заменяют стандартное отклонение объединенного среднего арифметического (или размах средних арифметических). Эту замену выполняют только для начального значения w^*, после чего процедуру продолжают согласно описанию.

Значение w^* вычисляют следующим образом:

$$\psi = \eta \cdot w^*. \quad (C.12)$$

Для каждого значения w_i ($i = 1, 2, \ldots, p$) вычисляют

$$w_i^* = \begin{cases} \psi, & \text{если } w_i > \psi, \\ w_i, & \text{в противном случае.} \end{cases} \quad (C.13)$$

Вычисляют новое значение w^*

$$w^* = \xi\sqrt{\sum_{i=1}^{p}(w_i^*)^2/p}. \quad (C.14)$$

Робастную оценку w^* получают итеративным методом, вычисляя значение w^* несколько раз, пока процесс не начнет сходиться. Сходимость считают достигнутой, если значения w^* в последовательных итерациях совпадают в трех знаках после запятой.

Примечание — Алгоритм *S* обеспечивает оценку стандартного отклонения генеральной совокупности, если оно получено по стандартным отклонениям из того же нормального распределения (и, следовательно, обеспечивает оценку стандартного отклонения повторяемости при выполнении предположений в соответствии с *ГОСТ Р ИСО 5725-2*).

Таблица С.1 — Коэффициенты, необходимые для проведения робастного анализа: алгоритм *S*

Число степеней свободы ν	Лимитирующий коэффициент η	Поправочный коэффициент ξ
1	1,645	1,097
2	1,517	1,054
3	1,444	1,039
4	1,395	1,032
5	1,359	1,027
6	1,332	1,024
7	1,310	1,021
8	1,292	1,019
9	1,277	1,018
10	1,264	1,017
Примечание — Значения ξ и η приведены в *ГОСТ Р ИСО 5725-5*.		

С.5 Сложные для вычислений робастные оценки: *Q*-метод и оценка Хампеля

С.5.1 Обоснование оценок

Робастные оценки среднего и стандартного отклонения генеральной совокупности, описанные в С.2 и С.3, используют в тех случаях, когда вычислительные ресурсы ограничены или когда требуется краткое обоснование статистических процедур. Эти процедуры оказались полезными в самых разных ситуациях, в том числе в программах проверки квалификации в новых областях исследований или при калибровке и в тех областях экономики, где проверка квалификации раньше не была доступна. Однако эти методы являются недостоверными в тех случаях, когда количество выбросов в результатах превышает 20 %, или в случае бимодального (или мультимодального) распределения данных, и некоторые из них могут стать неприемлемо изменчивыми для небольшого количества участников. Кроме того, ни один из этих методов не может работать с данными репликаций измерений участников. В соответствии с *ГОСТ ISO/IEC 17043* необходимо, чтобы эти ситуации были предусмотрены до проведения расчетов или выполнены в процессе анализа до проведения оценки функционирования участника, однако это не всегда возможно.

Кроме того, некоторые робастные методы, описанные в С.2 и С.3, имеют низкую статистическую эффективность. Если количество участников менее 50, а робастное среднее и/или стандартное отклонение используют для определения индексов, то существует значимый риск неверной классификации участников при применении неэффективных статистических методов.

Робастные методы, объединяющие высокую эффективность (то есть сравнительно низкую изменчивость) с возможностью работы с высокой долей выбросов в данных, обычно являются достаточно сложными и требуют серьезных вычислительных ресурсов, но эти методы представлены в литературе и международных стандартах. Некоторые из них обеспечивают получение дополнительных преимуществ, когда основное распределение данных является асимметричным или определенные результаты находятся ниже предела их обнаружения.

Ниже приведены некоторые высокоэффективные методы определения оценок стандартного отклонения и параметра положения (среднего), которые показывают более низкую изменчивость, чем простые оценки, и полезны при использовании для данных с большой долей выбросов. Одну из описанных оценок можно применять для оценки стандартного отклонения воспроизводимости, если участники сообщают о большом количестве наблюдений.

С.5.2 Определение робастного стандартного отклонения с использованием *Q*-метода и Q_n-метода

С.5.2.1 Оценка Q_n [15] является высокоэффективной оценкой стандартного отклонения генеральной совокупности с разбивкой, которая становится несмещенной для данных нормального распределения (при условии отсутствия выбросов).

Q_n-метод учитывает единственный результат для каждого участника (включающий среднее или медиану репликаций измерений). Расчет основан на использовании попарных различий в наборе данных и поэтому не зависит от оценки среднего или медианы.

Выполнение этого метода включает корректировки, позволяющие обеспечить несмещенность оценки для всех фактических объемов наборов данных.

При вычислении Q_n для набора данных (x_1, x_2, ..., x_p) с p результатами:

i) вычисляют $p(p–1)/2$ абсолютных разностей

$$d_{ij} = \left| x_i - x_j \right| \quad \text{для } i = 1,2,...,p, j = i+1, i+2, ..., p; \tag{C.15}$$

ii) для разностей d_{ij} используют обозначения

$$d_{\{1\}}, d_{\{2\}}, \ldots, d_{\{p(p-1)/2\}}; \tag{C.16}$$

iii) вычисляют

$$k = \frac{h(h-1)}{2}, \tag{C.17}$$

где k — количество различных пар, выбранных из h объектов,

где

$$h = \begin{cases} p/2, & \text{если } p \text{ — четное,} \\ (p-1)/2, & \text{если } p \text{ — нечетное;} \end{cases} \tag{C.18}$$

iv) вычисляют Q_n

$$Q_n = 2{,}2219 d_{(k)} b_p, \tag{C.19}$$

где b_p определяют по таблице С.2 для конкретного количества данных, если $p > 12$, b_p вычисляют по формуле

$$b_p = \frac{1}{r_p + 1}, \tag{C.20}$$

где

$$r_p = \begin{cases} \dfrac{1}{p}\left[1{,}6019 + \dfrac{1}{p}\left(-2{,}128 - \dfrac{5{,}172}{p}\right)\right], & \text{если } p \text{ — нечетное,} \\ \dfrac{1}{p}\left[3{,}6756 + \dfrac{1}{p}\left(1{,}965 + \dfrac{1}{p}\left(6{,}987 - \dfrac{77}{p}\right)\right)\right], & \text{если } p \text{ — четное.} \end{cases} \tag{C.21}$$

Примечание 1 — Коэффициент 2,2219 является поправочным, обеспечивающим несмещенность оценки стандартного отклонения для больших p. Поправочные коэффициенты b_p для небольших значений p определяют по таблице С.2, а при $p > 12$ эти коэффициенты устанавливают в соответствии с [15], используя экстенсивное моделирование и последующее применение регрессионного анализа.

Примечание 2 — Простой алгоритм, описанный выше, для больших наборов данных, например, при $p > 1000$, требует значительных вычислительных ресурсов. Для быстрой обработки опубликованы программы (см. [15]) для использования с более крупными наборами данных (на момент публикации приведена обработка данных с объемом выше 8000 за приемлемое время).

Таблица С.2 — Поправочный коэффициент b_p для $2 \le p \le 12$

p	2	3	4	5	6	7	8	9	10	11	12
b_p	0,9937	0,9937	0,5132	0,8440	0,6122	0,8588	0,6699	0,8734	0,7201	0,8891	0,7574

С.5.2.2 Q-метод позволяет получить высокоэффективную оценку стандартного отклонения результатов проверки квалификации, представленных различными лабораториями, с разбивкой. Q-метод не является устойчивым не только при наличии выбросов, но и в той ситуации, когда большая часть результатов испытаний равны между собой, например, когда результаты представляют собой дискретные числа или при округлении данных. В такой ситуации другие подобные методы не следует применять, поскольку многие разности равны нулю.

Q-метод можно использовать для проверки квалификации как в случае предоставления участником единственного результата (в виде среднего и медианы репликаций измерений), так и результатов репликаций. Прямое использование репликаций измерений в вычислениях повышает эффективность метода.

Расчет основан на использовании разностей пар в наборе данных, и таким образом оценка не зависит от оценки среднего или медианы данных. Метод называют Q-методом, или методом Хампеля, если его используют вместе с алгоритмом конечных шагов для определения оценки Хампеля, описанной в С.5.3.3.

Обозначим результаты измерений участников, сгруппированные по лабораториям

$$\underbrace{y_{11}, \ldots, y_{1n_1}}_{\text{лаборатория 1}}, \ \underbrace{y_{21}, \ldots, y_{2n_2}}_{\text{лаборатория 2}}, \ \ldots, \ \underbrace{y_{p1}, \ldots, y_{pn_p}}_{\text{лаборатория } p}.$$

Кумулятивная функция распределения абсолютных значений разностей результатов участников имеет следующий вид:

$$H_1(x) = \frac{2}{p(p-1)} \sum_{1 \le i < j \le p} \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{m=1}^{n_j} I\left\{\left|y_{ik} - y_{jm}\right| \le x\right\}, \qquad \text{(C.22)}$$

где $I\left\{\left|y_{ik} - y_{jm}\right| \le x\right\} = \begin{Bmatrix} 1, \text{ если } \left|y_{ik} - y_{jm}\right| \le x \\ 0, \text{ в противном случае} \end{Bmatrix}$ — индикаторная функция.

Обозначим точки разрыва функции $H_1(x)$:

$$x_1, \ldots, x_r, \text{ где } x_1 < x_2 < \ldots < x_r$$

Значения функции в точках $x_1, \ldots, x_r$

$$G_1(x_i) = \begin{cases} 0{,}5 \cdot \left(H_1(x_i) + H_1(x_{i-1})\right), & \text{если } i \ge 2, \\ 0{,}5 \cdot (H_1(x_1), & \text{если } i = 1; x_1 > 0. \end{cases} \qquad \text{(C.23)}$$

Пусть $G_1(0) = 0$.

Значения функции $G_1(x)$ для x вне интервала $[0, x_r]$ вычисляют с помощью линейной интерполяции между точками разрыва $0 \le x_1 < x_2 < \ldots < x_r$.

Робастное стандартное отклонение s^* результатов испытаний для различных лабораторий имеет вид:

$$s^* = \frac{G_1^{-1}\left(0{,}25 + 0{,}75 \cdot H_1(0)\right)}{\sqrt{2\Phi^{-1}\left(0{,}625 + 0{,}375 \cdot H_1(0)\right)}}, \qquad \text{(C.24)}$$

где $H_1(0)$ вычисляют аналогично формуле (C.22) и $H_1(0) = 0$ в случае точного совпадения данных, и $\Phi^{-1}(q)$ — квантиль стандартного нормального распределения уровня q.

Примечание 1 — Этот алгоритм не зависит от среднего, он может быть использован либо вместе со значением, полученным по объединенным результатам участников, или в соответствии с установленным опорным значением.

Примечание 2 — Другие варианты Q-метода, позволяющие получить робастную оценку стандартных отклонений воспроизводимости и повторяемости, приведены в [14], [15].

Примечание 3 — Теоретические основы Q-метода, включая его асимптотическую эффективность и разбивку на конечное число выборок, описаны в [16] и [15].

Примечание 4 — Если исходные данные участников представлены единственным результатом измерений, полученным с помощью одного установленного метода измерений, робастное стандартное отклонение является оценкой стандартного отклонения воспроизводимости, как и в (C.21).

Примечание 5 — Стандартное отклонение воспроизводимости не обязательно является наиболее подходящим стандартным отклонением для использования в проверке квалификации, так как это, как правило, оценка разброса единственных результатов, а не оценка разброса средних или медиан результатов репликаций каждого участника. Однако разброс средних или медиан результатов репликаций лишь немного менее разброса единственных результатов различных лабораторий, если отношение стандартного отклонения воспроизводимости к стандартному отклонению повторяемости более двух. Если это отношение менее двух, для определения оценок при проверке квалификации может быть использована замена стандартного отклонения воспроизводимости s_R скорректированным значением

$$\sqrt{s_R^2 - \frac{m-1}{m} s_r^2},$$

где m — количество репликаций;

s_r^2 — дисперсия повторяемости, вычисленная в соответствии с [17], или можно использовать среднее значение репликаций измерений участника Q-метода.

Примечание 6 — Примечание 5 применяют только в том случае, если индексы определяют на основе средних или медиан результатов репликаций. Если репликации проводят вслепую, индексы следует рассчитывать

для каждой репликации. В этом случае стандартное отклонение воспроизводимости является наиболее подходящим стандартным отклонением.

Примечание 7 — Пример применения Q-метода приведен в Е.3 приложения Е.

С.5.3 Определение робастного среднего, используемого в оценке Хампеля

С.5.3.1 Оценка Хампеля является высокоустойчивой высокоэффективной оценкой общего среднего всех результатов различных лабораторий. Поскольку формулы вычисления оценки Хампеля не существует, ниже приведены два алгоритма получения этой оценки. Первый из них является более простым, но может привести к отклонениям результатов при выполнении. Второй алгоритм обеспечивает получение однозначных результатов, зависящих только от базового стандартного отклонения.

С.5.3.2 Далее приведены вычисления, обеспечивающие получение итеративной взвешенной оценки Хампеля, для параметра положения.

i) Пусть $x_1, x_2, \ldots, x_p$ — данные.

ii) Пусть x^* — медиана $med(x)$ (см. С.2.1).

iii) Пусть s^* — соответствующая робастная оценка стандартного отклонения, например, $MADe$, Q_n или s^* в соответствии с Q-методом.

iv) Для каждой точки x_i вычисляют q_i

$$q_i = \left| \frac{x_i - x^*}{s^*} \right|.$$

v) Вычисляют вес w_i

$$w_i = \begin{cases} 0, & \text{если } |q| > 4,5, \\ (4,5 - q)/q, & \text{если } 3 < |q| \le 4,5, \\ 1,5/q, & \text{если } 1,5 < |q| \le 3,0, \\ 1, & \text{если } |q| \le 1,5. \end{cases}$$

vi) Пересчитывают x^*

$$x^* = \frac{\sum_{i=1}^{p} w_i x_i}{\sum_{i=1}^{p} w_i}.$$

vii) Повторяют действия в соответствии с перечислениями iv)—vi) до тех пор, пока значения x^* не начнут сходиться. Сходимость считают достаточной, если разность x^* в двух последних итерациях станет менее $0,1s^*/\sqrt{p}$, что соответствует приблизительно 1 % стандартной погрешности x^*. Могут быть использованы и другие более точные критерии сходимости.

Данный алгоритм получения оценки Хампеля не гарантирует получение единственной и наилучшей оценки, так как неудачный выбор начального значения x^* и/или s^* может привести к исключению важной части набора данных. Провайдеру следует предпринять соответствующие меры для проверки возможности получения неудачного результата или обеспечить однозначные правила выбора параметра положения. Наиболее общим правилом является выбор параметра положения, максимально близкого к медиане. Анализ результатов для подтверждения того, что большая часть данных не выходит за пределы области $|q| > 4,5$, может также помочь в принятии правильного решения.

Примечание 1 — Определение оценки Хампеля для данных из нормального распределения обладает эффективностью, приблизительно равной 96 %.

Примечание 2 — Примеры выполнения этого алгоритма приведены в Е.3 приложения Е.

Примечание 3 — Эффективность и устойчивость к выбросам оценки Хампеля могут быть повышены с помощью изменения весовой функции. Общая форма весовой функции имеет вид:

$$w_i = \begin{cases} 0, \quad \in & \text{если } |q| > c, \\ a(c - q)/[q(c - b)], & \text{если } b < |q| \le c, \\ a/q, & \text{если } a < |q| \le b, \\ 1, & \text{если } |q| \le a, \end{cases}$$

где a, b и c — регулируемые параметры. Для приведенного алгоритма a = 1,5, b = 3,0 и c = 4,5. Более высокая эффективность достигается за счет увеличения области изменений q. Повышения устойчивости к выбросам или изменениям режимов достигают за счет уменьшения области изменений q.

С.5.3.3 Ниже приведен алгоритм конечных шагов, позволяющий получить оценку Хампеля для параметра положения [14].

Вычисляют средние арифметические $y_1, y_2, ..., y_p$.

Вычисляют робастное среднее x^* как корень уравнения

$$\sum_{i=1}^{p} \psi\left(\frac{y_i - x^*}{s^*}\right) = 0, \qquad (C.25)$$

где

$$\psi(q) = \begin{cases} 0, & \text{если } q \le -4,5, \\ -4,5 - q, & \text{если } -4,5 < q \le -3, \\ -1,5, & \text{если } -3 < q \le -1,5, \\ q, & \text{если } -1,5 < q \le 1,5, \\ 1,5, & \text{если } 1,5 < q \le 3, \\ 4,5 - q, & \text{если } 3 < q \le 4,5, \\ 0, & \text{если } q > 4,5, \end{cases} \qquad (C.26)$$

s^* — робастное стандартное отклонение, полученное Q-методом.

Точное решение может быть получено за конечное число шагов, без итерации, используя свойство, при котором ψ как функция x^* является частично линейной, имея в виду точки интерполяции в левой стороне уравнения (С.25).

Вычисляют все точки интерполяции:

- для 1-го значения y_1:

$$d_1 = y_1 - 4,5 \cdot s^*,\ d_2 = y_1 - 3 \cdot s^*,\ d_3 = y_1 - 1,5 \cdot s^*,\ d_4 = y_1 + 1,5 \cdot s^*,\ d_5 = y_1 + 3 \cdot s^*,\ d_6 = y_1 + 4,5 \cdot s^*;$$

- для 2-го значения y_2:

$$d_7 = y_2 - 4,5 \cdot s^*,\ d_8 = y_2 - 3 \cdot s^*,\ d_9 = y_2 - 1,5 \cdot s^*,\ d_{10} = y_2 + 1,5 \cdot s^*,\ d_{11} = y_2 + 3 \cdot s^*,\ d_{12} = y_2 + 4,5 \cdot s^*;$$

- и так далее для всех $y_3, ..., y_p$.

Располагают $d_1, d_2, d_3, ..., d_{6 \cdot p}$ в порядке неубывания $d_{\{1\}}, d_{\{2\}}, d_{\{3\}}, ..., d_{\{6 \cdot p\}}$.

Затем для каждого $m = 1, ..., (6 \cdot p - 1)$ вычисляют

$$p_m = \sum_{i=1}^{p} \left(\frac{y_i - d_{\{m\}}}{s^*}\right)$$

и проверяют, являются ли следующие условия:

(i) если $p_m = 0$, то $d_{\{m\}}$ — решение уравнения (С.25);

(ii) если $p_{m+1} = 0$, то, $d_{\{m+1\}}$ — решение уравнения (С.25);

(iii) если $p_m \cdot p_{m+1} < 0$, то $x_m = d_{\{m\}} - \dfrac{p_m}{\dfrac{p_{m+1} - p_m}{d_{(m+1)} - d_{(m)}}}$ — решение уравнения (С.25).

Пусть S — множество всех решений уравнения (С.25).

Решением $x^* \in S$ является ближайшая медиана, используемая в качестве параметра положения x^*, то есть

$$\left|x^* - med(y_1, y_2, ..., y_p)\right| = \min\left\{\left|x - med(y_1, y_2, ..., y_p)\right|; x \in S\right\}.$$

Могут существовать несколько решений. Если существуют два решения, наиболее близких к медиане, или если не существует никакого решения вообще, то в качестве параметра положения x^* используют медиану.

Примечание 1 — Эта оценка Хампеля для данных из нормального распределения обладает эффективностью, приблизительно равной 96 %.

Примечание 2 — При использовании этого метода результаты лабораторий, отличающиеся от среднего более чем на 4,5 стандартных отклонений воспроизводимости, не оказывают никакого влияния на результат, то есть их рассматривают как выбросы.

С.5.4 Метод Q/Хампеля

Метод Q/Хампеля использует Q-метод, описанный в С.5.3.2, для вычисления робастного стандартного отклонения s^* и алгоритм конечных шагов для оценки Хампеля, описанный в С.5.3.3, для вычисления параметра положения x^*.

Если участники сообщают много наблюдений для вычисления робастного стандартного отклонения воспроизводимости s_R, используют Q-метод, описанный в С.5.3.2. Для вычисления робастного стандартного отклонения повторяемости s_r применяют 2-й алгоритм, использующий парные разности в пределах лаборатории.

Примечание — Веб-приложения для метода Q/Хампеля приведены в [18].

С.6 Другие робастные методы

Методы, описанные в настоящем приложении, не представляют собой целостную совокупность всех подходов. Ни один из них не является гарантированно оптимальным во всех ситуациях. По усмотрению провайдера могут быть использованы другие робастные методы при условии анализа их эффективности и всех остальных свойств, соответствующих определенным требованиям программы проверки квалификации.

Приложение D
(справочное)

Дополнительное руководство к статистическим процедурам

D.1 Процедуры в случае небольшого количества участников

D.1.1 Общие положения

В программах проверки квалификации обычно принимает участие небольшое количество лабораторий, или при наличии большого общего количества участников выполняют сопоставление групп с небольшим количеством участников. Это происходит, когда участников группируют и подсчитывают оценки с помощью единого метода, как, например, это бывает в медицинских лабораториях.

В том случае, если количество участников невелико, приписанное значение в идеале должно быть определено с использованием валидированной метрологической процедуры независимо от участников, например, путем расчета или по данным эталонной лаборатории.

Критерий оценки функционирования лаборатории также должен быть основан на внешних критериях, таких как выборы экспертов или соответствие целям. В идеальной ситуации, когда качество функционирования оценивают с использованием предварительно определенного приписанного значения и критерия функционирования, проверка квалификации может быть проведена при наличии одного участника. Такой тип межлабораторных сравнительных испытаний можно назвать билатеральным, или аудитом измерений, он может быть очень полезным во многих ситуациях, например, при калибровке.

Если эти идеальные условия не могут быть выполнены, приписанное значение или дисперсия, или то и другое должны быть получены по результатам участников. Если количество участников слишком мало для выполнения конкретной процедуры, то оценка функционирования лаборатории может стать недостоверной, поэтому важно рассмотреть вопрос об установлении минимального количества участников оценки функционирования.

Далее приведено руководство по получению критерия оценки качества работы при небольшом количестве участников, когда для получения критерия функционирования используют результаты участников.

D.1.2 Процедуры идентификации выбросов

Хотя для загрязненных выбросами генеральных совокупностей настоятельно рекомендуется использование робастных статистик, для очень небольших наборов данных их все же не рекомендуют (исключения приведены ниже). Проверка на наличие выбросов для очень маленьких наборов данных, однако, возможна. В случае очень маленьких программ или групп предпочтительным является отклонение выброса с последующим вычислением среднего или стандартного отклонения.

Различные критерии выявления выбросов применимы к различным размерам наборам данных. В *ГОСТ Р ИСО 5725-2* приведены таблицы теста Граббса для выявления единичного выброса и для двух одновременных выбросов в одном и том же направлении. В критерии Граббса и ряде других критериев необходимо установить заранее количество возможных выбросов, а при большом количестве выбросов эти критерии не выполняются, лучше всего они работают при $p > 10$ (в зависимости от возможной доли выбросов).

Примечание 1 — После исключения выбросов следует соблюдать осторожность при оценке дисперсии, так как оценка может быть смещена в меньшую сторону. Смещение обычно не очень большое, если исключают выбросы, выявляют с уровнем доверия 99 % и выше.

Примечание 2 — Большинство одномерных робастных оценок параметров положения и разброса приемлемы при $p \geq 12$.

D.1.3 Процедуры оценки параметра положения

D.1.3.1 Приписанные значения, полученные из небольших наборов данных участников, должны по возможности удовлетворять критерию неопределенности приписанного значения, приведенному в 9.2.1. В ситуации с использованием в качестве приписанного значения среднего и в качестве стандартного отклонения оценки стандартного отклонения результатов этот критерий не может быть использован для нормального распределения с $p \leq 12$ после удаления выбросов. При использовании медианы в качестве приписанного значения (с эффективностью 0,64) критерий не может быть использован для $p \leq 18$. Другие робастные оценки, такие как в алгоритме *A* (С.3), имеют промежуточную эффективность и могут соответствовать критерию при $p > 12$, если учтены положения примечания 2 к 7.7.3.

D.1.3.2 Существуют ограничения на объем набора данных, применяемых для определения некоторых оценок параметра положения. Рекомендуются несколько численных робастных оценок среднего набора данных

небольшого объема. Нижний предел, как правило, составляет $p \le 15$, хотя провайдеры могут иметь возможность продемонстрировать приемлемую работу с учетом установленных предположений для меньших наборов данных. Медиана применима для меньших объемов данных вплоть до $p = 2$ (если она равна среднему), но при $3 \le p \le 5$ медиана обладает небольшим преимуществом по сравнению со средним, за исключением тех случаев, когда существует необычно высокий риск получения плохих результатов.

D.1.4 Процедуры оценки дисперсии

D.1.4.1 Не рекомендуется использовать критерии функционирования, основанные на разбросе результатов участников для набора данных небольшого объема из-за очень высокой изменчивости всех оценок разброса. Например, при $p = 30$ оценки стандартного отклонения для данных из нормального распределения в среднем отклоняются от истинного значения не более чем на 25 % (с уровнем доверия 95 %). Для данных из нормального распределения не существует лучших оценок.

D.1.4.2 Если оценки разброса необходимы для других целей (например, как суммарные статистики или оценка разброса данных для робастной оценки параметра положения) или если программа проверки квалификации устойчива к высокой изменчивости оценок разброса данных, для небольших наборов данных следует выбирать оценки разброса с самой высокой доступной эффективностью.

Примечание 1 — Под высокой доступностью следует понимать наличие программного обеспечения и соответствующего опыта.

Примечание 2 — Оценка Q_n стандартного отклонения, описанная в С.5, является значительно более эффективной, чем *MADe* или *nIQR* из С.1.

Примечание 3 — При очень небольших наборах данных для робастных оценок разброса данных необходимо использовать следующие рекомендации [2]:

- для $p = 2$ необходимо использовать $|x_1 - x_2| / \sqrt{2}$;
- $p = 3$, если параметр положения и шкала неизвестны, необходимо использовать *MADe* для защиты от чрезмерно высоких оценок стандартного отклонения или среднего абсолютного значения отклонения, для защиты от слишком маленьких оценок стандартного отклонения, например, если из-за ошибок округления могут быть получены два одинаковых значения;
- $p \ge 4$ необходимо использовать установленную *M*-оценку стандартного отклонения, полученную на основе логарифмически взвешенной функции, рекомендуемой в [19], а также близкий эквивалент алгоритма *A* без итерации при определении параметра положения с использованием медианы в качестве оценки параметра положения.

Примечание 4 — Для получения оценки стандартного отклонения на основе абсолютного значения разности x_i и медианы используют следующую формулу:

$$s^* = \frac{1}{0{,}798 \cdot p} \sum_{i=1}^{p} |x_i - med(x)|. \qquad (D.1)$$

D.2 Эффективность и пороговые точки робастных процедур

D.2.1 Различные статистические оценки (робастные методы) можно сопоставлять по трем показателям:

- пороговая точка — доля значений в наборе данных, которые можно заменить сколь угодно большими значениями, без того чтобы оценка также стала сколь угодно большой;
- эффективность оценки — отношение дисперсии оценки к дисперсии минимальной оценки дисперсии для рассматриваемого распределения;
- устойчивость к противоречивым результатам — способность оценки быть устойчивой к небольшому количеству противоречивых результатов (как правило, менее 20 % набора данных).

Эти показатели в значительной степени зависят от распределения результатов участников и особенностей результатов, полученных некомпетентными участниками (участниками, которые не следуют инструкциям или методике измерений). Загрязнение данных может проявиться в виде наличия выбросов, результатов с большой дисперсией или результатов с различными средними (бимодальное распределение).

Пороговые точки и эффективность оценок различны для различных ситуаций, их тщательный анализ выходит за рамки настоящего стандарта. Однако в предположении о нормальном распределении данных могут быть сделаны простые сопоставления результатов лабораторий со средним, равным x_{pt}, и стандартным отклонением, равным σ_{pt}.

D.2.2 Пороговая точка

Пороговая точка — доля выбросов в наборе данных, которая не влияет на оценку неблагоприятным образом. Пороговая точка — это мера устойчивости к выбросам, высокое значение этой точки говорит об устойчивости к наличию большой доли выбросов. Пороговые точки и устойчивость к противоречивым результатам оценок, приведенных в приложении С, представлены в таблице D.1. Следует отметить, что для процедуры, приведенной в 6.3 и 6.4, необходимо проводить предварительный анализ данных и не использовать данные с большим количеством выбросов. Однако существуют ситуации, в которых визуальный анализ нецелесообразен.

Таблица D.1 — Пороговая точка для оценки среднего и стандартного отклонений (доля выбросов, которая может привести к несостоятельности оценки)

Статистическая оценка	Оцениваемый параметр совокупности	Пороговая точка, %	Устойчивость к противоречивым результатам
Выборочное среднее	Среднее	0	Плохая
Выборочное стандартное отклонение	Стандартное отклонение	0	Плохая
Выборочная медиана	Среднее	50	Хорошая
nIQR	Стандартное отклонение	25	Умеренная
MADe	Стандартное отклонение	50	Умеренно хорошая
Алгоритм *A*	Среднее и стандартное отклонения	25	Умеренная
Оценки Q_n и *Q*/Хампеля	Среднее и стандартное отклонения	50	Умеренная (очень хорошая для точки устойчивости, отстоящей более чем на 6 s^*)

Примечание — Определение пороговой точки, используемое здесь, сводится к определению доли большого набора данных из нормального распределения, которая может изменяться до бесконечности, без того чтобы оценка также двигалась к бесконечности. Например, если менее 50 % данных набора заменить на бесконечность, медиана останется конечной величиной.

Таким образом, выборочные среднее и стандартное отклонения могут дать недостоверную оценку при наличии единственного выброса. Робастные методы, использующие медиану *MADe* и *Q*/Хампеля, могут выдержать очень большую долю выбросов. Алгоритм *A* с итеративным стандартным отклонением и *nIQR* имеют пороговую точку 25 %. Надо помнить, что в любой ситуации при большой доле выбросов (> 20 %) как традиционные, так и робастные оценки могут дать смещенные оценки параметров положения и разброса, и это следует учитывать при интерпретации таких оценок.

D.2.3 Относительная эффективность

Все оценки имеют выборочную дисперсию, то есть оценки могут отличаться от раунда к раунду программы проверки квалификации, даже если все участники квалифицированные и нет выбросов или подгрупп участников с различными средними или дисперсиями. Робастные оценки видоизменяют представленные результаты, которые находятся слишком далеко от середины распределения, на основании теоретических предположений, и поэтому эти оценки имеют большую дисперсию, чем оценки с минимальной дисперсией, в том случае, когда набор данных фактически подчиняется нормальному распределению.

Выборочное среднее и стандартное отклонения являются оценками среднего и стандартного отклонений с минимальной дисперсией, и поэтому они имеют эффективность 100 %. Оценки с более низкой эффективностью имеют большую изменчивость, то есть они могут изменяться от раунда к раунду, даже если нет выбросов или различных подгрупп участников. В таблице D.2 приведена относительная эффективность оценок, представленных в приложении С.

Таблица D.2 — Относительная эффективность робастных оценок среднего и стандартного отклонений генеральной совокупности для нормально распределенного набора данных с *n* от 50 до 500 участников

В процентах

Статистическая оценка	Среднее *n* = 50	Среднее *n* = 500	SD *n* = 50	SD *n* = 500
Выборочное среднее и стандартное отклонения	100 %	100 %	100 %	100 %
Медиана и *nIQR*	66 %	65 %	38 %	37 %
Медиана и *MADe*	66 %	65 %	37 %	37 %
Алгоритм *A*	97 %	97 %	74 %	73 %
Q_n и *Q*/Хампеля	96 %	96 %	73 %	81 %

Согласно таблице D.2 становится очевидным, что не существует статистического метода, идеально подходящего во всех ситуациях. Выборочные среднее и стандартное отклонения являются оптимальными оценками

в случае нормального распределения данных, но неудачны в случае выбросов. Простые робастные методы, такие как медиана, *MADe* или *nIQR*, являются не очень хорошими для данных из нормального распределения, но могут быть эффективными при наличии выбросов или небольшом объеме данных.

D.3 Использование данных проверки квалификации для оценки воспроизводимости и повторяемости метода измерений

D.3.1 Во введении *ГОСТ ISO/IEC 17043—2013* установлено, что оценка свойств метода измерений, как правило, не является целью проверки квалификации. Тем не менее результаты программы проверки квалификации можно использовать для проверки и, возможно, установления повторяемости и воспроизводимости метода измерений [20], если программа проверки квалификации удовлетворяет следующим условиям:

a) образцы для проверки квалификации однородны и стабильны;

b) участники способны дать последовательные удовлетворительные результаты;

c) квалификация участников (или подгруппы участников) продемонстрирована до начала раунда проверки квалификации, и результаты проверки квалификации не ставят эту квалификацию под сомнение.

D.3.2 Для того чтобы обеспечить достаточное количество данных для оценки повторяемости и воспроизводимости метода испытаний в программе проверки квалификации, должны быть выполнены следующие условия:

a) при проведении исследований имеется достаточное количество участников, продемонстрировавших свою квалификацию в части выполнения метода измерений на предыдущих раундах программы проверки квалификации, которые приняли обязательства следовать методу измерений без изменений;

b) при оценке повторяемости в каждом раунде проверки квалификации должно быть использовано не менее двух образцов для проверки квалификации или выполнены репликации наблюдений;

c) по возможности участники должны быть обеспечены отдельно идентифицированными репликациями «вслепую», что является более предпочтительным, чем выполнение репликаций на одном и том же образце;

d) образцы, используемые в одном или нескольких раундах программы проверки квалификации, должны охватывать весь диапазон уровней и типов обычных образцов, для которых предназначен метод измерений;

e) процедуры анализа данных, применяемые для оценки повторяемости и воспроизводимости, должны соответствовать стандартам серии *ГОСТ Р ИСО 5725* или используемому совместному протоколу исследования.

Приложение Е
(справочное)

Иллюстративные примеры

В данном приложении приведены примеры, иллюстрирующие процедуры, установленные в настоящем стандарте, и позволяющие пользователям настоящего стандарта проверить правильность своих вычислений. Приведенные примеры не следует рассматривать как рекомендации для использования в конкретных программах проверки квалификации.

Е.1 Влияние цензурированных данных (см. 5.5.3.3)

В таблице Е.1 приведены 23 результата раунда программы проверки квалификации, из которых пять результатов представлены в виде «менее некоторого числа». Робастные среднее x^* и стандартное отклонение s^* в соответствии с алгоритмом *A* вычисляют тремя различными способами:

1) знак «<» игнорируют, а данные анализируют как обычные количественные данные;

2) данные со знаком «<» удаляют;

3) данные со знаками «<» и «>» заменяют половиной их значений, а результаты используют для определения оценки как количественные результаты.

В каждом варианте результаты, выпадающие за допустимые границы, помечены знаком «#». Это означает, что в результате будет получена неприемлемая оценка (сигнал к действиям) для любых результатов, количественное значение которых находится вне пределов ($x^* \pm 3s^*$). Провайдер может использовать альтернативные правила для действий с результатами со знаками «<» или «>».

Таблица Е.1 — Пример набора данных с цензурированными (<) результатами и три способа работы с такими результатами

Участник	Результат	Набор данных с прогнозированным знаком «<»	Набор данных после исключения данных со знаком «<»	Набор данных после результатов со знаком «>» и «<» на половинное значение
A	< 10	10	—	5
B	< 10	10	—	5
C	12	12	12	12
D	19	19	19	19
E	< 20	20	—	10
F	20	20	20	20
G	23	23	23	23
H	23	23	23	23
J	25	25	25	25
K	25	25	25	25
L	26	26	26	26
M	28	28	28	28
N	28	28	28	28
P	< 30	30	—	15
Q	28	28	28	28
R	29	29	29	29
S	30	30	30	30
T	30	30	30	30

Окончание таблицы Е.1

Участник	Результат	Набор данных с прогнозированным знаком «<»	Набор данных после исключения данных со знаком «<»	Набор данных после результатов со знаком «>» и «<» на половинное значение
U	31	31	31	31
V	32	32	32	32
W	32	32	32	32
Y	45	45	45 #	45
Z	< 50	50 #	—	25
Итого				
Количество результатов	23	23	18	23
x^*		26,01	26,81	23,95
s^*		7,23	5,29	8,60

Выбор способа обработки данных со знаком «<» оказывает существенное влияние на робастное среднее и стандартное отклонения, а также на оценку функционирования. Провайдер проверки квалификации должен определить соответствующий метод.

Е.2 Тест на однородность и стабильность. Содержание мышьяка As в шоколаде (см. 6.1)

Образцы для проверки квалификации подготавливают для использования при международной проверке квалификации и затем используют как стандартный образец. Изготовлено 1000 флаконов.

Проверка однородности: выбирают 10 образцов, используя стратифицированный случайный отбор образцов из различных порций производственного процесса. Две исследуемые порции извлекают из каждой бутылки и проверяют в случайном порядке. Данные приведены в таблице Е.2. Итоговые статистики получены в соответствии с процедурой, описанной в В.3 приложения В. Целевое значение σ_{pt} для As в шоколаде составляет 15 %, таким образом, оценку выборочной изменчивости сопоставляют при проверке с 0,3 σ_{pt}.

Таблица Е.2 — Данные для проверки однородности данных при контроле содержания мышьяка в шоколаде

Номер бутылки	Проба 1	Проба 2
3	0,185	0,194
111	0,187	0,189
201	0,182	0,186
330	0,188	0,196
405	0,191	0,181
481	0,188	0,180
599	0,187	0,196
704	0,177	0,186
766	0,179	0,187
858	0,188	0,196

Общее среднее: 0,18715.

SD среднего арифметического: 0,00398.

s_w: 0,00556.

s_{s^*}: 0,00060.

σ_{pt}: 0,18715 · 0,15 = 0,02807.

Проверочное значение: 0,3 σ_{pt} = 0,00842.

Значение s_s менее проверочного значения, следовательно, однородность является достаточной.

Проверка стабильности: два образца для проверки квалификации отбирают случайным образом и хранят при повышенной температуре (60 °С) в течение всего раунда программы проверки квалификации (6 нед.). Образцы протестированы в двух экземплярах (см. таблицу Е.3), и четыре результата проверены на однородность значений.

Таблица Е.3 — Данные для проверки стабильности образцов на содержание мышьяка в шоколада

Выбранный образец	Проба 1	Проба 2
164	0,191	0,198
732	0,190	0,196

Общее среднее: 0,19375.

Разность со средним при проверке однородности: 0,19375 – 0,18715 = 0,00660.

Проверочное значение: 0,3σ_{pt} = 0,00842.

Разность средних меньше проверочного значения, следовательно, стабильность является достаточной.

Е.3 Содержание атразина в питьевой воде

В программе проверки квалификации по определению содержания гербицида (атразина) в питьевой воде участвуют 34 участника. В таблице Е.4 представлены исходные данные, упорядоченные по возрастанию, а также значения робастных среднего и стандартного отклонений, рассчитанных в соответствии с алгоритмом *A* в процессе шести итераций, то есть до тех пор пока в робастных среднем и стандартном отклонениях перестают изменяться три значащих цифры после запятой. На рисунках Е.1, Е.2 и Е.3 представлены соответственно график упорядоченных данных, гистограмма и график плотности вероятностей.

В таблице Е.5 приведены оценки параметра положения среднего и стандартного отклонений, полученные с использованием классических и робастных методов. Неопределенность оценки параметра положения также приведена. Статистики для бутстреп-метода получены в соответствии с процедурами, описанными в [9], [10], а также с применением пакета программного обеспечения *R* [см. *R* 3.1.1]. На рисунке Е.4 показаны различные оценки параметра положения и оценки расширенной неопределенности $2u(x_{pt})$.

Таблица Е.4 — Вычисление робастных среднего и стандартного отклонений для содержания атразина в питьевой воде

	x_i	1-я итерация	2-я итерация	3-я итерация	4-я итерация	5-я итерация	6-я итерация
$x^*-\delta$		0,204163	0,199732	0,198466	0,198037	0,197865	0,197790
$x-\delta$		0,319837	0,315969	0,315871	0,316065	0,316185	0,316243
1	0,0400	0,2042	0,1997	0,1985	0,1980	0,1979	0,1978
2	0,0550	0,2042	0,1997	0,1985	0,1980	0,1979	0,1978
3	0,1780	0,2042	0,1997	0,1985	0,1980	0,1979	0,1978
4	0,2020	0,2042	0,2020	0,2020	0,2020	0,2020	0,2020
5	0,2060	0,2060	0,2060	0,2060	0,2060	0,2060	0,2060
6	0,2270	0,2270	0,2270	0,2270	0,2270	0,2270	0,2270
7	0,2280	0,2280	0,2280	0,2280	0,2280	0,2280	0,2280
8	0,2300	0,2300	0,2300	0,2300	0,2300	0,2300	0,2300
9	0,2300	0,2300	0,2300	0,2300	0,2300	0,2300	0,2300
10	0,2350	0,2350	0,2350	0,2350	0,2350	0,2350	0,2350
11	0,2360	0,2360	0,2360	0,2360	0,2360	0,2360	0,2360
12	0,2370	0,2370	0,2370	0,2370	0,2370	0,2370	0,2370
13	0,2430	0,2430	0,2430	0,2430	0,2430	0,2430	0,2430
14	0,2440	0,2440	0,2440	0,2440	0,2440	0,2440	0,2440

Окончание таблицы Е.4

	x_i	1-я итерация	2-я итерация	3-я итерация	4-я итерация	5-я итерация	6-я итерация
$x^*-\delta$		0,204163	0,199732	0,198466	0,198037	0,197865	0,197790
$x-\delta$		0,319837	0,315969	0,315871	0,316065	0,316185	0,316243
15	0,2450	0,2450	0,2450	0,2450	0,2450	0,2450	0,2450
16	0,2555	0,2555	0,2555	0,2555	0,2555	0,2555	0,2555
17	0,2600	0,2600	0,2600	0,2600	0,2600	0,2600	0,2600
18	0,2640	0,2640	0,2640	0,2640	0,2640	0,2640	0,2640
19	0,2670	0,2670	0,2670	0,2670	0,2670	0,2670	0,2670
20	0,2700	0,2700	0,2700	0,2700	0,2700	0,2700	0,2700
21	0,2730	0,2730	0,2730	0,2730	0,2730	0,2730	0,2730
22	0,2740	0,2740	0,2740	0,2740	0,2740	0,2740	0,2740
23	0,2740	0,2740	0,2740	0,2740	0,2740	0,2740	0,2740
24	0,2780	0,2780	0,2780	0,2780	0,2780	0,2780	0,2780
25	0,2811	0,2811	0,2811	0,2811	0,2811	0,2811	0,2811
26	0,2870	0,2870	0,2870	0,2870	0,2870	0,2870	0,2870
27	0,2870	0,2870	0,2870	0,2870	0,2870	0,2870	0,2870
28	0,2880	0,2880	0,2880	0,2880	0,2880	0,2880	0,2880
29	0,2890	0,2890	0,2890	0,2890	0,2890	0,2890	0,2890
30	0,2950	0,2950	0,2950	0,2950	0,2950	0,2950	0,2950
31	0,2960	0,2960	0,2960	0,2960	0,2960	0,2960	0,2960
32	0,3110	0,3110	0,3110	0,3110	0,3110	0,3110	0,3110
33	0,3310	0,3198	0,3160	0,3159	0,3161	0,3162	0,3162
34	0,4246	0,3198	0,3160	0,3159	0,3161	0,3162	0,3162
Среднее арифметическое	0,2512	0,2579	0,2572	0,2571	0,2570	0,2570	0,2570
SD	0,0672	0,0342	0,0345	0,0347	0,0348	0,0348	0,0348
δ		0,0578	0,0581	0,0587	0,0590	0,0592	0,0592
Новое x^*	0,2620	0,2579	0,2572	0,2571	0,2570	0,2570	0,2570
Новое s^*	0,0386	0,0387	0,0391	0,0393	0,0394	0,0395	0,0395

Таблица Е.5 — Итоговые статистики для примера Е.3

Процедура определения оценки	Параметр положения (среднее)	Стандартное отклонение	$u(x_{pt})$
Определение робастной оценки: Медиана *nIQR*(*MADe*)	0,2620	0,0402 (0,0386)	0,0086
Определение робастной оценки: в соответствии с алгоритмом *A* (x^*, s^*)	0,2570	0,0395	0,0085
Определение робастной оценки: в соответствии с методом *Q*/Хампеля	0,2600	0,0426	0,0091
Бутстреп метод для среднего	0,2503	0,0667	0,0113

Окончание таблицы Е.5

Процедура определения оценки	Параметр положения (среднее)	Стандартное отклонение	$u(x_{pt})$
Вычислительная процедура с исключением выбросов	0,2588	0,0337	0,0061
Вычислительная процедура без исключения выбросов	0,2512	0,0672	0,0115

Примечание — Различные коммерческие пакеты программ используют различные процедуры расчета квартилей, что может привести к заметным различиям в значениях *nIQR*. Незначительные отклонения от приведенных выше значений могут быть вызваны этими различиями или различиями при округлении.

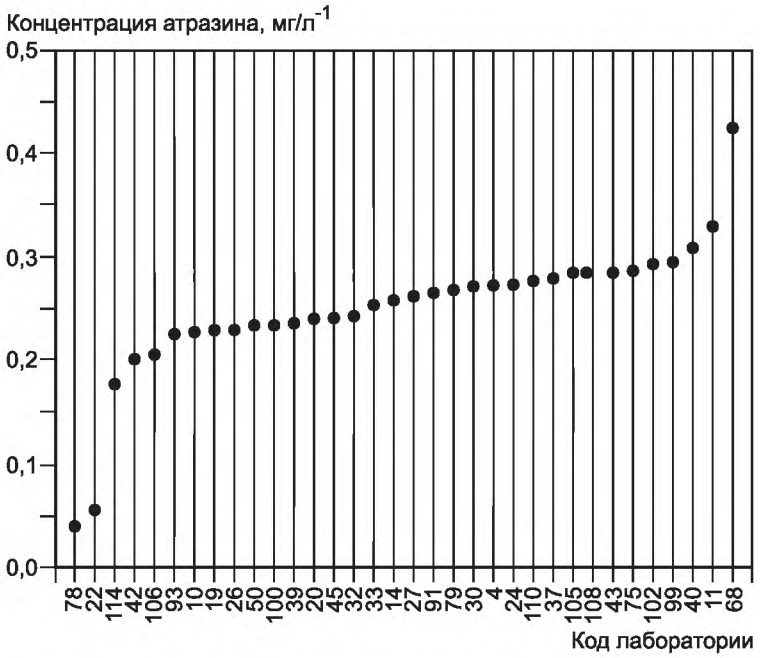


Рисунок Е.1 — Упорядоченные значения содержания атразина в питьевой воде (см. данные таблицы Е.4)

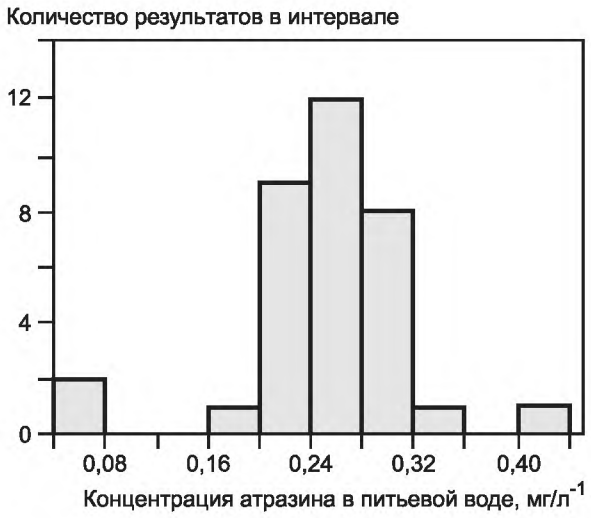


Рисунок Е.2 — Гистограмма результатов участников



Рисунок Е.3 — График плотности по результатам участников

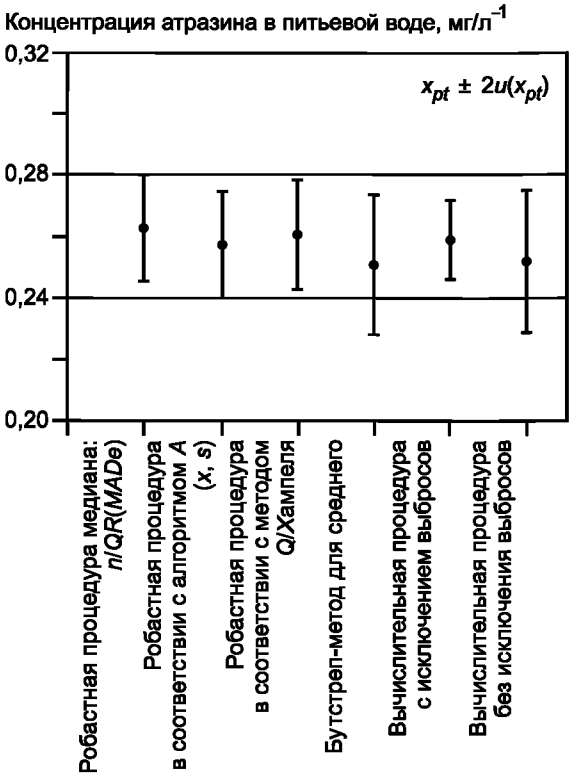


Рисунок Е.4 — Итоговые робастные статистики по данным таблицы Е.5

Е.4 Содержание ртути в корме для животных

В раунде программы проверки квалификации участников просили фиксировать свои результаты так, как они это обычно делают, а также расширенную неопределенность U_{lab} и коэффициент охвата k. Затем провайдер вычислял стандартную неопределенность u_{lab} в виде U_{lab}/k. Флажки присваивали полученным неопределенностям в соответствии с критериями 9.8. Данные, приведенные в таблицах Е.6 и Е.7, показывают общее содержание ртути в корме. Приведенная в таблице Е.6 стандартная неопределенность u_{lab} получена

на основе, указанной участниками расширенной неопределенности U_{lab} в виде U_{lab}/k и приведена с округлением. Для расчета статистик, приведенных в таблице Е.7, использованы неокругленные значения u_{lab}. Участник с кодом *L*23 не сообщил коэффициент охвата, поэтому использовано значение 1,732 (корень квадратный из 3, округленный).

Индексы работы вычислены с использованием методов, описанных в разделе 9. Для всех расчетов в качестве x_{pt} использовалось опорное значение, а в качестве σ_{pt} — значение функции пригодности назначению, определенное на основе предыдущего опыта. Неопределенность приписанного значения определена в виде суммы общей стандартной неопределенности опорного значения и неопределенности вследствие неоднородности (различия от бутылки к бутылке):

$$x_{pt} = 0{,}044 \text{ mg/kg}; \; U(x_{pt}) = 0{,}0082 \text{ mg/kg}; \; \sigma_{pt} = 0{,}0066 \text{ mg/kg} \; (= 15 \; \%).$$

На графике плотности (см. рисунок Е.6) показано бимодальное распределение, связанное с применением различных методов. Однако это не влияет на оценку функционирования, поскольку в качестве x_{pt} использовано опорное значение, а в качестве σ_{pt} значение функции пригодности назначению. При выполнении этого анализа результаты со знаком «<» удалены.

Таблица Е.6 — Результаты проверки квалификации для 24 участников при исследовании IMEP 111

Код лаборатории	Значение	U_{lab}	*k*	u_{lab}	Критерий	Метод
*L*04	0,013	0,003	2	0,002	b	AMA
*L*05	0,013	0,007	2	0,004	a	AMA
*L*23	0,0135	0,00108	1,732	0,00062	b	AMA
*L*02	0,014	0,004	2	0,002	b	AMA
*L*15	0,014	0,0005	2	0,0003	b	AMA
*L*17	< 0,015	—	—	—	—	CV—ICP—AES
*L*06	0,016	0,003	2	0,002	b	AMA
*L*09	0,017	0,008	2	0,004	a	AMA
*L*26	0,019	0,003	2	0,002	b	AAS
*L*12	0,0239	0,0036	2	0,0018	b	AMA
*L*13	< 0,034	—	—	—	—	TDA—AAS
*L*03	0,037	0,013	2	0,007	a	CV—AAS
*L*29	0,039	0,007	2	0,004	a	CV—AAS
*L*07	0,04	0,008	2	0,004	a	ICP—MS
*L*21	0,04	0,03	2	0,02	c	HG—AAS
*L*25	0,040	0,010	2	0,005	a	CV—AAS
*L*16	0,0424	0,008	2	0,004	a	CV—AAS
*L*08	0,044	0,007	2	0,004	a	CV—AAS
*L*10	0,045	0,007	2	0,004	a	ICP—MS
*L*24	0,045	0,005	2	0,003	a	HG—AAS
*L*18	0,046	0,007	2	0,004	a	CV—AAS
*L*28	0,049	0,0072	2	0,0036	a	CV—AAS

Окончание таблицы Е.6

Код лаборатории	Значение	U_{lab}	k	u_{lab}	Критерий	Метод
*L*01	0,053	0,007	2	0,004	a	CV—AAS
*L*14	< 0,1	—	—	—	—	ICP—MS



Рисунок Е.5 — Результаты участников и соответствующие неопределенности при определении концентрации IMEP 111 (данные таблицы Е.6)

Линии с крупным пунктиром показывают границы со значениями $x_{pt} \pm U(x_{pt})$, а линии с мелким пунктиром — границы со значениями $(x_{pt} \pm 2\ \sigma_{pt})$.

Вертикальные линии, заканчивающиеся открытым кружком, показывают значения со знаком «<».

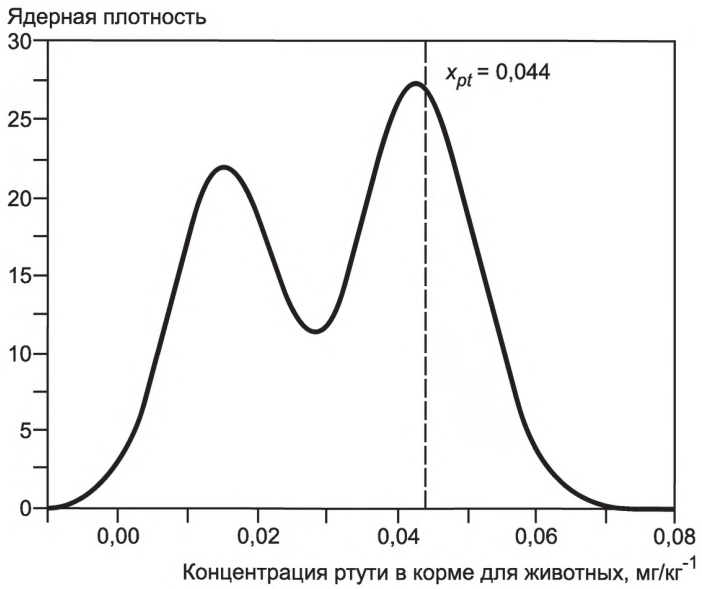


Рисунок Е.6 — График плотности распределения результатов участников

Таблица Е.7 — Статистики функционирования для различных методов

Код лаборатории	D, %	P_a	z	z'	ξ	E_n
*L*04	– 70,5 %	– 156,6 %	– 4,70	– 3,99	– 7,10	– 3,55
*L*05	– 70,5 %	– 156,6 %	– 4,70	– 3,99	– 5,75	– 2,88
*L*23	– 69,3 %	– 154,0 %	– 4,62	– 3,93	– 7,35	– 3,69
*L*02	– 68,2 %	– 151,5 %	– 4,55	– 3,86	– 6,58	– 3,29
*L*15	– 68,2 %	– 151,5 %	– 4,55	– 3,86	– 7,30	– 3,65
*L*17	—	—	—	—	—	—
*L*06	– 63,6 %	– 141,4 %	– 4,24	– 3,60	– 6,41	– 3,21
*L*09	– 61,4 %	– 136,4 %	– 4,09	– 3,47	– 4,71	– 2,36
*L*26	– 56,8 %	– 126,3 %	– 3,79	– 3,22	– 5,73	– 2,86
*L*12	– 45,7 %	– 101,5 %	– 3,05	– 2,59	– 4,49	– 2,24
*L*13	—	—	—	—	—	—
*L*03	– 15,9 %	– 35,4 %	– 1,06	– 0,90	– 0,91	– 0,46
*L*29	– 11,4 %	– 25,3 %	– 0,76	– 0,64	– 0,93	– 0,46
*L*07	– 9,1 %	– 20,2 %	– 0,61	– 0,51	– 0,70	– 0,35
*L*21	– 9,1 %	– 20,2 %	– 0,61	– 0,51	– 0,26	– 0,13
*L*25	– 9,1 %	– 20,2 %	– 0,61	– 0,51	– 0,62	– 0,31
*L*16	– 3,6 %	– 8,1 %	– 0,24	– 0,21	– 0,28	– 0,14
*L*08	0,0 %	0,0 %	0,00	0,00	0,00	0,00
*L*10	2,3 %	5,1 %	0,15	0,13	0,19	0,09
*L*24	2,3 %	5,1 %	0,15	0,13	0,21	0,10
*L*18	4,5 %	10,1 %	0,30	0,26	0,37	0,19
*L*28	11,4 %	25,3 %	0,76	0,64	0,92	0,46
*L*01	20,5 %	45,5 %	1,36	1,16	1,67	0,83
*L*14	—	—	—	—	—	—

Данный пример предоставлен Институтом эталонных материалов и измерений Объединенного исследовательского центра Европейской комиссии из Международной программы по оценке измерений (IMEP®), исследование 111.

Е.5 Опорное значение по данным единственной лаборатории: испытания по методу Лос-Анджелеса (см. 7.5)

В таблице Е.8 приведен пример данных, полученных в серии испытаний образцов для проверки квалификации и аналогичного сертифицированного стандартного образца *CRM*, который имеет сертифицированное значение показателя 21,62 единиц LA и соответствующую неопределенность 0,26 единиц LA. В примере показано, как опорное значение и неопределенность могут быть получены для образца проверки квалификации. Следует помнить, что неопределенность значения *CRM* включает в себя неопределенность, связанную с неоднородностью, транспортированием и долгосрочной стабильностью:

$$x_{pt} = 21{,}62 + 1{,}73 = 23{,}35,$$

$$u(x_{pt}) = \sqrt{0{,}26^2 + 0{,}24^2} = 0{,}35,$$

где 0,26 — стандартная неопределенность значения *CRM*, а 0,24 — стандартная неопределенность $\overline{d}$.

Таблица Е.8 — Вычисление разности средних арифметических, соответствующих *CRM* и образцам проверки квалификации, и стандартного отклонения этой разности

Выборка	Образец проверки квалификации		*CRM*		Разность средних испытуемого объекта и *CRM*, единицы LA
	Испытание 1, единицы LA	Испытание 2, единицы LA	Испытание 1, единицы LA	Испытание 2, единицы LA	
1	20,5	20,5	19,0	18,0	2,00
2	21,1	20,7	19,8	19,9	1,05
3	21,5	21,5	21,0	21,0	0,50
4	22,3	21,7	21,0	20,8	1,10
5	22,7	22,3	20,5	21,0	1,75
6	23,6	22,4	20,3	20,3	2,70
7	20,9	21,2	21,5	21,8	– 0,60
8	21,4	21,5	21,9	21,7	– 0,35
9	23,5	23,5	21,0	21,0	2,50
10	22,3	22,9	22,0	21,3	0,95
11	23,5	24,1	20,8	20,6	3,10
12	22,5	23,5	21,0	22,0	1,50
13	22,5	23,5	21,0	21,0	2,00
14	23,4	22,7	22,0	22,0	1,05
15	24,0	24,2	22,1	21,5	2,30
16	24,5	24,4	22,3	22,5	2,05
17	24,8	24,7	22,0	21,9	2,80
18	24,7	25,1	21,9	21,9	3,00
19	24,9	24,4	22,4	22,6	2,15
20	27,2	27,0	24,5	23,7	3,0
Разность средних арифметических $\bar{d}$					1,73
Стандартное отклонение					1,07
Стандартная неопределенность $\bar{d}$ (стандартное отклонение/$\sqrt{20}$)					0,24

Примечание — Данные представляют собой результат измерений механической прочности заполнителя, полученный при испытании методом Лос-Анджелеса.

Е.6 Пример бутстреп-метода для содержания бактерий группы кишечной палочки *Coliform* в образце пищи (см. 7.7.2)

В программе проверки квалификации на наличие бактерий группы кишечной палочки в пробе молока участвовали 35 лабораторий, которые выполняли по 5 репликаций измерений каждая. Среднее логарифмов CFU данных каждого участника было использовано для оценки приписанного значения и его неопределенности. Значение пригодности цели, равное 0,25 logCFU/ml, установлено в качестве σ_{pt}, в то время как стандартное отклонение функции ядерной плотности задано 0,75 σ_{pt}. График ядерной плотности (см. рисунок Е.7) имеет вид ассиметричного распределения. Бутстреп-метод (1000 репликаций) применен для определения оценки моды и соответствующей стандартной погрешности функции ядерной плотности распределения данных, обозначенных x_{pt} и $u(x_{pt})$, соответственно получены следующие значения:

$$x_{pt} = 3{,}79 \text{ и } u(x_{pt}) = 0{,}0922 \text{ (log CFU/мл)}.$$

Примечание — Поскольку $u(x_{pt}) > 0{,}3\ \sigma_{pt}$, оценка функционирования лаборатории определена с помощью z'-индекса.



Рисунок Е.7 — Плотность распределения результатов участников

Ниже приведена программа обработки приведенных данных на языке *R*.

```
################################
#LIBRARY TO DOWNLOAD AND TO USE
################################

library(boot)          #for bootstrap estimates
library(pastecs)       #for descriptive statistics


#DATA
#DATA
colif<-c(3.80, 3.90, 3.07, 3.64, 4.06, 3.40, 3.59, 3.39, 3.47, 3.47, 3.77, 3.53, 2.83,
2.75, 2.06, 3.75, 3.73, 3.82, 3.86, 3.88, 3.97, 3.96, 3.80, 3.88, 3.25, 3.45, 3.64, 2.86,
3.17, 3.19, 3.17, 4.22, 3.82, 3.82, 3.95)



#DESCRIPTIVE STATISTICS
options(digits = 3)  #number of decimal
stat.desc(colif)

#CONDITIONS
sigmat<-0.25          #standard deviation "fitness for purpose"
bw=0.75*sigmat        #standard deviation of kernel density


#HISTOGRAM AND KERNEL DENSITY GRAPH
hist(colif, freq=F,main="", cex.axis= 1.5,cex.lab=1.5, xlim=c(1,5) , ylim=c(0,1.5),
xlab="Coliforms (log10CFU/ml)",ylab="Kernel density", breaks=10)

lines(density(colif, kernel="gaussian", bw), col="black", lwd=3)


#FUNCTION TO DEFINE THE STATISTICS
theta<- function(y,i)
{
dens<-density(y[i],  kernel="gaussian",  bw=bw) mode<-dens$x[which.max(dens$y)]
}


#BOOTSTRAP MODE CALCULATION AND ITS UNCERTAINTY
set.seed(220)                              #START POINT OF BOOTSTRAP
boot.statistics<- boot(colif,theta,R=1000)
boot.statistics                            #MODE AND STANDARD ERROR
```

Е.7 Сопоставление опорного значения с согласованным средним (см. 7.8)

Для пояснения процедуры, приведенной в 7.8, сравнивают опорное значение с робастным средним по полученным результатам участников на основе примера Е.4 и данных таблицы Е.6.

В этом раунде программы проверки квалификации робастное среднее $x^* = 0,03161$ и робастное стандартное отклонение $s^* = 0,0164$ получены с помощью алгоритма *A*, после удаления трех результатов со знаком «<» ($n = 24$). Затем определена неопределенность робастного среднего

$$u(x^*) = 1,25(s^*/\sqrt{n}),$$

$$u(x^*) = 1,25(0,0164/\sqrt{24}) = 0,0042.$$

В соответствии с 7.8 неопределенность разности между x_{ref} и x^* имеет следующий вид:

$$u_{diff} = \sqrt{u^2(x_{ref}) + u^2(x^*)} = \sqrt{0,0041^2 + 0,0042^2} = 0,0059,$$

$$U_{diff} = 2(0,0059) = 0,012,$$

$$x_{diff} = x_{ref} - x^* = 0,044 - 0,032 = 0,012.$$

Таким образом, разность в два раза больше ее неопределенности.

Никаких действий не рекомендуется, так как в некоторых методах присутствует смещение.

Е.8 Определение критериев оценки на основании опыта предыдущих раундов: содержание токсафена в питьевой воде (см. 8.3)

Два провайдера организации программы проверки квалификации лаборатории проверяют содержание пестицида токсафена в питьевой воде.

В течение пяти лет проведено 20 раундов проверки квалификации, в которых каждый раз принимали участие 20 или более лабораторий, уровень токсафена в исследуемых пробах питьевой воды колебался от 3 до 20 мг/л. В таблице Е.9 представлены результаты 20 раундов проверки квалификации, упорядоченные в порядке убывания приписанных значений. На рисунках Е.8 и Е.9 приведены точечные диаграммы для относительного робастного стандартного отклонения *RSD*, %, и робастного стандартного отклонения *SD* для каждого раунда программы проверки квалификации по отношению к приписанному значению (рассчитанному по формуле). На каждом рисунке обозначена линия регрессии, полученная по методу наименьших квадратов. Линию регрессии по методу наименьших квадратов можно определить с помощью общедоступного программного обеспечения. (Полиномиальная модель 2-го порядка также проверена в качестве функции взаимосвязи стандартного отклонения и приписанного значения, но квадратичный член не признан статистически значимым, что указывает на отсутствие существенной кривизны линии регрессии для этой модели, следовательно, линейная модель более подходящая.)

Очевидно, что *RSD* является достаточно постоянным и составляет около 19 % для всех уровней, линия регрессии для стандартного отклонения достаточно достоверная ($R^2 = 0,82$). Регулирующий орган может потребовать, чтобы стандартное отклонение оценки составляло 19 % от приписанного значения (или, возможно, 20 %), или осуществить вычисления среднего стандартного отклонения на основе уравнения линии регрессии для стандартного отклонения.

Таблица Е.9 — Данные содержания токсафена в питьевой воде для $p \geq 20$ результатов

Код провайдера	Приписанное значение	Робастное среднее отклонение	Стандартное отклонение	Выполнение требований, %	*RSD*, %	*p*
*P*004	3,96	3,98	0,639	100,5	16,1	25
*P*001	4,56	5,18	0,638	113,6	14,0	23
*P*001	5,99	5,98	0,995	99,8	16,6	22
*P*004	6,08	5,80	1,48	95,4	24,3	20
*P*001	6,20	6,66	0,97	107,4	15,7	23
*P*001	6,72	7,13	1,43	106,1	21,3	22
*P*004	8,10	7,09	2,23	87,5	27,5	21
*P*001	8,73	8,15	1,80	93,4	20,6	22
*P*001	9,57	8,60	1,45	89,9	15,2	23
*P*001	12,1	12,4	1,44	102,5	11,9	23
*P*001	12,5	13,8	2,25	110,4	18,0	24

Окончание таблицы Е.9

Код провайдера	Приписанное значение	Робастное среднее отклонение	Стандартное отклонение	Выполнение требований, %	*RSD*, %	*p*
*P*004	13,1	12,0	2,41	91,6	18,4	20
*P*004	15,6	13,3	3,57	85,3	22,9	27
*P*004	15,9	13,6	2,44	85,5	15,3	28
*P*004	16,3	13,5	3,60	82,8	22,1	31
*P*004	16,3	14,2	3,09	87,1	19,0	40
*P*004	17,0	15,6	2,63	91,8	15,5	24
*P*004	17,4	16,0	2,85	92,0	16,4	23
*P*004	17,4	16,0	3,36	92,0	19,3	23
*P*004	19,0	16,4	3,20	86,3	16,8	27

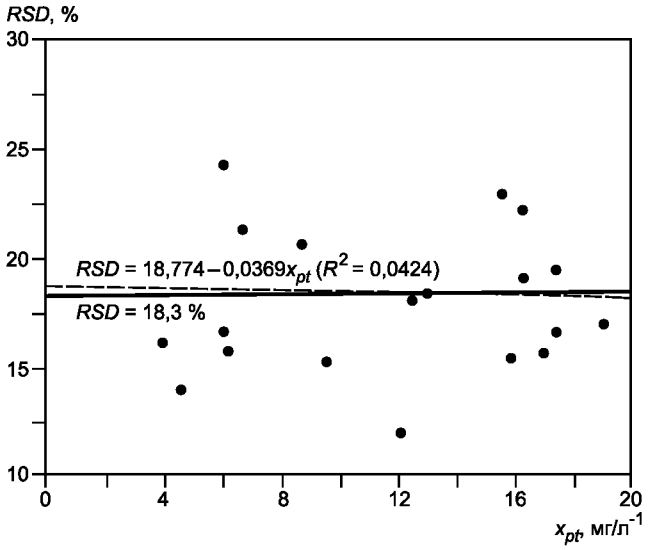


Рисунок Е.8 — Относительное стандартное отклонение результатов участников, %, и приписанное значение, мг/л

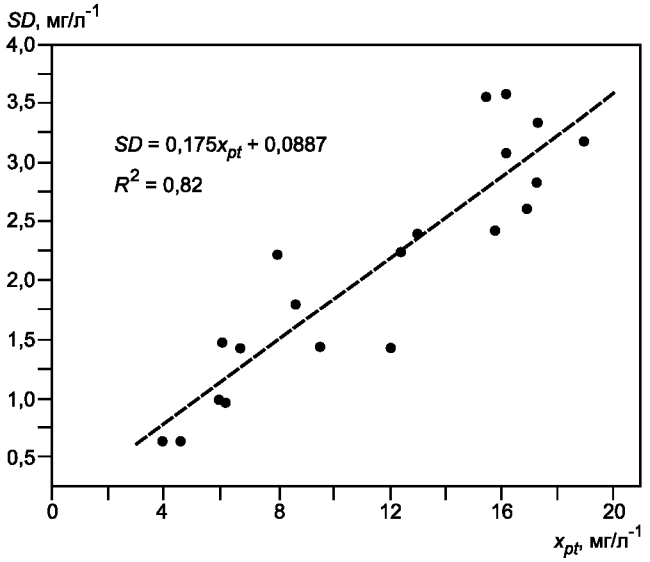


Рисунок Е.9 — Стандартное отклонение участников (мг/л) и приписанное значение (мг/л)

Е.9 Общая модель: уравнение Хорвица (см. 8.4)

Одна из общих моделей, применяемых в химии, описана Хорвицем [8], [7]. Этот подход дает общую модель воспроизводимости аналитических методов, которая может быть использована для получения следующего выражения для стандартного отклонения воспроизводимости:

$$\sigma_R = 0{,}02 \cdot c^{0{,}8495},$$

где c — содержание химических компонентов, определенное в массовых долях.

Например, в программе проверки квалификации по определению содержания меламина в сухом молоке использованы два образца с опорными уровнями A = 1,195 мг/кг и B = 2,565 мг/кг (0,000 001 195 и 0,000 002 565). Тогда среднее стандартное отклонение воспроизводимости имеет следующий вид для образца:

A (1,195 мг/кг): σ_R = 0,186 мг/кг или σ_R = 15,6 %,

B (2,565 мг/кг): σ_R = 0,356 мг/кг или σ_R = 13,9 %.

Е.10 Определение характеристик в экспериментах на прецизионность: определение содержания цемента в твердом бетоне.

Содержание цемента в бетоне обычно измеряют в единицах массы, кг/м^3. На практике бетон производят в соответствии с сортами, которые отличаются по содержанию цемента на 25 кг/м^3 друг от друга, и необходимо, чтобы участники имели возможность правильно определить сорт. По этой причине желательно, чтобы выбранное значение σ_{pt} составляло не более половины от 25 кг/м^3 (σ_{pt} < 12,5 кг/м^3).

В эксперименте на прецизионность получены следующие результаты для бетона со средним содержанием цемента 260 кг/м^3: σ_R = 23,2 кг/м^3 и σ_r = 14,3 кг/м^3. Предположим, что сделано m = 2 репликаций измерений.

Тогда в соответствии с формулой (9):

$$\sigma_{pt} = \sqrt{23{,}2^2 - 14{,}3^2(1-1/2)}\ \text{кг/м}^3 = 20{,}9\ \text{кг/м}^3.$$

Таким образом, цель σ_{pt} < 25/2 кг/м^3 = 12,5 кг/м^3 является неосуществимой на практике.

Примечание — В *ГОСТ Р ИСО 5725-2* $\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$, где σ_L — составляющая, характеризующая межлабораторную дисперсию.

В данном примере σ_L может быть вычислена следующим образом:

$$\sigma_L = \sqrt{\sigma_R^2 - \sigma_r^2} = \sqrt{(23{,}2^2 - 14{,}3^2)} = 18{,}3\ \text{кг/м}^3.$$

Е.11 Штриховые графики для нормированного смещения: концентрация антител (см. 10.4)

Значения z-индексов для раунда проверки квалификации с тремя связанными измеряемыми величинами (антителами) приведены на рисунке Е.10 на штриховом графике. Данные для двух из трех аллергенов приведены в таблице Е.10.

Из этого графика видно, что, например, лабораториям B и Z следует искать причину, которая влияет на все три уровня и дает примерно одинаковую величину смещения, в то время как для лабораторий K и P z-индекс зависит от типа антител.

Е.12 График Юдена: концентрация антител (см. 10.5)

В таблице Е.10 приведены данные, полученные с помощью исследования двух аналогичных образцов проверки квалификации при определении концентрации антител. Показанные на рисунке Е.11 индексы функционирования z основаны на робастных среднем и стандартном отклонениях, полученных в соответствии с алгоритмом A.

Из рисунка Е.11 видно, что значения двух участников (5 и 23) находятся в правом верхнем квадрате и, следовательно, могут иметь устойчивое положительное смещение. Лаборатория 26 имеет высокое значение z-индекса на образце аллергена B и отрицательное значение z-индекса (– 0,055) на образце аллергена A и поэтому может иметь низкую воспроизводимость.

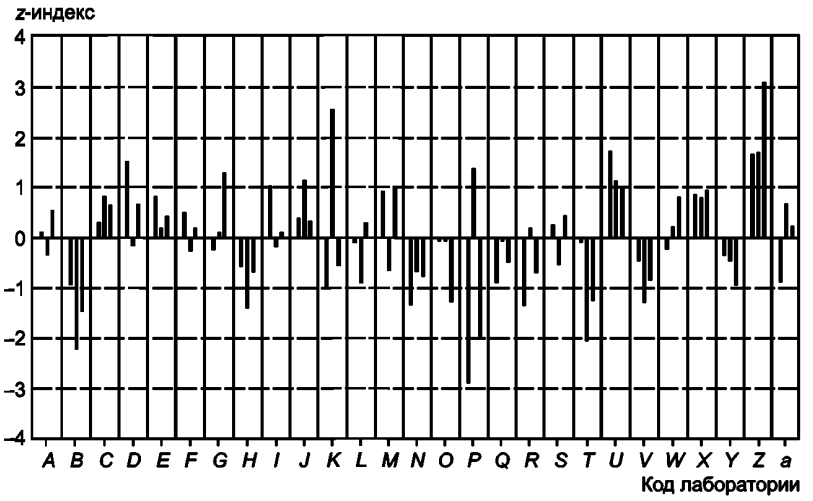


Рисунок Е.10 — Штриховой график z-индексов (от 4,0 до – 4,0)
для одного раунда программы проверки квалификации,
в которой участники определяли концентрации (IgE антител) для трех аллергенов

Результаты участников 5, 23 и 26 попадают в область сигнала «предупреждения». Эти участники должны проверить, в какую зону попадают их результаты в следующем раунде программы. Визуальный анализ и коэффициент корреляции указывают на тенденцию к последовательному изменению z-индекса (положительному или отрицательному), так что может существовать возможность улучшения метода измерений с более подробными инструкциями.

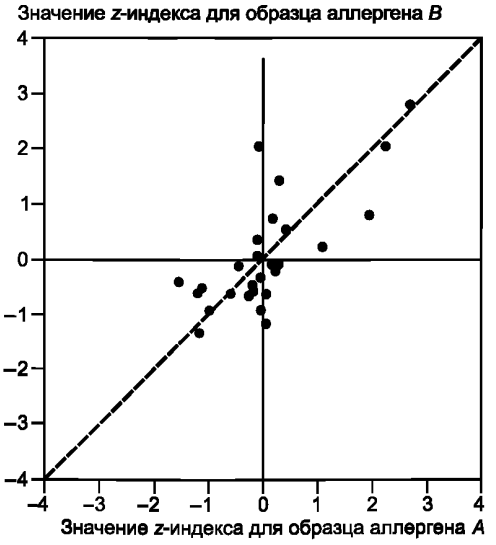


Рисунок Е.11 — График Юдена для z-индексов согласно таблице Е.10

Таблица Е.10 — Данные и вычисления по определению концентрации антител для двух аналогичных образцов аллергенов

Лаборатория	Данные		z-индексы	
i	Аллерген A $x_{A,i}$	Аллерген B $x_{B,i}$	Аллерген A $z_{A,i}$	Аллерген B $z_{B,i}$
1	12,95	9,15	0,427	0,515
2	6,47	6,42	– 1,540	– 0,428
3	11,40	6,60	– 0,043	– 0,366

Окончание таблицы Е.10

Лаборатория	Данные		z-индексы	
i	Аллерген A $x_{A,i}$	Аллерген B $x_{B,i}$	Аллерген A $z_{A,i}$	Аллерген B $z_{B,i}$
4	8,32	4,93	– 0,978	– 0,942
5	18,88	13,52	2,228	2,023
6	15,14	8,22	1,092	0,194
7	10,12	7,26	– 0,432	– 0,138
8	17,94	9,89	1,942	0,770
9	11,68	4,17	0,042	– 1,204
10	12,44	7,39	0,272	– 0,093
11	6,93	7,78	– 1,400	0,042
12	9,57	5,80	– 0,599	– 0,642
13	11,73	5,77	0,057	– 0,652
14	12,29	6,97	0,227	– 0,238
15	10,95	6,23	– 0,180	– 0,493
16	10,95	5,90	– 0,180	– 0,607
17	11,17	7,74	– 0,113	0,028
18	11,20	8,63	– 0,104	0,335
19	7,64	3,74	– 1,185	– 1,353
20	12,17	7,33	0,190	– 0,114
21	10,71	5,70	– 0,253	– 0,676
22	7,84	6,07	– 1,124	– 0,549
23	20,47	15,66	2,710	2,762
24	12,60	11,76	0,321	1,415
25	11,37	4,91	– 0,052	– 0,949
26	11,36	13,51	– 0,055	2,019
27	10,75	5,48	– 0,241	– 0,752
28	12,21	9,77	0,203	0,729
29	7,49	5,82	– 1,230	– 0,635
Среднее арифметическое	11,54	7,66	0,00	0,00
Стандартное отклонение	3,29	2,90	1,00	1,00
Коэффициент корреляции	0,706		0,706	

Примечание 1 — Данные выражаются в тысячах, к, единиц, U, на литр, л, образца, где единицы определяют концентрацию международного стандартного образца.

Примечание 2 — z-индексы в этой таблице вычислены с использованием неокругленных значений робастных среднего и стандартного отклонений, без использования округленных значений, приведенных в конце таблицы.

Е.13 График стандартных отклонений повторяемости: концентрации антител (см. 10.6)

В таблице Е.11 приведены результаты определения концентрации определенных антител в образце сыворотки крови. Каждый участник выполнил четыре репликации определения в условиях повторяемости. Для построения графика, представленного на рисунке Е.12, использована формула, приведенная выше. Согласно графику результаты некоторых лабораторий попадают в зону действия или предупреждения.

Таблица Е.11 — Концентрация антител в образце сыворотки крови (четыре репликации определения на каждом образце, выполненные каждым участником)

Номер лаборатории	Среднее арифметическое, kU/L	Стандартное отклонение, kU/L
1	2,15	0,13
2	1,85	0,21
3	1,80	0,08
4	1,80	0,24
5	1,90	0,36
6	1,90	0,32
7	1,90	0,14
8	2,05	0,26
9	2,35	0,39
10	2,03	0,53
11	2,08	0,25
12	1,25	0,24
13	1,13	0,72
14	1,00	0,26
15	1,08	0,17
16	1,20	0,32
17	1,35	0,4
18	1,23	0,36
19	1,23	0,33
20	0,90	0,43
21	1,48	0,40
22	1,20	0,55
23	1,73	0,39
24	1,43	0,30
25	1,28	0,22
Робастное среднее отклонение	1,57	
Робастное стандартное отклонение		0,34
Примечание — Данные представлены в тысячах единиц, kU, на литр, л, образца, где единицей является концентрация в международном стандартном образце.		

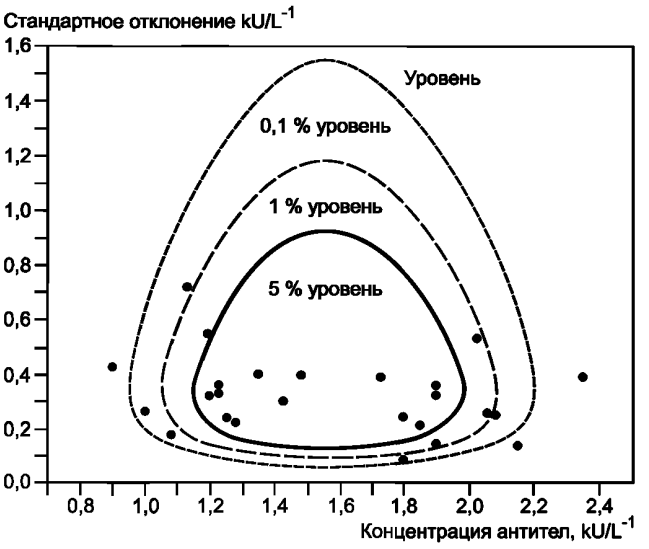


Рисунок Е.12 — График стандартного отклонения и среднего для 25 участников (см. данные таблицы Е.10)

Е.14 Графические методы отслеживания функционирования во времени (см. 10.8)

Участникам полезно отслеживать свою оценку работы с течением времени или иметь данные, подготовленные провайдером проверки квалификации. Самым простым методом контроля для этого является контрольная карта, или карта Шухарта. В связи с чем необходимо иметь стандартизованные индексы функционирования, такие как z-индекс или P_A-индекс, и участвовать в нескольких раундах. Данный пример относится к программе проверки квалификации медицинских лабораторий, в процессе которой участники определяют содержание калия в сыворотке крови.

Для приемки провайдер использовал фиксированный интервал ± 5 % с округлением значений до 0,1 ммоль/л, но не менее ± 0,2 ммоль/л приписанного значения. Провайдер использовал оценки P_A-индексов, а не z-индекс (см. таблицу Е.12).

Таблица Е.12 — Значения P_A-индексов за пять раундов программы проверки квалификации с тремя образцами сыворотки крови каждый

Код раунда	Образец для проверки квалификации	Результат	Приписанное значение	P_A-индекс	Среднее арифметическое значений P_A
101	*A*	6,4	6,2	75	42
101	*B*	4,2	4,1	50	—
101	*C*	4,1	4,1	0	—
102	*A*	6,0	5,9	25	8
102	*B*	4,3	4,4	– 33	
102	*C*	5,5	5,4	33	
103	*A*	4,1	4,2	– 33	– 28
103	*B*	3,6	3,7	– 50	—
103	*C*	4,2	4,2	0	—
104	*A*	5,7	5,8	– 25	11
104	*B*	3,9	4,0	– 50	—
104	*C*	6,3	5,9	110	—
105	*A*	3,6	3,7	– 50	– 19

Окончание таблицы Е.12

Код раунда	Образец для проверки квалификации	Результат	Приписанное значение	P_A-индекс	Среднее арифметическое значений P_A
105	*B*	4,5	4,6	– 33	—
105	*C*	5,3	5,2	25	—

Полученные результаты могут быть представлены на графике. Для визуального анализа рекомендуется использовать два вида графиков:

- контрольная карта для стандартизованного индекса функционирования для каждого раунда, показывающая результаты на нескольких образцах в одном и том же раунде проверки квалификации. Это позволяет показать характеристику функционирования во времени, в том числе выявить тенденции ее изменения (см. рисунок Е.13);
- точечная диаграмма стандартизированных индексов функционирования в зависимости от приписанных значений для анализа зависимости характеристики функционирования от уровня концентрации и выявления тенденций, связанных с уровнем измеряемой величины (см. рисунок Е.14).

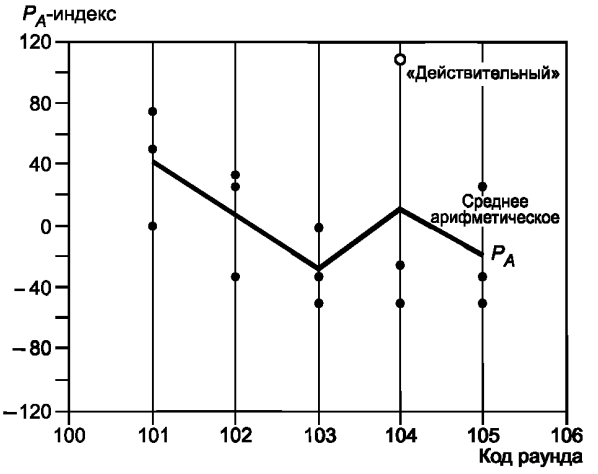


Рисунок Е.13 — Индексы функционирования для каждого раунда (см. данные таблицы Е.12)

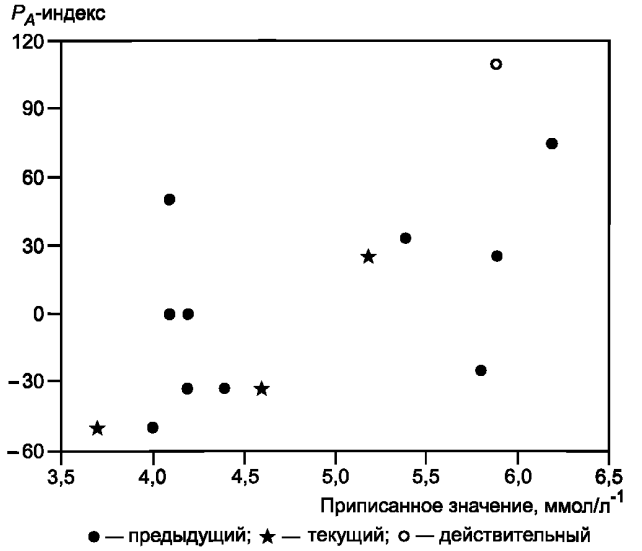


Рисунок Е.14 — Индексы функционирования для различных уровней измеряемой величины

Е.15 Качественный анализ данных: реакция кожи на косметическое средство (см. 11)

Программа проверки квалификации включает анализ реакции на продукцию, предназначенную для ухода за кожей, при ее применении к стандартному животному. Любую воспалительную реакцию оценивают по следующей шкале:

- 1 — отсутствие реакции;
- 2 — умеренное покраснение;
- 3 — значительное раздражение или отек;
- 4 — тяжелая реакция, в том числе нагноение или кровотечение.

Участникам представлены два образца, состоящие из двух различных продуктов, обозначенных как продукт *А* и продукт *В*. Каждый продукт исследуют 50 участников. Результаты участников приведены в таблице Е.13 и графически показаны на рисунке Е.15. Мода и медиана указаны по результатам участников для каждого образца проверки квалификации.

Таблица Е.13 — Результаты для двух образцов: раздражение кожи

Реакция	Продукт *А*	Продукт *В*
1	20 (40 %) #	8 (16 %)
2	18 (36 %) @	12 (24 %)
3	10 (20 %)	20 (40 %) # @
4	2 (4 %)	10 (20 %)
# — мода		
@ — медиана		

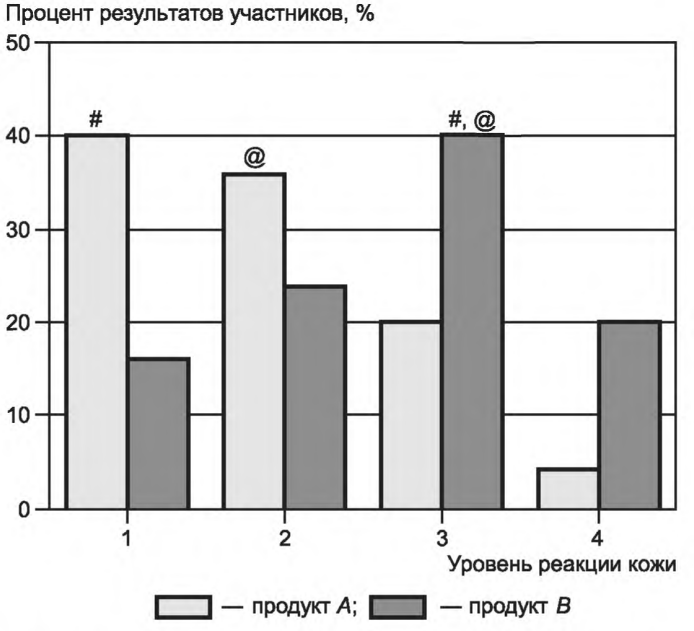


Рисунок Е.15 — Диаграмма процента результатов (появление раздражения) для двух образцов проверки квалификации

Следует учитывать, что медиана или мода могут быть использованы в качестве общей статистики для этих образцов, они указывают, что уровень реакции на продукт *В* является более тяжелым, чем реакция на продукт *А*. Провайдер может определить, что сигнал к действиям будет появляться для любого результата, который отстоит от медианы более чем на одну единицу измерения, в этом случае для продукта *А* — два результата «4» (4 %) попадают в зону сигнала к действиям, а для продукта *В* — восемь результатов «1» (16 %).

Приложение ДА
(справочное)

Сведения о соответствии ссылочных национальных и межгосударственных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте

Таблица ДА.1

Обозначение ссылочного национального и межгосударственного стандарта	Степень соответствия	Обозначение и наименование соответствующего международного стандарта
ГОСТ ISO/IEC 17043—2013	IDT	ISO/IEC 17043:2010 «Оценка соответствия. Основные требования к проведению проверки квалификации»
ГОСТ ISO Guide 35—2015	IDT	ISO Guide 35:2006 «Стандартные образцы. Общие и статистические принципы сертификации (аттестации)»
ГОСТ ИСО/МЭК 17025—2009	IDT	ISO/IEC 17025:2005 «Общие требования к компетентности испытательных и калибровочных лабораторий»
ГОСТ Р ИСО 5725-1—2002	IDT	ISO 5725-1:1994 «Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Основные положения и определения»
ГОСТ Р ИСО 5725-2—2002	IDT	ISO 5725-2:1994 «Точность (правильность и прецизионность) методов и результатов измерений. Часть 2. Основной метод определения повторяемости и воспроизводимости стандартного метода измерений»
ГОСТ Р ИСО 5725-5—2002	IDT	ISO 5725-5:1998 «Точность (правильность и прецизионность) методов и результатов измерений. Часть 5. Альтернативные методы определения прецизионности стандартного метода измерений»
ГОСТ Р ИСО 16269-4—2017	IDT	ISO 16269-4:2010 «Статистическое представление данных. Часть 4. Выявление и обработка выбросов»
ГОСТ Р 54500.3—2011/ Руководство ИСО/МЭК 98-3:2008	IDT	ISO/IEC Guide 98-3:2008 «Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения (GUM:1995)»
ГОСТ Р ИСО 7870-2—2015	IDT	ISO 7870-2:2013 «Контрольные карты. Часть 2. Контрольные карты Шухарта»
Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандартов: - IDT — идентичные стандарты.		

Библиография

[1] Thompson M., Ellison S.L.R., Wood R. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure Appl. Chem. 2006, 78 (1) pp. 145—196

[2] Kuselman I., Fajgelj A. IUPAC/CITAC Guide: Selection and use of proficiency testing schemes for a limited number of participants—chemical analyticallaboratories (IUPAC Technical Report). Pure Appl. Chem. 2010, 82 (5) pp. 1099—1135

[3] Helsel D.R. Nondetects and data analysis: statistics for censored environmental data. Wiley Interscience, 2005 Horwitz W. Evaluation of analytical methods used for regulations of food and drugs. Anal. Chem. 1982, 54 pp. 67A—76A

[4] CCQM Guidance note: Estimation of a consensus KCRV and associated Degrees of Equivalence. Version 10. Bureau International des Poids et Mesures, Paris (2013)

[5] Davison A.C., Hinkley D.V. Bootstrap Methods and Their Application. Cambridge University Press, 1997

[6] Efron B., Tibshirani R. An Introduction to the Bootstrap. Chapman & Hall, 1993

[7] Thompson M. Analyst (Lond.). 2000, 125 pp. 385—386

[8] Horwitz W. Evaluation of analytical methods used for regulations of food and drugs. Anal. Chem. 1982, 54 pp. 67A—76A

[9] Silverman B.W. Density Estimation. Chapman and Hall, London, 1986

[10] Sheather S.J., Jones M.C. A reliable data-based bandwidth selection method for kernel density estimation. J.R. Stat. Soc., B. 1991, 53 pp. 683—690

[11] van Nuland Y. ISO 9002 and the circle technique. Qual. Eng. 1992, 5 pp. 269—291

[12] Gower J.C. A general coefficient of similarity and some of its properties. Biometrics. 1971, 27 (4) pp. 857—871

[13] Thompson M., Willetts P., Anderson S., Brereton P., Wood R. Collaborative trials of the sampling of two foodstuffs, wheat and green coffee. Analyst (Lond.). 2002, 127 pp. 689—691

[14] Maronna R.A., Martin R.D., Yohai V.J. Robust Statistics: Theory and methods. John Wiley & Sons Ltd, Chichester, England, 2006

[15] Uhlig S. Robust estimation of variance components with high breakdown point in the 1-way random effect model. In: Kitsos C.P. and Edler L.; Industrial Statistics; Physica, S. 65—73, 1997

[16] Muller C.H., Uhlig S. Estimation of variance components with high breakdown point and high efficiency; Biometrika; 88: Vol. 2, pp. 353—366, 2001

[17] Uhlig S. Robust estimation of between and within laboratory standard deviation measurement results below the detection limit, Journal of Consumer Protection and Food Safety, 2015

[18] gost.e-ecolog.ru

[19] Rousseeuw P.J., Verboven S. Comput. Stat. Data Anal. 2002, 40 pp. 741—758

[20] Analytical Method Committee. Royal Society of Chemistry Accred Qual Assur. 2010, 15 pp. 73—79

УДК 658.562.012.7:65.012.122:006.354 ОКС 03.120.30 Т59

Ключевые слова: межлабораторные сравнительные испытания, проверка квалификации, приписанное значение, z-индекс, дзета-индекс, согласованная величина, выброс, образец для проверки квалификации, провайдер проверки квалификации, программа проверки квалификации, образец сравнения, сертифицированный образец сравнения *CRM*

БЗ 9—2017/217

Редактор *Л.С. Зимилова*
Технический редактор *В.Н. Прусакова*
Корректор *Е.И. Рычкова*
Компьютерная верстка *Ю.В. Поповой*

Сдано в набор 13.09.2017 Подписано в печать 19.10.2017. Формат 60 × 84$^{1}/_{8}$. Гарнитура Ариал.
Усл. печ. л. 10,23. Уч.-изд. л. 9,95. Тираж 28 экз. Зак. 2288.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ИД «Юриспруденция», 115419, Москва, ул. Орджоникидзе, 11.
gost.e-ecolog.ru gost.e-ecolog.ru

Издано и отпечатано во ФГУП «СТАНДАРТИНФОРМ», 123001, Москва, Гранатный пер., 4.
gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru