МЕЖГОСУДАРСТВЕННЫЙ СОВЕТ ПО СТАНДАРТИЗАЦИИ, МЕТРОЛОГИИ И СЕРТИФИКАЦИИ
(МГС)
INTERSTATE COUNCIL FOR STANDARDIZATION, METROLOGY AND CERTIFICATION
(ISC)

МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

ГОСТ ISO 13528— 2024

Статистические методы

ПРИМЕНЕНИЕ ПРИ ПРОВЕРКЕ КВАЛИФИКАЦИИ ПОСРЕДСТВОМ МЕЖЛАБОРАТОРНОГО СЛИЧЕНИЯ

(ISO 13528:2022, Statistical methods for use in proficiency testing by interlaboratory comparison, IDT)

Издание официальное

Москва
Российский институт стандартизации
2025

Предисловие

Цели, основные принципы и общие правила проведения работ по межгосударственной стандартизации установлены ГОСТ 1.0 «Межгосударственная система стандартизации. Основные положения» и ГОСТ 1.2 «Межгосударственная система стандартизации. Стандарты межгосударственные, правила и рекомендации по межгосударственной стандартизации. Правила разработки, принятия, обновления и отмены»

Сведения о стандарте

1 ПОДГОТОВЛЕН Закрытым акционерным обществом «Научно-исследовательский центр контроля и диагностики технических систем» (ЗАО «НИЦ КД») при участии Уральского научно-исследовательского института метрологии — филиала Федерального государственного унитарного предприятия «Всероссийский научно-исследовательский институт метрологии им. Д.И. Менделеева» (УНИИМ — филиал ФГУП «ВНИИМ им. Д.И. Менделеева») на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 5

2 ВНЕСЕН Межгосударственным техническим комитетом по стандартизации МТК 125 «Применение статистических методов»

3 ПРИНЯТ Межгосударственным советом по стандартизации, метрологии и сертификации (протокол от 27 декабря 2024 г. № 180-П)

За принятие проголосовали:

Краткое наименование страны по МК (ИСО 3166) 004—97	Код страны по МК (ИСО 3166) 004—97	Сокращенное наименование национального органа по стандартизации
Армения	AM	ЗАО «Национальный орган по стандартизации и метрологии» Республики Армения
Беларусь	BY	Госстандарт Республики Беларусь
Киргизия	KG	Кыргызстандарт
Россия	RU	Росстандарт
Таджикистан	TJ	Таджикстандарт
Узбекистан	UZ	Узбекское агентство по техническому регулированию

4 Приказом Федерального агентства по техническому регулированию и метрологии от 5 февраля 2025 г. № 53-ст межгосударственный стандарт ГОСТ ISO 13528—2024 введен в действие в качестве национального стандарта Российской Федерации с 1 июля 2025 г.

5 Настоящий стандарт идентичен международному стандарту ISO 13528:2022 «Статистические методы при проверке квалификации посредством межлабораторного сличения» («Statistical methods for use in proficiency testing by interlaboratory comparison», IDT).

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ 1.5 (подраздел 3.6).

При применении настоящего стандарта рекомендуется использовать вместо ссылочных международных стандартов соответствующие им межгосударственные стандарты, сведения о которых приведены в дополнительном приложении ДА.

Дополнительные сноски в тексте стандарта, выделенные курсивом, приведены для пояснения текста оригинала

6 ВВЕДЕН ВПЕРВЫЕ

Информация о введении в действие (прекращении действия) настоящего стандарта и изменений к нему на территории указанных выше государств публикуется в указателях национальных стандартов, издаваемых в этих государствах, а также в сети Интернет на сайтах соответствующих национальных органов по стандартизации.

В случае пересмотра, изменения или отмены настоящего стандарта соответствующая информация будет опубликована на официальном интернет-сайте Межгосударственного совета по стандартизации, метрологии и сертификации в каталоге «Межгосударственные стандарты»

© ISO, 2022
© Оформление. ФГБУ «Институт стандартизации», 2025

В Российской Федерации 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 Руководство по предварительному анализу образцов и результатов проверки квалификации 9
7 Определение приписанного значения и его стандартной неопределенности 13
8 Выбор критериев для оценивания функционирования участников 18
9 Расчет статистик функционирования 22
10 Графические методы для описания показателей функционирования 29
11 Разработка и анализ программ проверки квалификации по качественным данным (включая номинальные и порядковые свойства) 35
Приложение А (обязательное) Обозначения 40
Приложение В (справочное) Проверка образцов для проверки квалификации на однородность и стабильность 42
Приложение С (справочное) Робастный анализ 47
Приложение D (справочное) Дополнительные рекомендации по статистическим процедурам 55
Приложение Е (справочное) Иллюстративные примеры 59
Приложение F (справочное) Пример компьютерной программы для построения графика и анализа результатов проверки квалификации бутстреп-методом 80
Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов межгосударственным стандартам 81
Библиография 82

Введение

0.1 Цели проверки квалификации

Проверка квалификации включает использование межлабораторных сличений для определения способности выполнения участниками (которыми могут быть лаборатории, контролирующие органы или физические лица) конкретных испытаний или измерений, а также для проверки функционирования лабораторий. Основные цели проверки квалификации включают оценивание функционирования лабораторий, выявление проблем в лабораториях, установление результативности и сопоставимости методов испытаний или измерений, обеспечение лабораторией более высокого доверия заказчиков, подтверждение заявленной неопределенности, обучение участвующих лабораторий и др. Типичные задачи межлабораторных сличений детально описаны в ISO/IEC 17043. Применяемые план эксперимента и аналитические методы должны соответствовать заявленной(ым) цели (целям).

0.2 Обоснование определения показателей в программах проверки квалификации

Для проверки квалификации доступны и могут быть использованы различные стратегии определения показателей функционирования лаборатории. Несмотря на то, что конкретные способы вычислений отличаются друг от друга в различных программах проверки квалификации, в большинстве из них отклонение участника от приписанного значения сравнивают с числовым критерием, который используют для принятия решения о том, стоит ли «беспокоиться» по поводу полученного значения отклонения. Поэтому стратегии, используемые для определения приписанного значения и выбора критерия для оценки отклонений участников, имеют решающее значение. В частности, важно рассмотреть вопрос о том, следует ли определять приписанное значение и критерий для оценки отклонений независимо от результатов участников или его следует определять на основе представленных результатов. В настоящем стандарте предусмотрены обе стратегии. Однако необходимо обратить внимание на обсуждение (см. разделы 7 и 8) преимуществ и недостатков выбора приписанных значений и критериев для оценки отклонений, которые не выведены из результатов участников. В общем случае выбор приписанных значений и критериев независимо от результатов участников обладает преимуществом. Это частный случай критерия, используемого для оценки отклонения от приписанного значения, такого как стандартное отклонение для оценки квалификации или допустимое значение погрешности измерений, для которого особенно полезен последовательный выбор, основанный на пригодности результатов измерений, для конкретного использования.

0.3 ISO 13528 и ISO/IEC 17043

Настоящий стандарт обеспечивает основу для внедрения ISO/IEC 17043, в частности в отношении требований к статистическому плану эксперимента, валидации образцов для проверки квалификации, анализу результатов и представлению сводной статистики. В приложении B ISO/IEC 17043:2010 кратко описаны общие статистические методы, используемые в программах проверки квалификации. Настоящий стандарт дополняет ISO/IEC 17043 руководством по использованию статистических методов при экспериментальной проверке квалификации, отсутствующим в ISO/IEC 17043.

Определение проверки квалификации, приведенное в ISO/IEC 17043, также представлено в настоящем стандарте с примечаниями, описывающими различные виды проверки квалификации и набор планов эксперимента, которые могут быть использованы. Настоящий стандарт не охватывает все цели, планы эксперимента, матрицы (объекта) и измеряемые величины. Методы, представленные в настоящем стандарте, предназначены для широкого применения, особенно для вновь разработанных программ проверки квалификации. Предполагается, что статистические методы, используемые в конкретной программе проверки квалификации, будут развиваться по мере совершенствования этой программы; показатели, критерии оценивания и графические методы также будут усовершенствованы для лучшего удовлетворения конкретным потребностям целевой группы участников, органов аккредитации и регулирующих органов.

В настоящем стандарте приведено опубликованное руководство по проверке квалификации аналитических химических лабораторий [32], и дополнительно приведен более широкий набор процедур, позволяющих использовать валидированные методы измерений и качественной идентификации. Настоящее издание стандарта содержит большое количество статистических методов и руководств из первого издания, дополненных, по мере необходимости, ранее упомянутыми документами и более широкой областью применения ISO/IEC 17043. ISO/IEC 17043 включает методы проверки квалификации для отдельных лиц и проверяющих органов, включая ISO/IEC 17043:2010, приложение B, в котором рассмотрены качественные результаты.

Настоящий стандарт включает статистические методы, соответствующие другим международным стандартам, в частности стандартам, разработанным техническим комитетом ISO/TC 69/SC 6, а именно стандартам серии ISO 5725 «Точность: (правильность и прецизионность)». Методы также отражают требования других международных стандартов, где это уместно, и соответствуют ISO/IEC Guide 98-3 (GUM) и ISO/IEC Guide 99 (VIM).

0.4 Компетентность в области статистики

В ISO/IEC 17043 установлено, что для того, чтобы быть компетентным, провайдер проверки квалификации должен иметь доступ к проведению статистической экспертизы при оценке компетентности и должен уполномочить специальный персонал на проведение статистического анализа. Ни ISO/IEC 17043, ни настоящий стандарт не устанавливают, в чем состоит эта необходимая экспертиза. Для некоторых ситуаций полезно наличие ученой степени в области статистики, но, как правило, требованиям к компетентности могут удовлетворять лица, обладающие техническими знаниями в других областях, смежных с базовыми статистическими подходами, и имеющие опыт или подготовку в области общих методов, применимых при анализе данных в программах проверки квалификации. Если какой-либо специалист отвечает за статистический анализ и/или планирование эксперимента, очень важно, чтобы этот специалист имел опыт проведения межлабораторных сличений, даже если он имеет ученую степень в области статистики. Традиционное углубленное обучение в области статистики часто не включает в себя выполнение заданий по межлабораторным сличениям и истинные причины погрешностей измерений, возникающих при проверке квалификации, могут оказаться не выяснены. Руководство, приведенное в настоящем стандарте, не может представить всю необходимую информацию, обеспечивающую рассмотрение всех возможных ситуаций, и не может заменить опыт, полученный при проведении межлабораторных сличений.

0.5 Компьютерное программное обеспечение

Компьютерное программное обеспечение, необходимое для статистического анализа данных при проверке квалификации, может сильно различаться — от простой таблицы расчетов для небольших программ проверки квалификации с использованием известных опорных значений до сложного статистического программного обеспечения, используемого для статистических методов, основанных на итеративных вычислениях или других передовых численных методах. Для большей части методов, рассмотренных в настоящем стандарте, расчеты могут быть выполнены с применением обычных электронных таблиц, возможно, с использованием пользовательских процедур для конкретной программы проверки квалификации или анализа; некоторые методы требуют использования компьютерных программ, находящихся в свободном доступе. Во всех случаях предполагается, что пользователи проверяют достоверность и точность своих вычислений, особенно если пользователем введены специальные процедуры. Однако, даже если методы, приведенные в настоящем стандарте, уместны и правильно реализованы соответствующими компьютерными программами, они не могут быть применены без рассмотрения специалистом, обладающим квалификацией в технических и статистических вопросах, достаточной для понимания особенностей применения статистических методов и предположений, а также для выявления и рассмотрения отклонений, которые могут возникнуть на любом этапе программы проверки квалификации.

М Е Ж Г О С У Д А Р С Т В Е Н Н Ы Й С Т А Н Д А Р Т

Статистические методы

ПРИМЕНЕНИЕ ПРИ ПРОВЕРКЕ КВАЛИФИКАЦИИ ПОСРЕДСТВОМ МЕЖЛАБОРАТОРНОГО СЛИЧЕНИЯ

Statistical methods. Use in proficiency testing by interlaboratory comparison

Дата введения — 2025—07—01

1 Область применения

В настоящем стандарте приведено детальное описание статистических методов, которые провайдеры проверки квалификации могут использовать для разработки программ проверки квалификации и анализа данных, полученных в результате реализации этих программ. Настоящий стандарт включает рекомендации по интерпретации данных проверки квалификации, представленных участниками таких программ и органами по аккредитации.

Процедуры, описанные в настоящем стандарте, могут быть применены для демонстрации того, что результаты измерений, полученные лабораториями, контролирующими органами и физическими лицами, соответствуют указанным критериям приемлемого функционирования.

Настоящий стандарт применим к проверке квалификации, когда сообщаемые результаты являются либо количественными результатами измерений, либо качественными наблюдениями за исследуемыми образцами.

Примечание — Процедуры, установленные в настоящем стандарте, также могут быть применены для оценки мнений экспертов, если эти мнения или заключения представлены в форме, допускающей объективное сравнение с независимым опорным значением или согласованной статистикой. Например, при классификации образцов для проверки квалификации по известным категориям (при проверке того, что образцы для проверки квалификации происходят из одного и того же исходного источника) или при объективном сравнении результатов классификации, это позволяет применять положения настоящего стандарта по отношению к номинальным (качественным) свойствам.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты [для датированных ссылок применяют только указанное издание ссылочного стандарта, для недатированных — последнее издание (включая все изменения)]:

ISO 3534-1, Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability (Статистика. Словарь и условные обозначения. Часть 1. Общие статистические термины и термины, используемые в теории вероятностей)

ISO 3534-2, Statistics — Vocabulary and symbols — Part 2: Applied statistics (Статистика. Словарь и условные обозначения. Часть 2. Прикладная статистика)

ISO 5725-1, Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions (Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Общие принципы и определения

ISO/IEC 17043, Conformity assessment — General requirements for proficiency testing (Оценка соответствия. Основные требования к проведению проверки квалификации)

Издание официальное

ISO Guide 30, Reference materials — Selected terms and definitions (Стандартные образцы. Некоторые термины и определения)

ISO/IEC Guide 99, International vocabulary of metrology — Basic and general concepts and associated terms (VIM) (Международный словарь по метрологии. Основные и общие понятия и соответствующие термины (VIM))

3 Термины и определения

В настоящем стандарте применены термины по ISO 3534-1, ISO 3534-2, ISO 5725-1, ISO/IEC 17043, ISO/IEC Guide 99, ISO Guide 30, а также следующие термины с соответствующими определениями. В случае расхождений между этими ссылками в использовании терминов применяют определения по ISO 3534-1, ISO 3534-2. Математические обозначения приведены в приложении А.

ИСО и МЭК ведут терминологические базы данных для использования в области стандартизации по следующим адресам:

- платформа онлайн-просмотра ISO: доступна по адресу http://www.iso.org/obp;
- Электропедия IEC: доступна по адресу http://www.electropedia.org/.

3.1 **межлабораторное сличение** (interlaboratory comparison): Организация, выполнение и оценивание 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): Оценивание функционирования участника по заранее установленным критериям посредством межлабораторных сличений (3.1).

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

- количественная программа — когда целью является определение количественной оценки одной или более измеряемых величин каждого образца для проверки квалификации;
- качественная программа — когда целью является идентификация или описание одной или нескольких качественных характеристик образца для проверки квалификации;
- последовательная программа — когда один или несколько образцов для проверки квалификации последовательно распределяют для испытаний или измерений и возвращают провайдеру проверки квалификации через определенные интервалы времени;
- параллельная программа — когда образцы для проверки квалификации распределяют для выполнения одновременных испытаний или измерений в течение определенного периода времени;
- единоразовое задание — когда образцы для проверки квалификации предоставляют только один раз;
- непрерывная программа — когда образцы для проверки квалификации предоставляют через установленные интервалы времени;
- отбор проб — когда отбирают пробы для последующего анализа, а цель программы проверки квалификации включает оценивание отбора проб;
- интерпретация данных — когда наборы данных или другую информацию предоставляют и обрабатывают с целью 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 (изменчивости), используемая при оценивании результатов проверки квалификации (3.2).

Примечание 1 — Стандартное отклонение для оценки квалификации может быть интерпретировано как стандартное отклонение результатов гипотетической совокупности лабораторий, функционирующих в точном соответствии с установленными требованиями.

Примечание 2 — Оценку стандартного отклонения для оценки квалификации применяют только к результатам, полученным по интервальной шкале или шкале отношений.

Примечание 3 — Не во всех программах проверки квалификации оценивают функционирование на основе рассеяния результатов.

[ISO/IEC 17043:2010, изменено — в определении фраза «на основе имеющейся информации» исключена. Добавлено примечание 1 и отредактированы примечания 2 и 3]

3.5 **погрешность измерения** (measurement error): Разность между измеренным значением величины и опорным значением величины.

[ISO/IEC Guide 99:2007, изменено — исключены примечания]

3.6 **максимальная допускаемая погрешность измерения** (maximum permissible error): Максимальное значение погрешности измерения (3.5) относительно известного опорного значения величины, разрешенное спецификацией или нормативными документами для данного измерения, средства измерений или измерительной системы.

[ISO/IEC Guide 99:2007, изменено — исключены примечания]

3.7 **показатель *z*** (*z* score): Нормированный показатель функционирования участника, вычисленный с использованием результата участника, приписанного значения (3.3) и стандартного отклонения для оценки квалификации (3.4).

Примечание 1 — Распространенный вариант показателя *z*, иногда обозначаемый *z*′ (как правило, произносят *z*-штрих), формируют путем объединения неопределенности приписанного значения со стандартным отклонением для оценки квалификации до вычисления показателя *z*.

3.8 **показатель дзета** (zeta score): Нормированный показатель функционирования участника, вычисленный с использованием результата участника, приписанного значения (3.3) и комбинации стандартных неопределенностей результата и приписанного значения (3.3).

3.9 **доля допустимого предельного показателя** (proportion of allowed limit score): Нормированный показатель функционирования участника, вычисленный с использованием результата участника, приписанного значения (3.3) и критерия погрешности измерения (3.5) при проверке квалификации.

Примечание — Для единичных результатов показатель может быть выражен в виде отклонения от приписанного значения (*D* или *D* %).

3.10 **сигнал действия**[1)] (action signal): Указание на необходимость действий, вытекающее из результата проверки квалификации.

Пример — Показатель z, превышающий 2, как правило, принимают как указание на необходимость исследования возможных причин; показатель z, равный 3 или более, как правило, принимают как сигнал действия, указывающий на необходимость корректирующих действий.

3.11 **согласованное значение** (consensus value): Значение, полученное на основе совокупности результатов межлабораторного сличения (3.1).

Примечание — Термин «согласованное значение», как правило, используют для описания оценок параметра положения и дисперсии, полученных на основе результатов участников в туре программы проверки квалификации, он может также быть использован для значений, полученных на основе результатов определенного подмножества таких результатов или, например, ряда экспертных лабораторий.

3.12 **выброс** (outlier): Элемент совокупности значений, который несовместим с остальными элементами данной совокупности.

Примечание 1 — Выброс может возникнуть случайно в ожидаемой совокупности, принадлежать другой совокупности или быть результатом некорректной записи или другой ошибки.

Примечание 2 — Многие программы проверки квалификации используют термин «выброс» для обозначения результата, который вызывает сигнал действия. Это не предполагаемое использование термина. Несмотря на то, что выбросы обычно вызывают сигналы действия, возможно получение сигналов действия от результатов, которые не являются выбросами.

[ISO 5725-1:1994, изменено — добавлены примечания]

3.13 **участник** (participant): Лаборатория, организация или физическое лицо, которые получают образец для проверки квалификации и представляют результаты на рассмотрение провайдеру проверки квалификации (3.2).

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 оценки функционирования участника (3.13) при проверке квалификации (3.2).

Примечание — В большинстве случаев образцы для проверки квалификации соответствуют определению термина «стандартный образец», приведенному в ISO Guide 30 (3.17).

3.15 **провайдер проверки квалификации** (proficiency testing provider): Организация, которая несет ответственность за все задачи по разработке и выполнению программы проверки квалификации (3.2).

[1)] *Синонимом термина «сигнал действия» является термин «сигнал регулирования».*

3.16 **программа проверки квалификации** (proficiency testing scheme): Проверка квалификации (3.2), разработанная и реализованная за один или несколько туров 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;** СО (reference material; RM): Материал, достаточно однородный и стабильный по отношению к одному или нескольким определенным свойствам, которые были установлены для того, чтобы использовать его по назначению в измерительном процессе.

Примечание 1 — Стандартный образец — это общее понятие.

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

Примечание 3 — Применение может включать калибровку измерительной системы, оценивание методики измерений, приписывание значений свойств другим материалам и контроль качества.

[ISO Guide 30:2015, изменено — примечание 4 исключено]

3.18 **сертифицированный стандартный образец;** ССО (certified reference material; CRM): Стандартный образец (СО) (3.17), одно или несколько определенных свойств которого установлены метрологически обоснованной процедурой, сопровождаемый сертификатом СО, в котором приведено значение этого свойства, связанной с ним неопределенности, и утверждение о метрологической прослеживаемости.

Примечание — Понятие «значение» включает также номинальное свойство или такой качественный признак, как идентичность или последовательность. Неопределенности для таких признаков могут быть выражены в виде вероятности или уровня доверия.

[ISO Guide 30:2015, изменено — примечания 2, 3 и 4 исключены]

4 Основные принципы

4.1 Основные требования к статистическим методам

4.1.1 Применяемые статистические методы должны соответствовать цели проверки квалификации и быть статистически обоснованными. Все статистические предположения, на которых основаны методы или план проверки, должны быть установлены в плане программы проверки квалификации или в ее письменном описании, эти предположения должны быть обоснованы.

Примечание — Статистически достоверный метод имеет прочную теоретическую основу, известное выполнение в ожидаемых условиях использования и основан на предположениях или условиях, которые, как может быть показано, достаточно хорошо применимы к данным для поставленной цели.

4.1.2 Методы планирования эксперимента и статистического анализа данных должны соответствовать заявленным целям программы проверки квалификации.

4.1.3 Провайдер проверки квалификации должен представить участникам описание используемых методов вычислений, объяснение общей интерпретации результатов и изложение всех ограничений, связанных с интерпретацией результатов. Эта информация должна быть приведена либо в каждом отчете по каждому туру программы проверки квалификации, либо в отдельном кратком изложении процедур, доступном участникам.

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

4.2 Базовая модель

4.2.1 Для количественных результатов в программах проверки квалификации, если получен единственный результат для данного образца для проверки квалификации, базовую модель описывает формула (1)

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

где x_i — результат проверки квалификации участника i;

μ — истинное значение измеряемой величины;

ε_i — погрешность измерений i-го участника, подчиняющаяся соответствующему распределению.

Примечание 1 — Общие модели распределения для ε включают: нормальное распределение $\varepsilon_i \sim N(0, \sigma^2)$ со средним 0 и дисперсией σ^2, одинаковой или различной для всех лабораторий; или, что более распространено, нормальное распределение, загрязненное/засоренное (далее загрязненное) выбросами, состоящее из смеси нормального распределения с более широким распределением, представляющим совокупность ошибочных результатов.

Примечание 2 — Основой для оценивания квалификации с использованием показателя z и σ_{pt} является то, что в «идеальной» совокупности компетентных лабораторий стандартное отклонение результатов проверки квалификации должно быть равно σ_{pt} или меньше.

Примечание 3 — Данная модель отличается от исходной модели, установленной в ISO 5725, тем, что она не включает член «систематическая погрешность лаборатории» B_i. Это связано с тем, что члены «систематическая погрешность лаборатории» и «погрешность остатка» невозможно различить при наличии информации только об одном наблюдении. Однако при рассмотрении результатов участника по нескольким турам или образцам в программе проверки квалификации может быть полезно использовать отдельный член для систематической погрешности лаборатории.

4.2.2 Для порядковых или качественных результатов могут подходить другие модели или статистическая модель может отсутствовать.

4.3 Общие подходы к оцениванию функционирования

4.3.1 Существует три различных общих подхода к оцениванию функционирования участников в программе проверки квалификации. Эти подходы используют для достижения различных целей программы проверки квалификации.

Подходы перечислены ниже:

a) оценивание функционирования путем сравнения с независимо выведенными критериями;

b) оценивание функционирования путем сравнения с показателями других участников;

c) оценивание функционирования путем сравнения с заявленной неопределенностью измерений.

4.3.2 Общие подходы применяют по-разному для определения приписанного значения и определения критериев для оценивания функционирования участников, например, когда приписанное значение является робастным средним результатов участников, а оценивание функционирования участников основывают на σ_{pt} или δ_E, где δ_E — допускаемое значение, заранее определенное для погрешности измерения, и $\sigma_{pt} = \delta_E/3$; аналогично в некоторых ситуациях приписанное значение может быть опорным значением, а σ_{pt} может быть робастным стандартным отклонением результатов участников. В подходе c), использующем неопределенность измерений, приписанное значение, как правило, является соответствующим опорным значением.

5 Руководство по планированию эксперимента для программ проверки квалификации

5.1 Введение в статистическое планирование эксперимента для программ проверки квалификации

Проверка квалификации связана с оценкой функционирования участников и, как правило, не связана со смещением или прецизионностью (хотя они могут быть оценены с применением конкретных планов эксперимента). Функционирование участников оценивают с использованием статистического оценивания их результатов, полученных при измерениях или интерпретации, выполняемых участниками на образцах для проверки квалификации. Оценку функционирования часто выражают в виде показателей, это обеспечивает согласованную интерпретацию по ряду измеряемых величин и позволяет сравнивать результаты для различных измеряемых величин на единой основе. Показатели функционирования, как правило, выводят путем сравнения разности результата, сообщенного участником, и приписанного значения с допустимым отклонением или с оценкой неопределенности этой разности. Анализ показателей функционирования участника в течение нескольких туров программы проверки квалификации может дать информацию о том, демонстрируют ли отдельные лаборатории свидетельства устойчивых систематических эффектов («смещение») или низкой долгосрочной прецизионности.

В разделах 5—10 приведено руководство по разработке плана для программ количественной проверки квалификации и статистической обработке результатов, включая расчеты и интерпретацию различных показателей функционирования. Рассмотрение программ проверки квалификации по качественным данным (включая программы для порядковых величин) приведено в разделе 11.

5.2 Основы статистического планирования эксперимента

5.2.1 В соответствии с 4.4.4.1 ISO/IEC 17043:2010 статистический план эксперимента «должен быть разработан для достижения целей программы проверки квалификации на основе вида данных (количественные или качественные, включая порядковые и категоризированные данные), статистических предположений, погрешностей и ожидаемого количества результатов». Следовательно, программы проверки квалификации с различными целями и с различными источниками погрешностей могут иметь различные планы.

Планы для общих целей приведены ниже. Возможны и другие цели.

Пример 1 — Для программы проверки квалификации, позволяющей сравнивать результат участника с заранее определенным опорным значением в пределах границ, установленных до начала тура, план должен включать метод получения внешне заданного опорного значения, метод установления границ и метод расчета показателей.

Пример 2 — Для программы проверки квалификации, позволяющей сравнить результат участника с совокупными результатами группы в одном и том же туре, и ограничениями, заданными до начала тура, в плане должно быть указано, как приписанное значение определяют на основе объединенных результатов, а также метод установления ограничений и метод расчета показателей.

Пример 3 — Для программы проверки квалификации, позволяющей сравнить результат участника с объединенными результатами группы в одном и том же туре, с ограничениями, определяемыми вариабельностью результатов участников, в плане должен быть указан расчет приписанного значения и соответствующей меры рассеяния, а также метод расчета показателей.

Пример 4 — Для программы проверки квалификации, позволяющей сравнить результат участника с приписанным значением, при использовании собственной неопределенности измерений участника в плане должно быть указано, как должны быть получены приписанное значение и его неопределенность, а также как неопределенности измерений участников должны быть использованы при расчете показателей.

Пример 5 — Для программы проверки квалификации, целью которой является сравнение различных методов измерения, в плане должны быть указаны соответствующие обобщенные статистики и процедуры их расчета.

5.2.2 При проверке квалификации используют различные виды данных, включая количественные, номинальные (категоризированные) и порядковые. Среди количественных переменных некоторые результаты могут быть определены по интервальной шкале, относительной шкале, или шкале отношений. Некоторые результаты измерений, определяемые по количественной шкале, могут быть представлены только в виде дискретных значений и прерывистого набора значений (например, в случае последовательных разведений); однако во многих случаях эти результаты могут быть обработаны методами, применимыми к непрерывным количественным переменным.

Примечание 1 — Для количественных значений интервальная шкала представляет собой шкалу, на которой приведены интервалы (разности), но нельзя определить отношения величин, например, шкала температур Цельсия. Шкала отношений — это шкала, по которой можно определить как интервалы, так и отношения величин, например температурная шкала Кельвина или шкалы наиболее распространенных единиц измерения длины.

Примечание 2 — Для качественных величин применяют шкалу категорий, для которых упорядочивание не имеет смысла, например упорядочивание наименований видов бактерий. Значения в порядковой шкале имеют разумный порядок, но разности между ними не имеют смысла; например, можно упорядочить такую шкалу, как «большой, средний, маленький», но разности между этими значениями не могут быть определены, но можно определить количество промежуточных значений.

5.2.3 В дополнение к вышеуказанным программы проверки квалификации могут быть использованы для других целей (см. 0.1 ISO/IEC 17043). План эксперимента для конкретной программы проверки квалификации должен соответствовать всем установленным целям.

5.3 Анализ аспектов, связанных со статистическим распределением результатов

5.3.1 В соответствии с 4.4.4.2 ISO/IEC 17043:2010 методы статистического анализа должны соответствовать статистическим предположениям в отношении данных. В наиболее распространенных методах анализа для проверки квалификации предполагается, что набор результатов компетентных участников подчиняется приблизительно нормальному распределению или, по крайней мере, унимодальному и достаточно симметричному (после преобразования, при необходимости). Распространенное дополнительное предположение состоит в том, что распределение квалифицированно определенных результатов измерений представляет собой смесь («загрязнение»), включающую результаты

совокупности ошибочных значений, и может содержать выбросы. Как правило, интерпретация результатов основана на предположении о нормальности распределения, но только для базового предполагаемого распределения компетентных участников.

5.3.2 Как правило, нет необходимости в проверке того, что результаты подчиняются нормальному распределению, но важно проверить приблизительную симметрию плотности распределения, хотя бы визуально. Если симметричность распределения невозможно проверить, то провайдер проверки квалификации должен использовать методы, устойчивые к асимметрии распределения.

5.3.3 Если распределение, ожидаемое для результатов измерений программы проверки квалификации, недостаточно симметрично (допускает загрязнение выбросами), провайдер проверки квалификации должен выбрать методы анализа данных, которые должным образом учитывают ожидаемую асимметрию и устойчивы к выбросам, и методы расчета показателей, которые также учитывают особенности ожидаемого распределения результатов компетентных участников. Методы анализа данных могут включать:

- преобразование данных для обеспечения приблизительной симметрии распределения;
- методы оценки, устойчивые к асимметрии распределения;
- методы оценки, включающие соответствующие предположения о распределении (например, метод максимального правдоподобия, подходящий для выбора распределения, соответствующего предположениям о распределении результатов и, при необходимости, для исключения выбросов).

Пример 1 — Результаты, полученные на основе разведений, например при подсчетах 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 — Количество мелких частиц может подчиняться распределению Пуассона, и, следовательно, критерий для оценивания функционирования участника может быть определен с использованием таблицы вероятностей распределения Пуассона, на основе среднего количества участников в группе.

Примечание 1 — Преобразование данных может повлиять на обработку и интерпретацию неопределенностей, связанных с результатами участников и приписанным значением.

Примечание 2 — В приложении С приведена дополнительная информация об обработке асимметричных распределений, содержащих выбросы, а в Е.6 приведен пример метода оценки, устойчивого к асимметрии распределения.

5.3.4 В некоторых областях калибровки результаты участников при калибровке могут соответствовать статистическим распределениям, которые описаны в процедуре выполнения измерений (например, экспоненциальному распределению или распределению в форме волны); эти распределения должны быть указаны во всех протоколах оценивания.

5.3.5 В соответствии с 4.4.4.2 ISO/IEC 17043:2010 провайдер проверки квалификации должен установить все статистические предположения и продемонстрировать их обоснованность. Такая демонстрация может быть основана, например, на наблюдаемых данных, результатах предыдущих туров программы проверки квалификации или на анализе технической литературы.

Примечание — Демонстрация корректности предположения о виде распределения результатов является менее строгой, чем демонстрация обоснованности этого предположения.

5.4 Анализ аспектов, связанных с небольшим количеством участников

5.4.1 Статистический план эксперимента программы проверки квалификации должен учитывать минимальное необходимое количество участников в соответствии с целью эксперимента и устанавливать альтернативные подходы, которые могут быть использованы, если минимальное количество не достигнуто (ISO/IEC 17043:2010, 4.4.4.3 b)). Статистические методы, которые подходят для большого количества участников, могут быть неподходящими при ограниченном количестве участников. Проблемы связаны с тем, что результаты, определенные по небольшому количеству результатов участников, могут быть недостоверны и результат участника может быть оценен по сравнению с несоответствующей для сравнения группой.

Примечание — В техническом отчете IUPAC/CITAC «Выбор и использование программ проверки квалификации при ограниченном количестве участников» [24] приведены полезные рекомендации для программ проверки квалификации, при небольшом количестве участников. А именно: в отчете IUPAC/CITAC рекомендовано, чтобы приписанное значение было основано на достоверных независимых результатах измерений, например,

полученных с использованием сертифицированного стандартного образца, независимо установленного путем калибровки или национальным институтом метрологии, или полученного с помощью гравиметрического приготовления. В отчете указано, что стандартное отклонение для оценки квалификации может быть не основано на наблюдаемой дисперсии результатов участников одного тура программы проверки квалификации.

5.4.2 Минимальное количество участников, необходимое для различных статистических методов, зависит от множества факторов:

- используемых статистических методов, например, выбора конкретного робастного метода или стратегии исключения выбросов;
- опыта участников в конкретной программе проверки квалификации;
- опыта работы провайдера проверки квалификации с матрицей, измеряемой величиной, методами и группой участников;
- наличия (отсутствия) цели определения приписанного значения или стандартного отклонения (или и того, и другого).

Дополнительные рекомендации по методам работы с небольшим количеством участников приведены в D.1.

5.5 Рекомендации по выбору формы представления данных

5.5.1 Общие требования к форме представления данных

5.5.1.1 В соответствии с требованием 4.6.1.2 ISO/IEC 17043:2010 провайдеры проверки квалификации инструктируют участников по проведению измерений и представлению отчета о результатах измерений на образцах для проверки квалификации тем же самым способом, как и для большинства обычно выполняемых измерений, за исключением особых обстоятельств.

5.5.1.2 Это требование может, в некоторых ситуациях, затруднить получение точной оценки правильности и прецизионности или компетентности участников при выполнении процедуры измерений. Провайдер проверки квалификации должен принять согласованную форму представления данных для программы проверки квалификации, но должен, по возможности, использовать единицы измерений, знакомые большинству участников, и выбрать форму представления данных, которая сводит к минимуму неверные записи и другие ошибки. Отчет может включать автоматическое предупреждение о неподходящих единицах измерений, если известно, что участники регулярно отчитываются в единицах измерений, отличных от тех, которые требует программа проверки квалификации.

Примечание 1 — Для некоторых программ проверки квалификации целью является оценка способности участника следовать стандартному методу (например, официально требуемому или международному стандартизованному), который может включать использование определенной единицы измерения или определенного количества значащих цифр.

Примечание 2 — Ошибки провайдера проверки квалификации при сравнении результатов могут быть существенно уменьшены или устранены путем использования электронных систем отчетности, которые позволяют участникам напрямую вводить свои данные.

5.5.2 Представление данных о повторных измерениях

Если программа проверки квалификации требует повторения измерений на образцах для проверки квалификации, участник должен сообщить обо всех значениях результатов повторных измерений. Это может произойти, например, если цель состоит в том, чтобы оценить прецизионность участника по известным идентичным образцам для проверки квалификации, или если процедура измерения требует отдельного отчета о нескольких наблюдениях. В этих ситуациях провайдер проверки квалификации может также запросить у участника среднее значение (или другую оценку параметра положения) и неопределенность, что помогает провайдеру проверки квалификации при анализе данных.

5.5.3 Представление данных в виде «менее ...» или «более ...» (цензурированные данные)

5.5.3.1 В случае когда практикой представления данных является их запись в виде «менее или более какого-либо предельного значения» (такого как калибровочный уровень или предел количественного определения), а для вычислений необходимы числовые результаты, провайдер проверки квалификации должен определить способ обработки таких результатов.

5.5.3.2 Провайдер проверки квалификации должен либо принять валидированные процедуры обработки данных и расчета показателей, которые подходят для цензурированных данных (см. Е.1), либо потребовать от участников сообщать измеренные значения вместо обычно сообщаемых данных или в дополнение к ним.

Примечание 1 — Одним из вариантов процедуры расчета показателей может быть отказ от расчета показателя по таким данным и/или представление отчета о том, соответствует ли заявленный участником верхний (или нижний) предел приписанному значению.

Примечание 2 — Требование к участникам сообщать числовые значения, выходящие за границы диапазона обычно указываемых значений (например, ниже установленного для участника количественного предела), позволяет использовать статистические методы, которые требуют знания числовых значений, но могут привести к получению показателей, которые не отражают уровень обычного обслуживания участником своих потребителей.

5.5.3.3 При использовании согласованной статистики может оказаться невозможным оценить функционирование участника, если количество цензурированных значений достаточно велико, так как это влияет 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 [21], или результаты не должны быть использованы для оценки. При возникновении сомнений относительно выбранной процедуры провайдер проверки квалификации должен рассчитать суммарные статистики и оценки функционирования участников с использованием каждой из альтернативных статистических процедур, которые считаются применимыми в данных обстоятельствах, и исследовать значимость различий полученных оценок.

5.5.3.4 В тех случаях, когда ожидаются или наблюдались цензурированные результаты, такие как «менее …», план программы проверки квалификации должен включать положения по расчету показателей и/или другие действия с цензурированными данными, сообщаемыми участниками, участники должны быть уведомлены об этих положениях.

Примечание — В Е.1 приведен пример некоторых подходов к анализу цензурированных данных. В этом примере показана робастная согласованная статистика с использованием трех различных подходов; с исключением цензурированных значений, с сохранением этих значений, но удалением знака «<», и с заменой результатов на половину предельного значения.

5.5.4 Количество значащих цифр

5.5.4.1 Как правило, количество значащих цифр для отчета определено в плане программы проверки квалификации.

5.5.4.2 Количество значащих цифр, которые должны быть указаны в отчете, должно быть таким, чтобы ошибка округления была незначимой по сравнению с ожидаемыми различиями между участниками.

Примечание — В некоторых ситуациях правильное составление отчета является частью определения компетентности участника и количество значащих цифр и десятичных разрядов может изменяться.

5.5.4.3 В тех случаях, когда количество значащих цифр (разрядов), зафиксированных в обычных условиях измерения, оказывает заметное негативное влияние на обработку данных провайдером проверки квалификации (например, когда процедуры измерения требуют представления небольшого количества значащих цифр), провайдер проверки квалификации может указать количество значащих цифр (разрядов), которое должно быть в отчете.

Пример — Процедура измерения предусматривает представление данных с точностью до 0,1 г, что приводит к значительной доле (> 50 %) идентичных результатов и, в свою очередь, ставит под угрозу расчет робастных средних и стандартных отклонений. Поэтому провайдер проверки квалификации может потребовать от участников сообщать данные с точностью до двух или трех десятичных разрядов, чтобы получить достаточно достоверные оценки параметров положения и рассеяния.

5.5.4.4 Если допустимо, что разные участники сообщают результаты, используя различное количество значащих цифр, провайдер проверки квалификации должен учитывать это при формировании всех согласованных статистик (таких, как приписанное значение и стандартное отклонение для оценки квалификации).

6 Руководство по предварительному анализу образцов и результатов проверки квалификации

6.1 Однородность и стабильность образцов для проверки квалификации

6.1.1 Провайдер проверки квалификации для целей программы проверки квалификации должен обеспечить достаточную однородность и стабильность партии образцов для проверки квалификации в период реализации программы проверки квалификации. Провайдер должен оценить однородность и стабильность образцов, используя критерий, гарантирующий, что неоднородность и нестабильность

образцов для проверки квалификации не окажут неблагоприятного влияния на оценивание функционирования участников. При оценке однородности и стабильности следует использовать один или несколько подходов из перечисленных ниже:

a) экспериментальные исследования, описанные в приложении B, или альтернативные экспериментальные методы, обеспечивающие эквивалентную или бóльшую уверенность в однородности и стабильности;

b) порядок работы с очень похожими образцами для проверки квалификации в предыдущих турах программы проверки квалификации, верифицированный при необходимости для текущего тура;

c) оценку данных участников в текущем туре программы проверки квалификации для подтверждения соответствия предыдущим турам, для проверки наличия изменений по отношению ко времени представления отчета, изменений в процессе производства или наличия любого неожиданного рассеяния, влияющего на неоднородность или нестабильность.

Примечание 1 — Такие подходы могут быть приняты за основу в каждом конкретном случае с использованием соответствующих статистических методов и технического обоснования. Подход часто изменяется в процессе выполнения программы проверки квалификации, например, по мере того как накопленный опыт снижает первоначальные требования к необходимости экспериментальных исследований.

Примечание 2 — Полагаться на опыт (в соответствии с перечислением b)) разумно только до тех пор, пока:

a) в процессе производства партий образцов для проверки квалификации не возникают изменения, которые могут повлиять на однородность;

b) материалы, используемые при изготовлении образца (образцов) для проверки квалификации, не изменяются таким образом, что это может повлиять на однородность;

c) нет нарушения однородности, выявленного ни с помощью проверки однородности, ни с помощью обратной связи с участниками;

d) требования к однородности материала регулярно пересматривают с учетом предполагаемого использования материала на момент проверки, чтобы гарантировать, что однородность, достигаемая в процессе производства, остается соответствующей цели проверки квалификации.

Пример — Если в предыдущих турах программы проверки квалификации с теми же участниками использованы образцы для проверки квалификации, которые были исследованы и продемонстрировали достаточную однородность и стабильность, то, если стандартное отклонение результатов проверки квалификации в текущем туре не превышает стандартного отклонения в предыдущих турах, это свидетельствует о достаточной однородности и стабильности в текущем туре.

6.1.2 Для программ проверки квалификации при калибровке, когда один и тот же артефакт использует несколько участников, провайдер проверки квалификации должен обеспечить его стабильность на протяжении всего тура или иметь процедуры для выявления и учета его нестабильности в ходе выполнения тура программы проверки квалификации. Процедуры должны включать рассмотрение тенденций, таких как дрейф, для конкретных образцов для проверки квалификации и измеряемых величин. Там, где это уместно, для обеспечения стабильности следует учитывать влияние многократных перемещений одного и того же артефакта.

6.1.3 Все измеряемые величины (или свойства), как правило, необходимо проверять на однородность и стабильность. Однако, если может быть показано, что особенности подмножества свойств обеспечивают хорошие признаки стабильности и/или однородности для всех свойств, исследуемых в туре программы проверки квалификации, оценка, описанная в 6.1.1, может быть ограничена этим подмножеством свойств. Проверяемые измеряемые величины должны быть чувствительны к источникам неоднородности или нестабильности при работе с образцами для проверки квалификации. Некоторыми важными случаями являются ситуации:

a) когда измеряемая величина представляет собой малую долю компонента, поскольку в этом случае труднее достичь однородности и более высокой чувствительности при проверке однородности;

b) если образец для проверки квалификации нагревается во время обработки, то выбирают измеряемую величину, чувствительную к неравномерному нагреву;

c) если на измеряемое свойство могут повлиять отстаивание, выпадение осадка или другие зависящие от времени эффекты в процессе подготовки образцов для проверки квалификации, то это свойство следует проверять в течение всей процедуры.

Пример — В программе проверки квалификации на содержание токсичных металлов в почве на результат, в первую очередь, влияет содержание влаги. В этом случае проверку на постоянное содержание влаги можно считать достаточной для обеспечения надлежащей стабильности содержания токсичных металлов.

Примечание — Пример проверки однородности и стабильности с использованием статистических методов, рекомендованных в приложении B, приведен в E.2.

6.2 Особенности применения различных методов измерений

6.2.1 Если предполагается, что все участники сообщают значение одной и той же измеряемой величины, приписанное значение, как правило, должно быть одинаковым для всех участников. Однако, если участникам позволяют выбирать свой собственный метод измерений, возможно, что одно и то же приписанное значение для каждого компонента или свойства не будет подходить всем участникам. Это может произойти, например, когда различные методы измерения дают несопоставимые результаты. В этом случае провайдер проверки квалификации может использовать отдельное приписанное значение для каждого метода измерений.

Примеры

a) Медицинские испытания, когда известно, что различные утвержденные методы измерения по-разному реагируют на один и тот же тестируемый образец и используют разные референтные интервалы для диагностики.

b) Операционно определяемые измеряемые величины, такие как выщелачиваемые токсичные металлы в почвах, для которых доступны различные стандартные методы и не предполагается, что их будут напрямую сравнивать, а программа проверки квалификации определяет показатель без ссылки на конкретный метод испытаний.

6.2.2 При разработке программы проверки квалификации следует учитывать необходимость в различных приписанных значениях для подмножеств участников (например, для отчета о конкретных методах), а также при анализе данных для каждого тура.

6.3 Устранение грубых ошибок (промахов)

6.3.1 В ISO/IEC 17043:2010, B.2.5 и согласованном протоколе IUPAC рекомендовано удалять очевидные грубые ошибки из набора данных на ранней стадии анализа до использования любой робастной процедуры или проверки для выявления статистических выбросов. Как правило, эти результаты рассматривают отдельно (например, при общении с участником). Обычно некоторые ошибки удается исправить, но это следует делать только в соответствии с утвержденной политикой и процедурой.

Примечание — Очевидные грубые ошибки, такие как представление результатов в некорректных единицах измерения или результатов, принадлежащих различным образцам для проверки квалификации, происходят в большинстве туров проверки квалификации, эти результаты только ухудшают выполнение последующих статистических методов.

6.3.2 Если существуют сомнения относительно того, является ли результат грубой ошибкой, его следует сохранить в наборе данных и подвергнуть последующей обработке в соответствии с 6.4—6.6.

6.4 Визуальный анализ данных

6.4.1 В качестве первого этапа анализа данных провайдер проверки квалификации должен систематизировать данные для их визуального анализа, проводимого лицом, обладающим соответствующими техническими и статистическими знаниями. Этот анализ предназначен для подтверждения ожидаемого распределения результатов и выявления аномалий или непредвиденных источников изменчивости. Например, бимодальное распределение может свидетельствовать о наличии совокупности, представляющей собой смесь результатов, полученных различными методами, при использовании загрязненных образцов или плохо сформулированных инструкций. В этом случае проблему следует устранить до выполнения анализа или оценивания.

Примечание 1 — Гистограмма значений участников является полезной и широкодоступной процедурой анализа для поиска унимодального и симметричного распределения и выявления необычных выбросов (см. 10.2). Однако интервалы, используемые для объединения результатов в гистограмме, чувствительны к количеству результатов и точкам деления, и поэтому ее может быть сложно создать. Для выявления возможных бимодальности или отсутствия симметрии часто более полезен график плотности распределения (см. 10.3).

Примечание 2 — Могут быть полезны и другие методы анализа, такие как график кумулятивной функции распределения или диаграмма «стебель с листьями». Некоторые графические методы анализа данных представлены в E.3 и E.4.

6.4.2 Если невозможно проведение визуального анализа всех наборов исследуемых данных, должна быть предусмотрена процедура предупреждения о появлении неожиданной изменчивости в наборе данных; например, с помощью анализа неопределенности приписанного значения путем ее сравнения с критерием оценивания или сравнения с предыдущими турами программы проверки квалификации.

6.5 Робастные статистические методы

6.5.1 Робастные статистические методы могут быть использованы для описания центральной части нормально распределенного набора результатов без идентификации конкретных значений как выбросов и исключения их из последующего анализа. Многие используемые робастные методы основаны (на первом этапе) на оценке медианы и размаха для 50 % центральных результатов, представляющих собой показатели положения центра и рассеяния данных, аналогичные среднему и стандартному отклонению. В целом следует отдавать предпочтение робастным методам, а не методам, которые требуют исключения результатов, отмеченных как выбросы.

Примечание — Стратегии, которые применяют классические статистики, такие как стандартное отклонение, после исключения выбросов обычно приводят к заниженной оценке рассеяния для ненормальных данных; робастные статистики, как правило, приспособлены для получения несмещенных оценок рассеяния.

6.5.2 В качестве простых оценок применимы медиана, масштабированное абсолютное отклонение медианы (MAD_e) и нормализованный IQR (nIQR). Алгоритм A[1)] преобразует исходные данные с помощью процесса, называемого винсоризацией, для получения альтернативных оценок среднего и стандартного отклонения данных, близких к нормальному распределению, он наиболее полезен в ситуациях, когда ожидаемая доля выбросов составляет менее 20 %. Методы Q_n и Q (описанные в приложении C) для оценки стандартного отклонения особенно полезны в ситуациях, когда значительная доля (> 20 %) результатов может быть противоречивой или когда данные не могут быть достоверно проанализированы экспертами. Другие методы, описанные в приложении C, также хорошо работают, когда ожидаемая доля экстремальных значений превышает 20 % (см. D.2).

Примечание — Медиана, межквартильный размах и масштабированное абсолютное отклонение медианы имеют бóльшую изменчивость, чем среднее и стандартное отклонение, когда их применяют к приблизительно нормально распределенным данным. Более сложные робастные оценки обеспечивают более высокую эффективность для приблизительно нормально распределенных данных, сохраняя при этом бóльшую устойчивость к выбросам за счет использования медианы и межквартильного размаха.

6.5.3 Ответственность за выбор статистических методов лежит на провайдере проверки квалификации. Робастные среднее и стандартное отклонения могут быть использованы для различных целей, из которых оценивание функционирования участников является лишь одной. Робастные средние и стандартные отклонения также могут быть использованы в качестве объединенной статистики для различных групп участников или для конкретных методов.

Примечание — Подробная информация о робастных процедурах приведена в приложении C. В E.3 и E.4 приведены примеры, иллюстрирующие использование различных робастных статистических методов, представленных в приложении C.

6.6 Методы выявления выбросов для отдельных результатов

6.6.1 Проверка на наличие выбросов может быть использована либо для поддержки визуального анализа аномальных значений, либо, в сочетании с исключением выбросов, для обеспечения определенной устойчивости к экстремальным значениям при расчете объединенной статистики. При использовании методов выявления выбросов следует подтвердить выполнение предположений, лежащих в их основе, чтобы обеспечить применимость этих методов для целей программы проверки квалификации; в частности, в основе многих критериев выявления выбросов лежит предположение о нормальности распределения наблюдений.

Примечание — В ISO 16269-4 [10] и ISO 5725-2 [1] приведено несколько процедур выявления выбросов, применимых к межлабораторным данным.

6.6.2 В ситуациях, когда робастные методы неприменимы (см. 6.5.1), допустимо использование стратегий исключения выбросов, основанных на исключении выбросов, обнаруженных с помощью соответствующего критерия с высоким уровнем доверия, с последующим применением простых статистик, таких как среднее и стандартное отклонение. При использовании стратегий исключения выбросов провайдер проверки квалификации должен:

а) документировать критерии выявления выбросов и уровень доверия, необходимый для исключения соответствующего значения;

[1)] *См. приложение C.*

b) установить ограничения для доли данных, исключенных при последовательном применении критериев выявления выбросов, если их используют;

c) продемонстрировать, что полученные оценки параметра положения и (при необходимости) параметра масштаба соответствуют достаточным уровням свойств (включая эффективность и смещение) для целей программы проверки квалификации.

Примечание — В ISO 5725-2 приведены рекомендации по выбору уровня доверия для исключения выбросов в межлабораторных исследованиях по определению прецизионности методов испытаний. В частности, в ISO 5725-2 рекомендовано исключать выбросы только с уровнем доверия 99 %, если нет других веских причин для исключения конкретного результата.

6.6.3 Если исключение выбросов является частью процедуры обработки данных, и результат исключают как выброс, функционирование участника все равно необходимо оценивать в соответствии с критериями, используемыми для всех участников программы проверки квалификации.

Примечание 1 — Выбросы среди зафиксированных результатов участников часто выявляют с помощью критерия Граббса в соответствии с ISO 5725-2. При оценивании в этой процедуре используют стандартное отклонение всех участников, включая возможные выбросы. Поэтому данную процедуру лучше всего применять в случае, когда функционирование участников соответствует ожиданиям по предыдущим турам программы проверки квалификации с небольшим количеством выбросов (по одному или два выброса с каждой стороны от среднего). Обычные таблицы критерия Граббса предполагают одно применение для одного возможного выброса (или 2) при определении параметра положения, но не неограниченное последовательное применение этого критерия. Если критерий Граббса применяют последовательно, вероятности ошибки первого рода для критерия не применимы.

Примечание 2 — Если участники сообщают повторные результаты или в тур программы проверки квалификации включают идентичные образцы для проверки квалификации, обычно для исключения выбросов используют критерий Кохрена, описанный в ISO 5725-2.

Примечание 3 — Выбросы также могут быть идентифицированы с помощью робастных или непараметрических методов; например, если вычисляют робастные среднее и стандартное отклонение, значения, отклоняющиеся от робастного среднего более чем на 3 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 приведено пять методов определения приписанного значения x_{pt}. За выбор одного из этих методов отвечает провайдер проверки квалификации.

Примечание — Методы, приведенные в 7.3—7.6, схожи с подходами, используемыми для определения характеристик свойств сертифицированных стандартных образцов, описанными в ISO Guide 35 [13].

7.1.2 Альтернативные методы определения приписанного значения и его неопределенности могут быть использованы при условии, что они имеют статистическое обоснование, а используемый метод описан в документированном плане программы проверки квалификации и полностью известен участникам. Независимо от метода, используемого для определения приписанного значения, проверка обоснованности приписанного значения для данного тура программы проверки квалификации всегда является целесообразной. Более подробное обсуждение данного положения приведено в 7.8.

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

7.1.4 Метод определения приписанного значения и соответствующей ему неопределенности должен быть установлен 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 ISO/IEC Guide 98-3 [14] приведены рекомендации по оцениванию неопределенностей измерений. В ISO Guide 35 приведены рекомендации по оценке неопределенности приписанного значения для значений сертифицированных свойств, которые могут быть применены при разработке многих программ проверки квалификации.

7.2.2 Общую модель приписанного значения и его неопределенности описывают формулы (2) и (3). Модель приписанного значения может быть представлена следующим образом:

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

где x_{pt} — приписанное значение;

x_{char} — значение свойства, полученное при характеризации (определении приписанного значения);

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

δ_{trans} — погрешность, обусловленная нестабильностью в условиях транспортировки;

δ_{stab} — погрешность, обусловленная нестабильностью в течение периода проверки квалификации.

Соответствующая модель неопределенности приписанного значения может быть представлена следующим образом:

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

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

u_{char} — стандартная неопределенность, обусловленная характеризацией;

u_{hom} — стандартная неопределенность, обусловленная различиями образцов для проверки квалификации («неоднородность»);

u_{trans} — стандартная неопределенность, обусловленная нестабильностью, вызванной транспортировкой образцов для проверки квалификации;

u_{stab} — стандартная неопределенность, обусловленная нестабильностью в течение периода проверки квалификации.

Примечание 1 — Наличие ковариации между источниками неопределенности или незначимыми источниками неопределенности может привести к другой модели в конкретных ситуациях. В некоторых ситуациях любая из составляющих неопределенности может быть незначимой или равной нулю.

Примечание 2 — Если σ_{pt} рассчитывают как стандартное отклонение результатов участников, составляющие неопределенности, обусловленные неоднородностью, условиями транспортировки и нестабильностью в течение проверки квалификации, 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 образцов в условиях проверки квалификации или условиях транспортировки, незначимы по сравнению со стандартным отклонением для оценки квалификации; то есть гарантирует, что δ_{trans} и δ_{stab} незначимы. Если это требование выполнено, то u_{stab} и u_{trans} можно считать равными нулю.

7.2.3 Приписанное значение может включать смещение, которое не учтено в приведенном выше выражении. Смещение, по возможности, следует учитывать при разработке плана программы проверки квалификации. Если в приписанном значении смещение скорректировано, то неопределенность, соответствующая такой корректировке, должна быть включена в оценивание неопределенности приписанного значения.

7.3 Процедура приготовления образцов

7.3.1 Образец для проверки квалификации может быть подготовлен путем смешивания материалов с различными известными уровнями свойств в определенных пропорциях или путем добавления установленной пропорции вещества к основному материалу.

7.3.2 Приписанное значение x_{pt} выводят путем вычисления на основе содержания веществ в используемых материалах. Такой подход особенно важен, когда таким способом подготавливают отдельные образцы для проверки квалификации, и необходимо определить соотношение свойств в этих образцах.

7.3.3 Следует проявлять разумную осторожность для обеспечения того, чтобы:

a) основной материал был фактически свободен от добавленного компонента или доля компонента, добавленного в основной материал, была точно известна;

b) компоненты были смешаны друг с другом до однородности (если это необходимо);

c) были определены все существенные источники ошибок (например, не всегда учитывают, что стекло поглощает соединения ртути, так что концентрация водного раствора соединения ртути может быть изменена содержащим его сосудом);

d) отсутствовало взаимодействие между компонентами и матрицей;

e) свойства образцов для проверки квалификации, содержащих добавленный материал, были аналогичны свойствам образцов потребителей, которые регулярно исследуют. Например, чистые материалы, добавленные к природной матрице, часто легче извлечь, чем то же самое вещество из материала естественного происхождения. Если есть опасения по поводу того, что это произойдет, провайдер проверки квалификации должен обеспечить пригодность образцов для проверки квалификации для методов, которые будут использованы.

7.3.4 Если в соответствии с процедурой приготовления в образцах для проверки квалификации добавка находится в менее связанном состоянии, чем в образцах, исследуемых в обычной практике, или в другом виде, предпочтительнее использовать другой способ приготовления образцов для проверки квалификации.

7.3.5 Определение приписанного значения по процедуре приготовления образцов является одним из примеров общего подхода к характеризации сертифицированных стандартных образцов, описанного в ISO Guide 35, когда единственная лаборатория определяет приписанное значение, используя первичный метод измерения. Другие варианты использования первичного метода единственной лабораторией могут быть использованы для определения приписанного значения для программы проверки квалификации (см. 7.5).

7.3.6 Если приписанное значение вычисляют на основе процедуры приготовления образца для проверки квалификации, стандартную неопределенность характеризации (u_{char}) оценивают как комбинацию неопределенностей с использованием соответствующей модели. Например, при проверке квалификации в области химических измерений неопределенности обычно связаны с гравиметрическими измерениями, измерениями объема и чистотой всех материалов, используемых при приготовлении образца. Затем стандартную неопределенность приписанного значения ($u(x_{pt})$) вычисляют в соответствии с формулой (3).

7.4 Сертифицированный стандартный образец

7.4.1 Если образец для проверки квалификации является сертифицированным стандартным образцом (ССО), значение его сертифицированного свойства x_{CRM} используют в качестве приписанного значения x_{pt}.

Ограничения такого подхода заключаются в том, что:

- обеспечение каждого участника единицей аттестованного стандартного образца требует больших средств;
- ССО часто подвергают довольно интенсивной обработке для обеспечения долговременной стабильности, что может поставить под угрозу взаимозаменяемость элементов проверки квалификации;
- ССО может быть известен участникам и поэтому важно скрыть информацию, позволяющую идентифицировать образец для проверки квалификации.

7.4.2 При использовании сертифицированного стандартного образца в качестве образца для проверки квалификации стандартную неопределенность приписанного значения выводят из информации о неопределенности значения свойства, указанной в сертификате. Сертификат должен включать в себя сведения о составляющих неопределенности, указанных в формуле (3), и информацию о предполагаемом использовании образца для целей программы проверки квалификации.

7.5 Результаты, полученные одной лабораторией

7.5.1 Приписанное значение может быть определено единственной лабораторией с использованием референтного метода, такого как, например, первичный метод. Используемый референтный метод должен быть полностью описан и понятен с утверждением полной неопределенности и документированной метрологической прослеживаемости и соответствовать программе проверки квалификации. Референтный метод должен быть совместим со всеми методами измерений, используемыми участниками.

7.5.1.1 Приписанное значение должно быть средним арифметическим в соответствии с планом исследований при использовании более одного образца для проверки квалификации или нескольких условий измерения при достаточном количестве повторных измерений.

7.5.1.2 Неопределенность характеризации — это соответствующая оценка неопределенности референтного метода и соответствующих плану условий исследования.

7.5.2 Приписанное значение x_{pt} образца для проверки квалификации может быть выведено единственной лабораторией с использованием подходящего метода измерения по данным калибровки на соответствие опорному значению максимально приближенного сертифицированного стандартного образца. Такой подход предполагает, что ССО обладает коммутативностью по отношению ко всем методам измерений, используемым участниками.

7.5.2.1 Для такого определения необходимо выполнение серии испытаний, проводимых в одной лаборатории на образцах для проверки квалификации и ССО с использованием одного и того же метода измерений в условиях повторяемости.

Если x_{CRM} — сертифицированное значение свойства ССО;

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

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

$\overline{d}$ — среднее арифметическое разностей d_i, то

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

Примечание — x_{CRM} и $\overline{d}$ независимы, за исключением редкого случая, когда ССО произведен 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 способом, метрологически прослеживаемым до аттестованного значения ССО, со стандартной неопределенностью, которая может быть рассчитана по формуле (5).

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

Пример, приведенный в Е.5, показывает, как требуемую неопределенность можно рассчитать в простом случае, когда приписанное значение образца для проверки квалификации устанавливают прямым сравнением с единственным ССО.

7.5.3 Если опорное значение присваивают до начала тура последовательной программы проверки квалификации, а затем опорное значение последовательно проверяют с использованием одной и той же измерительной системы, разность полученных значений должна быть меньше удвоенной неопределенности этой разности (то есть результаты должны быть метрологически совместимы). 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 проверки квалификации должен исследовать причину различий и предпринять соответствующие меры, включая использование альтернативных методов определения приписанного значения и его неопределенности или отказ от тура программы проверки квалификации.

Примечание — При предположении о нормальном распределении данных можно ожидать, что разность случайно превысит удвоенную стандартную неопределенность примерно в одном случае из двадцати.

7.6 Согласованное значение по данным экспертных лабораторий

7.6.1 Приписанные значения могут быть определены с использованием межлабораторного сличения с привлечением экспертных лабораторий в соответствии с ISO Guide 35 по использованию межлабораторных сличений для характеризации ССО. Сначала подготавливают образцы для проверки квалификации, которые затем раздают участникам. Некоторые из этих образцов затем выбирают случайным образом, их анализирует группа экспертов с использованием правил, которыми установлено количество образцов для проверки квалификации и количество повторений, а также всех других соответствующих условий. Каждая экспертная лаборатория обязана указать стандартную неопределенность своих результатов.

7.6.2 В тех случаях, когда экспертные лаборатории сообщают единственный результат и протокол измерений не требует представления достаточной информации о неопределенности результатов или при наличии в результатах или других источниках свидетельств того, что данные о неопределенности недостаточно достоверны, согласованное значение, как правило, должно быть получено методами, приведенными в 7.7, применяемыми к набору результатов экспертной лаборатории. Если экспертные лаборатории сообщают более одного результата каждая (например, включая повторения), провайдер программы проверки квалификации должен установить альтернативный метод определения приписанного значения и соответствующей ему неопределенности, который является статистически обоснованным (см. 4.1.1) и допускает возможность выбросов или других отклоняющихся результатов, в соответствии с ожидаемым распределением.

7.6.3 В тех случаях, когда экспертные лаборатории сообщают неопределенности результатов, оценка значения путем согласования результатов является сложной проблемой, и для ее решения существует много подходов, включая, например, взвешенные средние арифметические, невзвешенные средние арифметические, процедуры, допускающие чрезмерное рассеяние данных, и процедуры, допускающие наличие выбросов или ошибочных результатов и оценок неопределенности [16]. Провайдер проверки квалификации должен соответствующим образом установить процедуру оценки, которая:

a) должна включать проверку достоверности представленных оценок неопределенности, например, путем проверки того, полностью ли указанные неопределенности учитывают наблюдаемое рассеяние результатов;

b) должна использовать процедуру с весовыми коэффициентами, соответствующими величине и достоверности сообщаемых неопределенностей, которая может включать равные весовые коэффициенты, если сообщаемые неопределенности либо близки, либо имеют низкую или неизвестную достоверность (см. 7.6.2);

c) должна допускать возможность того, что сообщаемые неопределенности могут не в полной мере учитывать наблюдаемое рассеяние («чрезмерную дисперсию»), например, путем включения дополнительного члена, учитывающего чрезмерное рассеяние;

d) должна допускать возможность неожиданного отклонения значений от заявленного результата или неопределенности;

e) должна иметь прочную теоретическую основу;

f) должна демонстрировать функционирование (например, на тестовых данных или при моделировании), соответствующую целям программы проверки квалификации.

7.7 Согласованное значение по результатам участников

7.7.1 При таком подходе приписанное значение x_{pt} образца для проверки квалификации, используемого в туре программы проверки квалификации, представляет собой оценку параметра положения (например, робастное среднее, медиана или среднее арифметическое), сформированную на основе результатов, представленных участниками тура, рассчитанную с использованием подходящей процедуры в соответствии с планом, описанным в приложении С. Методы, описанные в 6.2—6.6, следует использовать для подтверждения наличия достаточной согласованности результатов до их объединения.

7.7.2 В некоторых ситуациях провайдер проверки квалификации может использовать подгруппу участников, признанных надежными по некоторым заранее определенным критериям, таким как статус аккредитации, или на основе их предыдущих результатов. Методы, описанные в данном разделе, применимы к таким ситуациям, включая определение размера группы.

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

7.7.4 Преимущества такого подхода состоят в следующем:

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

b) подход может быть особенно полезен при использовании стандартизованной, операционно-определяемой измеряемой величины, поскольку часто не существует более надежного способа получения эквивалентных результатов.

7.7.5 Недостатки данного подхода состоят в следующем:

a) между участниками может отсутствовать достаточная согласованность;

b) согласованное значение может включать неизвестное смещение, вызванное общим использованием неверной методологии, 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. Хотя результат всегда можно проследить до результатов отдельных лабораторий, четкое установление прослеживаемости за их пределами может быть обеспечено только тогда, когда провайдер проверки квалификации располагает полной информацией об используемых калибровочных стандартах и контроле других условий метода всеми участниками, вносящими вклад в согласованное значение.

7.7.6 Стандартная неопределенность приписанного значения зависит от используемой процедуры. Если необходим наиболее общий подход, провайдер проверки квалификации должен рассмотреть возможность использования методов, основанных на повторных выборках («бутстреп-метода»), для оценки стандартной погрешности приписанного значения. Подробное описание «бутстреп-метода» приведено в [17], [18].

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

7.7.7 Если в качестве приписанного значения выбрано робастное среднее, рассчитанное с использованием процедур, описанных в С.2, С.3, оценку стандартной неопределенности приписанного значения x_{pt} можно определить по формуле

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

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

Примечание 1 — В этой модели, где приписанное значение и робастное стандартное отклонение определяют по результатам участников, предполагается, что неопределенность приписанного значения включает влияние неопределенностей, обусловленных неоднородностью, условиями транспортировки и нестабильностью.

Примечание 2 — Коэффициент 1,25 выбран 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 выше, чем методов, использующих медиану, что обосновывает значение корректирующего коэффициента меньше 1,25. Однако рекомендуется применять такой коэффициент, поскольку результаты проверки квалификации, как правило, не подчиняются строго нормальному распределению и содержат неизвестные доли результатов из других распределений («загрязненные результаты»). Коэффициент 1,25 обеспечивает определение гарантированной (завышенной) оценки, учитывающей возможное загрязнение. В зависимости от опыта и используемой робастной процедуры может быть обосновано применение меньшего коэффициента или другого уравнения.

Примечание 3 — Пример использования приписанного значения по результатам участников приведен в Е.3.

7.8 Сравнение приписанного значения с независимым опорным значением

7.8.1 Если методы, описанные в 7.7, используют для установления приписанного значения (x_{pt}) при наличии достоверной независимой оценки (обозначаемой x_{ref}), например, на основе сведений о подготовке образцов или об опорном значении, согласованное значение x_{pt} следует сравнить с x_{ref}.

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

Разность вычисляют в виде $x_{\text{diff}} = x_{ref} - x_{pt}$ (или $x^* - x_{pt}$), а стандартную неопределенность разности u_{diff} оценивают в виде:

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

где $u(x_{ref})$ — неопределенность опорного значения, используемого для сравнения;

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

Примечание — Пример сравнения опорного значения с согласованным значением приведен в Е.7.

7.8.2 Если разность больше удвоенной соответствующей стандартной неопределенности, следует выяснить причину. Возможными причинами являются:

- наличие смещения соответствующего метода измерений;
- наличие общего смещения результатов участников;
- неверное понимание ограничений метода измерений при использовании метода приготовления образцов, описанного в 7.3;
- наличие смещения результатов «экспертов» при использовании подходов в соответствии с 7.5 или 7.6;
- опорное значение и приписанное значение не прослеживаются до одного и того же метрологического эталона.

7.8.3 В зависимости от выявленной причины провайдер проверки квалификации должен решить, следует ли оценивать результаты и (для непрерывных программ проверки квалификации) нужно ли вносить изменения в план последующих программ проверки квалификации. В тех случаях, когда разность достаточно велика и может повлиять на оценку функционирования участников или можно предположить наличие существенного смещения в методах измерений, используемых участниками, разность должна быть указана в отчете по туру программы проверки квалификации. В таких случаях разность следует учитывать при разработке будущих программ проверки квалификации.

8 Выбор критериев для оценивания функционирования участников

8.1 Подходы к выбору критериев

8.1.1 Основной подход для всех целей состоит в сравнении результатов измерений, выполненных на образце для проверки квалификации (x_i), с приписанным значением (x_{pt}). Для оценивания разность сравнивают с допустимой погрешностью измерений. Это сравнение обычно выполняют с помощью

нормированной статистики (например, z, z', ζ, E_n), в соответствии с 9.4—9.7. Сравнение также можно выполнить, сравнивая разность с определенным критерием (D или D% по сравнению с δ_E), в соответствии с 9.3. Альтернативный подход к оцениванию состоит в сравнении разности с неопределенностью результатов, заявленной участником, в сочетании с неопределенностью приписанного значения (E_n и ζ).

8.1.2 Если обязательное требование или целевое значение заданы в виде стандартного отклонения, то оно может быть использовано непосредственно как σ_{pt}. Если требованием или целью является максимальная допускаемая погрешность результатов измерения, то для получения σ_{pt} этот критерий можно поделить на значение, соответствующее границе действия/регулирования (далее — действия). Установленная максимальная допускаемая погрешность может быть использована непосредственно в качестве δ_E для использования с D или D%. Преимущества такого подхода для непрерывных программ проверки квалификации состоят в следующем:

a) показатели функционирования имеют последовательную интерпретацию с точки зрения целевого значения от одного тура к следующему в рамках программы проверки квалификации;

b) показатели функционирования не подвержены изменениям, ожидаемым при оценке рассеяния зафиксированных результатов.

Пример — Если обязательный критерий установлен в виде максимальной допускаемой погрешности и 3,0 является границей действия для оценивания с использованием показателя z, то указанный критерий делят на 3,0 для определения σ_{pt}.

8.1.3 Если критерий для оценивания функционирования участников основан на согласованных статистиках текущего тура или предыдущих туров программы проверки квалификации, то предпочтительной статистикой является робастная оценка стандартного отклонения результатов участников. При использовании такого подхода, как правило, удобнее использовать оценку показателя функционирования в виде показателя z и установить в качестве вычисленной оценки стандартного отклонения стандартное отклонение для оценки квалификации (σ_{pt}).

8.2 На основании мнения экспертов

8.2.1 Значение максимальной допускаемой погрешности или стандартного отклонения для оценки квалификации может быть установлено как значение, соответствующее такому уровню функционирования участника, который, по мнению регулирующего органа, органа по аккредитации или технических экспертов провайдера проверки квалификации, является приемлемым для участников.

8.2.2 Установленная максимальная допускаемая погрешность может быть преобразована в стандартное отклонение для оценки квалификации с помощью деления этого допускаемого значения на то же число, на которое умножают σ_{pt} при определении границы сигнала действия (или неприемлемого результата). Аналогично: установленное значение σ_{pt} может быть преобразовано в δ_E.

8.3 На основании опыта предыдущих туров программы проверки квалификации

8.3.1 Стандартное отклонение для оценки квалификации (σ_{pt}) и максимальная допускаемая погрешность (δ_E) могут быть определены на основе опыта предыдущих туров программы проверки квалификации одной и той же измеряемой величины с сопоставимыми значениями свойств, если участники используют согласованные процедуры измерений. Данный подход полезен, если среди экспертов нет согласия относительно целевого значения. Преимущества такого подхода состоят в следующем:

- оценки основаны на разумных ожиданиях функционирования участников;
- критерии оценивания не изменяют от тура к туру программы проверки квалификации из-за случайных отклонений или изменений в составе участников;
- критерии оценивания не отличаются у различных провайдеров проверки квалификации, если для испытаний или калибровки утверждены два или более провайдеров проверки квалификации.

8.3.2 Анализ предыдущих туров программы проверки квалификации должен включать рассмотрение результатов функционирования, достижимых компетентными участниками, на которые не влияют новые участники или случайные отклонения, вызванные, например, меньшим объемом группы или другими факторами, уникальными для конкретного тура. Определения могут быть сделаны субъективно путем изучения предыдущих туров программы проверки квалификации на предмет согласованности или объективности с использованием выборочных средних или регрессионной модели, которая корректирует значение измеряемой величины. Уравнение регрессии может представлять собой прямую линию, а может быть кривой [31]. Следует рассматривать стандартные отклонения и относительные стандартные отклонения, делая выбор на основе того, что является более последовательным

в соответствующем диапазоне уровней измеряемой величины. Аналогично может быть получена соответствующая максимальная допускаемая погрешность.

8.3.3 Если критерий для оценивания функционирования участника основан на согласованной статистике предыдущих туров программы проверки квалификации, следует использовать робастные оценки стандартного отклонения.

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

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

8.4 Использование общей модели

8.4.1 Значение стандартного отклонения для оценки квалификации может быть выведено из общей модели воспроизводимости метода измерений. Преимущества этого метода состоят в объективности и согласованности измеряемых величин, а также в его эмпирической основе. В зависимости от используемой модели данный подход можно рассматривать как частный случай критерия соответствия целевому значению.

8.4.2 Любое среднее стандартное отклонение, выбранное с помощью общей модели, должно быть обосновано. Если для очень большой или очень маленькой доли участников установлены сигналы действия или предупреждения, провайдер проверки квалификации должен убедиться, что это соответствует цели программы проверки квалификации.

8.4.3 Конкретная оценка, учитывающая специфику задачи измерения, как правило, предпочтительнее общего подхода. Следовательно, прежде чем использовать общую модель, следует исследовать возможность использования подходов, описанных в 8.2, 8.3 и 8.5.

Пример — Кривая Хорвитца.

Общая модель для применения при выполнении химических исследований описана Хорвитцем [22] и модифицирована Томпсоном [31]. Такой подход позволяет построить общую модель воспроизводимости аналитических методов, которая может быть использована для получения следующего 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, подлежащих определению, $0 \le c \le 1$.

Примечание 1 — Модель Хорвитца является эмпирической, она основана на наблюдениях в ходе совместных испытаний многих параметров в течение длительного периода времени. Значения σ_R представляют собой средние верхние границы межлабораторной изменчивости, когда в совместном исследовании не было существенных проблем. Таким образом, значение σ_R может быть непоходящим критерием при определении квалификации в программе проверки квалификации.

Примечание 2 — Пример определения значений на основе модифицированной модели Хорвитца приведен в Е.9.

8.5 Использование стандартных отклонений повторяемости и воспроизводимости на основе данных предыдущего совместного исследования прецизионности метода измерений

8.5.1 Если метод измерений, используемый в программе проверки квалификации, стандартизован и доступна информация о стандартных отклонениях повторяемости (σ_r) и воспроизводимости (σ_R) метода, стандартное отклонение для оценки квалификации (σ_{pt}) может быть вычислено с использованием этой информации следующим образом:

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

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

Примечание — Формула (9) выведена на основе базовой модели случайных воздействий в соответствии с ISO 5725-2.

8.5.2 Если стандартные отклонения повторяемости и воспроизводимости зависят от среднего арифметического результатов испытаний, функциональные зависимости должны быть выведены методами, описанными в ISO 5725-2. Затем эти зависимости следует использовать для вычисления значений стандартных отклонений повторяемости и воспроизводимости, соответствующих приписанному значению, используемому в программе проверки квалификации.

8.5.3 Описанные выше методы следует применять только в том случае, если совместные исследования были проведены в соответствии с требованиями ISO 5725-2 или аналогичными способами.

Примечание — Пример приведен в Е.10.

8.6 На основании данных, полученных в текущем туре программы проверки квалификации

8.6.1 В данном подходе стандартное отклонение для оценки квалификации (σ_{pt}) вычисляют по результатам участников, полученным в одном и том же туре программы проверки квалификации. При использовании такого подхода, как правило, удобнее применять в качестве показателя функционирования показатель z. Для расчета σ_{pt} обычно следует использовать робастную оценку стандартного отклонения результатов, зафиксированных всеми участниками, рассчитанную с использованием метода, приведенного в приложении С. В программах проверки квалификации, использующих δ_E и D или D%, P_A (процент допустимого отклонения) можно использовать в качестве нормированного показателя для сравнения по измеряемым величинам и турам (см. 9.3.6).

8.6.2 Использование результатов участников может привести к определению неподходящих критериев для оценивания их функционирования. Провайдер проверки квалификации должен убедиться, что σ_{pt}, используемое для оценки функционирования, соответствует цели.

8.6.2.1 Провайдер проверки квалификации должен установить ограничение на наименьшее значение σ_{pt} для использования в случае, когда робастное стандартное отклонение очень мало. Этот предел должен быть выбран таким образом, чтобы, когда погрешность результатов измерений соответствует наиболее трудному предполагаемому использованию, показатель функционирования удовлетворял неравенству $z < 3{,}0$.

Пример — В программе проверки качества ткани одной из измеряемых величин является количество нитей на сантиметр. Робастное стандартное отклонение может быть небольшим в некоторых турах (менее 1 нити на сантиметр), а погрешности менее 4 нитей на сантиметр считаются незначительными. Провайдер проверки квалификации определил, что в качестве σ_{pt} следует использовать робастное стандартное отклонение, если только оно не составляет менее 1,3 нитей/см, в этом случае используют $\sigma_{pt} = 1{,}3$.

8.6.2.2 Провайдер проверки квалификации должен установить ограничение на наибольшее значение σ_{pt}, которое следует использовать, или на результаты измерений, которые могут быть оценены как «приемлемые» (отсутствие сигнала) в случае, если робастное стандартное отклонение очень велико. Этот предел должен быть выбран таким образом, чтобы результаты, которые не соответствуют цели, вызывали сигнал действия.

8.6.2.3 В некоторых случаях провайдер проверки квалификации может установить верхние или нижние пределы интервала результатов, которые могут быть оценены как «приемлемые» (без сигналов предупреждения и действия), когда симметричные интервалы включают результаты, не соответствующие цели.

Пример — В обязательной программе проверки квалификации для непитьевой воды установлены такие правила, что результаты должны находиться в границах $3\sigma_{pt}$ от робастного среднего результатов участников. Однако, поскольку в некоторых случаях размах приемлемых результатов может включать 0 мкг/л, любой результат, составляющий менее 10 % от указанного значения, должен вызывать сигнал действия («неприемлемый»). В состав образца для проверки квалификации входит 4,0 мкг/л регулируемого вещества. Робастное среднее участников составляет 3,2 мкг/л, а $\sigma_{pt} = 1{,}1$ мкг/л. Следовательно, участник может представить результат 0,0 мкг/л и быть в пределах $\pm 3\sigma_{pt}$, но любой результат менее 0,4 мкг/л оценивают как «неприемлемый».

8.6.3 Основными преимуществами такого подхода, благодаря успешному использованию во многих ситуациях, являются простота и распространенность. Иногда он может быть единственным выполнимым подходом.

8.6.4 У данного подхода существует несколько недостатков:

a) значение σ_{pt} может существенно изменяться от тура к туру при выполнении программы проверки квалификации, что затрудняет участнику использование значений показателя z для выявления тенденций, которые сохраняются в течение нескольких туров;

b) оценки стандартных отклонений могут быть недостоверными, когда количество участников программы проверки квалификации невелико или при объединении результатов, полученных с помощью различных методов. Например, если p = 20, оценка стандартного отклонения данных из нормального распределения может отличаться на ±30 % от своего истинного значения от одного тура программы проверки квалификации к другому;

c) использование показателей рассеяния (изменчивости), полученных на основе данных, может привести к приблизительно постоянной доле приемлемых показателей. Как правило, плохое функционирование лаборатории может быть не выявлено путем контроля показателей, и по результатам хорошего функционирования участники могут получить плохие показатели;

d) отсутствие полезной интерпретации с точки зрения пригодности результатов для любого конечного использования.

Примечание — Примеры использования данных участников приведены в комплексном примере в Е.3.

8.7 Мониторинг межлабораторной согласованности

8.7.1 Для проверки функционирования участников и оценки преимуществ программы проверки квалификации для участников провайдер проверки квалификации должен применять процедуру мониторинга межлабораторной согласованности для отслеживания изменений в функционировании участников и обеспечения правильности выполнения статистических процедур.

8.7.2 Результаты, полученные в каждом туре программы проверки квалификации, должны быть использованы для расчета оценок стандартных отклонений воспроизводимости метода измерений (и повторяемости, при необходимости) с использованием робастных методов, описанных в приложении С. Эти оценки следует последовательно наносить на графики или использовать в виде временных рядов вместе со значениями стандартных отклонений повторяемости и воспроизводимости, полученными при исследовании прецизионности в соответствии с ISO 5725-2 (при наличии), и/или σ_{pt}, при применении методов, описанных в 8.2—8.4.

8.7.3 Полученные графики должны быть проверены провайдером проверки квалификации. Если графики показывают, что значения прецизионности, полученные в конкретном туре проверки квалификации, в два или более раз превышают средние значения, полученные из предыдущих данных или из опыта, то провайдер проверки квалификации должен выяснить, почему согласованность в этом туре была хуже, чем раньше. Аналогично тенденция к улучшению или ухудшению значений прецизионности становится основанием для анализа наиболее вероятных причин их появления.

9 Расчет статистик функционирования

9.1 Общие положения для определения показателя функционирования

9.1.1 Статистики, используемые при определении оценки функционирования, должны соответствовать цели (целям) программы проверки квалификации.

Примечание — Статистики функционирования наиболее полезны, если статистические данные и их источники понятны участникам и другим заинтересованным сторонам.

9.1.2 Показатели функционирования должны быть легко анализируемы по уровням измерений и различным турам программы проверки квалификации.

9.1.3 Результаты участников должны быть проанализированы и определены как соответствующие предположениям, использованным при разработке программы проверки квалификации для получения значимых статистик показателя функционирования. Например, отсутствие свидетельств ухудшения образца для проверки квалификации, или смешения результатов групп участников, или серьезных нарушений каких-либо статистических предположений относительно особенностей данных.

9.1.4 В целом нецелесообразно использование методов оценивания, которые намеренно классифицируют фиксированную долю результатов как результаты, вызывающие сигнал действия.

9.2 Ограничение неопределенности приписанного значения

9.2.1 Если стандартная неопределенность $u(x_{pt})$ приписанного значения велика по сравнению с критерием для оценивания функционирования, то существует вероятность того, что некоторые участники получат сигналы действия и предупреждения из-за неточности в определении приписанного значения, а не по какой-либо другой причине, связанной с функционированием участника. Поэтому стандартная неопределенность приписанного значения должна быть определена и доведена до сведения участников (см. ISO/IEC 17043:2010, 4.4.5 и 4.8.2).

При выполнении следующего критерия неопределенность приписанного значения можно считать незначимой и ее не следует учитывать при интерпретации результатов тура программы проверки квалификации.

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

Примечание — Значение $0{,}3\sigma_{pt}$ эквивалентно $0{,}1\delta_E$, если сигнал действия формируется при $|z| \geq 3{,}0$.

9.2.2 Если этот критерий не выполняется, то провайдер проверки квалификации должен рассмотреть следующие меры, гарантируя, что все предпринятые действия соответствуют политике согласованной оценки функционирования в программе проверки квалификации:

a) выбор метода определения приписанного значения таким образом, чтобы его неопределенность удовлетворяла критерию формулы (10);

b) использование неопределенности приписанного значения при интерпретации результатов программы проверки квалификации (см. 9.5 для показателей z', 9.6 — для показателей ζ, 9.7 — для показателей E_n);

c) фиксирование отдельных значений и неопределенностей для каждой подгруппы участников (например, участников, использующих разные методы измерений), если приписанное значение выведено из результатов участников, а большая неопределенность является следствием различий между идентифицированными подгруппами участников.

Примечание — В гармонизированном протоколе IUPAC [32] описана конкретная процедура обнаружения бимодальности, основанная на проверке графика ядерной плотности с установленной «шириной полосы пропускания».

d) информирование участников о том, что неопределенность приписанного значения не является незначимой и это может повлиять на оценки.

Если ни одно из перечислений a)—d) не может быть применено, то участники должны быть проинформированы о том, что приписанное значение не может быть достоверно определено и не могут быть определены показатели функционирования участников.

Примечание — Методы, приведенные в данном разделе, представлены в E.3 и E.4.

9.3 Оценка отклонения (погрешности измерений)

9.3.1 Пусть x_i — результат (или среднее арифметическое при повторениях), указанный i-м участником при измерении свойства объекта для проверки квалификации в одном туре программы проверки квалификации. Тогда простым показателем функционирования участника является разность результата участника x_i и приписанного значения x_{pt}:

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

D_i можно интерпретировать как погрешность измерения этого результата в той же степени, в которой приписанное значение можно считать общепринятым или опорным значением.

Разность D_i может быть выражена в тех же единицах, что и приписанное значение, или в виде процентной разности:

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

Примечание — Формула (12) не может быть применена, если $x_{pt} = 0$.

9.3.2 Разность D_i или D_i %, как правило, сравнивают с критерием δ_E, основанным на целевом значении или данных предыдущих туров программы проверки квалификации; критерий обозначаемый δ_E, представляет собой допускаемое значение погрешности измерений. Если $-\delta_E < D_i < \delta_E$, то функционирование участника считают «приемлемым» («сигнала нет»). (Тот же критерий применим для D_i %, в зависимости от выражения δ_E.)

9.3.3 Значение δ_E связано с σ_{pt}, используемым для показателей z (см. 9.4), если σ_{pt} определяют в соответствии с целевым значением или средним по предыдущим турам программы проверки

квалификации. Соотношение определяют по критерию оценивания для показателей z. Например, если $z \geq 3$ формирует сигнал действия, то $\delta_E = 3\sigma_{pt}$ или, что эквивалентно, $\sigma_{pt} = \delta_E/3$. При проверке квалификации в области медицины и в технических спецификациях по методам измерений продукции общепринятыми являются различные выражения для δ_E.

9.3.4 Преимущество D как статистики функционирования участника и δ_E как критерия функционирования участника, состоит в том, что участники интуитивно понимают эти статистики, поскольку они напрямую связаны с погрешностью измерений и представляют собой общие критерии определения соответствия целевому значению. Преимущество D% состоит в том, что он понятен на интуитивном уровне, стандартизован для уровня измеряемой величины и связан с распространенными причинами погрешностей (например, неправильной калибровкой или смещением при разбавлении).

9.3.5 Недостатки состоят в том, что во многих странах или областях измерений проверка квалификации не является общепринятой, а также в том, что D не стандартизован и не допускает беглый просмотр отчетов на предмет сигналов действия в программах проверки квалификации при химическом анализе нескольких веществ или в случае, когда критерии соответствия целевому значению изменяются в зависимости от уровня измеряемой величины.

Примечание — Использование D и D% обычно предполагает симметричность распределения результатов участников в том смысле, что допустимый диапазон составляет $-\delta_E < D < \delta_E$.

9.3.6 Для целей сравнения уровней измеряемой величины, если критерии соответствия целевому значению могут изменяться, или в случае комбинации измеряемых величин от тура к туру, D и D% в программе проверки квалификации могут быть преобразованы в нормированный показатель функционирования участника для измеряемых величин. Для этого следует вычислить «процент допустимого отклонения» (P_A) для каждого результата следующим образом:

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

Следовательно, неравенства $P_A \geq 100$ % или $P_A \leq -100$ % указывают на сигнал действия (или «неприемлемое функционирование»).

Примечание 1 — Показатели P_A можно сравнивать по уровням и различным турам программы проверки квалификации или отслеживать с использованием карт. Эти показатели функционирования по использованию и интерпретации аналогичны показателям z, которые имеют общий критерий оценивания, такой как $z \leq -3$ или $z \geq 3$ для сигналов действия.

Примечание 2 — Вариации этой статистики обычно часто используют, особенно в области медицины, где характерны более высокая частота проверки квалификации и использование большого количества анализируемых веществ.

Примечание 3 — Иногда целесообразно использовать абсолютное значение P_A для отражения последовательно приемлемых (или неприемлемых) результатов относительно приписанного значения.

9.4 Показатель *z*

9.4.1 Значение показателя z для результата проверки квалификации x_i вычисляют по формуле (14)

$$z_i = \frac{x_i - x_{pt}}{\sigma_{pt}}, \tag{14}$$

где x_{pt} — приписанное значение;

σ_{pt} — стандартное отклонение для оценки квалификации.

9.4.2 Общепринятая интерпретация показателей z состоит в следующем (см. ISO/IEC 17043:2010, B.4.1.1):

- результат считают приемлемым, если $|z| \leq 2{,}0$;
- результат считают вызывающим сигнал предупреждения, если $2{,}0 < |z| < 3{,}0$;
- результат считают неприемлемым (или сигналом действия), если $|z| \geq 3{,}0$.

Участникам рекомендуется проверить свои процедуры измерений после получения сигналов предупреждения в случае, если они указывают на возникающую или повторяющуюся проблему.

Примечание 1 — В некоторых приложениях провайдеры проверки квалификации используют 2,0 в качестве сигнала действия для показателей z.

Примечание 2 — Выбор критерия σ_{pt} обычно осуществляют таким образом, чтобы он допускал приведенную выше интерпретацию, которую широко используют для оценки квалификации и которая похожа на знакомые границы контрольных карт.

Примечание 3 — Обоснование использования пределов 2,0 и 3,0 для показателей z состоит в следующем. Предполагается, что правильно выполненные измерения дают результаты, которые могут быть описаны (после преобразования, при необходимости) нормальным распределением со средним x_{pt} и стандартным отклонением σ_{pt}. Полученные затем показатели z подчиняются нормальному распределению с нулевым средним и стандартным отклонением 1,0. В этом случае в среднем только около 0,3 % показателей выйдут за пределы диапазона $-3,0 \le z \le 3,0$ и только около 5 % окажутся за пределами диапазона $-2,0 \le z \le 2,0$. Поскольку вероятность того, что z выйдет за пределы ±3,0, очень мала, то вероятность случайного появления сигнала действия очень низка, когда реальной проблемы не существует, поэтому вероятно, что при появлении сигнала действия существует идентифицируемая причина аномалии.

Примечание 4 — Предположение, на котором основана данная интерпретация, применимо только к гипотетическому распределению квалифицируемых лабораторий, но не к распределению наблюдаемых результатов. Не следует делать никаких предположений относительно самих результатов наблюдений.

Примечание 5 — Если истинная межлабораторная изменчивость (стандартное отклонение совокупности) меньше σ_{pt}, то вероятность ошибок снижается.

Примечание 6 — Если стандартное отклонение для оценки квалификации определено одним из методов, описанных в 8.2 или 8.4, оно может существенно отличаться от (робастного) стандартного отклонения результатов, и доли результатов за пределами ±2,0 и ±3,0 могут значительно отличаться от 5 % и 0,3 % соответственно.

9.4.3 Провайдер проверки квалификации должен определить требования к округлению сообщаемых показателей z путем указания количества значащих цифр в результате, а также в приписанном значении и стандартном отклонении для проверки квалификации. Правила округления должны быть включены в информацию, доступную участникам.

Примечание — Иногда бывает полезно иметь более двух цифр после запятой для оценок показателя z.

9.4.4 Если стандартное отклонение результатов участников используют в качестве σ_{pt}, а программы проверки квалификации включают очень большое количество участников, провайдер проверки квалификации может проверить нормальность распределения, используя фактические результаты или показатели z. С другой стороны, если количество участников слишком мало, сигнал действия может отсутствовать. В этом случае графические методы, которые объединяют показатели функционирования за несколько туров программы проверки квалификации, могут дать более удобный способ индикации качества функционирования участников, чем результаты отдельных туров.

9.5 Показатель z'

9.5.1 Если существуют сомнения по поводу неопределенности приписанного значения $u(x_{pt})$, например, когда $u(x_{pt}) > 0,3\sigma_{pt}$, то неопределенность может быть учтена путем увеличения знаменателя показателя функционирования. Эту статистику называют показателем z' и вычисляют следующим образом (с обозначениями, как в 9.4):

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

Примечание — Если x_{pt} и/или σ_{pt} вычисляют по результатам участников, то показатели функционирования коррелируют с результатами отдельных участников, поскольку индивидуальные результаты влияют как на робастное среднее, так и на стандартное отклонение. Корреляция для отдельного участника зависит от весового коэффициента, присвоенного этому участнику в комбинированной статистике. По этой причине показатели функционирования, включающие неопределенность приписанного значения без учета корреляции, представляют собой заниженные оценки показателей по сравнению с показателями, формула расчета которых включает ковариацию. Например, если $u(x_{pt}) = 0,3\sigma_{pt}$, то недооценка составляет приблизительно 10 % от оценки z'. Следовательно, формулу (15) можно использовать, когда x_{pt} и/или σ_{pt} определены по результатам участников.

9.5.2 Показатели D и D% также могут быть изменены с учетом неопределенности приписанного значения с помощью формулы (16) для расширения δ_E до δ'_E

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

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

9.5.3 Показатели z' могут быть интерпретированы так же, как и показатели z (см. 9.4), с использованием тех же критических значений 2,0 и 3,0 — в зависимости от плана программы проверки квалификации. Аналогично показатели D и D% затем сравнивают с δ'_E (см. 9.3).

9.5.4 Сравнение формул для показателей z и z' в 9.4 и 9.5 показывает, что показатели z' для тура программы проверки квалификации всегда меньше соответствующих показателей z на постоянный коэффициент, заданный формулой (17)

$$\frac{\sigma_{pt}}{\sqrt{\sigma_{pt}^2 + u^2\left(x_{pt}\right)}}. \tag{17}$$

При выполнении рекомендации по ограничению неопределенности приписанного значения в 9.2.1 этот коэффициент изменяется в диапазоне неравенства (18)

$$0{,}96 < \frac{\sigma_{pt}}{\sqrt{\sigma_{pt}^2 + u^2\left(x_{pt}\right)}} < 1{,}00. \tag{18}$$

В этом случае показатель z' близок по значению к показателю z, и может быть сделан вывод, что неопределенность приписанного значения для оценивания функционирования является незначимой.

Если требование 9.2.1 в отношении неопределенности приписанного значения не выполняется, то разность значений показателя z' и показателя z может быть такой, что некоторые показатели z становятся больше критического значения 2,0 или 3,0 и таким образом дают «сигнал предупреждения» или «сигнал действия», в то время как соответствующие показатели z' не превышают этих критических значений и поэтому не дают этих сигналов.

В общем, для ситуаций, когда приписанное значение и/или σ_{pt} не определены по результатам участников, z' может быть предпочтительнее, поскольку при выполнении критерия, указанного в 9.2.1, разность между z и z' является незначимой.

9.6 Показатель дзета (ζ)

9.6.1 Показатель дзета может быть полезен, если целью программы проверки квалификации является оценка способности участника получать результаты, близкие к приписанному значению с заявленной неопределенностью.

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

$$\zeta_i = \frac{x_i - x_{pt}}{\sqrt{u^2\left(x_i\right) + u^2\left(x_{pt}\right)}}, \tag{19}$$

где $u(x_i)$ — собственная оценка стандартной неопределенности, выполненная участником по его результату x_i;

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

Примечание 1 — Если приписанное значение x_{pt} вычисляют как согласованное значение по результатам участников, тогда x_{pt} коррелирует с результатами отдельных участников. Корреляция для отдельного участника зависит от весового коэффициента, присвоенного этому участнику в приписанном значении и, в меньшей степени, от неопределенности приписанного значения. По этой причине показатели функционирования, включающие неопределенность приписанного значения без учета корреляции, представляют собой заниженные значения показателей, которые могли бы быть получены, при учете ковариации. Недооценка является несущественной, если неопределенность приписанного значения невелика; если используют робастные методы, это наименее чувствительно для участников, демонстрирующих крайние результаты, которые с наибольшей вероятностью получат неприемлемые показатели функционирования. Таким образом, формулу (19) можно использовать с согласованной статистикой без корректировки корреляции.

Примечание 2 — Показатели ζ отличаются от показателей E_n (см. 9.7) за счет использования стандартных неопределенностей $u(x_i)$ и $u(x_{pt})$, а не расширенных неопределенностей $U(x_i)$ и $U(x_{pt})$. Показатель ζ может быть особенно полезен, если участники используют различные процедуры измерений, приводящие к очень разным неопределенностям измерений. Показатели ζ выше 2 или ниже минус 2 могут быть вызваны систематической погрешностью методов или плохой оценкой участником неопределенности измерений. Таким образом, показатели ζ обеспечивают грубую оценку окончательного результата, представленного участником.

9.6.2 Использование показателей ζ позволяет непосредственно оценить способность лаборатории давать правильные результаты, т. е. результаты, согласованные с x_{pt} в пределах ее неопределенности измерений. Показатели ζ могут быть интерпретированы с использованием тех же критических значений 2,0 и 3,0, что и для показателей z, или с использованием коэффициентов охвата участника, используемых при оценке расширенной неопределенности. Однако неприемлемый показатель ζ может

указывать либо на большое отклонение x_i от x_{pt}, либо на заниженную оценку неопределенности со стороны участника, либо на сочетание того и другого.

Примечание — Провайдеру проверки квалификации может быть полезно представить дополнительную информацию о достоверности сообщенных неопределенностей. Полезные рекомендации по такой оценке приведены в 9.8.

9.6.3 Показатели ζ могут быть использованы в сочетании с показателями z в качестве вспомогательного средства для улучшения результатов участников следующим образом. Если участник получает показатели z, которые неоднократно превышают критическое значение 3,0, ему может быть полезно изучить свою процедуру испытаний этап за этапом и рассчитать оценку неопределенности этой процедуры. Оценивание неопределенности позволит определить этапы процедуры, на которых возникают наибольшие неопределенности, так что участник может видеть, на что следует направить усилия для достижения улучшения. Если показатели ζ участника также неоднократно превышают критическое значение 3,0, это означает, что оценивание неопределенности участника не включает все значимые источники неопределенности (т. е. он упускает что-то важное). И наоборот, если участник повторно получает показатели $z \geq 3$, но показатели $\zeta \leq 2$, это показывает, что участник, возможно, точно оценил неопределенность своих результатов, но его результаты не соответствуют средним значениям показателя функционирования программы проверки квалификации. Это может иметь место, например, для участника, который использует метод скрининга в процедурах измерений, когда другие участники применяют количественные методы. Никаких действий не требуется, если участник считает, что неопределенность его результатов достаточна.

Примечание — Если используют только показатель ζ, он может быть интерпретирован только как проверка соответствия неопределенности участника конкретному наблюдаемому отклонению и не может быть интерпретирован как признак соответствия результатов конкретного участника целевому значению. Определение соответствия целевому значению может быть выполнено отдельно (например, участником или органом по аккредитации) с помощью проверки отклонения x-x_{pt} или комбинированных стандартных неопределенностей по сравнению с целевой неопределенностью.

9.7 Число E_n

9.7.1 Использование числа E_n может быть полезно, если целью программы проверки квалификации является оценка способности участника получать результаты, близкие к приписанному значению, в пределах заявленной им расширенной неопределенности. Эта статистика является обычной для проверки квалификации при калибровке, но она может быть использована и для других видов проверки квалификации.

Эту статистику вычисляют по формуле (20)

$$(E_n)_i = \frac{x_i - x_{pt}}{\sqrt{U^2(x_i) + U^2(x_{pt})}}, \tag{20}$$

где x_{pt} — приписанное значение, определенное в референтной лаборатории;

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

$U(x_i)$ — расширенная неопределенность результата участника x_i.

Примечание — Прямое объединение расширенных неопределенностей не соответствует требованиям ISO/IEC Guide 98-3 и неэквивалентно вычислению комбинированной расширенной неопределенности, если коэффициенты охвата и значения эффективного числа степеней свободы неидентичны для $U(x_i)$ и $U(x_{pt})$.

9.7.2 Число E_n следует интерпретировать с осторожностью, поскольку оно представляет собой отношение двух отдельных (но взаимосвязанных) показателей функционирования. Числитель представляет собой отклонение результата от приписанного значения и имеет интерпретацию, приведенную в 9.3. Знаменатель представляет собой комбинированную расширенную неопределенность, которая не должна быть больше числителя, если участник правильно определил $U(x_i)$ и если провайдер проверки квалификации правильно определил $U(x_{pt})$. Следовательно, оценки $E_n > 1{,}0$ или $E_n < -1{,}0$ могут указывать на необходимость пересмотра оценок неопределенности или исправления результата измерения; аналогично $-1{,}0 \leq E_n \leq 1{,}0$ следует понимать как указание на успешное функционирование только в том случае, если неопределенности обоснованы и отклонение $x_i - x_{pt}$ меньше, чем необходимо потребителям участника.

Примечание — Несмотря на то, что интерпретация числа E_n может быть достаточно трудной, это не мешает его использованию. Включение информации о неопределенности в интерпретацию результатов проверки квалификации может сыграть важную роль в улучшении понимания участниками неопределенности измерений и ее оценивании.

9.8 Оценивание неопределенностей участников при испытаниях

9.8.1 С расширением применения ISO/IEC 17025 улучшается понимание неопределенности измерений. Использование лабораторных оценок неопределенности при оценивании функционирования участников распространено в программах проверки квалификации в различных областях калибровки, например с использованием числа E_n, но это не было распространено при проверке квалификации испытательных лабораторий. Показатель ζ, описанный в 9.6, и число E_n, описанное в 9.7, являются вариантами оценивания результатов с учетом заявленной неопределенности.

9.8.2 Некоторые провайдеры проверки квалификации признали полезность запросов у лабораторий неопределенности результатов при проверке квалификации. Это может быть полезно даже тогда, когда неопределенности не используют при расчете показателей. Существует несколько целей сбора такой информации:

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

b) участники могут анализировать сообщенную ими неопределенность вместе с неопределенностями других участников, чтобы оценить их согласованность (или несогласованность) и таким образом получить возможность определить, все ли составляющие неопределенности учитывает их оценивание или оно завышает вклады некоторых составляющих неопределенности;

c) проверка квалификации может быть использована для подтверждения заявленной неопределенности, это проще сделать, когда о неопределенности сообщают вместе с результатом.

Примечание — Пример анализа данных, когда участники сообщают о неопределенности, приведен в Е.4.

9.8.3 Если x_{pt} определяют, используя процедуры, установленные в 7.3—7.6, и $u(x_{pt})$ соответствует критерию, указанному в 9.2.1, то маловероятно, что результат участника будет иметь меньшую стандартную неопределенность, чем $u(x_{pt})$, поэтому $u(x_{pt})$ может быть использована в качестве нижнего предела для скрининга, обозначаемого $u_{\min}$. Если приписанное значение определяют по результатам участников (см. 7.7), то провайдер проверки квалификации должен определить практические пределы для скрининга $u_{\min}$.

Примечание — Если $u(x_{pt})$ включает вклады изменчивости, связанной с неоднородностью или нестабильностью, $u(x_i)$ участника может быть меньше $u_{\min}$.

9.8.4 Также маловероятно, что сообщенная стандартная неопределенность всех участников будет в 1,5 раза выше робастного стандартного отклонения участников (1,5s^*), поэтому это значение может быть использовано в качестве практического верхнего предела для скрининга сообщенных неопределенностей, обозначаемого $u_{\max}$.

Примечание 1 — Коэффициент 1,5 является верхним пределом вариабельности стандартных отклонений, который можно ожидать для согласованного стандартного отклонения с 10 и более результатами, основанным на квадратном корне из процентилей F-распределения. Провайдер проверки квалификации, применяющий данную процедуру, может использовать другой множитель.

Примечание 2 — Применение коэффициентов более 1,5 возможно, если участники используют широкий спектр методов.

9.8.5 Если $u_{\min}$ или $u_{\max}$ или другие критерии используют для выявления отклоняющихся неопределенностей, провайдер проверки квалификации должен объяснить это участникам и дать им понять, что заявленная неопределенность $u(x_i)$ может быть обоснованной, даже если она ниже $u_{\min}$ или больше $u_{\max}$, и, если это происходит, участники и все заинтересованные стороны должны проверить результат или оценку неопределенности. Аналогично: заявленная неопределенность может быть больше $u_{\min}$ и меньше $u_{\max}$ и все равно быть необоснованной. Это только информативный индикатор.

9.8.6 Провайдеры проверки квалификации могут также обратить внимание на необычно высокие или низкие неопределенности, использующие, например:

- установленные квантили сообщаемых неопределенностей (например, ниже 5-го процентиля и выше 95-го процентиля сообщаемой стандартной или расширенной неопределенности);

- пределы, основанные на предполагаемом распределении со шкалой, основанной на рассеянии сообщаемых неопределенностей;
- требуемую неопределенность измерений.

Примечание — Поскольку неопределенности скорее всего не подчиняются нормальному распределению, необходимо преобразование при использовании пределов, которое предполагает приближенно нормальное или нормальное распределение, например, границы на диаграмме «ящик с усами» основаны на межквартильном размахе и имеют вероятностную интерпретацию только в случае, если распределение данных близко к нормальному.

9.9 Комбинированные показатели функционирования

9.9.1 Как правило, в рамках одного тура программы проверки квалификации результаты должны быть получены более чем по одному образцу для проверки квалификации или более чем по одной измеряемой величине. В таком случае результаты для каждого образца для проверки квалификации и для каждой измеряемой величины необходимо интерпретировать в соответствии с 9.3—9.7, т. е. результаты для каждого образца для проверки квалификации и каждой измеряемой величины следует оценивать отдельно.

9.9.2 Существуют ситуации, когда два или более образцов для проверки квалификации со специально разработанными уровнями включают в программу проверки квалификации для измерения других аспектов функционирования, таких как исследование повторяемости, систематической погрешности или линейности модели. Например, два схожих образца для проверки квалификации могут быть использованы в программе проверки квалификации для обработки с применением диаграммы Юдена, в соответствии с 10.5. В таких случаях провайдер проверки квалификации должен представить участникам полное описание статистического плана эксперимента и используемых процедур.

9.9.3 Графические методы, описанные в разделе 10, следует использовать, если результаты получены более чем для одного объекта для проверки квалификации или для нескольких измеряемых величин, при условии, что они тесно связаны и/или получены одним и тем же методом. Эти процедуры комбинируют показатели функционирования таким образом, что не скрывают высокие значения индивидуальных показателей и могут выявить дополнительную информацию о функционировании участников, в частности корреляцию между результатами измерений различных измеряемых величин, которую не видно в таблицах индивидуальных показателей.

9.9.4 В программах проверки квалификации, включающих большое количество измеряемых величин, для оценки функционирования может быть использовано количество или доля сигналов действия и предупреждения.

9.9.5 Комбинированные показатели функционирования или показатели, связанные с поощрением или штрафом, следует использовать с большой осторожностью, поскольку может быть трудно описать статистические предположения, лежащие в основе показателей. В то время как комбинированные показатели функционирования для результатов измерений одной и той же измеряемой величины в различных образцах для проверки квалификации могут иметь ожидаемые распределения и могут быть полезны для выявления стойкого смещения, усредненные или суммированные показатели по разным измеряемым величинам одного и того же или различных образцов для проверки квалификации могут скрывать наличие смещения в результатах по единственной измеряемой величине. Таким образом, метод расчета, интерпретация и ограничения всех используемых комбинированных показателей или показателей, связанных с поощрением или штрафом, должны быть доведены до сведения участников.

10 Графические методы для описания показателей функционирования

10.1 Применение графических методов

Провайдеру проверки квалификации, как правило, следует использовать показатели функционирования, полученные в каждом туре программы проверки квалификации, для составления графиков, описанных в 10.2 и 10.3. Преимущество использования графиков показателей, таких как P_A, z, z', ζ или числа E_n, состоит в том, что эти графики могут быть изображены с использованием нормированных осей, что упрощает их представление и интерпретацию. Участникам должны быть представлены графики, позволяющие каждому участнику видеть, как его собственные результаты соотносятся с результатами, полученными другими участниками. Для того чтобы каждый участник мог идентифицировать свои собственные результаты, но не мог определить, какой участник получил какой-либо другой

результат, могут быть использованы специальные буквенные или цифровые коды. Кроме того, графики дают возможность провайдеру проверки квалификации и любому аккредитующему органу сделать вывод об общей результативности программы проверки квалификации и необходимости пересмотра критериев, используемых для оценки функционирования.

Примечание — В следующих подразделах приведен неполный перечень выбранных методов, которые признаны полезными при проведении проверки квалификации. Могут быть полезны и другие графические методы, в том числе диаграмма «ящик с усами» и графики результатов участников с учетом заявленной неопределенности. Диаграмма «ящик с усами» описана в ISO 16269-4 [38]. Полезный график результатов участников и неопределенностей описан в [39].

10.2 Гистограммы результатов или показателей функционирования

10.2.1 Гистограмма является общим статистическим методом и полезна на двух различных этапах анализа результатов проверки квалификации. Гистограмма полезна на этапе предварительного анализа при проверке обоснованности используемых статистических предположений или при наличии отклонений, которые невозможно предвидеть, таких как бимодальное распределение, большая доля выбросов или неожидаемая необычная асимметрия.

Приведение гистограмм может быть полезно в отчетах о программе проверки квалификации, для описания показателей функционирования или для сравнения результатов, например, полученных различными методами или на различных образцах для проверки квалификации. Гистограммы особенно полезны в индивидуальных отчетах о небольших или средних программах проверки квалификации (где менее 100 участников), чтобы участники могли оценить свои результаты по сравнению с результатами других участников, например, выделив в вертикальной полосе блок для указания результата участника или, в небольших программах проверки квалификации (где менее 50 участников), используя индивидуальную гистограмму для каждого участника.

10.2.2 Гистограммы могут быть подготовлены с использованием фактических результатов участника или показателей функционирования. Преимущество приведения результатов участников состоит в том, что они напрямую связаны с представленными данными и могут быть оценены без дальнейших вычислений или преобразования показателя функционирования в погрешность измерений. Преимущество гистограмм, составленных на основе показателей функционирования, состоит в том, что они напрямую связаны с оценками функционирования и их можно легко сравнивать по измеряемым величинам и турам программы проверки квалификации.

Ширина и высота столбца, используемые в гистограмме, должны быть определены для каждого набора данных на основе изменчивости и количества результатов. Часто это можно сделать, опираясь на данные предыдущих проверок квалификации, но в большинстве случаев после первичного анализа необходимо скорректировать группировку данных. Если в гистограмме использованы показатели функционирования, полезно иметь шкалу, основанную на стандартном отклонении для оценки квалификации, и выделить точки сигналов предупреждения и действия.

10.2.3 Шкала и интервалы гистограммы должны быть выбраны так, чтобы можно было обнаружить бимодальность (если она присутствует) без создания ложных сигналов предупреждения из-за разрешения результатов измерений или небольшого количества результатов.

Примечание 1 — Внешний вид гистограммы зависит от выбранной ширины столбца и расположения его границ (при постоянной ширине столбца это в значительной степени зависит от начальной точки). Если ширина столбца слишком мала, график отражает много несущественных состояний (мелких мод); если ширина столбца слишком велика, то значимые моды вблизи основной могут быть неразличимы. Внешний вид относительной ширины и высоты соседних полос заметно изменяется при изменении начальной точки или ширины столбца, особенно в тех случаях, когда набор данных небольшой и/или демонстрирует некоторое разделение на классы.

Примечание 2 — Пример построения гистограммы приведен в Е.3.

10.3 График ядерной плотности

10.3.1 График ядерной плотности, часто сокращенно называемый «графиком плотности», представляет собой плавную кривую, описывающую общую форму плотности распределения набора данных. Идея, лежащая в основе оценки ядерной плотности, состоит в том, что каждую точку данных заменяют заданным распределением (как правило, нормальным), центрированным в данной точке со стандартным отклонением σ_k (σ_k обычно называют «шириной полосы»). Все распределения объединяют, и результирующее распределение, масштабированное к единичной площади, дает «оценку плотности», которую можно изобразить в виде сглаженной кривой.

10.3.2 Ниже приведены действия, которые следует выполнить для подготовки графика ядерной плотности. Предполагается, что график должен быть построен по набору данных X, состоящему из p значений x_1, x_2, …, x_p. Обычно это результаты участников, но могут быть и показатели функционирования, полученные на основе этих результатов.

i) Выбирают соответствующую ширину полосы σ_k. Для этого существуют два подходящих способа:

a) для общей проверки устанавливают $\sigma_k = 0{,}9\ s^*/p^{0,2}$, где s^* — робастное стандартное отклонение значений x_1, …, x_p, вычисленное с использованием процедур, описанных в С.2 или С.3;

b) при проверке набора данных на наличие больших мод, значимых по сравнению с критерием для определения оценки функционирования, устанавливают $\sigma_k = 0{,}75\ \sigma_{pt}$, если используют оценки показателя z или показателя ζ, или $\sigma_k = 0{,}25\delta_E$, если используют D и D%.

Примечание 1 — Вариант a), приведенный выше, предложен в [30], где рекомендовано s^*, основанное на нормализованном межквартильном размахе (nIQR). Другие правила выбора ширины полосы, которые обеспечивают аналогичные результаты, включают правило Скотта [29], которое заменяет множитель 0,9 на 1,06. В [29] описан почти оптимальный, но гораздо более сложный метод выбора ширины полосы. На практике различия при визуальном контроле незначительны, и выбор зависит от доступности программного обеспечения.

Примечание 2 — Вариант b), приведенный выше, соответствует рекомендациям IUPAC [32].

ii) Устанавливают область построения графика от $q_{\min}$ до $q_{\max}$ так, чтобы $q_{\min} \le \min(x_1, ..., x_p) - 3\sigma_k$ и $q_{\max} \ge \max(x_1, ..., x_p) + 3\ \sigma_k$.

iii) Выбирают количество точек n_k для построенной кривой. Обычно достаточно $n_k = 200$, если только в пределах области построения графика нет выбросов.

iv) Вычисляют положение точек на графике от q_1 до q_{n_k} по формуле (21)

$$q_i = q_{\min} + (i-1)\frac{\left(q_{n_k} - q_1\right)}{n_k - 1}. \qquad (21)$$

v) Вычисляют n_k плотностей от h_1 до h_{nk} по формуле (22)

$$h_i = \frac{1}{p}\sum_{j=1}^{p}\varphi\left(\frac{x_j - q_i}{\sigma_k}\right),\ i = 1, \ldots, n_k, \qquad (22)$$

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

vi) Строят график h_i в зависимости от q_i.

Примечание 1 — Кривая плотности обычно выходит за пределы диапазона данных; область построения кривой $q_1 - q_{nk}$ обычно выбирают так, чтобы она была по меньшей мере на $3\sigma_k$ больше экстремальных значений набора данных. В случаях, когда кривая выходит за пределы диапазона данных (например, ниже нуля), график может быть отсечен на границе допустимого диапазона, чтобы избежать неправильного толкования.

Примечание 2 — Может быть полезно добавить местоположения отдельных точек данных на графике. Это обычно делают путем нанесения местоположений ниже построенной кривой плотности в виде коротких вертикальных маркеров (иногда называемых «ковриком»), но это также может быть сделано путем нанесения на график точек данных в соответствующих местах вдоль рассчитанной кривой плотности.

Примечание 3 — Графики плотности лучше всего строить с помощью программного обеспечения. Приведенный выше поэтапный расчет может быть выполнен в электронной таблице для небольших объемов набора данных. Собственное и приобретаемое статистическое программное обеспечение часто включает графики плотности, основанные на аналогичном выборе ширины полосы. Продвинутые методы программного построения графиков плотности могут использовать этот алгоритм или более быстрые вычисления, основанные на методах свертки.

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

10.3.3 Форма кривой является индикатором вида распределения, которому подчиняются данные. Различные моды на графике соответствуют отдельным пикам. Выбросы отображены на графике в виде отдельных пиков, хорошо отделенных от основного массива данных.

Примечание 1 — График плотности чувствителен к выбранной ширине полосы σ_k. Если ширина полосы слишком мала, график показывает множество небольших мод, а слишком большие и заметные моды вблизи основной могут быть недостаточно четкими.

Примечание 2 — Как и гистограммы, графики плотности лучше всего использовать с наборами данных от умеренного до большого объема, поскольку небольшие наборы данных (десять и менее) могут содержать незначительное количество выбросов или видимые моды, особенно если в качестве основы для ширины полосы используют робастное стандартное отклонение.

10.4 Столбчатые диаграммы нормированных показателей функционирования

10.4.1 Столбчатые диаграммы подходят для представления показателей функционирования ряда аналогичных характеристик на одном графике. Они показывают наличие общих свойств показателей участника, например, если участник получил несколько высоких показателей z, указывающих на неудовлетворительное функционирование, это может означать, что участник может иметь положительное смещение.

10.4.2 Чтобы подготовить столбчатую диаграмму результатов, собирают нормированные показатели функционирования в столбчатую диаграмму, показанную на рисунке Е.10, в которой показатели каждого участника сгруппированы вместе. Для той же цели могут быть построены диаграммы других нормированных показателей функционирования, таких как показатели D% или P_A.

10.4.3 Если в туре программы проверки квалификации выполняют повторные определения, результаты могут быть использованы для расчета графика показателей прецизионности, например, статистики k, в соответствии с ISO 5725-2, или соответствующей меры, масштабируемой по отношению к стандартному отклонению робастного среднего — такому, как определено в алгоритме S (см. С.4).

Примечание 1 — Столбчатый или другой график показателей функционирования в порядке возрастания показателей может помочь участникам быстро сравнить свои показатели с показателями других участников.

Примечание 2 — Пример столбчатого графика с показателями z приведен в Е.11.

10.5 Диаграмма Юдена

10.5.1 Если в туре программы проверки квалификации исследуют два схожих образца для проверки квалификации, то диаграмма Юдена представляет собой очень информативный графический метод анализа результатов. Этот график может быть полезен для демонстрации наличия (или отсутствия) корреляции результатов по различным образцам для проверки квалификации и для исследования причин сигналов действия.

10.5.2 На график наносят результаты участника или показатели z, полученные по одному из образцов для проверки квалификации, по отношению к результатам участника или показателям z, полученным по другому образцу для проверки квалификации. Вертикальные и горизонтальные линии обычно рисуют для создания четырех квадрантов значений, это помогает в интерпретации. Линии проходят через приписанные значения или медианы для двух распределений результатов или через 0 на графике показателей z.

Примечание — Для надлежащей интерпретации диаграммы Юдена важно, чтобы два образца для проверки квалификации имели аналогичные (или идентичные) уровни измеряемой величины; это необходимо для того, чтобы свойства любой систематической погрешности измерений были одинаковыми в данной области измерений. Диаграммы Юдена могут быть полезны для самых разных уровней измеряемой величины при наличии постоянной систематической погрешности, но они могут вводить в заблуждение, если погрешность калибровки не является последовательно положительной или отрицательной во всем диапазоне уровней измеряемой величины.

10.5.3 При построении диаграммы Юдена интерпретация состоит в следующем:

a) проверяют график на наличие точек, заметно отстоящих от остальных данных. Если участник некорректно выполняет метод измерений, из-за чего его результаты включают систематическую погрешность, точка будет расположена далеко в нижнем левом или верхнем правом квадрантах. Точки, расположенные вдали от остальных в верхнем левом и нижнем правом квадрантах, представляют участников, повторяемость у которых выше, чем у большинства других участников, чьи методы измерений демонстрируют разную чувствительность к компонентам образца для проверки квалификации, или — иногда — участников, которые случайно перепутали образцы для проверки квалификации;

b) проверяют график для выявления свидетельств общей взаимосвязи между результатами по двум образцам для проверки квалификации (например, того, что они лежат приблизительно вдоль наклонной прямой). Если есть свидетельства такой взаимосвязи, это указывает на наличие смещения в результатах участников, которое аналогичным образом влияет на различные образцы для проверки квалификации. Если по результатам визуальной проверки очевидно отсутствие взаимосвязи между результатами (например, точки приблизительно равномерно распределены внутри области, имеющей форму эллипса или круга, обычно с возрастающей плотностью по направлению к центру), то погрешности измерений для двух образцов для проверки квалификации с большой вероятностью независимы. Это можно проверить с помощью ранговой статистики корреляции, если визуальной проверки недостаточно;

c) проверяют график на наличие групп близких результатов участников, либо по диагоналям, либо в других местах. Четко выраженные группы, скорее всего, указывают на различия методов.

Примечание 1 — В исследованиях, когда все участники используют один и тот же метод измерений или графики результатов получены с помощью единственного метода измерений, если результаты лежат вдоль прямой, это может быть свидетельством того, что метод измерений не был надлежащим образом установлен. Исследование метода может затем позволить улучшить воспроизводимость метода в целом.

Примечание 2 — Пример диаграммы Юдена приведен в Е.12.

Примечание 3 — Общий метод построения доверительных эллипсов, который может быть использован для облегчения интерпретации диаграммы Юдена, приведен в [36]. Рассмотрение и сравнение устойчивых к выбросам оценок коэффициентов корреляции и ковариации для доверительных эллипсов при наличии выбросов приведено в [40].

10.6 Графики стандартных отклонений повторяемости

10.6.1 Если в туре программы проверки квалификации участники выполняют повторные измерения, результаты могут быть использованы для построения графика, позволяющего выявить всех участников с необычными значениями среднего и стандартного отклонения.

10.6.2 График строят путем изображения для каждого участника стандартного отклонения s_i с соответствующим средним x_i. В качестве альтернативы вместо стандартного отклонения можно использовать размах повторяемых результатов. Пусть

x^* — робастное среднее x_1, x_2, ..., x_p, вычисленное с применением алгоритма A;

w^* — робастное объединенное среднее s_1, s_2, ..., s_p, вычисленное с применением алгоритма S.

Предположим, что данные подчиняются нормальному распределению. В соответствии с нулевой гипотезой о том, что между участниками нет различий в совокупности ни по средним участников, ни по их стандартным отклонениям, статистика (23) приближенно может быть описана распределением χ^2 с двумя степенями свободы.

$$\left(\sqrt{m}\,\frac{x_i - x^*}{w^*}\right)^2 + \left(\sqrt{2(m-1)}\,ln\left(\frac{s_i}{w^*}\right)\right)^2. \tag{23}$$

Следовательно, критическая область с уровнем значимости приблизительно 1 % может быть нанесена на график путем построения кривых

$$s = w^* exp\left\{\pm\frac{1}{\sqrt{2(m-1)}}\sqrt{\chi^2_{2;0,99} - \left[\sqrt{m}\,\frac{x - x^*}{w^*}\right]^2}\right\} \tag{24}$$

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

$$\text{от } x = x^* - w^*\sqrt{\frac{\chi^2_{2;0,99}}{m}} \text{ до } x = x^* + w^*\sqrt{\frac{\chi^2_{2;0,99}}{m}}. \tag{25}$$

Примечание — Эта процедура основана на методе, предложенном ван Нуландом [36]. Описанный метод использует обычное нормальное приближение для распределения стандартного отклонения и может дать критическую область, содержащую отрицательные стандартные отклонения. Приведенный в настоящем стандарте метод использует приближенное распределение стандартного отклонения, которое позволяет избежать этой проблемы, но критическая область уже не является кругом, как в оригинале. Кроме того, в качестве центральной точки использованы робастные значения вместо выборочных средних, как в исходном методе.

10.6.3 График может выявлять участников, у которых смещения больше, чем смещения, обусловленные их повторяемостью. При наличии большого количества повторений данный метод также может идентифицировать участников с очень низкой повторяемостью. Однако, поскольку обычно используют небольшое количество повторений, такие интерпретации затруднены.

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

10.7 Разделенные образцы

10.7.1 Разделенные образцы используют в случае, когда необходимо детальное сравнение двух участников или когда проверка квалификации недоступна и требуется некоторая внешняя верификация. Получают образцы нескольких материалов, представляющих широкий диапазон уровней исследуемого свойства, каждый образец делят на две части, и каждая лаборатория выполняет некоторое количество (по крайней мере — два) повторных определений на каждой части образца.

Иногда может быть более двух участников, в этом случае одного участника рассматривают в качестве эталона, а остальных сравнивают с ним, используя описанные здесь методы.

Примечание 1 — Такой вид исследования является распространенным, но его часто по-разному называют, например «парная выборка» или «взаимное сравнение».

Примечание 2 — План разделенного образца не следует путать с планом «разделенного уровня», используемым в ISO 5725, который включает использование двух образцов для проверки квалификации с немного различными уровнями, поставляемых всем участникам.

10.7.2 Данные плана с разделенными образцами могут быть использованы для построения графиков, отображающих различия между результатами повторных измерений двух участников и различия между их средними арифметическими результатов по каждому образцу для проверки квалификации. Взаимные графики, использующие полный диапазон концентраций, могут иметь масштаб, который затрудняет выявление важных различий между участниками, поэтому графики различий результатов двух участников или различий результатов в процентах могут быть более полезны. Дальнейший анализ зависит от выводов, сделанных на основе этих графиков.

10.8 Графические методы для объединения показателей функционирования в нескольких турах программы проверки квалификации

10.8.1 Если нормированные показатели функционирования объединяют по нескольким турам программы проверки квалификации, провайдер проверки квалификации может рассмотреть подготовку графиков в соответствии с 10.8.2 или 10.8.3. Использование таких графиков, в которых объединены показатели функционирования по нескольким турам программы проверки квалификации, позволяет выявить тенденции и другие особенности результатов, которые неочевидны, при рассмотрении показателей функционирования по каждому туру отдельно.

Примечание — При использовании «текущих показателей» или «кумулятивных показателей», в которых показатели функционирования, полученные участником, суммируют в течение нескольких туров программы проверки квалификации, графическое представление показателей функционирования может облегчить интерпретацию. Участник может иметь недостаток, который выявляет образец для проверки квалификации, использованный в одном туре, но не использованный в других турах; текущий показатель может скрыть этот недостаток. Однако в некоторых обстоятельствах (например, при частых турах) «сглаживание» произошедших выбросов может быть полезно для более четкой демонстрации того, что лежит в основе функционирования.

10.8.2 Контрольная карта Шухарта является эффективным методом выявления проблем, которые приводят к большим ошибочным значениям показателя z. Для получения рекомендаций по построению карт Шухарта и правил построения границ действия см. ISO 7870-2 [6].

10.8.2.1 Для подготовки такой карты нормированные показатели, такие как показатели z или показатели P_A, для участника отображают в виде отдельных точек, с границами действия и предупреждения в соответствии с планом программы проверки квалификации. Если в каждом туре программы проверки квалификации измеряют несколько характеристик, показатели функционирования для различных характеристик могут быть нанесены на один график, но точки для различия характеристик должны быть нанесены с использованием различных символов и/или различных цветов. Если в один и тот же тур программы проверки квалификации включено несколько образцов для проверки квалификации, показатели функционирования могут быть нанесены на график вместе с несколькими точками в каждый период времени. На график также могут быть добавлены линии, соединяющие средние показателей в каждой точке времени.

10.8.2.2 Общепринятые правила интерпретации контрольной карты Шухарта состоят в том, что сигнал о выходе из управляемого состояния возникает тогда, когда

a) единственная точка находится за границами действия (±3,0 для показателей z или 100 % для показателей P_A);

b) две из трех последовательных точек находятся за пределами любой из границ предупреждения (±2,0 для показателей z или 70 % для показателей P_A);

c) шесть последовательных результатов положительны или отрицательны.

10.8.2.3 Если в соответствии с контрольной картой Шухарта возникает сигнал о выходе из управляемого состояния, участник должен исследовать возможные причины этого явления.

Примечание — Стандартное отклонение для оценки квалификации σ_{pt} обычно не является стандартным отклонением разностей $(x_i - x_{pt})$, поэтому уровни вероятности, соответствующие границам действия и предупреждения на контрольной карте Шухарта, применимы не всегда.

10.8.3 Если уровень свойства меняется от одного тура программы проверки квалификации к другому, графики нормированных показателей функционирования, таких как z и P_A, по отношению к приписанному значению показывают изменение смещения результатов участника в зависимости от уровня. Если в один тур программы проверки квалификации включено более одного образца для проверки квалификации, все показатели функционирования могут быть изображены независимо.

Примечание 1 — Может быть полезно использовать на графике другой символ или другой цвет для результатов текущего тура программы проверки квалификации, чтобы отличать его от предыдущих туров.

Примечание 2 — Пример такого графика приведен в Е.14 с использованием оценок показателей P_A. На этом графике также можно использовать и показатели z, изменив только масштаб по вертикальной оси.

11 Разработка и анализ программ проверки квалификации по качественным данным (включая номинальные и порядковые свойства)

11.1 Вид качественных данных

Большое количество проверок квалификации проводят для свойств, которые измеряют или идентифицируют по качественной шкале. Они включают в себя следующее:

- программы проверки квалификации, которые требуют выполнения записей (и сообщения результатов) по категориальной шкале[1)] (иногда называемой «номинальной»), где значение свойства не имеет величины (например, тип вещества или организма);
- программы проверки квалификации на наличие или отсутствие свойства — независимо от того, выполняют определение по субъективным критериями или по величине сигнала, полученного при измерении. Эту ситуацию можно рассматривать как частный случай категориальной или порядковой шкалы, имеющей только два значения (также называемой «дихотомической» или бинарной);
- программы проверки квалификации, требующие представления результатов в соответствии с порядковой шкалой, на которой результаты могут быть упорядочены, но арифметической зависимости между различными результатами не существует. Например, порядковая шкала может иметь значения «высокий», «средний», «низкий».

Такие программы проверки квалификации требуют особого внимания при разработке плана, выборе приписанного значения и оценивании функционирования (определения показателя), поскольку:

- приписанные значения очень часто основаны на мнении эксперта;
- статистическая обработка для непрерывных и счетных данных неприменима к качественным данным. Например, не имеет смысла вычисление средних и стандартных отклонений результатов, определенных по порядковой шкале, даже если они могут быть ранжированы.

Соответственно, в следующих разделах приведены рекомендации по разработке плана, определению приписанного значения и оцениванию функционирования участника для программ проверки квалификации по качественным данным.

Примечание — Рекомендации по порядковым данным не применимы к результатам измерений, определенным по количественной шкале с дискретными значениями (такими, как разведения или титры); см. 5.2.2.

11.2 Статистическое планирование эксперимента

11.2.1 Для программ проверки квалификации, в которых мнение эксперта является главным, либо для назначения приписанного значения, либо для оценки отчетов участников, обычно необходимо собрать группу экспертов соответствующей квалификации и предоставить им время для обсуждения и достижения консенсуса по назначению соответствующего приписанного значения. При необходимости, чтобы быть уверенным в отдельных экспертах при оценке показателей или присвоении приписанного значения, провайдер проверки квалификации должен дополнительно обеспечить оценку и контроль согласованности мнений различных экспертов.

Пример — В программе проверки клинической квалификации, включающей для постановки диагноза исследования с применением микроскопа, мнение экспертов используют для оценки предметных стекол микроскопа, представляемых участникам, и установления соответствующего клинического диагноза по образцам для проверки квалификации. Провайдер проверки квалификации может по своему выбору распространять образцы для проверки квалификации «вслепую» среди членов экспертной группы для обеспечения согласованности диагноза или проведения периодических оценок согласованности мнений членов экспертной группы.

---

1) *Синонимом термина «категориальная шкала» является термин «категорийная шкала».*

11.2.2 Для программ проверки квалификации, в которых используют простые, однозначные категориальные или порядковые результаты, провайдер проверки квалификации должен предусмотреть:

- представление двух или более образцов для проверки квалификации на тур;
- запрос результатов нескольких повторных наблюдений по каждому образцу для проверки квалификации с заранее указанным количеством повторений.

Любая из этих стратегий позволяет рассчитать результаты для каждого участника, которые могут быть использованы либо для анализа данных, либо для вычисления показателей. Предоставление двух или более образцов для проверки квалификации обеспечивает возможность получения дополнительной информации об особенностях погрешностей, а также более достоверного определения показателей функционирования при проверке квалификации.

Пример 1 — В программе проверки квалификации, предназначенной для определения наличия или отсутствия загрязняющего вещества, представление образцов для проверки квалификации, содержащих диапазон уровней загрязняющего вещества, позволяет провайдеру проверки квалификации проверять количество успешных выявлений загрязняющего вещества на каждом уровне в зависимости от фактического уровня загрязняющего вещества. Это может быть использовано, например, для представления участникам информации о способности обнаружения выбранного ими метода или для получения средней вероятности обнаружения, которая, в свою очередь, позволяет назначать показатели функционирования участникам на основе оцененных вероятностей конкретных моделей отклика.

Пример 2 — Для проверки квалификации при криминалистических сравнениях часто необходимо сопоставление образцов для проверки квалификации с целью определения происхождений их из одного источника или из разных источников (например, отпечатков пальцев, ДНК, гильз от пуль, следов и т. д.). Во многих случаях ответ «не определено» является допустимым. Программа проверки квалификации может включать несколько образцов для проверки квалификации из разных источников, участников просят для каждой пары указать, какие из них принадлежат «одному источнику», «другому источнику» или соответствующий источник не определен. Это позволяет объективно оценивать количество (или %) правильных или неправильных ответов, или количество (%) правильных совпадений, или правильных отклонений. Затем могут быть определены критерии для оценки функционирования, в зависимости от пригодности для использования или степени сложности задачи.

11.2.3 Однородность должна быть продемонстрирована с помощью анализа соответствующей выборки образцов для проверки квалификации, все образцы из которой должны демонстрировать среднее значение свойства. Для некоторых качественных показателей, например наличия или отсутствия свойства, может быть возможна верификация однородности с помощью количественных измерений, например, микробиологического подсчета или спектрального поглощения выше порогового значения. В этих ситуациях может быть целесообразно проведение испытаний на однородность или демонстрация того, что все результаты находятся выше или ниже порогового значения.

11.3 Приписанные значения для программ проверки квалификации по качественным данным

11.3.1 Для образцов для проверки квалификации приписанные значения могут быть установлены:

a) на основе заключений экспертов;

b) при использовании стандартных образцов в качестве образцов для проверки квалификации;

c) исходя из знаний о происхождении или подготовке образцов для проверки квалификации;

d) используя моду или медиану результатов участников (медиана подходит только для порядковых значений).

Также может быть использован любой другой метод определения приписанного значения, который (это должно быть доказано) обеспечивает достоверные результаты. Ниже рассмотрена каждая из указанных выше стратегий.

Примечание — Обычно нецелесообразно давать количественную информацию относительно неопределенности приписанного значения в программах проверки квалификации по качественным данным. Тем не менее в 11.3.2—11.3.5 описано представление базовой информации о достоверности приписанного значения, что позволяет участникам делать выводы о возможности обоснованного объяснения появления плохого результата погрешностью приписанного значения.

11.3.2 Приписанные значения, установленные на основе заключений экспертов, обычно должны быть выбраны в результате консенсуса группы экспертов соответствующей квалификации. Все существенные разногласия членов комиссии должны быть зафиксированы в отчете о туре программы проверки квалификации. Если комиссия не может достичь консенсуса по конкретному образцу для проверки квалификации, провайдер проверки квалификации может рассмотреть альтернативный метод присвоения приписанного значения из перечисленных в 11.3.1. Если ни один из них не подходит, образцы для проверки квалификации не следует использовать при оценке функционирования участников.

Примечание — В некоторых случаях возможно определение приписанного значения единственным экспертом.

11.3.3 Если в качестве образца для проверки квалификации участникам представлен стандартный образец, в качестве приписанного значения для тура программы проверки квалификации следует использовать соответствующее опорное или сертифицированное значение. Вся информация, представленная вместе со стандартным образцом, относящаяся к достоверности приписанного значения, должна быть доступна участникам после тура проверки квалификации.

Примечание — Недостатки данного подхода приведены в 7.4.1.

11.3.4 Если образцы для проверки квалификации получены из известного источника, приписанное значение может быть определено на основе данных об источнике. Провайдер проверки квалификации должен хранить записи о происхождении, транспортировке и обращении с используемым материалом (материалами). Необходимо проявлять должную осторожность, чтобы предотвратить загрязнение материалов, которое может привести к искажению результатов участников. Свидетельства происхождения и/или детали подготовки материалов должны быть доступны участникам после тура программы проверки квалификации либо по запросу, либо как часть отчета об этом туре.

Пример — Образцы вина, распространяемые для проверки подлинности в программе проверки квалификации, должны быть доставлены непосредственно от изготовителя в указанном регионе происхождения или через коммерческого поставщика, способного обеспечить гарантию подлинности.

11.3.4.1 По возможности рекомендуется проведение подтверждающих испытаний или измерений, особенно, если загрязнение может поставить под сомнение использование материала в качестве образца для проверки квалификации. Например, образец для проверки квалификации, идентифицированный как образец одного вида микроорганизмов, растений или животных, обычно должен быть проверен на реакции для других соответствующих видов. Такие проверки должны быть максимально чувствительными, чтобы гарантировать, что загрязняющие вещества либо отсутствуют, либо имеется количественная оценка уровня загрязнения.

11.3.4.2 Провайдер проверки квалификации должен представить информацию обо всех обнаруженных загрязнениях или сомнениях в отношении происхождения материала, которые могут поставить под сомнение использование образца для проверки квалификации.

Примечание — Более подробная информация о характеристиках таких образцов для программы проверки квалификации выходит за область применения настоящего стандарта.

11.3.5 Мода (наиболее распространенное наблюдение) может быть использована в качестве приписанного значения для результатов, определяемых по категориальной или порядковой шкале, в то время как медиана может быть использована в качестве приписанного значения для результатов, полученных по порядковой шкале. При использовании таких статистических данных отчет для тура программы проверки квалификации должен включать установление доли результатов, использованных при определении приписанного значения, которые соответствуют приписанному значению. Никогда не следует вычислять средние или стандартные отклонения для результатов проверки квалификации по качественным свойствам, включая порядковые значения. Это вызвано отсутствием числовой зависимости между различными значениями на такой шкале.

11.3.6 Если определение приписанного значения основано на результатах измерений (например, наличие или отсутствие свойства), приписанное значение обычно может быть определено окончательно, т. е. с низкой неопределенностью. Статистические расчеты неопределенности могут быть применены для таких уровней измеряемой величины как «неопределенный» или «сомнительный».

11.4 Оценивание функционирования и показателей для программ проверки квалификации по качественным данным

11.4.1 Оценивание функционирования участника в программе проверки квалификации по качественным данным частично зависит от особенностей требуемого отчета. В некоторых программах проверки квалификации, где от участников требуют значительного количества оцениваний, а выводы требуют тщательного рассмотрения и формулировки, отчеты участников могут быть переданы экспертам для оценки и могут получить общую оценку. С другой стороны, участников могут оценивать исключительно по точному совпадению их результата с приписанным значением для соответствующего образца для проверки квалификации. Соответственно, ниже приведены рекомендации по оценке функционирования и подсчету показателей в целом ряде ситуаций.

11.4.2 При оценке экспертами отчетов участников необходимо, чтобы один или несколько отдельных экспертов проанализировали отчет каждого участника по каждому образцу для проверки квалификации и присвоили оценку или показатель функционирования. В такой программе проверки квалификации провайдер проверки квалификации должен гарантировать, что:

- конкретный участник эксперту неизвестен. В частности, отчет, переданный эксперту(ам), не должен содержать никакой информации, которая могла бы идентифицировать участника;
- анализ, маркировка и оценка функционирования соответствуют набору ранее согласованных критериев, которые являются настолько объективными, насколько это возможно;
- соблюдены требования, установленные в 11.3.2 в отношении согласованности действий экспертов;
- там, где это возможно, участник может обжаловать заключение конкретного эксперта и/или заключение вторичного рассмотрения мнений, близких к какому-либо важному порогу функционирования.

11.4.3 Для определения показателя единственного сообщенного качественного результата, основанного на приписанном значении, могут быть использованы две системы.

a) Каждый результат отмечают как приемлемый (или дают позитивный показатель), если он точно соответствует приписанному значению, а в противном случае отмечают как неприемлемый или дают ему неблагоприятный показатель функционирования.

Пример — В программе проверки квалификации для определения наличия или отсутствия загрязняющего вещества правильным результатам присваивают 1 балл, а неправильным — 0.

b) Результаты, точно соответствующие приписанному значению, отмечают как приемлемые, им присваивают соответствующий балл; результатам, которые не совсем соответствуют приписанному значению, приписывают оценку, которая зависит от особенностей несоответствия. В таких системах оценки следует присваивать более низкие баллы лучшему функционированию, чтобы обеспечить соответствие другим типам показателей функционирования (например, показателям z, P_A, ζ и числу E_n).

Пример 1 — В программе проверки медицинской квалификации провайдер проверки квалификации присваивает оценку «0» за абсолютно правильную идентификацию микробиологического вида и оценку «1» — за неверный результат, который не изменяет лечение (например, идентификация другого, но родственного микробиологического вида, требующего аналогичного лечения), и оценку «3» — за неправильную идентификацию, которая приводит к неправильному лечению пациента. Эта система оценок, как правило, требует заключения экспертов о свойствах несоответствия, по возможности получаемого до присвоения оценки.

Пример 2 — В программе проверки квалификации, в которой возможны шесть откликов, ранжированных по порядковой шкале, результату, соответствующему приписанному значению, присваивают оценку 0, и оценку увеличивают на 2 за каждый ранг различий, пока оценка не достигнет максимального значения (таким образом, результат, ближайший к приписанному значению, получит оценку «2»).

Участникам должны быть представлены их индивидуальные показатели функционирования по каждому образцу для проверки квалификации. При выполнении повторных наблюдений может быть представлена сводная информация о показателях функционирования для каждого результата.

11.4.4 Если указаны результаты нескольких повторений для каждого образца для проверки квалификации или если каждому участнику представлено несколько образцов для проверки квалификации, провайдер проверки квалификации может рассчитать и использовать комбинированные показатели функционирования или сводки показателей при оценке функционирования. Комбинированные показатели функционирования или сводные данные могут быть рассчитаны, например, в виде:

- простой суммы показателей функционирования по всем образцам для проверки квалификации;
- количества баллов для каждого уровня функционирования;
- доли правильных результатов — баллов;
- показателя, характеризующего различия результатов и приписанных значений.

Пример — Очень общей мерой различия, которую иногда используют для качественных данных, является коэффициент Гауэра [20]. Он может сочетать количественные и качественные переменные на основе комбинации аналогичных показателей. Для категориальных или бинарных данных показателю присваивают значение 1 для точно совпадающих категорий и 0 в противном случае; для порядковых шкал показателю присваивают значение, равное 1 минус разность рангов, деленная на количество рангов, а для данных интервальной шкалы или шкалы отношений показателю присваивают значение, равное 1 минус абсолютная величина разности, деленная на наблюдаемый размах всех значений. Эти показатели принимают значения в интервале от 0 до 1, их суммируют, и сумму делят на количество используемых переменных. Также можно использовать вариант с весовыми коэффициентами.

Комбинированные показатели функционирования могут быть связаны с итоговой оценкой функционирования. Например, функционирование можно считать «приемлемым» при наличии определенной (обычно высокой) доли удовлетворительных показателей, если это соответствует целям программы проверки квалификации.

11.4.5 Для представления участникам информации об их функционировании или представления сводной информации в отчете для тура программы проверки квалификации могут быть использованы графические методы.

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

Приложение А
(обязательное)

Обозначения

d_i — разность измеряемой величины для образца, используемого в испытаниях при проверке квалификации, и приписанного значения для ССО;

$\overline{d}$ — среднее арифметическое разности измеренных величин и приписанного значения для ССО;

D — разность результата участника и приписанного значения (x—x_{pt});

D % — разность результата участника и приписанного значения, выраженная в процентах от x_{pt};

δ_E — критерий максимальной допускаемой погрешности для разностей;

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

δ_{stab} — погрешность, обусловленная нестабильностью в течение тура проверки квалификации;

δ_{trans} — погрешность, обусловленная нестабильностью в условиях транспортирования;

E_n — «нормализованная погрешность», показатель, который включает неопределенности результатов участников и приписанного значения;

g — количество образцов для проверки квалификации, используемых при проверке однородности;

m — количество повторных измерений, которые должен выполнить каждый участник на образце для проверки квалификации;

p — количество участников в туре 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 межэкземплярного стандартного отклонения;

s^* — робастная оценка стандартного отклонения участника;

$s_{\overline{x}}$ — стандартное отклонение средних арифметических результатов;

s_w — внутриэкземплярное или внутрилабораторное стандартное отклонение;

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

σ_L — стандартное отклонение между лабораториями (или участниками);

σ_{pt} — стандартное отклонение для оценки квалификации;

σ_r — стандартное отклонение повторяемости;

σ_R — стандартное отклонение воспроизводимости;

u_{diff} — стандартная неопределенность разности между независимым опорным значением или робастным средним и приписанным значением;

U_{diff} — расширенная неопределенность разности между независимым опорным значением или робастным средним и приписанным значением;

u_{hom} — стандартная неопределенность, соответствующая различиям между образцами для проверки квалификации («неоднородности»);

u_{max} — верхний предел стандартной неопределенности, используемый для скрининга неопределенности измерений, которую сообщают участники;

u_{min} — нижний предел стандартной неопределенности, используемый для скрининга неопределенности измерений, которую сообщают участники;

u_{stab} — стандартная неопределенность, связанная с нестабильностью образца в течение периода проверки квалификации;

u_{trans} — стандартная неопределенность, связанная с нестабильностью образца в условиях 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 результатов i-го участника;

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

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

w_t — область результатов измерений на различных экземплярах одного образца;

w^* — робастная оценка повторяемости участников;

x — результат измерений (общий);

x_{char} — значение характеристики, полученное при определении приписанного значения;

x_{CRM} — сертифицированное значение характеристики для СОО;

x_{diff} — разность между независимым опорным значением или робастным средним и приписанным значением;

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 — показатель, используемый для оценки квалификации;

z' — модифицированный показатель z, включающий неопределенность приписанного значения;

ζ — показатель ζ, модифицированный показатель z, включающий неопределенность результата участника и приписанного значения.

Примечание — Вычисление таких параметров, как среднее, стандартное отклонение и т. д., в настоящем стандарте следует понимать как определение выборочных оценок соответствующих параметров генеральной совокупности. Однако слова «оценка» или «оцениваемый» иногда опущены для краткости.

Приложение В
(справочное)

Проверка образцов для проверки квалификации на однородность и стабильность

В.1 Общая процедура проверки однородности

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

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

Для проверки однородности выбирают лабораторию и метод измерений. Метод измерений должен иметь достаточно малое стандартное отклонение повторяемости s_r, чтобы любая значимая 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 повторных измерений.

Подготавливают и упаковывают образцы для проверки квалификации для тура программы проверки квалификации, убедившись, что количества образцов достаточно для программы проверки квалификации и для проверки 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 gm исследуемых порций и определяют результаты измерений на каждой порции, получая серию результатов измерений в условиях повторяемости.

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

В.1.2 Если невозможно провести повторные измерения, например, при разрушающих испытаниях, то в качестве s_s можно использовать стандартное отклонение результатов. В этой ситуации важно иметь метод с достаточно низким стандартным отклонением повторяемости s_r.

В.2 Критерии проверки однородности

В.2.1 Чтобы убедиться в пригодности данных для проверки однородности следует использовать следующие три проверки:

a) проверить результаты измерений для каждой исследуемой порции в порядке выполнения измерений 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) проверить результаты средних арифметических результатов, соответствующих образцам для проверки квалификации в порядке приготовления, если существует явный тренд, который может привести к тому, что результат образца превысит критерий, установленный в В.2.2, или возникнут обстоятельства, препятствующие использованию образца, то выполняют одно из следующих действий:

1) присваивают индивидуальные значения каждому образцу для проверки квалификации;
2) отбрасывают подмножество существенно искаженных образцов для проверки квалификации, и повторно проверяют оставшиеся образцы на достаточную однородность;
3) действуют в соответствии с В.2.4, если тренд охватывает все образцы;

c) сравнить разности результатов повторных измерений (или размахи при выполнении более двух повторных измерений) и, при необходимости, использовать критерий Кохрена для проверки наличия статистически значимых различий между результатами повторных измерений (см. ISO 5725-2). Если разность результатов повторных измерений велика для всех пар, следует провести анализ, 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.

Если $m > 2$ и одно наблюдение исключено, последующие вычисления s_w и s_s должны учитывать возникший дисбаланс.

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

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

Примечание 1 — Обоснованием коэффициента 0,3 является то, что при выполнении этого критерия квадрат межэкземплярного стандартного отклонения составляет менее 10 % дисперсии оценивания функционирования участника, поэтому маловероятно, что оценивание функционирования повлияет на оценку функционирования.

Примечание 2 — Аналогично s_s можно сравнить с δ_E:

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

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

где s_w — внутриэкземплярное стандартное отклонение, вычисляемое в соответствии с B.3;

F_1 и F_2 — значения из таблицы B.1 для выбранного количества образцов для проверки квалификации с двумя исследованиями каждого образца [32].

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

g	20	19	18	17	16	15	14	13	12	11	10	9	8	7
F_1	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
F_2	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 следует заменить на $F_m = (F_{g-1,\,g(m-1),\,0,95} - 1)/m$, где $F_{g-1,g(m-1),0,95}$ — значение, которое случайная величина, подчиняющаяся F-распределению с $(g-1)$ и $g(m-1)$ степенями свободы, превышает с вероятностью 0,05.

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

$F_1 = \chi^2_{0,95(g-1)}/(g-1)$, 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 проверки квалификации недостаточно однородна.

B.2.4 Если значение σ_{pt} заранее не известно, например, когда σ_{pt} является робастным стандартным отклонением результатов участника, провайдер проверки квалификации должен выбрать другие критерии определения достаточной однородности. Такие процедуры могут включать:

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

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

c) использование данных эксперимента на прецизионность (таких как стандартное отклонение воспроизводимости по ISO 5725-2);

d) принятие риска распределения недостаточно однородных 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 межэкземплярного стандартного отклонения в стандартное отклонение для оценки квалификации путем вычисления σ'_{pt} в соответствии с формулой (B.3). Важно, чтобы эта информация была доведена до участников

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

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

c) если σ_{pt} — робастное стандартное отклонение результатов участников, то неоднородность между образцами для проверки квалификации 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 перечислениях a)—c), не может быть применено, следует отказаться от применения образца для проверки квалификации и повторить его подготовку после устранения причины неоднородности.

B.3 Формулы для проверки однородности

Оценки внутриэкземплярного стандартного отклонения s_w и межэкземплярного стандартного отклонения s_s могут быть вычислены с использованием дисперсионного анализа, как показано ниже. Метод применяют для выбранного количества образцов g с m повторными измерениями каждого.

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

k — соответствует исследуемой порции ($k = 1, 2, \ldots, m$).

Вычисляют среднее арифметическое и оценку дисперсии между исследуемыми порциями образца для проверки квалификации:

$$\bar{x}_t = \frac{1}{m}\sum_{k=1}^{m} x_{t,k}, \tag{B.4}$$

$$s_t^2 = \frac{1}{(m-1)}\sum_{k=1}^{m}\left(x_{t,k} - \bar{x}_t\right)^2. \tag{B.5}$$

Вычисляют общее среднее:

$$\bar{\bar{x}} = \frac{1}{g}\sum_{t=1}^{g}\bar{x}_t, \tag{B.6}$$

оценку дисперсии средних арифметических:

$$s_{\bar{x}}^2 = \frac{1}{(g-1)}\sum_{t=1}^{g}\left(\bar{x}_t - \bar{\bar{x}}\right)^2, \tag{B.7}$$

оценку внутриэкземплярной дисперсии

$$s_w^2 = \frac{1}{g}\sum_{t=1}^{g} s_t^2. \tag{B.8}$$

Определяют оценку объединенной дисперсии $s_{s,w}^2$ для s_s и s_w

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

Определяют оценку межэкземплярной дисперсии:

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

Примечание 2 — В случае, если $s_s^2 < 0$, следует использовать $s_s = 0$.

В общем случае, если $m = 2$, следует выполнить следующие вычисления.
Определить средние арифметические по порциям:

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

и размах между исследуемыми порциями

$$w_t = |x_{t,1} - x_{t,2}|. \tag{B.12}$$

Вычислить общее среднее:

$$\bar{\bar{x}} = \frac{1}{g}\sum_{t=1}^{g}\bar{x}_t. \tag{B.13}$$

Вычислить оценку стандартного отклонения средних арифметических $\bar{x}_t$:

$$s_{\bar{x}} = \sqrt{\sum_{t=1}^{g}\left(\bar{x}_t - \bar{\bar{x}}\right)^2 / (g-1)} \tag{B.14}$$

и внутриэкземплярное стандартное отклонение по всем образцам:

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

где суммирования в формулах (B.13), (B.14) и (B.15) выполняют по образцам ($t = 1, 2, ..., g$).

Вычислить оценку межэкземплярного стандартного отклонения:

$$s_s = \sqrt{\max\left(0,\ s_{\bar{x}}^2 - s_w^2 / 2\right)}. \tag{B.16}$$

Примечание 1 — Если s_s меньше s_w, то оценка межэкземплярной дисперсии s_s^2 будет отрицательной. Это возможно при высокой однородности. В этом случае принимают $s_s = 0$.

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

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

Примечание 3 — Пример приведен в Е.2.

B.4 Процедуры проверки стабильности

B.4.1 Общие рекомендации по проверке стабильности

B.4.1.1 Ниже приведены общие рекомендации по проверке выполнения требований к стабильности, установленных в 6.1. Положения, установленные в 6.1.3 в отношении исследуемых свойств, применимы к любой экспериментальной проверке на стабильность в течение тура проверки квалификации и стабильности в процессе транспортирования.

B.4.1.2 Если предыдущие экспериментальные исследования, опыт и знания обеспечивают уверенность в том, что нестабильность маловероятна, экспериментальную проверку стабильности можно проводить только в случае появления значительных изменений в процессе тура проверки квалификации и после тура. Если такой уверенности нет, исследования влияния транспортирования на стабильность и стабильности в течение обычного тура проверки квалификации могут иметь форму плановых исследований, проводимых до распространения образцов для проверки квалификации, либо для каждого тура, 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.

B.4.1.3 При проверке стабильности выполняют следующие действия:

- все свойства и характеристики, используемые в программе проверки квалификации, должны быть проверены на стабильность (или их стабильность должна быть подтверждена другим способом). Это может быть сделано на основе предыдущего опыта и технических заключений, основанных на знании 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 повторных измерений 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.

B.4.2 Процедура проверки стабильности при проведении тура проверки квалификации

B.4.2.1 Удобным способом проверки стабильности при проведении тура проверки квалификации являются испытания небольшого количества образцов для проверки квалификации после окончания тура и сравнение их с образцами до проведения тура, чтобы гарантировать, что никаких изменений образцов за время проведения тура 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 тура проверки квалификации, с образцами, сохраняемыми в контролируемых условиях.

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

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

B.4.2.2 Если провайдер включает в процедуру оценки стабильности поставленные образцы в соответствии с B.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 B.6.

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

a) случайным образом для тура отбирают $2g$ образцов для проверки квалификации, где $g \geq 2$;

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

c) выполняют измерения на g образцах для проверки квалификации до плановой даты представления образцов участникам. Повторные измерения должны быть выполнены случайным образом;

d) оставшиеся g образцов для проверки квалификации сохраняют в условиях аналогичных предполагаемым условиям хранения у участников;

e) в максимально сжатые сроки после окончательной даты поступления результатов участников выполняют измерения на оставшихся g образцах в той же лаборатории, используя тот же метод измерений и то же количество повторных измерений, все повторные измерения проводят случайным образом;

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

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

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

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

c) вторая 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.

Примечание — Также могут быть применены процедуры, использующие наблюдения через регулярные промежутки времени от начала до окончания тура проверки квалификации, они могут быть полезны, если изменчивость измерительной системы с течением времени достаточно велика и может поставить под сомнение оценку, описанную в В.5.

В.5 Критерий оценки для проверки стабильности

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

$$|\bar{y}_1 - \bar{y}_2| \le 0{,}3\sigma_{pt} \text{ или } |\bar{y}_1 - \bar{y}_2| \le 0{,}1\delta_E. \quad (В.17)$$

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

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

b) увеличить неопределенность приписанного значения для учета возможной нестабильности;

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

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

Примечание 1 — Коэффициент 2 в формуле (В.18) является коэффициентом охвата расширенной неопределенности разности, обеспечивающим приближенно уровень доверительной вероятности 95 %, кроме того, при расчете суммарной неопределенности намеренно предполагалось, что $\bar{y}_1$ и $\bar{y}_2$ независимы.

Примечание 2 — Следует понимать, что неопределенности $u^2(\bar{y}_1)$ и $u^2(\bar{y}_2)$ в формуле (В.18) включают изменчивость измерительной системы за прошедшее время, а также ее повторяемость.

В.5.3 Если критерий, приведенный в формулах (В.17) или (В.18), не выполнен, следует рассмотреть следующие варианты:

- количественно оценить влияние нестабильности и учесть его при оценивании (например, с помощью показателя z');
- проверить процедуры подготовки и хранения образцов для выявления возможности их улучшений;
- не определять оценки функционирования участников.

В.5.4 Критерий, приведенный в В.5.1 или В.5.2, может быть заменен подходящим статистическим критерием, позволяющим установить различие двух наборов данных при условии, что этот критерий учитывает повторные измерения и обеспечивает идентификацию стабильности, по крайней мере не менее, чем в формуле (В.18).

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

В.6 Стабильность в условиях транспортирования

В.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) сравнить образцы, сохраняемые в помещении провайдера, с образцами, поставляемыми и возвращаемыми участниками. Также могут быть проведены исследования, основанные на наблюдении за образцами, например, под воздействием разумно предсказуемых условий транспортирования.

В.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 в условиях транспортирования включает сравнение результатов для двух групп образцов, то 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 для оценки квалификации определяют по результатам участников (например, с помощью робастных методов), то среднее и стандартное отклонение для оценки квалификации отражают любое смещение и увеличение изменчивости (соответственно), вызванные условиями транспортирования образцов.

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

Приложение С
(справочное)

Робастный анализ

С.1 Общие положения

Межлабораторное сличение представляет собой особый анализ данных. В то время как большинство межлабораторных сличений представляют данные, подчиняющиеся унимодальному и приблизительно симметричному распределению, большая часть наборов данных оценки квалификации включает некоторую 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 и стандартного отклонения, 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 методов, используемых в задачах проверки квалификации и имеющих различные возможности в отношении определения устойчивости оценок при наличии данных из загрязненных совокупностей (например, по эффективности и пороговой точке) и различный уровень сложности при использовании. Методы представлены в порядке возрастания сложности (первый — самый простой, последний — самый сложный), и в порядке неубывания эффективности, поскольку более сложные оценки разработаны, как правило, с целью повышения их эффективности.

Примечание 1 — В приложении D приведена дополнительная информация об эффективности, пороговых точках и чувствительности к небольшим модам — трех важных показателях выполнения различных робастных методов определения оценок.

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 они означают оценки среднего и стандартного отклонения, полученные в соответствии с робастным алгоритмом.

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

С.2.1 Медиана

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

a) расположить p данных в порядке неубывания:

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

b) 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 Масштабированное абсолютное отклонение от медианы MAD_e

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

a) абсолютные значения разностей d_i (i = 1,…, p)

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

b) значение $MAD_e(x)$

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

Если у половины или большего количества участников результаты совпадают, то $MAD_e(x) = 0$, 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 $MAD_e(x)$. 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 (С.4):

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

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 $MAD_e(x)$], 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 MAD_e, но *nIQR* имеет пороговую точку в 25 % (см. приложение D), в то время как у MAD_e пороговой точкой является 50 %. Поэтому MAD_e устойчива при значительно более высокой доле содержания выбросов, чем *nIQR*.

Примечание 2 — При $p < 30$ обе оценки обладают заметным отрицательным смещением, которое неблагоприятно влияет 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 простых робастных оценок приведен в Е.3.

С.3 Робастный анализ: алгоритм *A*

С.3.1 Алгоритм *A* с итеративной процедурой

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

Обозначим p данных, расположенных в порядке неубывания:

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

Обозначим робастное среднее и робастное стандартное отклонение этих данных x^* и s^*.

Для выполнения алгоритма *A* вычисляют начальные значения x^* и s^*:

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

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

Примечание 1 — Алгоритмы *A* и *S*, приведенные в данном приложении, соответствуют ISO 5725-5 с добавлением критерия остановки: при совпадении оценок среднего и стандартного отклонения до трех значащих цифр вычисления прекращают.

Примечание 2 — В некоторых случаях более половины результатов x_i идентичны (например, количество нитей в образцах ткани или количество электролитов в образцах сыворотки крови). В этом случае начальное значение $s^* = 0$ и робастная процедура становится некорректной. Если начальное значение $s^* = 0$, допустимо его заменить выборочным стандартным отклонением 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.

Затем вычисляют новые значения x^* и s^*. Для этого вычисляют:

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

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^*:

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

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

где суммирование выполняют по i.

Робастные оценки x^* и s^* получают на основе итеративных, т. е. повторяющихся несколько раз, вычислений x^* и s^* в соответствии с (С.7)—(С.10) до тех пор, пока процесс не начнет сходиться, т. е. предыдущие и последующие значения x^* и s^* станут совпадать до третьей значащей цифры. Альтернативные критерии сходимости могут быть определены в соответствии с требованиями к плану эксперимента и к отчету о результатах проверки квалификации.

Примечание 3 — Примеры использования алгоритма *A* приведены в Е.1 и Е.3.

С.3.2 Варианты алгоритма *A*

Итеративный алгоритм *A*, приведенный в С.3.1, имеющий умеренную пороговую точку (примерно 25 % для больших наборов данных [25]) и начальную точку для s^*, предложенную в С.3.1, для наборов данных, где $MAD_e(x) = 0$, может серьезно ухудшить устойчивость к выбросам 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:

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

b) если в вычислениях робастное стандартное отклонение не используют, следует применять MAD_e [исправленное в соответствии с a)], и не изменять s^* в процессе итерации. Если в вычислениях используют робастное стандартное отклонение, s^* заменяют в соответствии с оценкой *Q*, описанной в С.5, и не изменяют s^* в процессе итерации.

Примечание — Вариант, приведенный в перечислении b), улучшает пороговую точку алгоритма *A* до 50 % [25], что позволяет применять алгоритм при наличии высокой доли выбросов.

С.4 Робастный анализ: алгоритм *S*

Данный алгоритм применим к стандартным отклонениям (или размахам), которые вычисляют, если участники представляют результаты с *m* повторными измерениями измеряемой величины образца для проверки квалификации 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

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

Обозначим робастное объединенное значение w^*, а ν — число степеней свободы, соответствующих каждому w_i. (Если w_i — размах, $\nu = 1$. Если w_i — стандартное отклонение *m* результатов испытаний, $\nu = m - 1$.) Определяют значения ξ и η по таблице С.1.

Вычисляют начальное значение w^*:

$$w^* = med(w_i),\ (i = 1, 2, ..., p). \qquad (С.11)$$

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 среднее арифметическое, его заменяют стандартным отклонением объединенного среднего арифметического (или размахом средних арифметических). Эту замену выполняют только для начального значения w^*, после чего процедуру продолжают согласно описанию.

Значение w^* вычисляют следующим образом:

$$\psi = \eta \cdot w^*. \qquad (С.12)$$

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

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

Вычисляют новое значение w^*:

$$w^* = \xi \sqrt{\sum_{i=1}^{p} \left(w_i^*\right)^2 / p}. \qquad (С.14)$$

Робастную оценку w^* получают итеративным методом, вычисляя значение w^* несколько раз, до тех пор, пока процесс не начнет сходиться. Сходимость считают достигнутой, если значения w^* в последовательных итерациях совпадают в трех значащих цифрах.

Примечание — Алгоритм *S* обеспечивает оценку стандартного отклонения совокупности при использовании стандартных отклонений из единственного нормального распределения и, следовательно, обеспечивает оценку стандартного отклонения повторяемости при применении предположений ISO 5725-2.

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

Число степеней свободы ν	Ограничивающий коэффициент η	Корректирующий 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 выведены в ISO 5725-5:1998, приложение B.		

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 направлениях исследований или при калибровке и в тех областях, где проверка квалификации ранее была недоступна. Однако эти методы являются недостоверными в тех случаях, когда количество выбросов в результатах превышает 20 %, или в случае бимодального (или мультимодального) распределения данных, а некоторые из них могут стать неприемлемо изменчивыми для небольшого количества участников. Кроме того, ни один из этих методов не может работать с данными повторных измерений участников. В соответствии с ISO/IEC 17043 необходимо, чтобы эти ситуации были предусмотрены планом эксперимента или выявлены в процессе анализа до оценивания функционирования участников, 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 для определения показателей используют робастное среднее и/или робастное стандартное отклонение, то существует значимый риск неверной классификации участников при применении неэффективных статистических методов.

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

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

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

С.5.2.1 Оценка Q_n [34] является высокоэффективной оценкой с высокой пороговой точкой для стандартного отклонения генеральной совокупности, которая является несмещенной для данных нормального распределения (при условии отсутствия выбросов). Q_n-метод учитывает единственный результат для каждого участника (включающий среднее или медиану повторных измерений). Расчет основан на использовании попарных разностей 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.

Чтобы вычислить Q_n для набора данных (x_1, x_2, … x_p) с p результатами:

a) вычисляют $p(p-1)/2$ абсолютных разностей

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

b) для упорядоченных разностей d_{ij} используют обозначения

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

c) вычисляют

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

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

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

d) 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 корректирующим, обеспечивающим несмещенность оценки стандартного отклонения для больших p. Корректирующие коэффициенты b_p для небольших значений p определяют по таблице С.2, а при $p > 12$ эти коэффициенты устанавливают в соответствии с [41], 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 более крупных наборов данных опубликованы программы (см. [42]) (на момент публикации приведена обработка данных с объемом выше 8000 за приемлемое время).

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

p	2	3	4	5	6	7	8	9	10	11	12
b_p	0,3994	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 с высокой пороговой точкой для стандартного отклонения результатов проверки квалификации, представленных различными лабораториями. Q-метод является не только устойчивым к выбросам, но и в ситуации, когда большая часть результатов испытаний равны между собой, например, когда результаты представляют собой дискретные числа или округленные данные. В такой ситуации другие аналогичные методы не следует применять, поскольку многие разности равны нулю.

Q-метод можно использовать для проверки квалификации как в случае представления участником единственного результата (включая среднее или медиану повторных измерений), так и при предоставлении результатов повторных измерений. Прямое использование повторных измерений в вычислениях повышает эффективность метода.

Расчет основан на использовании попарных разностей из набора данных, таким образом оценка не зависит от оценки среднего или медианы данных. Метод называют методом Q/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

$$\underbrace{y_{11},\ldots,y_{1n_1}}_{\text{лаборатория 1}}, \underbrace{y_{21},\ldots,y_{2n_2}}_{\text{лаборатория 2}}, \ldots, \underbrace{y_{p1},\ldots,y_{pn_p}}_{\text{лаборатория } p}. \tag{C.22}$$

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

$$H_1(x) = \frac{2}{p(p-1)} \sum_{1 \le i < j \le p} \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{m=1}^{n_j} \mathbf{I}\left\{\left|y_{ik} - y_{jm}\right| \le x\right\}, \tag{C.23}$$

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

Обозначим точки разрыва функции $H_1(x)$:

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

Можно вычислить для всех положительных точек разрыва x_1, …, x_r:

$$G_1(x_i) = \begin{cases} 0{,}5 \cdot \left(H_1(x_i) + H_1(x_{i-1})\right), \text{ если } i \geq 2, \\ 0{,}5 \cdot H_1(x_i), \text{ если } i = 1;\ x_1 > 0. \end{cases} \quad \text{(C.24)}$$

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 $0 \leq x_1 < x_2 < \ldots < x_r$.

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

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

где $H_1(0)$ вычисляют аналогично формуле (C.23), $H_1(0) = 0$ в случае отсутствия точного совпадения данных, $\Phi^{-1}(q)$ — 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, либо в соответствии с установленным опорным значением.

Примечание 2 — Другие варианты *Q*-метода, позволяющие получить робастную оценку стандартных отклонений воспроизводимости и повторяемости, приведены в [25], [34].

Примечание 3 — Теоретические основы *Q*-метода, включая его асимптотическую эффективность и разбивку на конечное число выборок, описаны в [26] и [34].

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 формуле (C.21).

Примечание 5 — Стандартное отклонение воспроизводимости не обязательно является наиболее подходящим стандартным отклонением для использования при проверке квалификации, так как это, как правило, оценка рассеяния единственных результатов, а не оценка рассеяния средних или медиан результатов повторных измерений каждого участника. Однако рассеяние средних или медиан результатов повторных измерений лишь немного меньше рассеяния единственных результатов различных лабораторий, если отношение стандартного отклонения воспроизводимости к стандартному отклонению повторяемости более 2. Если это отношение менее двух, для определения показателей 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 повторных измерений;

s_r^2 — дисперсия повторяемости, вычисленная в соответствии с [35], или при использовании среднего значения повторных измерений вместо результатов повторных измерений участника для *Q*-метода.

Примечание 6 — Примечание 5 применяют только в том случае, если показатели определяют на основе средних или медиан результатов повторных измерений. Если повторные измерения образцов для проверки квалификации проводят вслепую, показатели следует рассчитывать для каждого повторения. В этом случае стандартное отклонение воспроизводимости является наиболее подходящим стандартным отклонением.

Примечание 7 — Пример применения *Q*-метода приведен в E.3.

C.5.3 Определение робастного среднего, используемого в оценке Хампеля

C.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 определения этой оценки. Первый из них является более простым, но может привести к отклонениям результатов при его выполнении. Второй алгоритм обеспечивает получение однозначных результатов, зависящих только от базового стандартного отклонения.

C.5.3.2 Далее приведены вычисления, обеспечивающие определение итеративной взвешенной оценки Хампеля для параметра положения.

a) Обозначим данные x_1, x_2 … x_p.

b) Пусть x^* — медиана $med(x)$ (см. C.2.1).

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

d) Для каждой точки данных x_i вычисляют q_i по формуле (C.26):

$$q_i = \left| \frac{x_i - x^*}{s^*} \right|. \quad \text{(C.26)}$$

e) Вычисляют весовые коэффициенты w_i по формуле (С.27):

$$w_i = \begin{cases} 0, \text{ если } |q| > 4,5, \\ (4,5 - q)/q, \text{ если } 3 < |q| \le 4,5, \\ 1,5/q, \text{ если } 1,5 < |q| \le 3,0, \\ 1, \text{ если } |q| \le 1,5. \end{cases} \quad (С.27)$$

f) Пересчитывают x^* по формуле (С.28)

$$x^* = \frac{\sum_{i=1}^{p} w_i x_i}{\sum_{i=1}^{p} w_i}. \quad (С.28)$$

g) Повторяют действия d)—f) до тех пор, пока значения x^* не начнут сходиться. Сходимость считают достаточной, если разность x^* в двух последних итерациях станет менее $0,01 s^* / \sqrt{p}$, что соответствует приблизительно 1 % стандартной погрешности x^*. Могут быть использованы и другие более точные критерии сходимости.

Данный алгоритм определения 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 данного алгоритма приведены в Е.3.

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

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

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

С.5.3.3 Ниже приведен алгоритм конечных шагов, позволяющий получить оценку Хампеля без использования итеративных весовых коэффициентов [25].

Вычисляют средние арифметические для каждой лаборатории $y_1, y_2, \ldots, y_p$.

Вычисляют робастное среднее x^* как корень уравнения (С.29):

$$\sum_{i=1}^{p} \Psi\left(\frac{y_i - x^*}{s^*}\right) = 0, \quad (С.29)$$

где

$$\Psi(q) = \begin{cases} 0, \text{ если } q \le -4,5, \\ -4,5 - q, \text{ если } -4,5 < q \le -3, \\ -1,5, \text{ если } -3 < q \le -1,5, \\ q, \text{ если } -1,5 < q \le 1,5, \\ 1,5, \text{ если } 1,5 < q \le 3, \\ 4,5 - q, \text{ если } 3 < q \le 4,5, \\ 0, \text{ если } q > 4,5, \end{cases} \quad (С.30)$$

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

Точное решение может быть получено за конечное число шагов, без итерации, используя свойство, при котором ψ функция x^* является частично линейной, имея в виду точки интерполяции в левой стороне уравнения (С.29), представленного здесь как функция x^*.

Вычисляют все точки интерполяции:
- для 1-го значения y_1:

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

- для 2-го значения y_2:

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

- 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 значения $m = 1,\ldots,(6 \cdot p - 1)$ вычисляют

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

и проверяют выполнение следующих условий:
a) если $p_m = 0$, то $d_{\{m\}}$ — решение уравнения (C.29);
b) если $p_{m+1} = 0$, то $d_{\{m+1\}}$ — решение уравнения (C.29);
c) 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 (C.29).

Пусть S — множество всех решений уравнения (C.29).
Решением $x^* \in S$ является ближайшая медиана, используемая в качестве параметра положения x^*, т. е.

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

Может существовать несколько решений. Если существует 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, т. е. их рассматривают как выбросы.

C.5.4 Метод *Q*/Хампеля

Метод *Q*/Хампеля использует *Q*-метод, описанный в C.5.3.2, для вычисления робастного стандартного отклонения s^* и алгоритм конечных шагов для оценки Хампеля, описанный в C.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 C.5.3.2. Для вычисления робастного стандартного отклонения повторяемости s_r применяют второй алгоритм, использующий попарные разности в пределах лаборатории.

Примечание — Веб-приложения для метода *Q*/Хампеля приведены в [37].

C.6 Другие робастные методы

Методы, описанные в данном приложении, не представляют собой полную совокупность всех обоснованных подходов. Ни один из них не является гарантированно оптимальным во всех ситуациях. По усмотрению провайдера могут быть использованы другие робастные методы при условии подтверждения их эффективности, пороговых точек и всех остальных свойств, соответствующих требованиям программы проверки квалификации.

Приложение D
(справочное)

Дополнительные рекомендации по статистическим процедурам

D.1 Процедуры в случае небольшого количества участников

D.1.1 Общие положения

В программах проверки квалификации обычно принимает участие небольшое количество участников или, даже при наличии большого общего количества участников, выполняют сопоставление групп с небольшим количеством участников. Часто это происходит, когда участников группируют и подсчитывают показатели с помощью единого метода, как, например, это бывает в медицинских лабораториях.

В том случае, если количество участников не велико, приписанное значение в идеале должно быть определено с использованием валидированной метрологической процедуры независимо от участников, например по процедуре приготовления образцов или по данным референтной лаборатории. Критерий для оценивания функционирования также должен быть основан на внешних критериях, таких как заключение экспертов или соответствие целевому значению. В этой идеальной ситуации качество функционирования оценивают с использованием предварительно определенного приписанного значения и критерия функционирования, поэтому проверка квалификации может быть проведена при наличии лишь одного участника. Такой тип межлабораторного сличения 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, может возникнуть необходимость определения приписанного значения или дисперсии или и того и другого по результатам участников. Если количество участников слишком мало для выполнения конкретной процедуры, то оценивание функционирования может стать недостоверным, поэтому для оценивания функционирования важно рассмотреть вопрос об установлении минимального количества участников.

Далее приведены рекомендации для ситуаций с небольшим количеством участников, когда критерий для оценивания функционирования определяют, используя результаты участников.

D.1.2 Процедуры идентификации выбросов

Несмотря на то, что для загрязненных выбросами генеральных совокупностей настоятельно рекомендуется использование робастных статистик, для очень небольших наборов данных они все же не подходят (исключения приведены ниже). Проверка на наличие выбросов для очень небольших наборов данных, однако, возможна. В случае очень маленьких программ проверки квалификации или групп предпочтительным является исключение выброса с последующим вычислением среднего или стандартного отклонения.

Различные тесты на наличие выбросов применимы к различным размерам наборов данных. В ISO 5725-2 приведены таблицы теста Граббса для одиночного выброса и для двух одновременных выбросов в одном направлении. Тест Граббса и другие тесты требуют, чтобы количество возможных выбросов было указано заранее, и могут завершиться неудачей при наличии большего количества выбросов, эти тесты наиболее полезны при $p > 10$ (в зависимости от вероятной доли выбросов).

Примечание 1 — После исключения выбросов следует соблюдать осторожность при оценке дисперсии, так как оценка может быть смещена в меньшую строну. Смещение обычно не очень большое, если исключают выбросы с уровнем доверия 99 % и выше.

Примечание 2 — Большинство одномерных робастных оценок положения и дисперсии приемлемы при $p \geq 12$.

D.1.3 Процедуры оценки параметра положения

D.1.3.1 Приписанные значения, полученные по небольшим наборам данных участников, должны по возможности удовлетворять критерию неопределенности приписанного значения, приведенному в 9.2.1. Для ситуации с использованием в качестве приписанного значения среднего и в качестве стандартного отклонения для оценки квалификации стандартного отклонения результатов этот критерий не может быть применен для нормального распределения с $p \leq 12$ после устранения 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 А (С.3), имеют промежуточную эффективность и могут соответствовать критерию при $p > 12$, если учтены положения примечания 2 к 7.7.7.

D.1.3.2 Существуют ограничения на объем набора данных, применяемых для определения некоторых оценок параметра положения. Рекомендуется несколько численных робастных оценок среднего для набора данных небольшого объема. Нижний предел, как правило, составляет $p \geq 15$, хотя провайдеры могут демонстрировать приемлемое функционирование с учетом установленных предположений для меньших наборов данных. Медиана применима для меньших объемов данных вплоть до $p = 2$ (если она равна среднему), но при $3 \leq p \leq 5$ 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 для оценки функционирования, основанные на рассеянии результатов участников, для набора данных небольшого объема из-за очень высокой изменчивости всех оценок рассеяния. Например, при p = 30 оценки стандартного отклонения для данных из нормального распределения в среднем отклоняются от истинного значения на 25 % (с уровнем доверия 95 %). Для данных из нормального распределения не существует лучших оценок.

D.1.4.2 Если оценки рассеяния необходимы для других целей (например, как суммарные статистики или оценка рассеяния данных для робастной оценки параметра положения) или если программа проверки квалификации устойчива к высокой изменчивости оценок рассеяния данных, для небольших наборов данных следует выбирать оценки рассеяния 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 MAD_e или $nIQR$ из С.1.

Примечание 3 — При очень небольших наборах данных для робастных оценок рассеяния данных необходимо использовать следующие рекомендации [24]:

- если p = 2, необходимо использовать $|x_1 - x_2| / \sqrt{2}$;
- если p = 3, параметры положения и масштаба неизвестны, необходимо использовать MAD_e для защиты от чрезмерно высоких оценок стандартного отклонения или среднее абсолютное значение отклонения (см. примечание 4) для защиты от слишком маленьких оценок стандартного отклонения, например, если из-за ошибок округления могут быть получены два одинаковых значения;
- если $p \geq 4$, необходимо использовать установленную M-оценку стандартного отклонения, полученную на основе логарифмически взвешенной функции, рекомендуемой в [27], а также близкий эквивалент алгоритма A без итерации при определении параметра положения с использованием медианы в качестве оценки параметра положения.

Примечание 4 — Для определения оценки стандартного отклонения на основе абсолютного значения разности x_i и медианы используют следующую формулу:

$$s^* = \frac{1}{0{,}798 \cdot p} \sum_{i=1}^{p} |x_i - med(x)|. \quad \text{(D.1)}$$

Примечание 5 — Коэффициент 0,798 в формуле (D.1) определяет абсолютное значение расстояния среднего от нуля в стандартном нормальном распределении.

D.2 Эффективность и пороговые точки робастных процедур

D.2.1 Различные статистические методы оценки (например, робастные методы) можно сопоставлять по трем показателям:

- пороговая точка — доля значений в наборе данных, которые можно заменить сколь угодно большими значениями, при этом оценка не станет сколь угодно большой;
- эффективность оценки — отношение минимальной оценки дисперсии к оценке дисперсии 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 в общей совокупности компетентных 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 для различных ситуаций различны, их тщательный анализ выходит за рамки настоящего стандарта. Однако в предположении о нормальном распределении данных могут быть сделаны простые сравнения 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 процедур, приведенных 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 отклонения (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
MAD_e	Стандартное отклонение	50	Умеренно хорошая
Алгоритм *A*	Среднее и стандартное отклонение	25	Умеренная
Оценки Q_n и *Q*/Хампеля	Среднее и стандартное отклонение	50	Умеренная (очень хорошая для устойчивости к противоречивым данным в точках, отстоящих более чем на 6 s^*)

Примечание — Используемое здесь определение пороговой точки сводится к определению доли большого набора данных из нормального распределения, значения данных в которой могут изменяться до бесконечности, без одновременного увеличения оценки до бесконечности. Например, если менее 50 % данных набора заменить на бесконечность, медиана останется конечной величиной.

Таким образом, внутриэкземплярные среднее и стандартное отклонения могут дать недостоверную оценку при наличии единственного выброса. В то же время робастные методы, использующие медиану, MAD_e и оценку *Q*/Хампеля, могут выдержать очень большую долю выбросов. Алгоритм *A* с итеративной оценкой стандартного отклонения 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 рассеяния, и это следует учитывать при интерпретации таких оценок.

D.2.3 Относительная эффективность

Все оценки имеют выборочную дисперсию, т. е. оценки могут отличаться от тура к туру 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.

Внутриэкземплярные среднее и стандартное отклонения являются оценками среднего и стандартного отклонения с минимальной дисперсией, и поэтому они имеют эффективность 100 %. Оценки с более низкой эффективностью, имеют большую изменчивость, т. е. они могут больше изменяться от тура к туру, 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 отклонения генеральной совокупности для нормально распределенного набора данных с $n = 50$ или $n = 500$ участников

Статистическая оценка	Среднее $n = 50$	Среднее $n = 500$	SD $n = 50$	SD $n = 500$
Выборочное среднее и стандартное отклонение	100 %	100 %	100 %	100 %
Медиана и *nIQR*	66 %	65 %	38 %	37 %
Медиана и MAD_e	66 %	65 %	37 %	37 %
Алгоритм *A*	97 %	97 %	74 %	73 %
Q_n и *Q*/Хампеля	96 %	96 %	73 %	81 %

Эти результаты показывают, что не существует статистического метода, идеального для всех ситуаций. Выборочные среднее и стандартное отклонения являются оптимальными оценками в случае нормального распределения данных, но не подходят при наличии выбросов. Простые робастные методы, такие как использование медианы, MAD_e 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 к ISO/IEC 17043 установлено, что оценивание характеристик метода измерений, как правило, не является целью программы проверки квалификации. Тем не менее результаты программы проверки квалификации можно использовать для проверки и, возможно, установления повторяемости и воспроизводимости метода измерений [15], если программа проверки квалификации удовлетворяет следующим условиям:

a) образцы для проверки квалификации однородны и стабильны;

b) участники способны к постоянному удовлетворительному функционированию;

c) компетентность участников (или подгруппы участников) продемонстрирована до начала тура программы проверки квалификации, и результаты тура не ставят под сомнение их компетентность.

D.3.2 Для получения достаточных данных для оценивания повторяемости и воспроизводимости метода испытаний на основе программы проверки квалификации должны быть выполнены следующие условия:

a) при проведении исследований имеется достаточное количество участников, продемонстрировавших свою компетентность в части выполнения метода измерений на предыдущих турах программы проверки квалификации, которые приняли обязательства следовать методу измерений без изменений;

b) при оценке повторяемости в каждом туре программы проверки квалификации должно быть использовано не менее двух образцов для проверки квалификации или выполнено необходимое количество повторных наблюдений;

c) по возможности участники должны быть обеспечены отдельно идентифицированными повторными измерениями «вслепую», что является более предпочтительным, чем выполнение повторных измерений на одном и том же образце для проверки квалификации;

d) образцы для проверки квалификации, используемые в одном или нескольких турах 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 ISO 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 значений (см. 5.5.3.3)

В таблице Е.1 приведены 23 результата тура программы проверки квалификации, из которых 5 результатов представлены в виде «менее некоторого значения». 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 оценивание будет неприемлемым (сигнал действия) для всех 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 (со знаком «<») результатами и три способа работы с такими результатами

Участник	Результат	Данные, в которых знак «<» проигнорирован	Данные со знаком «<» исключены	Данные со знаками «<» и «>» заменены 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
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

Окончание таблицы Е.1

Участник	Результат	Данные, в которых знак «<» проигнорирован	Данные со знаком «<» исключены	Данные со знаками «<» и «>» заменены на половинное значение
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 отклонение, а также на оценивание функционирования участника. Провайдер проверки квалификации должен выбрать подходящий метод.

Примечание — В [21] приведены некоторые методы, основанные на оценке максимального правдоподобия, которые могут должным образом учитывать результаты, указанные как «меньше верхнего предела».

Е.2 Проверка однородности и стабильности. 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, полученных из 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. Целевое значение σ_{pt} для As в шоколаде составляет 15 %. Поскольку приписанное значение для тура проверки квалификации невозможно определить до анализа однородности, оценку выборочной изменчивости сравнивают с предварительной оценкой σ_{pt}, рассчитанной как 15 % от среднего значения, полученного при проверке однородности.

Таблица Е.2 — Данные для проверки однородности образцов при контроле содержания мышьяка в шоколаде

Номер бутылки	Повторное измерение 1	Повторное измерение 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: $0{,}3\sigma_{pt}$ = 0,00842.

Вывод: 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 тура программы проверки квалификации (6 нед.). Образцы исследованы дважды (см. таблицу Е.3), и четыре результата сравнили с результатами при проверке однородности.

Таблица Е.3 — Данные для проверки стабильности образцов на содержание мышьяка в шоколаде

Выбранный образец	Повторное измерение 1	Повторное измерение 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: $0{,}3\sigma_{pt}$ = 0,00842.
Вывод: 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 отклонения, рассчитанных в соответствии с алгоритмом *A* в процессе шести итераций, т. е. до тех пор, пока в робастных среднем и стандартном отклонении перестают изменяться три значащих цифры. На рисунках Е.1, Е.2 и Е.3 представлены соответственно график упорядоченных данных, гистограмма и график ядерной плотности.

Примечание — Как гистограмма, так и график плотности показывают очевидные второстепенные моды в обеих крайних точках. Это связано скорее с небольшим количеством выбросов, чем с особенностью базового распределения достоверных результатов.

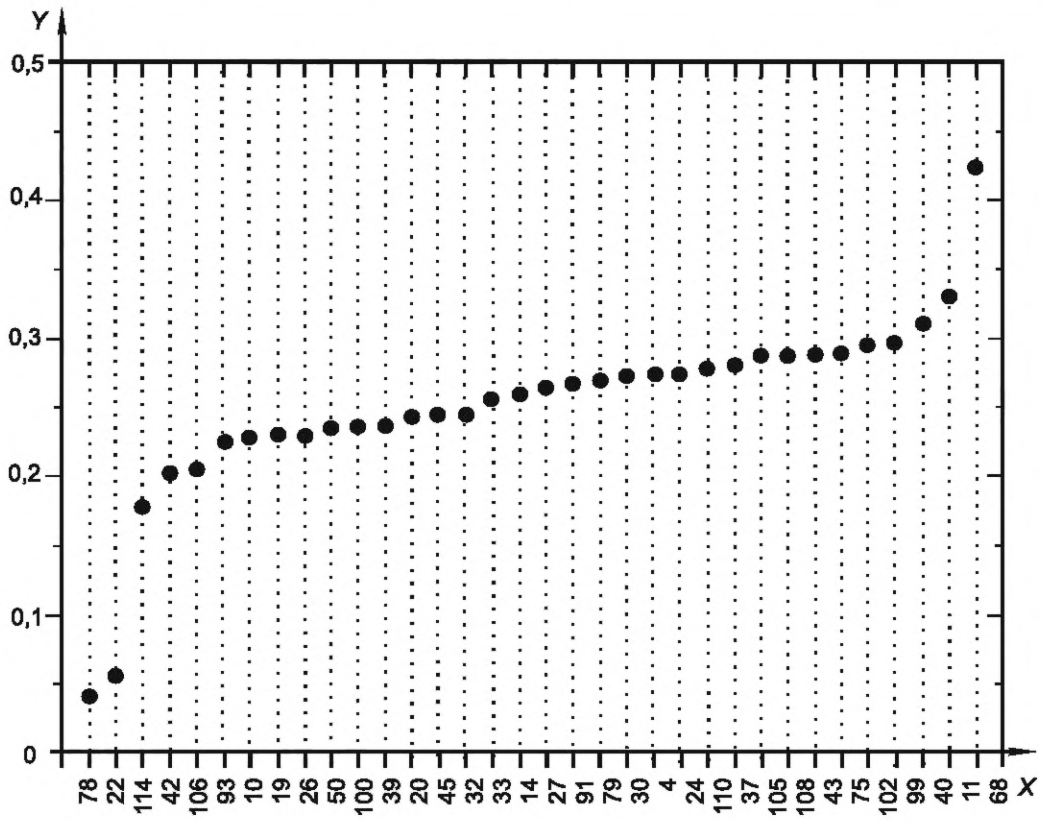
В таблице Е.5 приведены оценки параметра положения (среднего) и стандартного отклонения, полученные с использованием классических и робастных методов. Неопределенность оценки параметра положения также приведена. Статистики для бутстреп-метода получены в соответствии с процедурами, описанными в [17], [18], а также с применением пакета программного обеспечения *R* (см. пример Е.6 и приложение F). На рисунке Е.4 показаны различные оценки параметра положения и оценки расширенной неопределенности $2u(x_{pt})$.

Таблица Е.4 — Вычисление робастных среднего и стандартного отклонения для оценки 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
$x^* - \delta$		0,204 163	0,199 732	0,198 466	0,198 037	0,197 865	0,197 790
$x^* + \delta$		0,319 837	0,315 969	0,315 871	0,316 065	0,316 185	0,316 243
1	0,040 0	0,204 2	0,199 7	0,198 5	0,198 0	0,197 9	0,197 8
2	0,055 0	0,204 2	0,199 7	0,198 5	0,198 0	0,197 9	0,197 8
3	0,178 0	0,204 2	0,199 7	0,198 5	0,198 0	0,197 9	0,197 8
4	0,202 0	0,204 2	0,202 0	0,202 0	0,202 0	0,202 0	0,202 0
5	0,206 0	0,206 0	0,206 0	0,206 0	0,206 0	0,206 0	0,206 0
6	0,227 0	0,227 0	0,227 0	0,227 0	0,227 0	0,227 0	0,227 0
7	0,228 0	0,228 0	0,228 0	0,228 0	0,228 0	0,228 0	0,228 0
8	0,230 0	0,230 0	0,230 0	0,230 0	0,230 0	0,230 0	0,230 0
9	0,230 0	0,230 0	0,230 0	0,230 0	0,230 0	0,230 0	0,230 0
10	0,235 0	0,235 0	0,235 0	0,235 0	0,235 0	0,235 0	0,235 0
11	0,236 0	0,236 0	0,236 0	0,236 0	0,236 0	0,2360	0,236 0
12	0,237 0	0,237 0	0,237 0	0,237 0	0,237 0	0,237 0	0,237 0
13	0,243 0	0,243 0	0,243 0	0,243 0	0,243 0	0,243 0	0,243 0
14	0,244 0	0,244 0	0,244 0	0,244 0	0,244 0	0,244 0	0,244 0
15	0,245 0	0,245 0	0,245 0	0,245 0	0,245 0	0,245 0	0,245 0
16	0,255 5	0,255 5	0,255 5	0,255 5	0,255 5	0,255 5	0,255 5
17	0,260 0	0,260 0	0,260 0	0,260 0	0,260 0	0,260 0	0,260 0
18	0,264 0	0,264 0	0,264 0	0,264 0	0,264 0	0,264 0	0,264 0
19	0,267 0	0,267 0	0,267 0	0,267 0	0,267 0	0,267 0	0,267 0
20	0,270 0	0,270 0	0,270 0	0,270 0	0,270 0	0,270 0	0,270 0
21	0,273 0	0,273 0	0,273 0	0,273 0	0,273 0	0,273 0	0,273 0
22	0,274 0	0,274 0	0,274 0	0,274 0	0,274 0	0,274 0	0,274 0
23	0,274 0	0,274 0	0,274 0	0,274 0	0,274 0	0,274 0	0,274 0

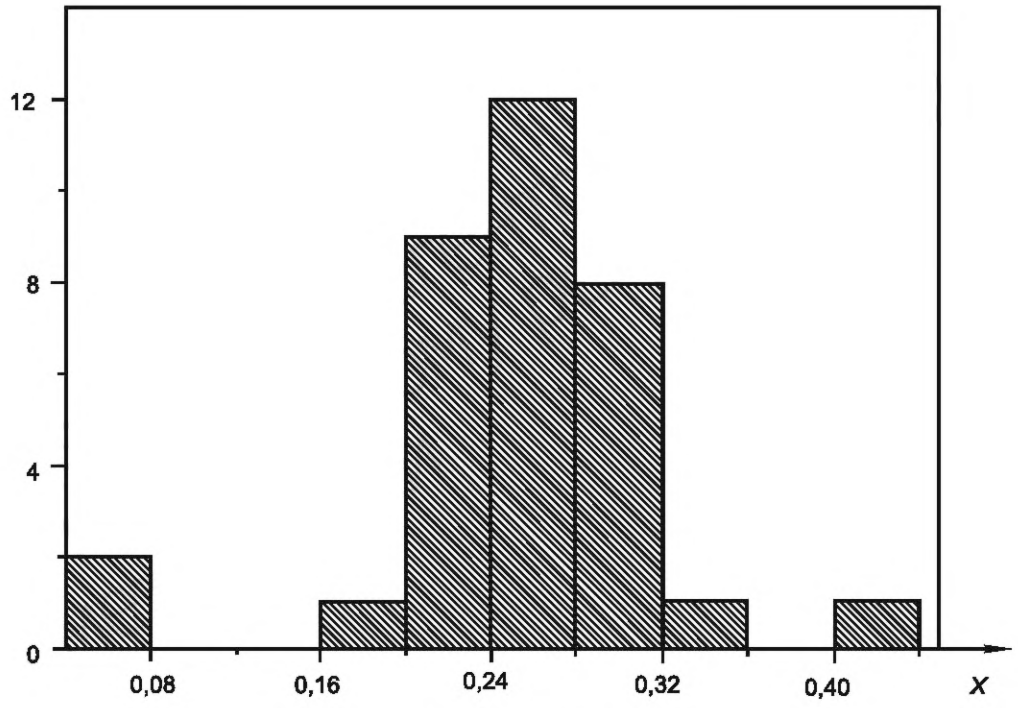
Окончание таблицы Е.4

x_i		1-я итерация	2-я итерация	3-я итерация	4-я итерация	5-я итерация	6-я итерация
$x^* - \delta$		0,204 163	0,199 732	0,198 466	0,198 037	0,197 865	0,197 790
$x^* + \delta$		0,319 837	0,315 969	0,315 871	0,316 065	0,316 185	0,316 243
24	0,278 0	0,278 0	0,278 0	0,278 0	0,278 0	0,278 0	0,278 0
25	0,281 1	0,281 1	0,281 1	0,281 1	0,281 1	0,281 1	0,281 1
26	0,287 0	0,287 0	0,287 0	0,287 0	0,287 0	0,287 0	0,287 0
27	0,287 0	0,287 0	0,287 0	0,287 0	0,287 0	0,287 0	0,287 0
28	0,288 0	0,288 0	0,288 0	0,288 0	0,288 0	0,288 0	0,288 0
29	0,289 0	0,289 0	0,289 0	0,289 0	0,289 0	0,289 0	0,289 0
30	0,295 0	0,295 0	0,295 0	0,295 0	0,295 0	0,295 0	0,295 0
31	0,296 0	0,296 0	0,296 0	0,296 0	0,296 0	0,296 0	0,296 0
32	0,311 0	0,311 0	0,311 0	0,311 0	0,311 0	0,311 0	0,311 0
33	0,331 0	0,319 8	0,316 0	0,315 9	0,316 1	0,316 2	0,316 2
34	0,424 6	0,319 8	0,316 0	0,315 9	0,316 1	0,316 2	0,316 2
Среднее	0,251 2	0,257 9	0,257 2	0,257 1	0,257 0	0,257 0	0,257 0
SD	0,067 2	0,034 2	0,034 5	0,034 7	0,034 8	0,034 8	0,034 8
δ		0,057 8	0,058 1	0,058 7	0,059 0	0,059 2	0,059 2
Новое x^*	0,262 0	0,257 9	0,257 2	0,257 1	0,257 0	0,257 0	0,257 0
Новое s^*	0,038 6	0,038 7	0,039 1	0,039 3	0,039 4	0,039 5	0,039 5



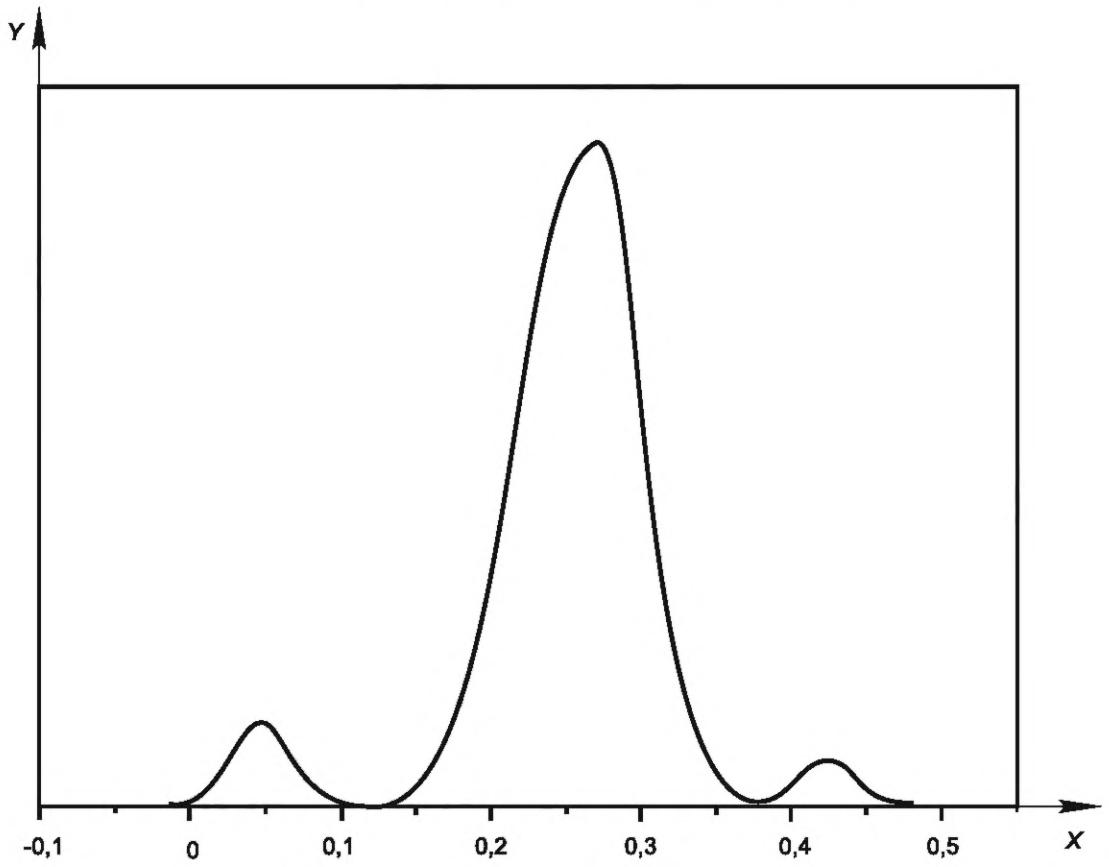
X — код лаборатории; *Y* — концентрация атразина (мг/л)

Рисунок Е.1 — Ранжированные результаты участников при анализе концентрации атразина (данные таблицы Е.4)



X — концентрация атразина (мг/л)

Рисунок Е.2 — Гистограмма результатов участников



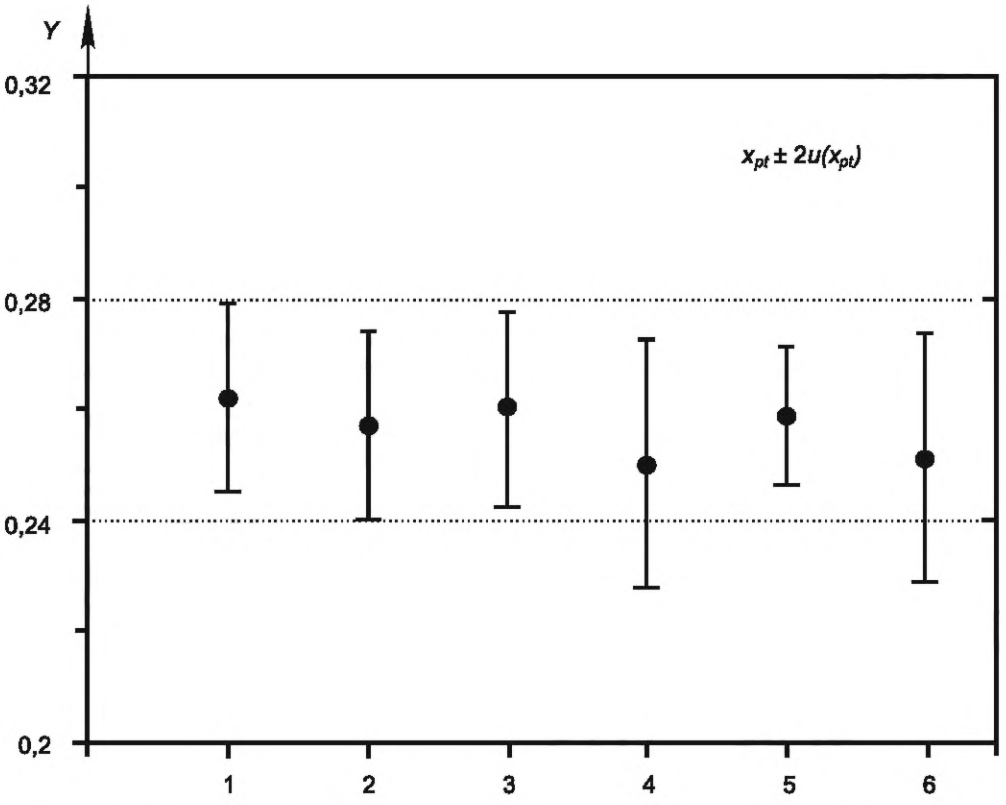
X — концентрация атразина (мг/л); *Y* — ядерная плотность

Рисунок Е.3 — График ядерной плотности по результатам участников

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* (MAD_e)	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
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.



Y — концентрация атразина (мг/л); *1* — робастный метод: медиана, *nIQR* (MAD_e); *2* — робастный метод: алгоритм *A* (x^*, s^*); *3* — робастный метод: оценка *Q*/Хампеля; *4* — бутстреп-метод (для среднего); *5* — арифметический метод: выбросы исключены; *6* — арифметический метод: выбросы включены

Рисунок Е.4 — Итоговые робастные статистики по данным таблицы Е.5

Е.4 Содержание ртути в корме для животных

В туре 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 для животных. 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).

Показатели функционирования вычислены с использованием методов, описанных в разделе 9. Для всех расчетов в качестве x_{pt} использовалось опорное значение, а в качестве σ_{pt} — значение, соответствующее назначению, определенное на основе предыдущего опыта. Неопределенность приписанного значения определена в виде суммы общей стандартной неопределенности опорного значения и неопределенности вследствие неоднородности:

$$x_{pt} = 0{,}044 \text{ мг/кг}; \; U(x_{pt}) = 0{,}0082 \text{ мг/кг}; \; \sigma_{pt} = 0{,}0066 \text{ мг/кг} \; (=15 \; \%).$$

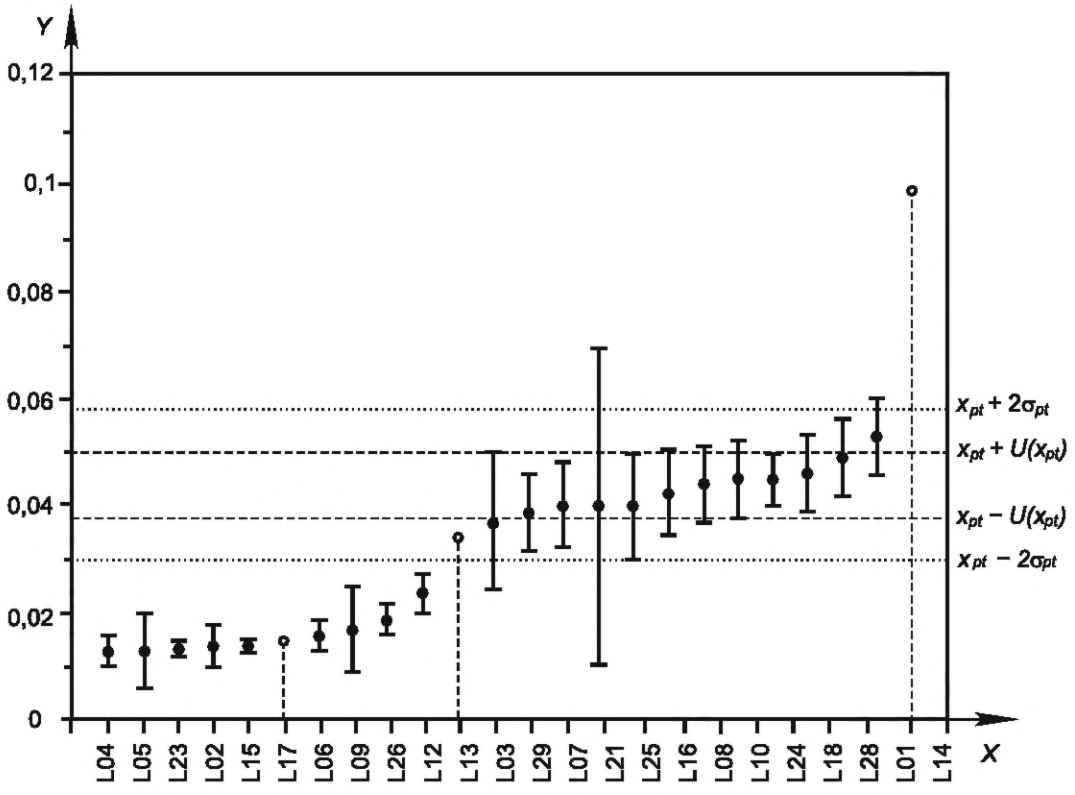
На графике ядерной плотности (см. рисунок Е.6) показано бимодальное распределение, связанное с применением различных методов. Однако это не влияет на оценивание функционирования, поскольку в качестве x_{pt} использовано опорное значение, а в качестве σ_{pt} — целевое значение. При выполнении этого анализа результаты со знаком «<» исключены.

Таблица Е.6 — Результаты проверки квалификации 24 участников исследования IMEP 111

Код лаборатории	Значение содержания ртути	U_{lab}	k	u_{lab}	Критерий	Метод
L04	0,013	0,003	2	0,002	b	AMA
L05	0,013	0,007	2	0,004	a	AMA
L23	0,0135	0,00108	1,732	0,00062	b	AMA
L02	0,014	0,004	2	0,002	b	AMA
L15	0,014	0,0005	2	0,0003	b	AMA
L17	<0,015					CV-ICP-AES
L06	0,016	0,003	2	0,002	b	AMA
L09	0,017	0,008	2	0,004	a	AMA
L26	0,019	0,003	2	0,002	b	AAS
L12	0,0239	0,0036	2	0,0018	b	AMA
L13	<0,034					TDA-AAS
L03	0,037	0,013	2	0,007	a	CV-AAS
L29	0,039	0,007	2	0,004	a	CV-AAS
L07	0,04	0,008	2	0,004	a	ICP-MS
L21	0,04	0,03	2	0,02	c	HG-AAS
L25	0,040	0,010	2	0,005	a	CV-AAS
L16	0,0424	0,008	2	0,004	a	CV-AAS
L08	0,044	0,007	2	0,004	a	CV-AAS
L10	0,045	0,007	2	0,004	a	ICP-MS
L24	0,045	0,005	2	0,003	a	HG-AAS
L18	0,046	0,007	2	0,004	a	CV-AAS
L28	0,049	0,0072	2	0,0036	a	CV-AAS
L01	0,053	0,007	2	0,004	a	CV-AAS
L14	<0,1					ICP-MS

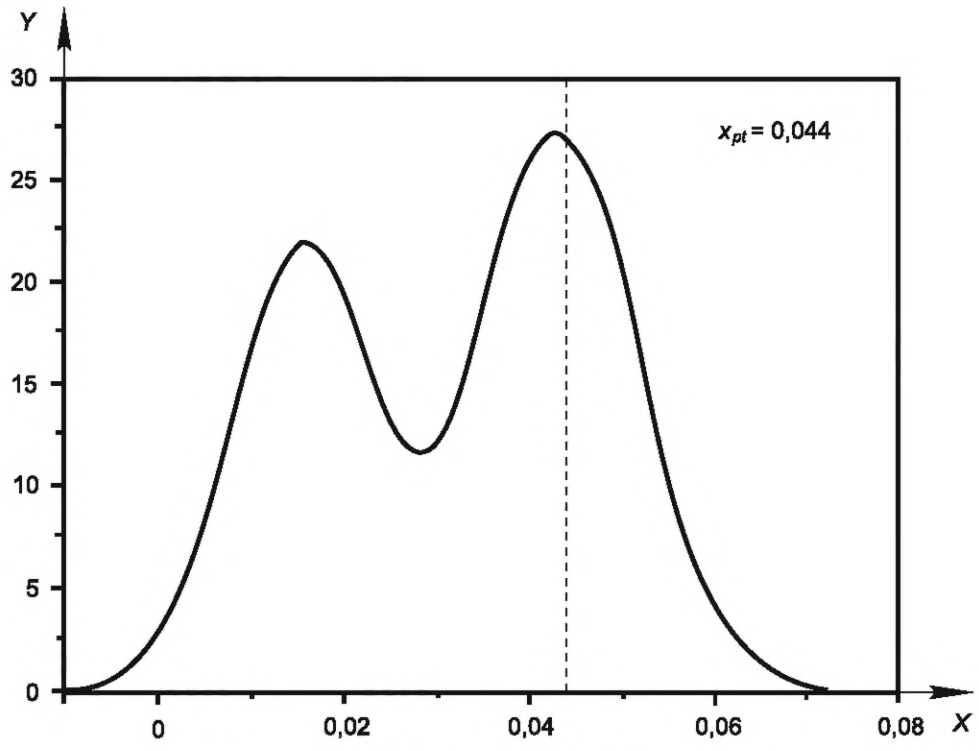
Линии с крупным пунктиром показывают границы со значениями $x_{pt} \pm U(x_{pt})$, а линии с мелким пунктиром — границы со значениями $(x_{pt} \pm 2\sigma_{pt})$.

Вертикальные линии, заканчивающиеся открытым кружком, показывают значения со знаком «<».



X — код лаборатории; *Y* — общее содержание ртути в кормах (мг/кг)

Рисунок Е.5 — Результаты и неопределенности результатов участников при определении концентрации IMEP 111 (данные таблицы Е.6)



X — концентрация ртути (мг/кг); *Y* — ядерная плотность

Рисунок Е.6 — График ядерной плотности для результатов участников

Таблица Е.7 — Статистика функционирования участников с использованием различных методов

Код лаборатории	D %	P_A	z	z'	ζ	E_n
L04	–70,5 %	–156,6 %	–4,70	–3,99	–7,10	–3,55
L05	–70,5 %	–156,6 %	–4,70	–3,99	–5,75	–2,88
L23	–69,3 %	–154,0 %	–4,62	–3,93	–7,35	–3,69
L02	–68,2 %	–151,5 %	–4,55	–3,86	–6,58	–3,29
L15	–68,2 %	–151,5 %	–4,55	–3,86	–7,30	–3,65
L17						
L06	–63,6 %	–141,4 %	–4,24	–3,60	–6,41	–3,21
L09	–61,4 %	–136,4 %	–4,09	–3,47	–4,71	–2,36
L26	–56,8 %	–126,3 %	–3,79	–3,22	–5,73	–2,86
L12	–45,7 %	–101,5 %	–3,05	–2,59	–4,49	–2,24
L13						
L03	–15,9 %	–35,4 %	–1,06	–0,90	–0,91	–0,46
L29	–11,4 %	–25,3 %	–0,76	–0,64	–0,93	–0,46
L07	–9,1 %	–20,2 %	–0,61	–0,51	–0,70	–0,35
L21	–9,1 %	–20,2 %	–0,61	–0,51	–0,26	–0,13
L25	–9,1 %	–20,2 %	–0,61	–0,51	–0,62	–0,31
L16	–3,6 %	–8,1 %	–0,24	–0,21	–0,28	–0,14
L08	0,0 %	0,0 %	0,00	0,00	0,00	0,00
L10	2,3 %	5,1 %	0,15	0,13	0,19	0,09
L24	2,3 %	5,1 %	0,15	0,13	0,21	0,10
L18	4,5 %	10,1 %	0,30	0,26	0,37	0,19
L28	11,4 %	25,3 %	0,76	0,64	0,92	0,46
L01	20,5 %	45,5 %	1,36	1,16	1,67	0,83
L14						

Данный пример предоставлен Институтом эталонных материалов и измерений Объединенного исследовательского центра Европейской комиссии Международной программы оценивания 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 (ССО), который имеет сертифицированное значение 21,62 единиц LA и соответствующую неопределенность 0,26 единиц LA. В примере показано, как опорное значение и неопределенность могут быть получены для образца для проверки квалификации. Следует помнить, что неопределенность значения ССО включает в себя неопределенность, связанную с неоднородностью, транспортированием и долгосрочной стабильностью:

$$x_{pt} = 21{,}62 + 1{,}73 = 23{,}35 \text{ (единиц LA)},$$

$$u\left(x_{pt}\right) = \sqrt{0{,}26^2 + 0{,}24^2} = 0{,}35 \text{ (единиц LA)},$$

где 0,26 — стандартная неопределенность значения ССО, а 0,24 — стандартная неопределенность $\overline{d}$.

Таблица Е.8 — Вычисление разности средних арифметических, соответствующих ССО и образцу для проверки квалификации и стандартного отклонения этой разности

Номер образца	Образец для проверки квалификации		ССО		Разность средних образца и ССО, 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	3,00
Разность средних арифметических $\overline{d}$					1,73
Стандартное отклонение					1,07
Стандартная неопределенность $\overline{d}$ (стандартное отклонение/$\sqrt{20}$)					0,24

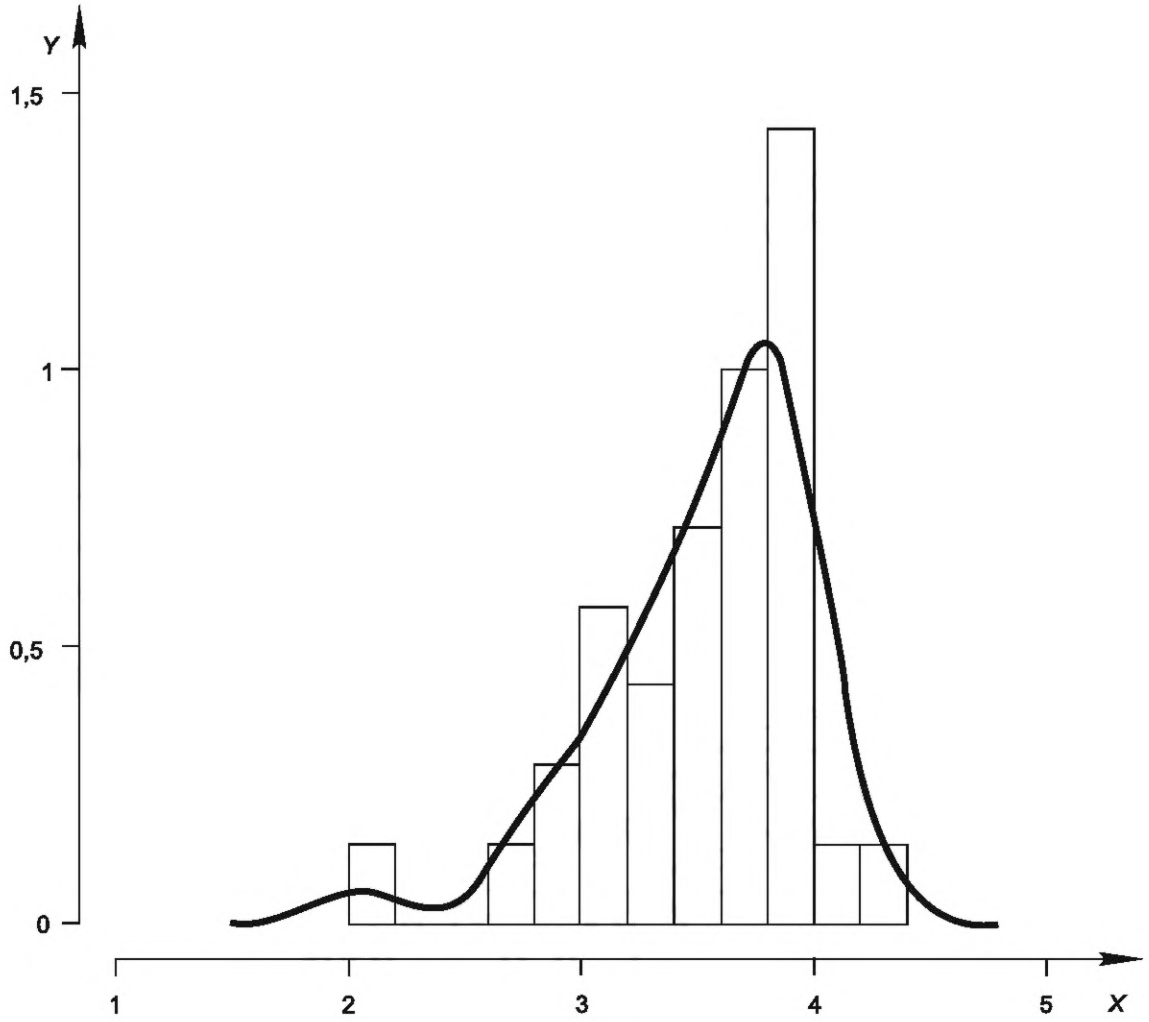
Примечание — Данные представляют собой результат измерений механической прочности заполнителя, полученные при испытании методом Лос-Анджелеса.

Е.6 Пример применения бутстреп-метода для определения содержания бактерий группы кишечной палочки *Coliform* в образце пищи (см. 7.7.6)

В программе проверки квалификации на наличие бактерий группы кишечной палочки в пробе молока участвовали 35 лабораторий, которые выполняли по 5 повторных измерений каждая. Среднее логарифмов CFU данных каждого участника было использовано для оценки приписанного значения и его неопределенности. Целевое значение, равное 0,25 log CFU/мл, установлено в качестве σ_{pt}, в то время как стандартное отклонение функции ядерной плотности задано $0{,}75\sigma_{pt}$. График ядерной плотности (см. рисунок Е.7) имеет вид ассиметричного распределения. Для определения оценки моды и соответствующей стандартной погрешности функции ядерной плотности распределения данных, обозначенных x_{pt} и $u(x_{pt})$, соответственно применен бутстреп-метод (1000 повторений). Результаты вычислений получены с помощью компьютерной программы, текст которой приведен в приложении F. Получены следующие значения:

$$x_{pt} = 3{,}79 \text{ и } u(x_{pt}) = 0{,}0922 \text{ (log CFU/мл)}.$$

Примечание — Поскольку $u(x_{pt}) > 0{,}3\sigma_{pt}$, функционирование лабораторий оценено с использованием показателей z'.



X — количество кишечных палочек (log10CFU/мл); *Y* — ядерная плотность

Рисунок Е.7 — График ядерной плотности по результатам участников

Е.7 Сравнение опорного значения с согласованным средним (см. 7.8)

Для демонстрации процедуры, приведенной в 7.8, проведено сравнение опорного значения с робастным средним по полученным результатам участников на основе примера Е.4 и данных таблицы Е.6.

В этом туре программы проверки квалификации робастное среднее $x^* = 0,03161$ и робастное стандартное отклонение $s^* = 0,0164$ получены с помощью алгоритма *A*, после исключения трех результатов со знаком «<» ($n = 21$ после исключения упомянутых результатов). Затем определена неопределенность робастного среднего

$$u\left(x^*\right) = 1,25\left(s^* / \sqrt{n}\right),$$

$$u\left(x^*\right) = 1,25\left(0,0164 / \sqrt{21}\right) = 0,0045.$$

В соответствии с 7.8 неопределенность разности между x_{ref} и x^* имеет следующий вид:

$$u_{diff} = \sqrt{u^2\left(x_{ref}\right) + u^2\left(x^*\right)} = \sqrt{0,0041^2 + 0,0045^2} = 0,0061,$$

$$U_{diff} = 2(0,0061) = 0,012,$$

$$x_{diff} = x_{ref} - x^* = 0,044 - 0,032 = 0,012.$$

Таким образом, разность в два раза превышает неопределенность.

Никаких действий предпринимать не рекомендуется, поскольку очевидно, что в некоторых методах присутствует смещение.

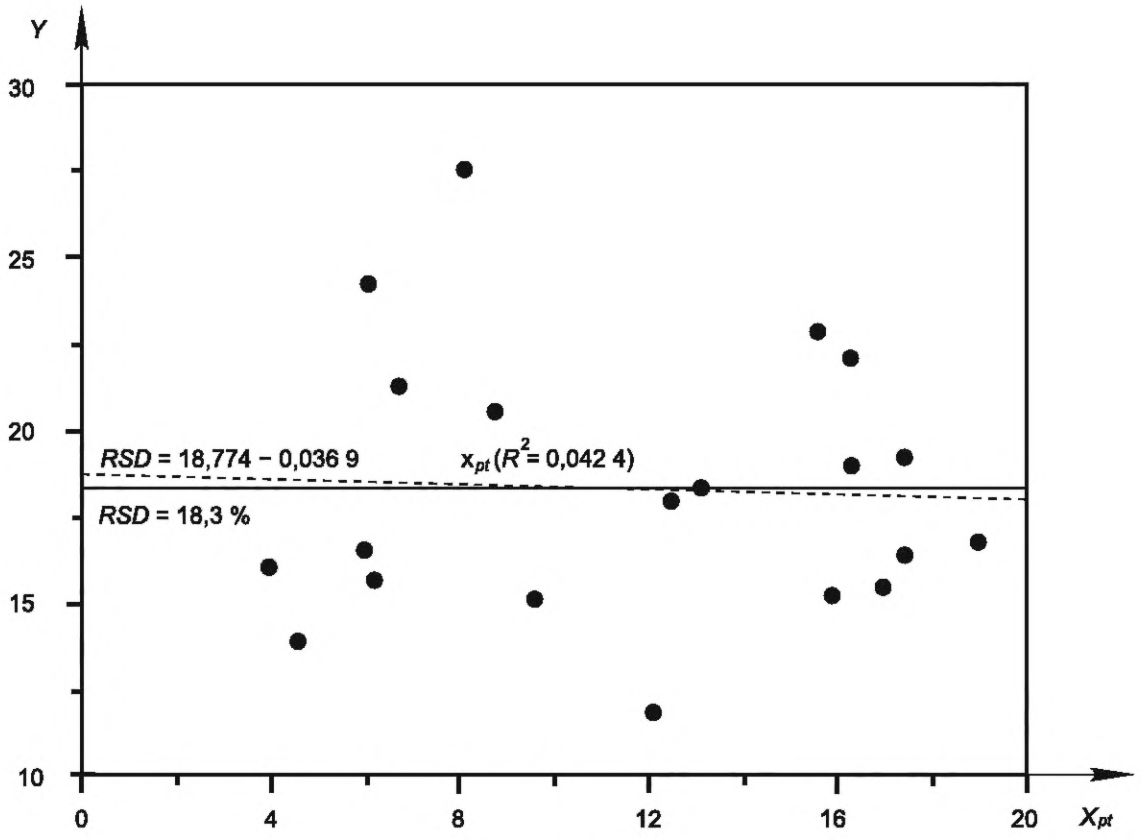
Е.8 Определение критериев для оценивания на основании опыта предыдущих туров: содержание токсафена в питьевой воде (см. 8.3)

Два провайдера организации программы проверки квалификации проверяют содержание пестицида токсафена в питьевой воде. В течение пяти лет проведено 20 туров проверки квалификации, в которых каждый раз принимали участие 20 или более участников, уровень токсафена в исследуемых пробах питьевой воды колебался от 3 до 20 мг/л. В таблице Е.9 представлены результаты 20 туров проверки квалификации, упорядоченные в порядке возрастания 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 тура программы проверки квалификации по отношению к приписанному значению (рассчитанному по процедуре приготовления). На каждом рисунке обозначена линия регрессии, полученная по методу наименьших квадратов. Линию регрессии по методу наименьших квадратов можно определить с помощью общедоступного программного обеспечения. (Также проверена полиномиальная модель 2-го порядка в качестве функции связи 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 является более подходящей.)

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 для оценки квалификации составляло 19 % от приписанного значения (или, возможно, 20 %), или осуществить вычисление среднего стандартного отклонения на основе уравнения линии регрессии для стандартного отклонения.

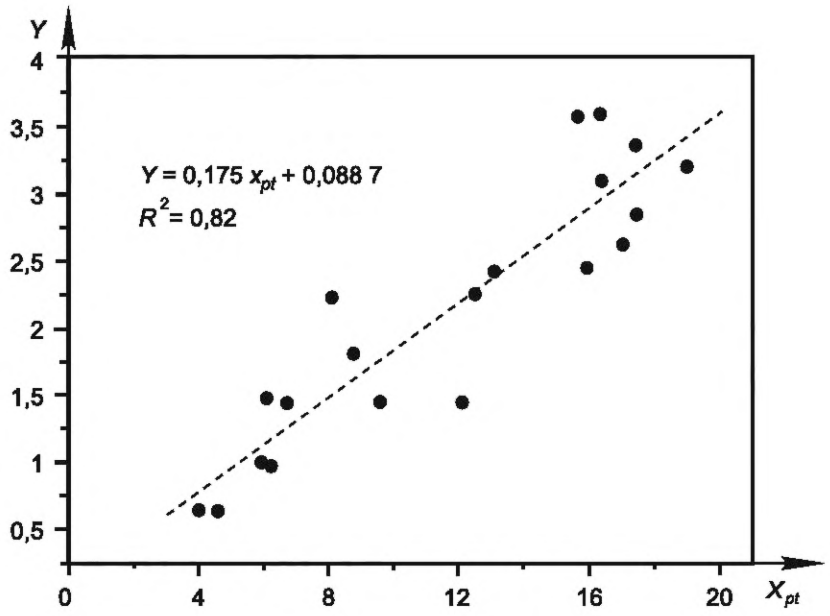
Таблица Е.9 — Данные по определению содержания токсафена в питьевой воде для $p \geq 20$

Код провайдера проверки квалификации	Приписанное значение	Робастное среднее	Стандартное отклонение	Среднее найденное	*RSD* (% от приписанного значения)	*p*
P004	3,96	3,98	0,639	100,5 %	16,1 %	25
P001	4,56	5,18	0,638	113,6 %	14,0 %	23
P001	5,99	5,98	0,995	99,8 %	16,6 %	22
P004	6,08	5,80	1,48	95,4 %	24,3 %	20
P001	6,20	6,66	0,97	107,4 %	15,7 %	23
P001	6,72	7,13	1,43	106,1 %	21,3 %	22
P004	8,10	7,09	2,23	87,5 %	27,5 %	21
P001	8,73	8,15	1,80	93,4 %	20,6 %	22
P001	9,57	8,60	1,45	89,9 %	15,2 %	23
P001	12,1	12,4	1,44	102,5 %	11,9 %	23
P001	12,5	13,8	2,25	110,4 %	18,0 %	24
P004	13,1	12,0	2,41	91,6 %	18,4 %	20
P004	15,6	13,3	3,57	85,3 %	22,9 %	27
P004	15,9	13,6	2,44	85,5 %	15,3 %	28
P004	16,3	13,5	3,60	82,8 %	22,1 %	31
P004	16,3	14,2	3,09	87,1 %	19,0 %	40
P004	17,0	15,6	2,63	91,8 %	15,5 %	24
P004	17,4	16,0	2,85	92,0 %	16,4 %	23
P004	17,4	16,0	3,36	92,0 %	19,3 %	23
P004	19,0	16,4	3,20	86,3 %	16,8 %	27



x_{pt} — (мг/л); Y — RSD (%)

Рисунок Е.8 — Относительное стандартное отклонение результатов участников (%) от приписанного опорного значения (мг/л)



x_{pt} — (мг/л); Y — SD (мг/л)

Рисунок Е.9 — Стандартное отклонение участников (мг/л) по отношению к приписанному значению (мг/л)

Е.9 Общая модель: уравнение Хорвица (см. 8.4)

Одна из общих моделей, применяемых в химии, описана Хорвицем [22], [31]. Такой подход формирует общую модель воспроизводимости аналитических методов, которая может быть использована для вывода 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 использовано 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 стандартное отклонение воспроизводимости имеет следующий вид:

Образец A (1,195 мг/кг): σ_R = 0,186 мг/кг или σ_R = 15,6 %,

Образец B (2,565 мг/кг): σ_R = 0,356 мг/кг или σ_R = 13,9 %.

Е.10 Определение показателя функционирования в экспериментах на прецизионность: определение содержания цемента в твердом бетоне (см. 8.5)

Содержание цемента в бетоне обычно измеряют в единицах массы на кубический метр бетона кг/м^3. На практике бетон производят в соответствии с сортами, которые отличаются по содержанию цемента на 25 кг/м^3 друг от друга, 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 повторных 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 практически неосуществимой.

Примечание — В ISO 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 вычислено следующим образом:

$$\sigma_L = \sqrt{\sigma_L^2 - \sigma_r^2} = \sqrt{\left(23{,}2^2 - 14{,}3^2\right)} = 18{,}3\ \text{кг/м}^3.$$

Е.11 Штриховые графики для нормированного смещения: концентрация антител (см. 10.4)

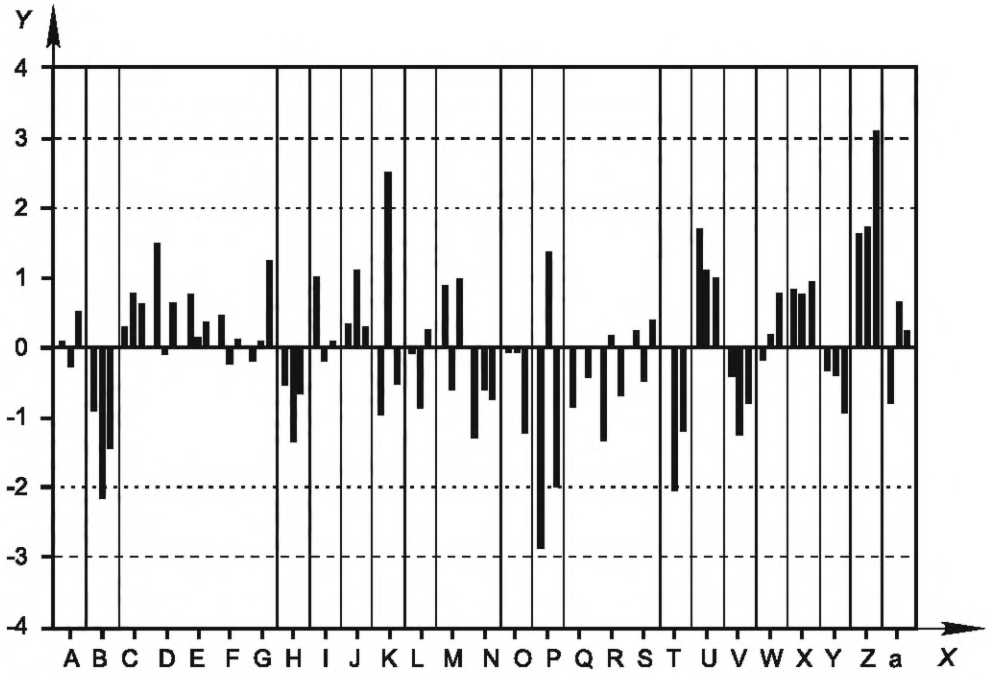
Значения показателей z для тура процедуры проверки квалификации с тремя связанными измеряемыми величинами (антителами) приведены на рисунке Е.10 в виде штрихового графика. Данные для двух из трех аллергенов приведены в таблице Е.10. Из графика видно, что, например, лабораториям B и Z следует искать причину, которая влияет на все три уровня и дает примерно одинаковую величину смещения, в то время как для лабораторий K и P показатель z зависит от типа антител.

Е.12 Диаграмма Юдена: концентрация антител (см. 10.5)

В таблице Е.10 приведены данные, полученные при исследовании двух схожих образцов для проверки квалификации при определении концентрации антител. Показанные на рисунке Е.11 показатели z функционирования основаны на робастных среднем и стандартном отклонении, полученных в соответствии с алгоритмом A (см. рисунок Е.11).

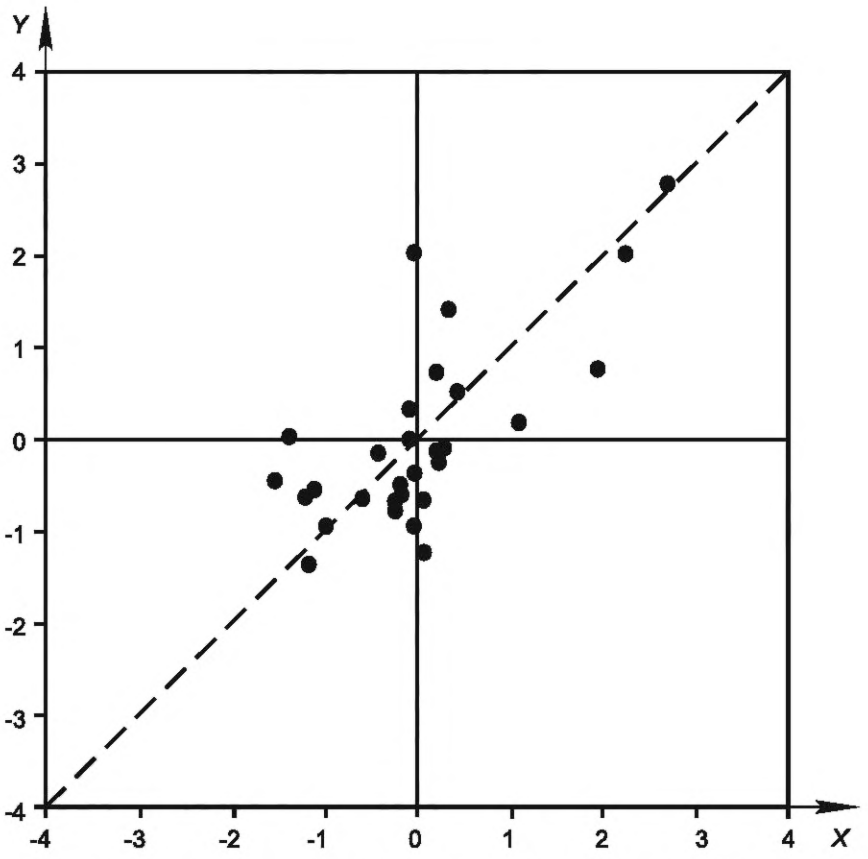
Из рисунка Е.11 видно, что значения двух участников (5 и 23) находятся вверху правого верхнего квадранта и, следовательно, могут иметь устойчивое положительное смещение. Лаборатория 26 имеет высокое значение показателя z на образце аллергена B и отрицательное значение показателя z (–0,055) на образце аллергена A, поэтому она может иметь низкую повторяемость.

Результаты участников 5, 23 и 26 попадают в область «сигнала предупреждения». Эти участники должны проверить, в какую зону попадут их результаты в следующем туре программы. Визуальный анализ и коэффициент корреляции указывают на тенденцию к последовательному изменению показателя z (положительному или отрицательному), так что может существовать возможность улучшения метода измерений с более подробными инструкциями.



X — код лаборатории; *Y* — показатель z

Рисунок Е.10 — Штриховой график показателей z (от 4,0 до –4,0) для одного тура программы проверки квалификации, в ходе которого участники определяли концентрации трех аллергенспецифичных IgE–антител



X — показатель *z* по аллергену *A*; *Y* — показатель *z* по аллергену *B*

Рисунок Е.11 — Диаграмма Юдена для показателей *z* из таблицы Е.10

Таблица Е.10 — Данные и расчеты по определению концентрации антител для двух схожих образцов аллергенов

Лаборатория	Данные		Показатели z	
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
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	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 представляют собой количество единиц (U) в тысячах (k) на литр (l) пробы, где единица определяется концентрацией международного эталонного материала.

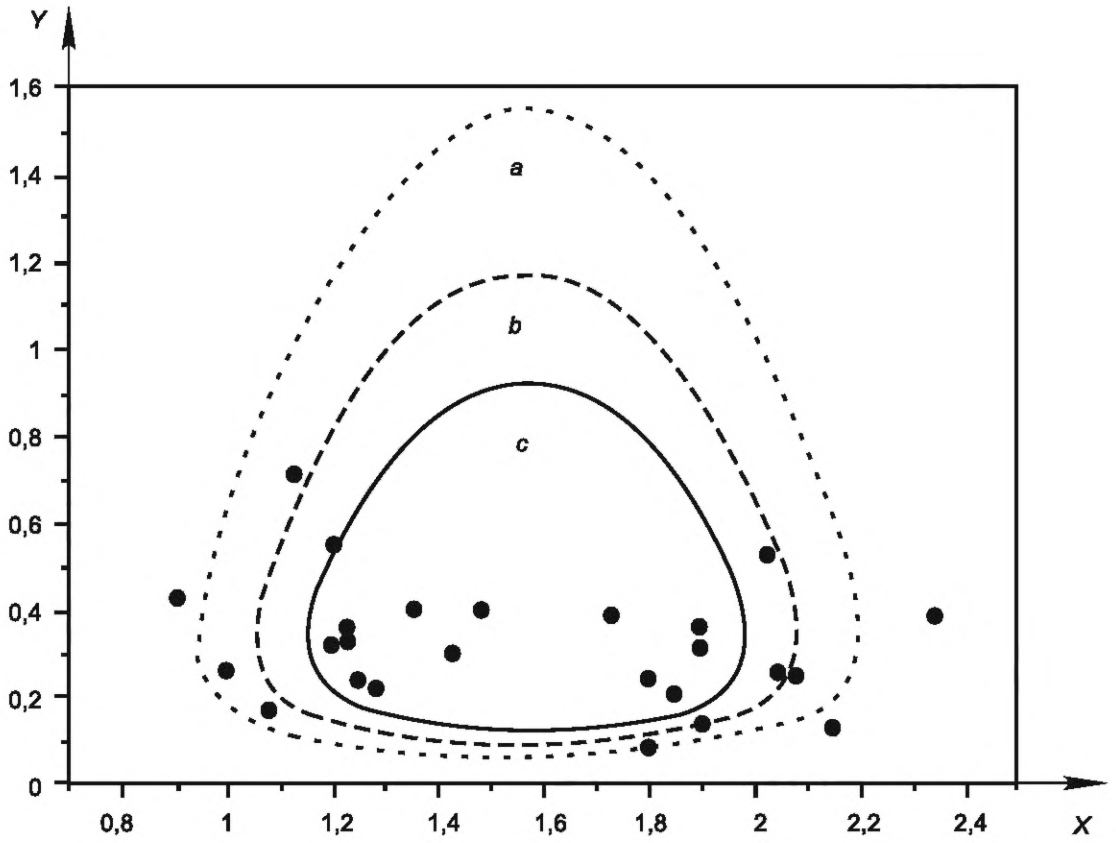
Примечание 2 — Показатели z в данной таблице рассчитаны с использованием неокругленных значений робастных среднего и стандартного отклонения, без использования округленных значений, приведенных в конце таблицы.

Е.13 График стандартных отклонений повторяемости: концентрации антител (см. 10.6)

В таблице Е.11 приведены результаты определения концентрации определенных антител в образцах сыворотки крови. Каждый участник выполнил четыре повторных 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 образцах сыворотки крови (четыре повторных определения на каждом образце, выполненных 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	1,57	
Робастное стандартное отклонение		0,34
Примечание — Данные представлены в тысячах единиц, kU, на литр, l, образца, где единицей является концентрация в международном стандартном образце.		



X — среднее арифметическое концентрации (kU/l); *Y* — стандартное отклонение (kU/l); *a* — уровень 0,1 **%;** *b* — уровень 1 %; *c* — уровень 5 %

Рисунок Е.12 — График стандартного отклонения и среднего для 25 участников (см. данные таблицы Е.11)

Е.14 Графические методы отслеживания функционирования участников во времени (см. 10.8)

Участникам полезно отслеживать свое функционирование во времени или иметь такие данные, подготовленные провайдером проверки квалификации. Самым простым методом контроля для этого является контрольная карта или карта Шухарта. В связи с чем необходимо иметь нормированные показатели функционирования, такие как показатели z или P_A, и участвовать в нескольких турах программы проверки квалификации. Данный пример относится к программе проверки квалификации медицинских лабораторий, в процессе которой участники определяют содержание калия в сыворотке крови.

Провайдер проверки квалификации использовал фиксированный интервал ±5 % с округлением значений до 0,1 ммоль/л, но не менее ±0,2 ммоль/л. Провайдер использовал показатели P_A (см. таблицу Е.12).

Таблица Е.12 — Значения показателя P_A за пять туров программы проверки квалификации с тремя образцами сыворотки крови каждый

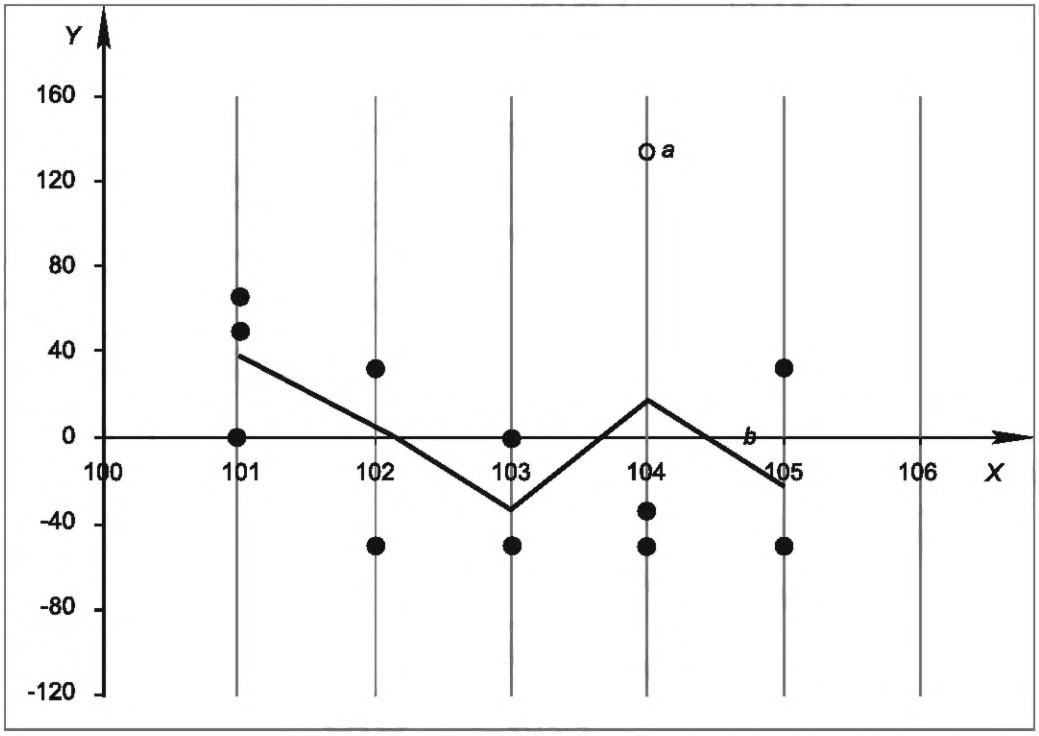
Код тура	Образец для проверки квалификации	Результат	Приписанное значение	Показатель P_A	Среднее арифметическое значений P_A
101	A	6,4	6,2	67	39
101	B	4,2	4,1	50	
101	C	4,1	4,1	0	
102	A	6,0	5,9	33	6
102	B	4,3	4,4	–50	
102	C	5,5	5,4	33	
103	A	4,1	4,2	–50	–33

Окончание таблицы Е.12

Код тура	Образец для проверки квалификации	Результат	Приписанное значение	Показатель P_A	Среднее арифметическое значений P_A
103	B	3,6	3,7	–50	
103	C	4,2	4,2	0	
104	A	5,7	5,8	–33	17
104	B	3,9	4,0	–50	
104	C	6,3	5,9	133	
105	A	3,6	3,7	–50	–22
105	B	4,5	4,6	–50	
105	C	5,3	5,2	33	

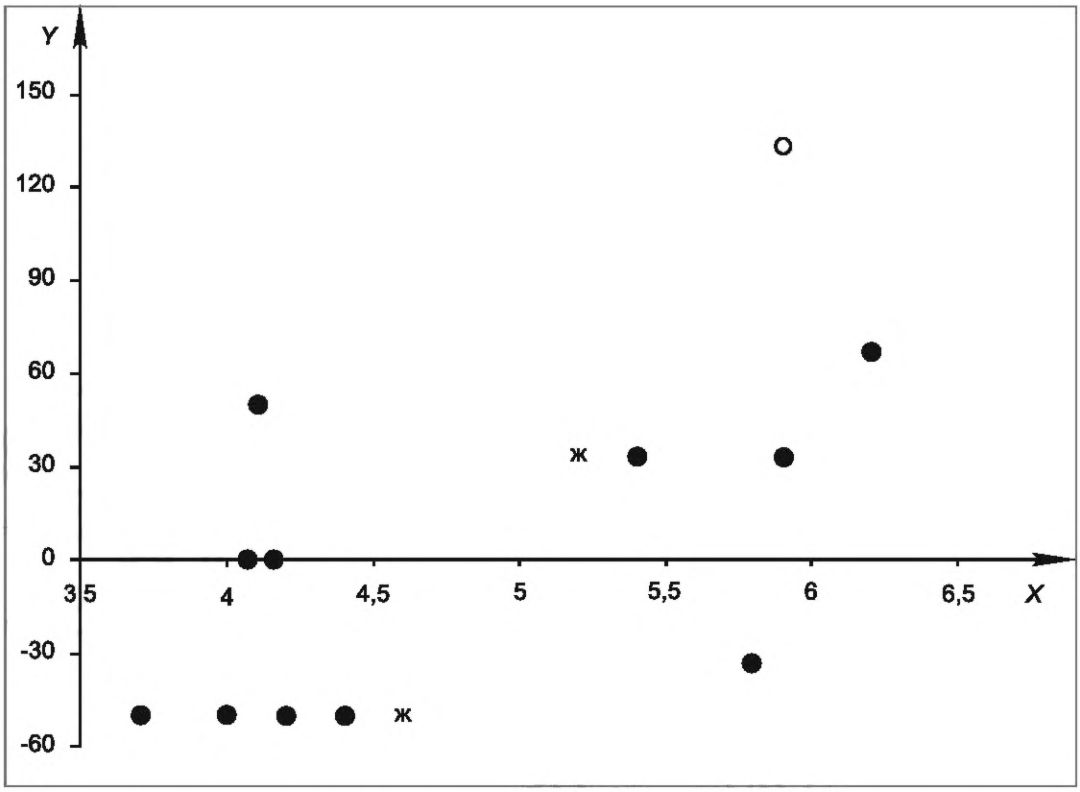
Полученные результаты могут быть представлены на графике. Для визуального анализа рекомендуется использовать два вида графиков:

- контрольная карта для нормированного показателя функционирования для каждого тура, показывающая результаты на нескольких образцах в одном и том же туре программы проверки квалификации. Это позволяет показать изменение показателя функционирования во времени, в том числе выявить тенденции его изменения (см. рисунок Е.13);
- точечная диаграмма нормированных показателей функционирования в зависимости от приписанных значений для анализа зависимости показателя функционирования от уровня концентрации и выявления тенденций, связанных с уровнем измеряемой величины (см. рисунок Е.14).



X — тур проверки квалификации; *Y* — показатель P_A; *a* — сигнал «действия»; *b* — среднее арифметическое P_A

Рисунок Е.13 — Показатели функционирования для каждого тура программы проверки квалификации (данные таблицы Е.12)



X — приписанное значение (ммоль/л); *Y* — показатель P_A; • — предыдущее значение; ж — текущее значение; ○ — сигнал действия

Рисунок Е.14 — Показатели функционирования для различных уровней измеряемой величины

Е.15 Качественный анализ данных: пример порядковой величины: реакция кожи на косметическое средство (см. раздел 11)

Программа проверки квалификации включает анализ реакции на продукцию, предназначенную для ухода за кожей, при ее применении к стандартному живому существу. Любую воспалительную реакцию оценивают по следующей шкале:

a) отсутствие реакции;

b) умеренное покраснение;

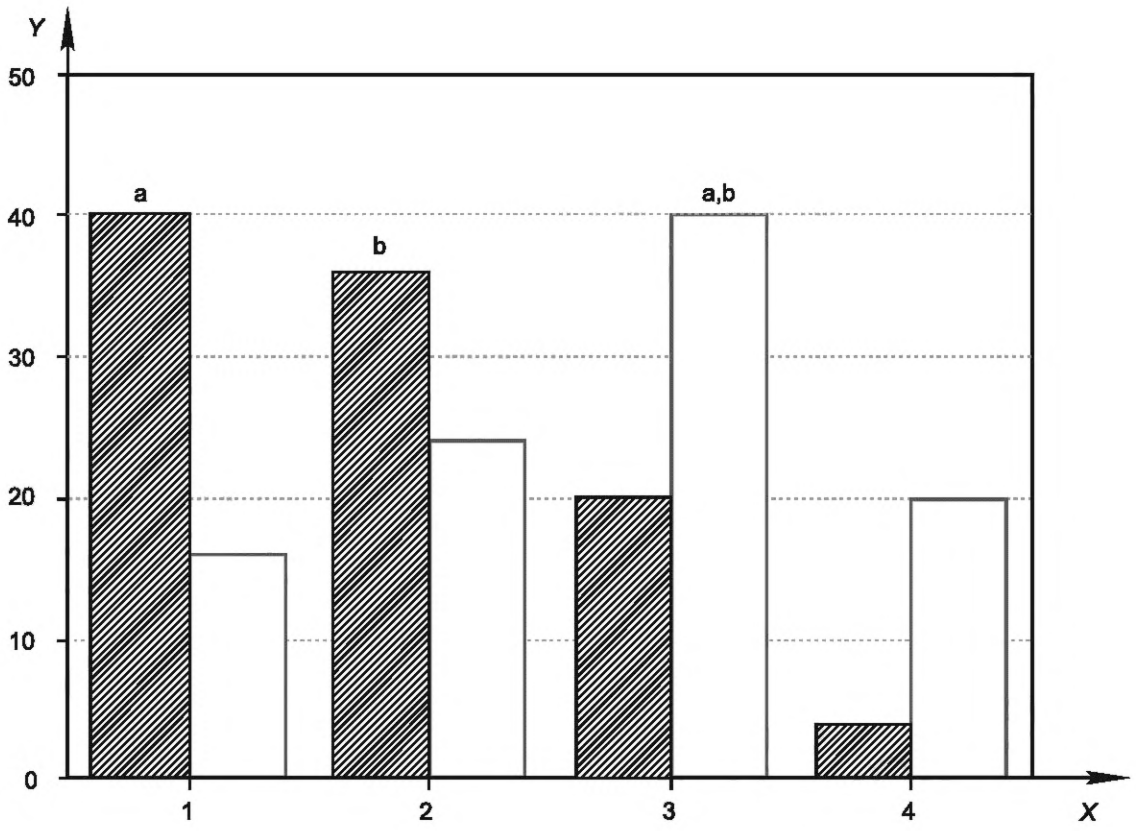
c) значительное раздражение или отек;

d) тяжелая реакция, включая 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 для проверки квалификации.

Таблица Е.13 — Результаты (реакция кожи) для двух образцов

Реакция	Продукт *A*	Продукт *B*
1	20 (40 %) #	8 (16 %)
2	18 (36 %) @	12 (24 %)
3	10 (20 %)	20 (40 %) # @
4	2 (4 %)	10 (20 %)
# — мода		
@ — медиана		



X — уровень реакции кожи; *Y* — процент результатов (%); *a* — #; *b* — @; *a*, *b* — #,@; — процент результатов по *A*; — процент результатов по *B*

Рисунок Е.15 — Диаграмма процента результатов (появление раздражения) для двух образцов проверки квалификации; # — мода, @ — медиана

Следует отметить, что медиана или мода могут быть использованы в качестве общей статистики для этих образцов, они указывают, что уровень реакции на продукт *B* является более тяжелым, чем реакция на продукт *A*. Провайдер может определить, что «сигнал действия» появляется для любого результата, который отстоит от медианы более чем на одну единицу измерения, в этом случае для продукта *A* — два результата «4» (4 %) попадают в зону сигнала действия, а для продукта *B* — восемь результатов «1» (16 %).

Приложение F
(справочное)

Пример компьютерной программы для построения графика и анализа результатов проверки квалификации бутстреп-методом

В следующей программе использована версия R 3.1.1 для построения рисунков и представления результатов примера E.6.

```
################################
#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
```

Разработчики благодарят Экспериментальный Зоопрофилактический институт Венеции — пищевая микробиология РТ «AQUA».

Приложение ДА
(справочное)

Сведения о соответствии ссылочных международных стандартов межгосударственным стандартам

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего межгосударственного стандарта
ISO 3534-1	—	*
ISO 3534-2	—	*
ISO 5725-1	—	*
ISO/IEC 17043	IDT	ГОСТ ISO/IEC 17043—2013 «Оценка соответствия. Основные требования к проведению проверки квалификации»
ISO Guide 30	IDT	ГОСТ ISO Guide 30—2019 «Стандартные образцы. Некоторые термины и определения»
ISO/IEC Guide 99	—	*

* Соответствующий межгосударственный стандарт отсутствует. До его принятия рекомендуется использовать перевод на русский язык данного международного стандарта. Официальный перевод данного международного стандарта находится в Федеральном информационном фонде стандартов.

Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандартов:

- IDT — идентичные стандарты.

Библиография

[1] ISO 5725-2 Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method

[2] ISO 5725-3 Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method

[3] ISO 5725-4 Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method

[4] ISO 5725-5 Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method

[5] ISO 5725-6 Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values

[6] ISO 7870-2, (2013) Control charts — Part 2: Shewhart control charts

[7] ISO 11352 Water quality — Estimation of measurement uncertainty based on validation and quality control data

[8] ISO 11843-1 Capability of detection — Part 1: Terms and definitions

[9] ISO 11843–2 Capability of detection — Part 2: Methodology in the linear calibration case

[10] ISO 16269-4 Statistical interpretation of data — Part 4: Detection and treatment of outliers

[11] ISO/IEC 17011 Conformity assessment — Requirements for accreditation bodies accrediting conformity assessment bodies

[12] ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories

[13] ISO Guide 35 Reference materials — Guidance for characterization and assessment of homogeneity and stability

[14] ISO/IEC Guide 98-3 Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

[15] Analytical Method Committee, Royal Society of Chemistry Accred Qual Assur. 2010, 15 pp. 73—79

[16] CCQM Guidance note: Estimation of a consensus KCRV and associated Degrees of Equivalence. Version 10. Bureau International des Poids et Mesures, Paris (2013)

[17] Davison A.C., Hinkley D.V., Bootstrap Methods and Their Application. Cambridge University Press, 1997

[18] Efron B., Tibshirani R., An Introduction to the Bootstrap. Chapman & Hall, 1993

[19] Lamberty A, Schimmel H, Pauwels J, The study of the stability of reference materials by isochronous measurements. Fres J, Anal Chem. 1998, 360 pp. 359—361

[20] Gower J.C., A general coefficient of similarity and some of its properties. Biometrics. 1971, 27 (4) pp. 857—871

[21] Helsel D.R., Nondetects and data analysis: statistics for censored environmental data. Wiley Interscience, 2005

[22] Horwitz W., Evaluation of analytical methods used for regulations of food and drugs. Anal. Chem. 1982, 54 pp. 67A—76A

[23] Jackson J.E., Quality control methods for two related variables. Industrial Quality Control. 1956, 7 pp. 2—6

[24] 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

[25] Maronna R.A., Martin R.D., Yohai V.J., Robust Statistics: Theory and methods. John Wiley & Sons Ltd, Chichester, England, 2006

[26] Müller C.H., Uhlig S. Estimation of variance components with high breakdown point and high efficiency; Biometrika; 88: Vol. 2, pp. 353—366, 2001.

[27] Rousseeuw P.J., Verboven S., Comput. Stat. Data Anal. 2002, 40 pp. 741—758

[28] Scott D.W., Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, 1992

[29] 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

[30] Silverman B.W., Density Estimation. Chapman and Hall, London, 1986

[31] Thompson M., Analyst (Lond.). 2000, 125 pp. 385—386

[32] 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

[33] 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

[34] 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

[35] 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

[36] van Nuland Y., ISO 9002 and the circle technique. Qual. Eng. 1992, 5 pp. 269—291

[37] https://quodata.de/en/web-services/QHampel.html

[38] ISO 16269-4 Statistical interpretation of data — Part 4: Detection and treatment of outliers

[39] Robouch P., Naji Y., Vermaercke P. The «Naji Plot», a simple graphical tool for the evaluation of inter-laboratory comparisons, in Richter D., Wöger W., Hässelbarth W., (eds.), Data analysis of key comparisons, Braunschweig and Berlin, 2003, ISBN 3-89701-933-3

[40] Ellison S. L. R., Applications of robust estimators of covariance in examination of interlaboratory study data. Analytical methods 2019, 11, 2639—2649, https://doi.org/10.1039/C8AY02724B

[41] Maechler M., Rousseeuw P., Croux C., Todorov V., Ruckstuhl A., Salibian-Barrera M. et al., c(«Eduardo», «L. T.») Conceicao and Maria Anna di Palma (2021). robustbase: Basic Robust Statistics R package version 0.93-7. URL http://CRAN.R-project.org/package=robustbase

[42] Christophe Croux and Peter J. Rousseeuw, Time-Efficient Algorithms for Two Highly Robust Estimators of Scale, in Computational Statistics, Volume 1, eds. Y. Dodge and J. Whittaker, Heidelberg: Physika-Verlag, 41 1-428, 1992

УДК 658.562.012.7:65.012.122:006.354 МКС 03.120.30

Ключевые слова: межлабораторное сличение, проверка квалификации, приписанное значение, показатель z, показатель дзета, выброс, образец для проверки квалификации, провайдер проверки квалификации, программа проверки квалификации, стандартный образец, сертифицированный стандартный образец

Технический редактор *И.Е. Черепкова*
Корректор *И.А. Королева*
Компьютерная верстка *Е.О. Асташина*

Сдано в набор 06.02.2025. Подписано в печать 24.02.2025. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 10,23. Уч.-изд. л. 8,49.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
www.gostinfo.ru info@gostinfo.ru