ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 50779.80 — 2013/ISO/TR 14468:2010

Статистические методы

АНАЛИЗ СОВПАДЕНИЯ РЕЗУЛЬТАТОВ ПРОВЕРОК ПО АЛЬТЕРНАТИВНОМУ ПРИЗНАКУ

ISO/TR 14468:2010
Selected illustrations of attribute agreement analysis
(IDT)

Издание официальное

Москва
Стандартинформ
2014

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Автономной некоммерческой организацией «Научно-исследовательский центр контроля и диагностики технических систем» (АНО «НИЦ КД») на основе собственного аутентичного перевода стандарта, указанного в пункте 4

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

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 22 ноября 2013 г. № 1664-ст

4 Настоящий стандарт идентичен международному документу ISO/TR 14468:2010 «Примеры анализа совпадения результатов проверок по альтернативному признаку» (ISO/TR 14468:2010 «Selected illustrations of attribute agreement analysis»).

Наименование настоящего стандарта изменено относительно наименования указанного международного документа для приведения в соответствие с ГОСТ Р 1.5.

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

5 ВВЕДЕН ВПЕРВЫЕ

Правила применения настоящего стандарта установлены в ГОСТ Р 1.0–2012 (раздел 8). Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок – в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования – на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (gost.ru)

© Стандартинформ, 2014

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 1
4 Обозначения и сокращения 2
5 Общая методика анализа соответствия на основе результатов проверок по альтернативному признаку 3
6 Описание приложений A – E 6
Приложение A (справочное) Контроль жидкокристаллических дисплеев 8
Приложение B (справочное) Диагностика проблем с программным обеспечением 13
Приложение C (справочное) Определение вида воды на основе дегустации 19
Приложение D (справочное) Контроль дефектов термистора 24
Приложение E (справочное) Определение уровня трудоспособности после инсульта 31
Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов ссылочным национальным стандартам Российской Федерации (и действующим в этом качестве межгосударственным стандартам) 37

Библиография 38

Введение

Стандарты по статистическим методам и методологии «Шесть сигм» обеспечивают применение подхода постоянного улучшения и методов анализа качества.

Настоящий стандарт посвящен анализу совпадения результатов проверок или контроля по альтернативному признаку и иллюстрирует его применение на нескольких примерах (см. приложения A – E).

В этих примерах показан не только общий подход к применению данной методики, но и особенности его использования в различных ситуациях.

ISO/TR 14468:2010 разработан ISO/TC 69 «Применение статистических методов».

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Статистические методы

АНАЛИЗ СОВПАДЕНИЯ РЕЗУЛЬТАТОВ ПРОВЕРОК ПО АЛЬТЕРНАТИВНОМУ ПРИЗНАКУ

Statistical methods. Attribute agreement analysis

Дата введения — 2014—12—01

1 Область применения

В настоящем стандарте установлен метод анализа процесса, когда результатами определения значений контролируемой характеристики качества являются данные проверки или контроля по альтернативному признаку (в том числе неупорядоченные и упорядоченные данные).

В настоящем стандарте приведены примеры анализа совпадения результатов проверок по альтернативному признаку (AAA[1)]) с различными выводами относительно совпадения оценок аттестующими исследуемого объекта в процессе испытаний (далее оценок аттестующих), в том числе совпадения оценок одного аттестующего, совпадения оценок, полученных несколькими аттестующими, совпадение оценок каждого аттестующего со стандартной оценкой и совпадения оценок всех аттестующих со стандартной оценкой.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

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

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

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

3 Термины и определения

В настоящем стандарте применены термины по ISO 3534-1, ISO 3534-2, а также следующие термины с соответствующими определениями.

3.1 **система определения значений (характеристики)** (measurement system): Набор действий, процедур, устройств и оборудования, программного обеспечения и персонала, используемый для определения значения характеристики.

[IWA 1:2005, 3.1.9]

Примечание –В настоящем стандарте персонал относят к аттестующим.

3.2 **неупорядоченные данные** (nominal data): Значения переменных, принимающих два или более значений без естественного упорядочивания.

[1)] AAA – attribute agreement analysis.

3.3 **упорядоченные данные** (ordinal data): Значения переменных, принимающих три или более значений с естественным упорядочиванием.

3.4 **двоичные данные** (binary data): Значения переменных, принимающих два значения без естественного упорядочивания.

3.5 **согласованность оценок одного аттестующего** (agreement within appraiser): Степень совпадения оценок, полученных одним и тем же аттестующим во всех испытаниях, если аттестующий участвует в двух или более испытаниях.

3.6 **согласованность оценок нескольких аттестующих** (agreement between appraisers): Степень совпадения оценок, полученных различными аттестующими, если два и более аттестующих участвуют в одном или нескольких испытаниях.

3.7 **согласованность оценок одного аттестующего со стандартной оценкой** (agreement of each appraiser vs. standard): Степень совпадения оценок, полученных одним аттестующим со стандартной оценкой, если она установлена.

3.8 **согласованность оценок всех аттестующих со стандартной оценкой** (agreement of all appraisers vs. standard): Степень совпадения результатов, полученных всеми аттестующими во всех испытаниях, между собой и со стандартной оценкой, если она установлена.

3.9 **процент совпадения оценок *P*, %** (percentage of agreement): Процент совпадения оценок, полученных одним аттестующим при многократном повторении испытаний или оценок различных аттестующих.

3.10 **κ-статистика** (kappa): Статистика, указывающая на степень совпадения упорядоченных или неупорядоченных результатов, полученных несколькими аттестующими по одной и той же выборке.

Примечание – κ-статистику обычно используют в исследовании AAA.

3.11 **κ-статистика Флейса** (Fleiss's kappa): Статистика, используемая для проверки согласованности оценок аттестующих, если аттестующие выбраны случайным образом из группы аналогичных аттестующих.

3.12 **κ-статистика Кохена** (Cohen's kappa): Статистика, используемая для проверки согласованности оценок аттестующих, если аттестующие выбраны и фиксированы до проведения испытаний.

3.13 ***p* – значение** (*p*-value): Вероятность появления наблюдаемого значения тестовой статистики (или другой величины), используемой для принятия решения при проверке гипотез, неблагоприятного по отношению к нулевой гипотезе.

[ИСО 3534-1:2006, 1.49]

Примечание – Эту величину используют при проверке гипотез для принятия решений об отклонении или принятии нулевой гипотезы.

3.14 **Z-статистика** (Z-statistic): Статистика, используемая при проверке гипотез, подчиняющаяся нормированному нормальному распределению

4 Обозначения и сокращения

95 % CI – доверительный интервал с уровнем доверия 95 %;

AAA – анализ совпадения результатов проверок по альтернативному признаку;

MSA – анализ измерительных систем;

σ_κ – стандартная ошибка κ-статистики (стандартное отклонение κ-статистики);

n – объем выборки;

P, % – процент совпадения оценок;

Z – Z-статистика.

5 Общая методика анализа совпадения результатов проверок по альтернативному признаку

5.1 Краткое описание анализа

В настоящем стандарте приведены общие принципы планирования эксперимента, проведения и анализа результатов испытаний для исследования совпадения оценок, данных аттестующими при отнесении результатов испытаний в две или более категорий (например, «хороший» и «плохой»). Ниже приведена процедура, состоящая из пяти этапов исследования, которая показана на пяти примерах, приведенных в приложениях А – Е.

Этапы анализа, приведенные в таблице 1, являются общими и могут быть применены к разработке и анализу исследований ААА. Для каждого из пяти этапов, а так же для общей методологии исследования ААА в 5.2 – 5.7 даны разъяснения. Пояснения к содержанию этих этапов приведены также в примерах (см. приложения А – Е).

Таблица 1 – Основные этапы исследования ААА

Номер этапа исследования	Наименование этапа исследования
1	Определение основных целей исследования
2	Описание процесса определения значений контролируемой характеристики
3	Разработка плана выборочного контроля
4	Анализ результатов контроля
5	Подготовка заключения и рекомендаций

5.2 Основные цели исследования ААА

В методологии «Шесть Сигм» и в программах повышения качества часто используют исследование ААА. Цель исследования ААА должна быть четко установлена и согласована всеми заинтересованными сторонами. Основной целью исследования ААА является анализ возможностей системы определения значений контролируемой характеристики на основе данных проверки (контроля) по альтернативному признаку для принятия решения о приемлемости (или неприемлемости) системы. Исследование ААА позволяет определить согласованность оценок одного аттестующего, нескольких аттестующих между собой, а также их согласованность со «стандартной оценкой».

Поводом для проведения исследования ААА могут быть (перечень может быть дополнен):

a) недостаточная согласованность результатов проверки части или единицы продукции, полученных одним аттестующим в процессе различных исследований;

b) недостаточная согласованность результатов проверки части или единицы продукции различными аттестующими;

c) недостаточная согласованность результатов проверки части или единицы продукции, полученных одним аттестующим или несколькими аттестующими с установленной стандартной оценкой;

d) выполнение требований стандартов по менеджменту качества, например [5].

5.3 Описание процесса определения значений контролируемой характеристики

Настоящий стандарт посвящен процессам, в которых результатами определения значений контролируемой характеристики являются данные проверки (контроля) по альтернативному признаку.

Процесс определения значений контролируемой характеристики должен быть четко описан до проведения исследований ААА с указанием аттестующих, используемых процедур определения значений контролируемой характеристики качества, условий выполнения процедур, типа данных (упорядоченные, неупорядоченные или двоичные).

5.4 Методология исследования ААА

Для определения значений контролируемых характеристик продукции на производстве применяют приборы, шаблоны, мерительный инструмент и т.д. Существует много ситуаций, в которых характеристики качества определить достаточно трудно, например, при ранжировании показателей качества автомобилей, определении качества ткани в виде «хорошая» / «плохая», оценке цвета, аромата и вкуса вина в баллах от 1 до 10.

В случаях, когда физическое определение значений контролируемой характеристики невозможно, используют субъективные классификации или экспертные оценки. В этом случае применение исследования ААА позволяет после получения оценок несколькими аттестующими (двумя и более) провести анализ согласованности оценок участвующих в испытаниях аттестующих.

Если оценки аттестующих совпадают, возможно, что полученные оценки являются верными. Если оценки различных аттестующих не совпадают, правильность оценок может вызывать сомнения.

Полученные оценки могут быть неупорядоченными, упорядоченными или двоичными. Неупорядоченные данные представляют собой значения переменных, принимающих два или более значений без естественного упорядочивания. Например, при дегустации продуктов, значениями оценок могут быть, например, такие как хрустящий, мягкий. Упорядоченные данные представляют собой значения переменной, принимающей три или более значений с естественным упорядочиванием, например, значениями такой переменной могут быть: «категорически не согласен», «не согласен», «воздерживаюсь», «согласен» и «абсолютно согласен». Однако дистанция между значениями переменной в этом случае неизвестна. Двоичные данные представляют собой значения переменных, принимающих только два значения. Например, аттестующие могут классифицировать единицы продукции на «годные/негодные» или «функционирующие /не функционирующие». Следует отметить, что двоичные данные фактически являются частным случаем неупорядоченных данных с двумя значениями. Двоичные данные широко используют в производстве при наличии установленного требования, когда проверенные единицы продукции признают соответствующими или несоответствующими этому требованию. Двоичные данные также используют для оценки функционирования системы определения значений контролируемой характеристики. Система, когда контролируемая характеристика принимает только два значения, рассмотрена в приложении А. Таким образом, в настоящем стандарте неупорядоченные данные отнесены к переменной, принимающей три или более возможных значений.

Независимо от типа используемых данных при проведении исследования ААА в качестве показателя применяют процент совпадения оценок. Процент совпадения оценок показывает количество совпадений оценок при многократных испытаниях для одного аттестующего и нескольких аттестующих. Процент совпадения оценок P является фактически точечной оценкой доли совокупности

$$P = \frac{n_{\text{match}}}{n} \cdot 100\ \%,$$

где n_{match} – количество совпадающих оценок при n повторениях испытаний;

n – количество проверяемых объектов (количество повторений испытаний).

Для анализа неупорядоченных данных более всего подходит κ-статистика. Она представляет собой долю неслучайных совпадений оценок.

$$\kappa = \frac{P_{\text{obs}} - P_{\text{exp}}}{1 - P_{\text{exp}}},$$

где P_{obs} – наблюдаемая доля совпадающих оценок;

P_{exp} – математическое ожидание доли совпадающих оценок, если совпадения являются случайными.

Каппа принимает значения от минус 1 до +1. Чем выше значение κ, тем больше доля совпадающих оценок. Если κ принимает значение 1, оценки полностью совпадают. Если κ принимает значение 0, совпадение оценок является случайным. Значения κ выше 0,9 считают превосходными.

Значения κ менее 0,7 указывают на то, что система определения значений контролируемой характеристики нуждается в улучшении, а значение κ менее 0,4 указывает на то, что эта система является недостоверной. Обычно требуют, чтобы значение κ было не менее 0,7 ($\kappa \geq 0{,}7$).

Двумя наиболее часто применяемыми κ-статистиками являются κ-статистика Кохена и κ-статистика Флейса. При использовании этих статистик оценки аттестующих обрабатывают по-разному, и по-разному вычисляют вероятность случайного совпадения данных. При вычислении κ-статистики Кохена предполагается, что аттестующие выбраны и фиксированы, а при вычислении κ-статистики Флейса предполагается, что аттестующие выбраны случайным образом из группы аттестующих. Это дает два различных метода определения вероятности совпадения оценок. Таким образом, κ-статистика и ее стандартная ошибка σ_κ (SE[1)]) могут быть вычислены методом Флейса или методом Кохена. В основе критерия лежит статистика

$$Z = \frac{\kappa}{\sigma_\kappa},$$

используемая для проверки нулевой гипотезы H_0: $\kappa = 0$ против альтернативной гипотезы H_1: $\kappa > 0$.

Это односторонний критерий. Нулевой гипотезе соответствует нормированное нормальное распределение Z. Нулевую гипотезу отклоняют, если p-значение меньше назначенного, обычно равного 0,05.

[1)] SE – Standard Error.

Так как двоичные данные представляют собой частный случай неупорядоченных данных, когда переменная принимает только два значения, κ-статистика может также быть использована и для двоичных данных.

Значения κ-статистики не учитывают разности между значениями в упорядоченных данных. Они характеризуют лишь абсолютное совпадение оценок. Поэтому для исследования упорядоченных данных более всего подходят коэффициенты Кендалла. В настоящем стандарте использованы два вида коэффициентов Кендалла: коэффициент совпадений Кендалла *W* и коэффициент корреляции Кендалла (названный т-Кендалла). Оба эти коэффициента являются непараметрическими статистиками. Первый коэффициент изменяется от 0 до 1 и отражает степень совпадения оценок при многократном повторении испытаний. Второй коэффициент изменяется от минус 1 до + 1 и отражает степень совпадения оценок с известной стандартной оценкой. Таким образом, коэффициент совпадения Кендалла следует использовать для анализа совпадения оценок одного аттестующего и нескольких аттестующих. Если известна истинная оценка, может быть использован коэффициент корреляции Кендалла для анализа совпадающих оценок каждого аттестующего со стандартной оценкой и согласованности результатов всех аттестующих со стандартной оценкой.

5.5 План выборочного контроля для исследования ААА

В плане выборочного контроля для исследования ААА должны быть определены объем выборки исследуемых объектов, количество аттестующих и количество проверок. При проверке более 20 объектов обычно выбирают от трех до пяти аттестующих, две или три проверки (при использовании нескольких альтернативных признаков необходимо использовать большее количество выборок). Следует отметить, что отобранные выборки должны представлять весь производственный процесс. Для неупорядоченных данных метод выбора аттестующего также определяет какую κ-статистику следует применять. Если аттестующие выбраны и фиксированы, более подходящей является κ-статистика Кохена. Если аттестующие выбраны случайным образом из группы аналогичных аттестующих, более предпочтительно применять κ-статистику Флейса. Статистика Кохена основана на двухсторонней таблице сопряженности. Если стандартная оценка неизвестна, κ-статистика Кохена может быть вычислена только если данные удовлетворяют следующим условиям:

a) для анализа оценок одного аттестующего имеются результаты двух проверок аттестующего;

b) для анализа оценок нескольких аттестующих имеется два аттестующих, каждый из которых выполнил по одной проверке.

В процессе определения значений контролируемой характеристики для исследования ААА рандомизация является очень важной частью. Рандомизация означает, что аттестующие выполняют проверки исследуемых объектов в случайном порядке.

В таблице 2 показана форма представления данных для исследования ААА при участии трех аттестующих при выполнении трех проверок и контроле 20 единиц продукции каждым аттестующим.

Таблица 2 – План эксперимента исследования ААА

Номер исследуемого объекта	Аттестующий А			Аттестующий В			Аттестующий С			Стандартная оценка
	Проверка № 1	Проверка № 2	Проверка № 3	Проверка № 1	Проверка № 2	Проверка № 3	Проверка № 1	Проверка № 2	Проверка № 3	
1										
2										
3										
. . .										
20										

5.6 Анализ данных

В процессе исследования могут быть рассмотрены следующие четыре типа согласованности:

a) совпадение оценок одного аттестующего (определение совпадающих оценок среди всех оценок каждого аттестующего во всех проверках);

b) совпадение оценок нескольких аттестующих (определение совпадающих оценок среди всех оценок аттестующих во всех проверках);

c) совпадение оценок каждого аттестующего со стандартной оценкой (определение оценок, совпадающих между собой и со стандартной оценкой, среди всех оценок каждого аттестующего);

d) совпадение оценок нескольких аттестующих со стандартной оценкой (определение оценок, совпадающих друг с другом и со стандартной оценкой, среди оценок всех аттестующих во всех проверках).

Очевидно, что тип согласованности, установленный в перечислении c), является более сложным, чем установленный в перечислении a), так как требования перечисления a) дополнены требованием совпадения оценок со стандартной оценкой. Эти требования аналогичны для четвертого и второго перечислений. Очевидно, последний тип согласованности оценок является условием наименее сложным среди всех. Для каждого типа согласованности обычно применяют два типа κ-статистики: κ-статистику Кохена и κ-статистику Флейса. Кроме того, для неупорядоченных данных с тремя или более значениями контролируемой переменной могут быть вычислены два типа коэффициентов. Во-первых, может быть вычислена общая κ-статистика, характеризующая совпадение оценок по всем значениям контролируемой переменной. Во-вторых, могут быть вычислены значения κ-статистики для каждого значения контролируемой переменной отдельно. Это позволяет выявить значения контролируемой переменной, для которых совпадения оценок встречаются реже.

В дополнение к отчету об исследовании ААА используют также графическое отображение согласованности оценок. Графики позволяют наглядно отразить согласованность оценок. Обычно для процента совпадения оценок одного аттестующего и нескольких аттестующих составляют таблицы κ-коэффициентов и коэффициентов Кендалла (только для упорядоченных данных). Кроме того, составляют график процента совпадающих оценок для каждого аттестующего и со стандартной оценкой, если каждым аттестующим выполнено более одной проверки. Дополнительно может быть построен также график доли совпадающих оценок разных аттестующих по отношению к известному значению стандартной оценки.

5.7 Выводы и предложения

На основе результатов исследований ААА может быть принято решение об адекватности определения значений контролируемой характеристики на основе данных контроля по альтернативному признаку. Обычно противоречивость оценок одного аттестующего указывает на то, что аттестующий не может получить совпадающие оценки (возможно потому, что аттестующий не точно выполняет процедуру проверки). Противоречивость оценок нескольких аттестующих указывает на то, что либо не совпадают процедуры, выполняемые аттестующими в процессе проверки, либо способности аттестующих различны (вследствие различного опыта, квалификации или физических возможностей, например, остроты зрения при визуальном контроле). Должны быть предприняты действия по устранению причин таких несоответствий.

После выполнения действий по улучшению системы определения значений контролируемой характеристики, например, обучения операторов, исследование ААА должно быть проведено повторно для проверки наличия достаточности улучшений системы определения значений контролируемой характеристики.

6 Описание приложений A - E

В приложениях A – E приведено пять примеров исследования ААА, общий обзор которых приведен в таблице 3.

Таблица 3 – Обзор примеров исследования ААА, приведенных в приложениях А – Е

Обозначение приложения	Наименование примера	Примечания
Приложение А	Контроль жидкокристаллических дисплеев	Три контролера случайным образом выбраны из группы аналогичных контролеров для визуального контроля качества ЖК-дисплеев по выборкам из 20 дисплеев. Проверка проведена дважды. Результатами контроля являются двоичные данные. Для выполнения исследования ААА был использован пакет программ Minitab
Приложение В	Диагностика проблем с программным обеспечением	Проблема может быть отнесена к 6 категориям: категорию проблемы определяю 4 аттестующих, каждый проверяет 48 проблем. Для анализа использован пакет программ SAS. Истинная проблема известна
Приложение С	Определение видов воды по дегустации	Проводится дегустация четырех видов воды тремя дегустаторами по три чашки каждого вида с дегустацией двенадцати чашек воды каждым дегустатором. Для исследования ААА использован пакет программ SAS, истинный вид воды известен
Приложение D	Контроль термистора	Три контролера, выбранные случайным образом из группы контролеров, проверили 20 термисторов дважды с помощью визуального контроля. Результаты контроля представляют собой 8 категорий без естественного упорядочивания
Приложение Е	Определение уровня трудоспособности после инсульта	Исходными данными являются упорядоченные данные с пятью категориями значений. Два доктора дают оценки трудоспособности 46 пациентам, не осматривая одного и того же пациента дважды. Использован пакет программ SAS. «Истинная» трудоспособность пациента известна

Приложение А
(справочное)
Контроль жидкокристаллических дисплеев

А.1 Общие положения

При производстве жидкокристаллических дисплеев (ЖК-дисплеев) качество дисплея оценивают на основе визуального контроля. Все ЖК-дисплеи в выборке проверяют в видеографическом (VGA[1)]) режиме. Результаты наблюдений принимают два значения: нормальный цвет (соответствие) или измененный цвет (не соответствие). На этапе проверки визуальный контроль, несмотря на субъективность оценок, главным образом используют для определения соответствующих и несоответствующих дисплеев в выборке. Поэтому опыт и навыки контролеров имеют огромное значение. Цель исследования состоит в анализе непротиворечивости и правильности результатов системы определения значений контролируемой характеристики.

А.2 Наблюдаемая переменная

Наблюдаемая переменная является двоичной и принимает два значения без естественного упорядочивания.

А.3 Стандартная оценка

Стандартная оценка (истинное состояние ЖК-дисплея) в данном случае известна.

А.4 Возможные причины ошибочных решений

Причиной ошибочных решений при контроле может быть нарушение инструкции по эксплуатации дисплея. Другими причинами могут быть недостаточные опыт и обучение контролеров.

А.5 План выборочного контроля

Для анализа непротиворечивости и правильности оценок трех контролеров проведена проверка качества ЖК-дисплеев по выборке из двадцати дисплеев (модель: LCD40b66) на основе визуального контроля. Проверка проведена дважды. Жидкокристаллические дисплеи из выборки в произвольном порядке были представлены трем контролерам, которых отбирали случайным образом из группы контролеров, имеющих одинаковую квалификацию.

Результаты контроля являются двоичными.

А.6 Исходные данные

Исходные данные, используемые для исследования ААА, приведены в таблице А.1.

Таблица А.1 – Результаты контроля ЖК-дисплеев и стандартная оценка

Номер дисплея	Стандартная оценка	Результат контролера № 1		Результат контролера № 2		Результат контролера № 3	
		Проверка № 1	Проверка № 2	Проверка №1	Проверка № 2	Проверка № 1	Проверка №2
8	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0
5	0	0	0	1	1	0	0
6	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1
14	1	1	1	1	1	0	0
15	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1
19	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0
Примечание – 1 – дисплей соответствует установленным требованиям; 0 – дисплей не соответствует установленным требованиям.							

[1] VGA – video graphics array.

А.7 Исследование ААА

Исследование ААА с применением пакета программного обеспечения Minitab 15 выполняют для анализа непротиворечивости и правильности субъективных оценок одного контролера и нескольких контролеров по отношению к стандартной оценке. Выходные данные исследования ААА состоят из текстовых и графических данных.

Текстовые результаты анализа позволяют сделать выводы по последующим типам согласованности:

a) совпадению оценок одного контролера: показана непротиворечивость оценок контролера в процессе контроля;

b) совпадение оценок нескольких контролеров: показана согласованность оценок нескольких контролеров по одной и той же выборке.

Так как стандартная оценка (истинное значение наблюдаемой переменной) в данном случае известна, таблица результатов контроля позволяет получить информацию еще по двум типам согласованности:

c) совпадение оценок каждого контролера со стандартной оценкой: показана доля совпадений оценок каждого контролера в каждой выборке со стандартной оценкой (согласованность оценок одного и того же контролера со стандартной оценкой);

d) совпадение оценок всех контролеров со стандартной оценкой: показана доля совпадения оценок всех контролеров с известной стандартной оценкой.

Для каждого типа согласованности в таблице результатов контроля для анализа непротиворечивости и правильности оценок контролеров указано значение κ-статистики Флейса.

А.7.1 Согласованность оценок одного контролера

В таблице исследования ААА контролером указаны результаты контроля каждого ЖК-дисплея.

Из таблицы А.2 видно, что каждый контролер проверил 20 ЖК-дисплеев (количество проверенных объектов). Контролеры № 1, № 2, № 3 проверили 20 одних и тех же жидкокристаллических дисплеев. Определено количество совпадающих оценок и их процент. Доверительный интервал уровня 95 % для процента совпадений оценок составляет [86,09, 100].

Таблица А.2 – Процент совпадения оценок для одного контролера

Номер контролера	Количество проверенных объектов	Количество совпадений оценок	Процент совпадений	Доверительный интервал уровня 95 %
№ 1	20	20	100,00	(86,09, 100,00)
№ 2	20	20	100,00	(86,09, 100,00)
№ 3	20	20	100,00	(86,09, 100,00)

Для анализа непротиворечивости оценок каждого контролера в процессе проверки может быть использована κ-статистика.

Существует два основных вида κ-статистики: κ-статистика Кохена и κ-статистика Флейса. При использовании этих статистик оценки контролеров обрабатывают по-разному и по-разному вычисляют вероятность случайного совпадения оценок. При использовании κ-статистики Кохена предполагают, что контролеры определены и фиксированы, а при вычислении κ-статистики Флейса предполагают, что контролеры выбраны случайным образом из группы аналогичных контролеров. В данном случае три контролера выбраны случайным образом из группы контролеров. Поэтому использована κ-статистика Флейса.

Чем выше значение κ, тем сильнее согласованность оценок. Значение $\kappa = 1$ указывает на полное совпадение оценок. Значение $\kappa = -1$, указывает на полное отсутствие совпадающих оценок. Значение $\kappa = 0$ указывает на то, что совпадение оценок является чисто случайным. Значения κ-статистики выше 0,9 считают превосходными. Значение κ менее 0,7 указывает на то, что система определения значений контролируемой характеристики (или качество обслуживания) нуждается в улучшении, а значение κ менее 0,4 указывает на возможность того, что система определения значений контролируемой характеристики является неадекватной. Обычно требуют, чтобы значение κ было не менее 0,70, но предпочтительнее значения $\kappa = 0,90$.

На основе *p*-значения принимают решение о справедливости нулевой гипотезы ($\kappa = 0$, т.е. совпадения оценок одного контролера являются случайными). Если *p*-значение для тестовой

статистики меньше заданного уровня значимости (α) (обычно используют значение α=0,05), нулевую гипотезу отклоняют. Поскольку *p*-значение для κ-статистики Флейса меньше 0,05, нулевую гипотезу отклоняют. Таким образом, совпадения оценок существенно отличаются от случайных совпадений. В таблице А.3 приведены *p*-значения для значений наблюдаемой переменной и контролеров.

Таблица А.3 – Значения k-статистики Флейса для контролеров

Номер контролера	Результат контроля	κ	Стандартная ошибка, σ_κ	*Z*	*p*-значение (>0)
№1	0	1	0,223607	4,47214	0,0000
	1	1	0,223607	4,47214	0,0000
№2	0	1	0,223607	4,47214	0,0000
	1	1	0,223607	4,47214	0,0000
№3	0	1	0,223607	4,47214	0,0000
	1	1	0,223607	4,47214	0,0000
Во второй колонке использованы следующие обозначения: 1 – ЖК-дисплей соответствует установленным требованиям; 0 – ЖК-дисплей не соответствует установленным требованиям.					

А.7.2 Совпадение оценок каждого контролера со стандартной оценкой

Необходимо определить степень совпадения оценок контролера в каждой выборке со стандартной оценкой (см. таблицу А.4).

Таблица А.4 – Совпадения и несовпадения оценок каждого контролера со стандартной оценкой

Номер контролера	Количество проверенных объектов	Количество совпадений оценок контролера со стандартной оценкой		Доверительный интервал уровня 95 % для количества совпадений оценок
		шт.	%	
№ 1	20	20	100,00	(86,09, 100,00)
№ 2	20	19	95,00	(75,13, 99,87)
№ 3	20	19	95,00	(75,13, 99,87)

Номер контролера	Отношение количества оценок, совпадающих со стандартными оценками к количеству оценок, не совпадающих со стандартными оценками в процессе проверки		Отношение количества ошибочных оценок к количеству оценок, совпадающих со стандартной оценкой		Количество ошибочных оценок	
	шт.	%	шт.	%	шт.	%
№ 1	0	0,00	0	0,00	0	0,00
№ 2	1	20,00	0	0,00	0	0,00
№ 3	0	0,00	1	6,67	0	0,00
Во второй колонке использованы следующие обозначения: 1 – ЖК-дисплей соответствует установленным требованиям; 0 – ЖК-дисплей не соответствует установленным требованиям.						

Данные таблицы А.5, показывают, что для каждого контролера κ-статистика больше 0,7, т.е. оценки контролеров хорошо согласуются со стандартной оценкой.

Таблица А.5 – Значения κ-статистика Флейса (оценки контролеров по отношению к стандартной оценке)

Номер контролера	Оценка	κ	Стандартная ошибка, σ_κ	*Z*	*p*- значение (>0)
№1	0	1,00000	0,158114	6,32456	0,0000
	1	1,00000	0,158114	6,32456	0,0000
№2	0	0,85663	0,158114	5,41781	0,0000
	1	0,85663	0,158114	5,41781	0,0000
№3	0	0,87461	0,158114	5,53151	0,0000
	1	0,87461	0,158114	5,53151	0,0000

А.7.3 Совпадение оценок нескольких контролеров

Результатам, приведенным в таблице А.6, соответствует значение κ-статистики 0,858. В соответствии с таблицей А.7 совпадение оценок различных контролеров является приемлемым.

Таблица А.6 – Процент совпадения оценок нескольких различных контролеров

Количество объектов	Количество совпадения оценок всех контролеров		Доверительный интервал для процента совпадения оценок с уровнем доверия 95 %
	шт.	%	
20	18	90,00	(68,30, 98,77)

Таблица А.7 – Значения κ-статистики Флейса для нескольких контролеров

Оценка	κ	Стандартная ошибка, σ_κ	*Z*	*p*- значение (>0)
0	0,857778	0,0577350	14,8571	0,0000
1	0,857778	0,0577350	14,8571	0,0000
Во второй колонке использованы следующие обозначения: 1 – ЖК-дисплей соответствует установленным требованиям; 0 –ЖК-дисплей не соответствует установленным требованиям.				

А.7.4 Совпадение оценок всех контролеров со стандартной оценкой

Данные таблиц А.8 и А.9 показывают согласованность оценок со стандартной оценкой.

Таблица А.8 – Процент совпадения оценок контролеров со стандартной оценкой

Количество проверок	Количество совпадений оценок всех контролеров со стандартной оценкой		Доверительный интервал для процента совпадения оценок с уровнем доверия 95 %
	шт.	%	
20	18	90,00	(68,30, 98,77)

Таблица А.9 – Значения κ-статистики Флейса для всех контролеров по отношению к стандартной ошибке

Оценка	κ	Стандартная ошибка, σ_κ	*Z*	*p*- значение (>0)
0	0,910413	0,0912871	9,97307	0,0000
1	0,910413	0,0912871	9,97307	0,0000
Во второй колонке использованы следующие обозначения: 1 – ЖК-дисплей соответствует установленным требованиям; 0 – ЖК-дисплей не соответствует установленным требованиям.				

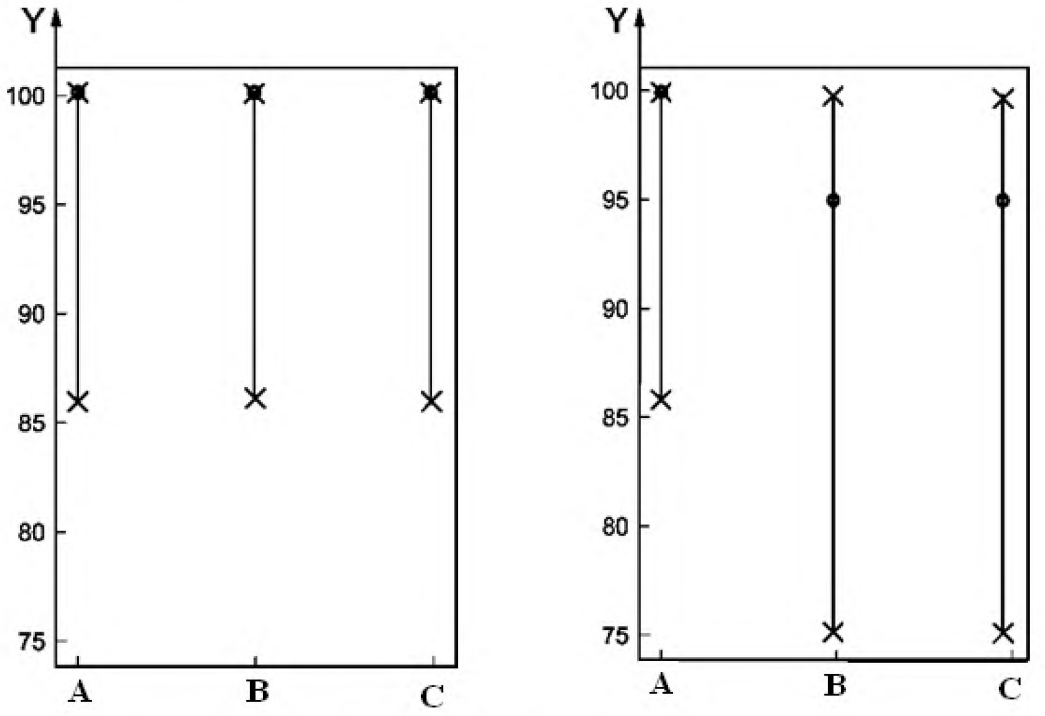
А.7.5 Графики согласованности оценок

Для наглядного представления согласованности оценок используют графическое отображение процента совпадения оценок и доверительного интервала с уровнем доверия 95 % для процента совпадения оценок для каждого контролера, и доверительного интервала с уровнем доверия 95 % для всех контролеров по отношению к стандартной оценке (см. рисунок А.1).

На рисунке А.1 a) показано совпадение оценок для каждого контролера, на рисунке А.1 b) – совпадение оценок контролера со стандартной оценкой. Черными кружочками указаны точечные оценки процента совпадения оценок, а вертикальными отрезками – соответствующий доверительный интервал уровня 95,0 %.

А.8 Выводы

Так как значения κ-статистик Флейса больше 0,7, система определения значений контролируемой характеристики при контроле по альтернативному признаку является приемлемой. Для непрерывного улучшения необходимо рассмотреть результаты контролеров № 1, № 2, № 3. Необходимо обеспечить больше внимания обучению контролеров.



a) Совпадения оценок одного контролера b) Совпадения оценок контролера со стандартной оценкой

Y – процент совпадения оценок;
A – контролер № 1;
B – контролер № 2;
C – контролер № 3.

Рисунок А.1 – Процент совпадения оценок и соответствующий доверительный интервал уровня 95 %

Приложение В
(справочное)

Диагностика проблем с программным обеспечением

В.1 Общие положения

Центру сопровождения программного обеспечения (ПО) необходимо проанализировать способность служащих первой телефонной линии правильно диагностировать проблемы с программным обеспечением, возникающие у пользователя, на основе информации, сообщаемой пользователям по телефону.

Обращение пользователя по телефону в службу сопровождения программного обеспечения сначала попадает к специалистам первой линии телефонной связи. Служащих этой линии обучают задавать вопросы и на основе ответов пользователей правильно относить проблему пользователя к одной из нескольких категорий. Затем вызов переводят на эксперта по сопровождению ПО для более точной диагностики проблемы. Если служащий первой телефонной линии неправильно классифицирует проблему пользователя, эксперт по сопровождению ПО может попытаться диагностировать проблему или вернуть пользователя обратно на первую линию. Необходимость повторного отнесения проблемы к одной из категорий требует дополнительных затрат времени, что может вызвать неудовлетворенность пользователя. Обычно телефонные переговоры по первой телефонной линии записывают для исследования и использования при обучении персонала.

Целью исследования является определение того, насколько правильно служащие первой телефонной линии классифицируют проблемы пользователей и идентифицируют их проблемы. Для служащих проводят обучение по определению категории проблемы на основе ответов пользователя на ряд стандартных вопросов.

В.2 Описание эксперимента

Эксперимент включает проверку классификации проблем пользователя четырьмя служащими первой линии, выбранными случайным образом. Все служащие первой линии имеют один и тот же уровень обучения и опыт работы в телефонном информационном центре (от 1 года до 1,5 лет). Каждый из четырех служащих первой линии прослушал 48 записей телефонных звонков потребителей с сообщением о различных проблемах. Служащим не сообщали правильное решение. Прослушивая записи телефонных сообщений пользователей, служащие должны отнести проблему пользователя (на основе их ответов на стандартные вопросы) к одной из шести категорий. Истинная категория проблемы для каждого из телефонных сообщений известна экспериментаторам.

В.3 Наблюдаемая переменная

Наблюдаемой переменной является проблема потребителя. Таким образом, исходные данные представляют собой неупорядоченные данные с шестью значениями – активация, вычисления, отображение данных, графика, электронные таблицы, Windows.

В.4 Стандартная оценка

Стандартная оценка известна и представляет собой правильное отнесение проблемы пользователя к определенной категории на основе телефонного сообщения.

В.5 Метод определения значений контролируемой характеристики

Каждый из служащих первой телефонной линии прослушал все записи телефонных переговоров один раз. Служащим не позволяли повторно воспроизводить записи (ни частично ни полностью). В течение 30 с после прослушивания записи служащий должен отнести проблему пользователя к одной из шести категорий.

В.6 Возможные причины неправильных оценок

У служащего первой телефонной линии может быть недостаточно знаний для определения категории некоторых типов проблем пользователей.

В.7 План выборочного контроля

Четверо служащих первой телефонной линии (№ 1, № 2, № 3, № 4) были случайным образом отобраны в качестве испытуемых из группы служащих, имеющих одинаковую подготовку и опыт (от 1 года до 1,5 лет).

Экспериментаторы объяснили испытуемым, что каждый из них прослушает 48 записей телефонных звонков потребителей с сообщением об определенной проблеме. После прослушивания записи каждого телефонного сообщения испытуемые должны в течение 30 с отнести указанную в телефонном сообщении проблему пользователя к одной из шести категорий. Записи телефонных звонков не повторялись. Для каждой категории проблем пользователя было записано восемь телефонных звонков.

Категории проблем пользователей, назначенные испытуемыми и истинные, приведены в таблице В.1.

В.8 Исходные данные

Результаты эксперимента приведены в таблице В.1.

Таблица В.1 – Результаты эксперимента

Номер телефонного сообщения	Истинная категория проблемы пользователя	Категория проблемы пользователя, назначенная испытуемым № 1	Категория проблемы пользователя, назначенная испытуемым № 2	Категория проблемы пользователя, назначенная испытуемым № 3	Категория проблемы пользователя, назначенная испытуемым №4
1	Активация	Windows	Активация	Активация	Активация
2	Активация	Активация	Windows	Активация	Активация
3	Графика	Графика	Windows	Windows	Графика
4	Вычисления	Вычисления	Вычисления	Вычисления	Вычисления
5	Вычисления	Вычисления	Вычисления	Вычисления	Вычисления
6	Windows	Windows	Windows	Windows	Windows
7	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы
8	Отображение данных	Электронная таблица	Отображение данных	Отображение данных	Вычисления
9	Отображение данных	Отображение данных	Отображение данных	Отображение данных	Отображение данных
10	Графика	Графика	Графика	Графика	Графика
11	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы
12	Активация	Активация	Активация	Активация	Активация
13	Графика	Графика	Графика	Графика	Графика
14	Активация	Активация	Активация	Активация	Активация
15	Электронная таблица	Электронная таблица	Электронная таблица	Электронная таблица	Электронная таблица
16	Графика	Графика	Графика	Графика	Графика
17	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы
18	Активация	Активация	Активация	Windows	Windows
19	Графика	Графика	Windows	Графика	Графика
20	Windows	Графика	Windows	Windows	Windows
21	Активация	Активация	Активация	Активация	Активация
22	Отображение данных	Windows	Отображение данных	Отображение данных	Отображение данных
23	Windows	Активация	Windows	Windows	Windows
24	Отображение данных	Отображение данных	Отображение данных	Отображение данных	Отображение данных
25	Активация	Активация	Активация	Активация	Активация
26	Графика	Графика	Графика	Графика	Графика
27	Windows	Windows	Windows	Windows	Windows
28	Вычисления	Электронные таблицы	Вычисления	Вычисления	Вычисления
29	Отображение данных	Отображение данных	Отображение данных	Отображение данных	Вычисления
30	Windows	Windows	Windows	Windows	Windows
31	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы

Окончание таблицы В.1

Номер телефонного сообщения	Истинная категория проблемы пользователя	Категория проблемы пользователя, назначенная испытуемым № 1	Категория проблемы пользователя, назначенная испытуемым № 2	Категория проблемы пользователя, назначенная испытуемым № 3	Категория проблемы пользователя, назначенная испытуемым №4
32	Windows	Windows	Windows	Windows	Windows
33	Электронные таблицы	Отображение данных	Электронные таблицы	Электронные таблицы	Электронные таблицы
34	Windows	Windows	Windows	Windows	Windows
35	Отображение данных	Электронные таблицы	Электронные таблицы	Электронные таблицы	Электронные таблицы
36	Вычисления	Отображение данных	Вычисления	Вычисления	Вычисления
37	Вычисления	Вычисления	Вычисления	Вычисления	Вычисления
38	Windows	Windows	Windows	Windows	Windows
39	Электронные таблицы	Электронные таблицы	Вычисления	Вычисления	Электронные таблицы
40	Электронные таблицы	Электронные таблицы	Электронные таблицы	Отображение данных	Электронные таблицы
41	Графика	Графика	Графика	Графика	Графика
42	Отображение данных	Электронные таблицы	Отображение данных	Отображение данных	Отображение данных
43	Вычисления	Вычисления	Вычисления	Вычисления	Вычисления
44	Активация	Активация	Активация	Активация	Активация
45	Вычисления	Отображение данных	Отображение данных	Вычисления	Вычисления
46	Вычисления	Отображение данных	Вычисления	Вычисления	Вычисления
47	Отображение данных	Отображение данных	Отображение данных	Отображение данных	Отображение данных
48	Графика	Графика	Графика	Графика	Графика

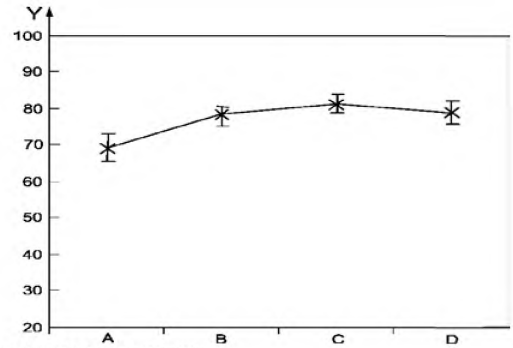
В.9 Анализ согласованности результатов одного испытуемого

В. 9.1 Общие положения

Исследование AAA проводилось на основе проблем с пакетом программ SAS для анализа согласованности и правильности субъективных решений служащих первой телефонной линии. При этом исследованы результаты каждого испытуемого, нескольких испытуемых, а также по отношению к стандартной оценке.

В.9.2 Согласованность результатов каждого испытуемого в процессе проверки

На рисунке В.1 и в таблице В.2 представлены данные совпадения результатов в процентах для каждого испытуемого.



Y – процент совпадения результатов;
A – испытуемый № 1;
B – испытуемый № 2;
C – испытуемый № 3;
D – испытуемый № 4.

Рисунок В.1 – Процент совпадения результатов для каждого испытуемого

Таблица В.2 – Процент совпадения результатов для каждого испытуемого с соответствующим доверительным интервалом

Номер испытуемого	Процент совпадения результатов, %	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %
№ 1	69,4444	66,5549	72,1887
№ 2	78,4722	76,1382	80,6360
№ 3	81,2500	79,1355	83,1957
№ 4	79,1667	76,8840	81,2788

В.9.3 Совпадение результатов каждого испытуемого по отношению к стандартной оценке

Данные таблицы В.3 позволяют сделать выводы о способности каждого испытуемого правильно относить проблему пользователя к одной из шести указанных категорий.

Таблица В.3 – Количество совпадений результатов каждого испытуемого

Номер испытуемого	Результативность	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %
№ 1	75,0000	61,2156	85,0794
№ 2	85,4167	72,8328	92,7518
№ 3	89,5833	77,8326	95,4678
№ 4	89,5833	77,8326	95,4678

Значения κ-статистики для совпадения результатов каждого испытуемого со стандартной оценкой приведены в таблице В.4. Следует отметить, что все испытуемые достаточно правильно определяют категорию проблемы по телефонному сообщению. Кроме испытуемого № 1.

Таблица В.4 – Значения κ-статистики каждого испытуемого по отношению к стандартной оценке

Номер испытуемого	Сопоставление оценки испытуемого со стандартной оценкой	κ	Стандартная ошибка, σ_κ
№ 1	верно	0,7000	0,0744
№ 2	верно	0,8250	0,0608
№ 3	верно	0,8750	0,0528
№ 4	верно	0,8750	0,0527

В.10 Анализ согласованности решений нескольких испытуемых

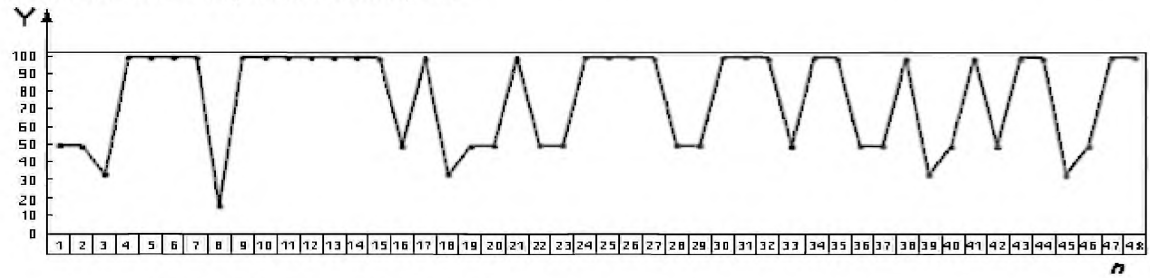
В.10.1 Согласованность результатов нескольких испытуемых (без сопоставления со стандартной оценкой)

В соответствии с данными таблицы В.5 приблизительно 58 % проблем пользователей, указанных в телефонных сообщениях, были правильно отнесены к соответствующей категории всеми испытуемыми.

Таблица В.5 – Согласованность результатов каждого испытуемого

Количество проверок	Количество совпадений	Процент совпадений, %	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %
48	28	58,333	44,281	71,150

На рисунке В.2 показаны совпадения результатов нескольких испытуемых в процентах для каждой записи телефонного сообщения.



Y – процент совпадения результатов;

n – номер записи телефонного сообщения

Рисунок В.2 – Совпадения результатов, полученных несколькими испытуемыми в процентах

В таблице В.6 приведены значения κ-статистики для каждой пары испытуемых. У испытуемого № 1 наименьшее количество совпадений результатов с результатами других испытуемых. У других испытуемых совпадений значительно больше.

Таблица В.6 – Значение κ-статистики для каждого испытуемого относительно другого испытуемого

Совпадение результатов		κ	Стандартная ошибка, σ_κ
испытуемого	с испытуемым		
№1	№2	0,6264	0,0788
№1	№3	0,6266	0,0790
№1	№4	0,6502	0,0772
№2	№3	0,8491	0,0575
№2	№4	0,7493	0,0702
№3	№4	0,8498	0,0572

В таблице В.7 приведены значения κ-статистики по категориям проблем пользователей. Категориям «Активация» и «Графические проблемы» соответствует наибольшее количество совпадений. Наименьшее количество совпадений соответствует проблемам представления данных.

Таблица В.7 – Значения κ-статистики по категориям проблем

Категория проблем пользователей	κ	Стандартная ошибка, σ_κ
Активация	0,8240	0,0589
Вычисление	0,6281	0,0589
Отображение данных	0,5786	0,0589
Графика	0,8240	0,0589
Электронные таблицы	0,7787	0,0589
Windows	0,7157	0,0589
Все	0,7244	0,0265

В. 10.2 Согласованность результатов испытуемых со стандартной оценкой

В таблице В.8 приведено количество ошибок испытуемых при определении категории проблемы пользователя. Приведенные результаты показывают, что проблемы отображения данных часто принимают за проблемы электронных таблиц. Проблемы вычислений часто принимают за проблемы отображения данных. Кроме того, проблемы активации иногда неправильно относят к проблемам Windows.

Таблица В.8 – Количество ошибок при классификации проблем пользователей

Истинная категория проблемы	Ошибочная категория					
	Активация	Вычисление	Отображение данных	Графика	Электронные таблицы	Windows
Активация	—	0	0	0	0	1
Вычисление	0	—	2	1	2	0
Отображение данных	0	5	—	0	2	0
Графика	0	0	0	—	0	1
Электронные таблицы	0	1	6	0	—	0
Windows	4	0	1	3	0	—
Другая	0	0	0	0	0	0

В.11 Заключение

Исследование ААА показывает, что возможно некоторое специальное обучение может помочь служащим первой телефонной линии улучшить диагностику проблем пользователя на основе информации, сообщаемой пользователем по телефону. Обучение помогает служащим дифференцировать проблемы пользователя между вычислением, отображением данных и проблемами электронных таблиц. Также полезно обучать служащих первой телефонной линии распознаванию проблем Windows и активации. В целом служащие первой телефонной линии должны адекватно распознавать проблемы потребителя по информации телефонного сообщения.

Приложение С
(справочное)

Определение вида воды на основе дегустации

С.1 Общие положения

Обучающая организация для демонстрации преимуществ исследования ААА проводит эксперимент по дегустации различных видов бутилированной воды, предполагая, что человек способен обнаружить различия воды по вкусу и запаху.

Цель исследования состоит в обучении студентов:

a) выявлению необходимости применения исследования ААА;
b) планированию эксперимента;
c) сбору данных;
d) анализу и интерпретации результатов исследования ААА;
e) разработке рекомендаций, основанных на информации, полученной из анализа.

Студентам был задан вопрос: «Могут ли люди обнаружить различия во вкусе нескольких видов бутилированной воды и воды из водопровода?» Чтобы ответить на этот вопрос, необходимо проверить способность людей различать вкус воды из этих источников.

С.2 Описание эксперимента

Эксперимент состоит из выбора трех различных видов бутилированной воды (видов А, В и С) и заполнения немаркированных чашек этой водой и водопроводной водой. Затем дегустаторы пьют воду из этих чашек и записывают вид воды (А, В, С или водопроводная вода) полагаясь на свой вкус.

Целью анализа является определение количества дегустаторов, способных распознать по вкусу различные виды бутилированной воды или водопроводной воды.

С.3 Наблюдаемая переменная

Наблюдаемой переменной является вид воды, определенный дегустатором. В данном эксперименте вид воды представляет собой неупорядоченные данные с четырьмя значениями А, В, С и «водопроводная вода»).

С.4 Стандартная оценка

Стандартная оценка известна и является фактическим видом воды в чашке.

С.5 Метод определения значений контролируемой характеристики

В начале эксперимента каждому дегустатору дают маркированные образцы каждого из четырех видов воды. Таким образом, дегустаторы обучаются распознавать каждый вид воды. В любое время в процессе эксперимента дегустатор может воспользоваться маркированными образцами и повторно их попробовать, прежде чем записать результат дегустации воды в исследуемой чашке.

С.6 Возможные причины ошибочных оценок

Появление ошибочных решений может быть вызвано недостаточными различиями во вкусе или запахе воды различных видов, что не позволяет человеку правильно идентифицировать источник воды, даже при наличии идентифицированных образцов воды.

С.7 План выборочного контроля

Из совокупности студентов, случайным образом отобраны три студента в качестве дегустаторов для участия в эксперименте.

После обучения с помощью дегустации маркированных образцов воды дегустаторам выдано 12 чашек воды. Им необходимо указать вид воды в каждой чашке, основываясь на вкусе и запахе воды. Каждый из четырех видов воды был представлен в исследовании трижды. Таким образом, каждый участник продегустировал 12 чашек воды

Результаты дегустации воды каждым из трех дегустаторов приведены в таблице С.1.

С.8 Исходные данные

Результат эксперимента, включающего дегустацию трех образцов четырех видов воды каждым дегустатором, приведены в таблице С.1.

Таблица С.1 – Результаты эксперимента

Номер чашки с водой	Истинный вид воды в чашке	Дегустатор № 1	Дегустатор № 2	Дегустатор № 3	Процент правильных решений, %
1	Вода вида А	Вода вида А	Водопроводная вода	Вода вида В	33
2	Водопроводная вода	Вода вида С	Вода вида А	Водопроводная вода	33
3	Вода вида В	Вода вида В	Вода вида В	Вода вида А	67
4	Вода вида С	Вода вида В	Вода вида В	Вода вида А	0
5	Вода вида В	Вода из-под крана	Вода из-под крана	Вода вида С	0
6	Водопроводная вода	Вода вида С	Вода вида С	Водопроводная вода	33
7	Вода вида А	Вода вида В	Вода вида В	Вода вида С	0
8	Вода вида С	Водопроводная вода	Вода вида А	Вода вида В	0
9	Водопроводная вода	Водопроводная вода	Водопроводная вода	Вода вида С	67
10	Вода вида А	Вода вида А	Вода вида А	Вода вида А	100
11	Вода вида В	Вода вида А	Вода вида С	Вода вида С	0
12	Вода вида С	Вода вида В	Вода вида В	Вода вида С	33

С.9 Анализ совпадения оценок одного дегустатора

С.9.1 Общие положения

Исследование AAA с использованием SAS проводилось для анализа непротиворечивости и правильности оценок дегустаторов. При этом исследованы результаты каждого дегустатора, нескольких дегустаторов и по отношению к стандартной оценке.

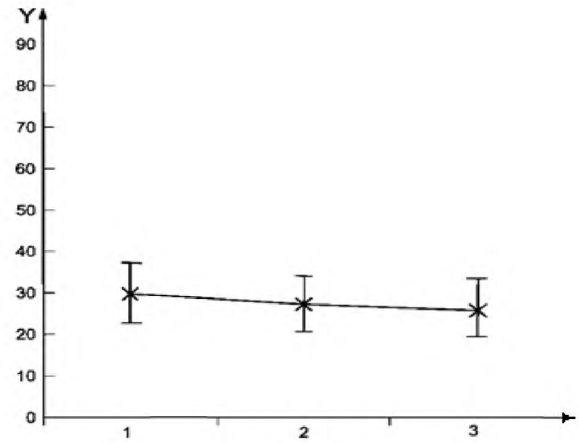
С.9.2 Непротиворечивость оценок одного дегустатора

Данные таблицы С.2 показывают отсутствие уверенного распознавания дегустаторами каждого из четырех видов воды.

Таблица С.2 – Процент совпадения оценок для каждого дегустатора и соответствующий доверительный интервал уровня доверия 95 %

Номер дегустатора	Количество проверок	Количество проверок, оценки в которых совпадают с истинным видом воды	Баллы дегустатора	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %
№ 1	4	0	0,0000	0,0000	48,9891
№ 2	4	0	0,0000	0,0000	48,9891
№ 3	4	0	0,0000	0,0000	48,9891

На рисунке С.1 и в таблице С.3 представлены данные о проценте совпадений для каждого дегустатора. Ни один из дегустаторов не смог безошибочно идентифицировать ни один из видов воды.



Y – процент совпадений оценок дегустации;
1 – дегустатор № 1;
2 – дегустатор № 2;
3 – дегустатор № 3.

Рисунок С.1 – Процент совпадений для каждого дегустатора

Таблица С.3 – Процент совпадений для каждого дегустатора со стандартной оценкой и соответствующий доверительный интервал уровня 95 %.

Номер дегустатора	Процент совпадений оценки дегустатора со стандартной оценкой, %	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %
№ 1	29,7619	23,3337	37,1040
№ 2	27,3810	21,6493	33,9722
№ 3	26,1905	20,6194	32,6477

С.9.3 Результативность каждого дегустатора в процессе испытаний (по сравнению со стандартной оценкой дегустатора)

Способность к распознаванию видов воды отражают данные таблицы С.4.

Таблица С.4 – Результативность каждого дегустатора

Номер дегустатора	Результативность дегустатора	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %	Стандартная ошибка, σ_κ
№ 1	33,3333	13,8120	60,9378	0,6667
№ 2	25,0000	8,8942	53,2305	0,7500
№ 3	33,3333	13,8120	60,9378	0,6667

В таблице С.5 приведены значения κ-статистики для сопоставления оценок каждого дегустатора со стандартной оценкой. Ни один из дегустаторов не продемонстрировал устойчивого распознавания видов воды по вкусу и запаху.

Таблица С.5 – Значения κ -статистики для каждого дегустатора при сопоставлении со стандартной оценкой

Номер дегустатора	Стандартная оценка	κ	Стандартная ошибка, σ_κ
№ 1	фактический	0,1111	0,1775
№ 2	Фактический	0,0000	0,1620
№ 3	фактический	0,1111	0,1848

С.10 Анализ совпадения оценок нескольких дегустаторов

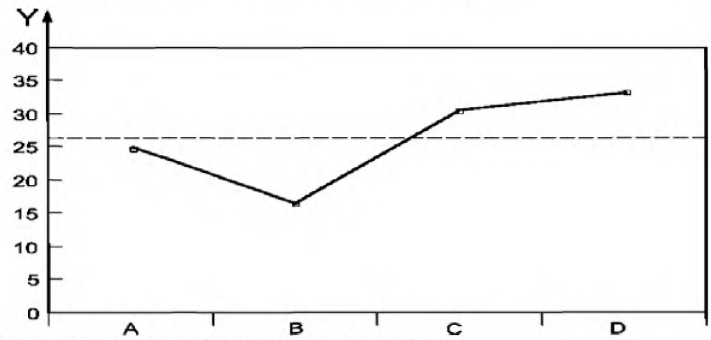
С.10.1 Совпадение оценок нескольких дегустаторов (без учета стандартной оценки)

Данные таблицы С.6 указывают на отсутствие четкого распознавания ни одним из дегустаторов ни одного из четырех видов воды. Это совпадает с данными таблицы С.2.

Таблица С.6 – Совпадение результатов нескольких дегустаторов

Количество проверок	Количество совпадений оценок дегустаторов	Процент совпадений оценок дегустаторов	Нижняя граница доверительного интервала с уровнем доверия 95 %	Верхняя граница доверительного интервала с уровнем доверия 95 %
4	0	0,0000	0,0000	48,989

На рисунке С.2 показан процент совпадения оценок для каждого дегустатора при идентификации исследуемых видов воды. Самый высокий процент совпадения оценок соответствует водопроводной воде, но даже он является достаточно низким (менее 35 %). Средний процент совпадения результатов для всех видов воды составляет 26 %.



Y – процент совпадения оценок дегустаторов;
A – вода вида A;
B – вода вида B;
C – вода вида C;
D – водопроводная вода

Рисунок С.2 – Процент совпадения оценок дегустаторов

В таблице С.7 приведены значения κ-статистик для пар дегустаторов. Оценки дегустаторов № 1 и № 2 совпадают для восьми чашек воды из 12, что дает значение κ-статистики 0,5472. Это лучше чем κ-статистика для дегустатора № 3, однако достаточно плохо.

Таблица С.7 – Значения κ-статистики для каждого дегустатора по отношению друг к другу

Совпадение оценок		κ	Стандартная ошибка, σ_κ
дегустатора	с дегустатором		
№ 1	№ 2	0,5472	0,1770
№ 2	№ 3	–0,189	0,1115
№ 3	№ 1	–0,081	0,1314

В таблице С.8 приведены значения κ-статистики для анализа совпадения оценок дегустаторов в зависимости от вида воды. У всех дегустаторов отсутствует четкое распознавание хотя бы одного вида воды.

Таблица С.8 – Значения κ-статистики для каждого дегустатора по отношению друг к другу

Вид воды	κ	Стандартная ошибка, σ_κ
A	–0,037	0,0833
B	0,1000	0,0833
C	0,0000	0,0833
Вода из-под крана	–0,004	0,0833
Все виды воды	0,0165	0,0482

С.10.2 Совпадение оценок дегустаторов со стандартной оценкой

В таблице С.9 приведено количество ошибок дегустаторов при определении вида воды в процессе испытаний. Воду вида А и водопроводную воду дегустаторы распознают лучше чем воду вида В и вида С, однако для всех видов воды характерен высокий уровень ошибочных оценок дегустаторов.

Таблица С.9 – Ошибки дегустаторов при распознавании вида воды

Вид воды	Количество ошибочных решений о виде воды			
	А	В	С	Вода из-под крана
А	—	2	2	1
В	3	—	5	0
С	1	3	—	4
Водопроводная вода	1	2	1	—
Другой	0	0	0	0

С.11 Выводы

Исследование ААА показывает, что дегустаторы не в состоянии четко распознавать по вкусу и запаху водопроводную воду и воду одного из видов воды в бутылках (А, В, С). Этот вывод помогает получить ответ на вопрос, поставленный в исследовании ААА, а именно, могут ли люди обнаруживать различие вкуса и запаха воды в бутылках и водопроводной воды. Если виды воды в бутылках, выбранной для эксперимента, являются «представительными» для всей совокупности воды в бутылках, то маловероятно, что люди могут отличить воду в бутылках от водопроводной воды.

Приложение D
(справочное)

Контроль дефектов термистора

D.1 Общие положения

Для защиты термистора[1)] от воздействия окружающей среды (температуры, влажности, химических воздействий, коррозии) контакты микросхемы термистора часто покрывают стекловидным защитным покрытием. При инкапсуляции контактов возможно семь категорий дефектов: покрытие не попало на контакты (LO[2)]); наличие в покрытии трещин (GC[3)]); использование слишком высокой температуры (HT[4)]); микросхема перевернута (CT[5)]); наличие воздушных пузырей в покрытии (Bub[6)]); неполная изоляция (PS[7)]); наличие загрязнений в покрытии (Con[8)]). На этапе контроля контролеры осматривают микросхемы термистора и субъективно относят выявленные дефекты к одному из перечисленных видов. Поэтому опыт и обучение контролера имеют большое значение. Целью исследования является анализ непротиворечивости и правильности системы определения значений контролируемой характеристики.

D.2 Наблюдаемая переменная

Наблюдаемая переменная представляет собой неупорядоченные данные, которые могут быть отнесены к восьми категориям и не имеют естественного упорядочивания.

D.3 Стандартная оценка

В данном случае стандартная оценка (истинное состояние термистора) известна.

D.4 Возможные причины ошибочных оценок

Поскольку выявление дефекта основано на визуальном контроле, иногда контролеры путают близкие виды дефектов.

D.5 План выборочного контроля

Для анализа непротиворечивости и правильности системы контроля из группы контролеров с одинаковым опытом работы случайным образом отбирают трех контролеров и предлагают им проверить качество термисторов на двух сериях из 20 термисторов каждая. Выборки термисторов отобраны случайным образом.

D.6 Исходные данные

В таблице D.1 приведены исходные данные, используемые в исследовании AAA.

D.7 Анализ совпадения оценок

Исследование AAA проводят для анализа непротиворечивости и правильности субъективных оценок каждого контролера по отношению к другим контролерам и стандартной оценке. Результаты исследования AAA с применением программного средства Minitab 15 представлены в виде табличных и графических результатов.

В виде таблицы приведены результаты проверки согласованности оценок:

a) для каждого контролера: показано совпадение оценок одного контролера при повторении выборки;

b) нескольких контролеров между собой: показано совпадение оценок контролеров при проверке одной и той же выборки.

[1)] Термистор – Терморезистор с отрицательным температурным коэффициентом сопротивления.

[2)] LO – lead outside	[3)] GC – glass crack	[4)] HT – high temperature	[5)] CT – chip turnover	[6)] Bub – bubble inside glass	[7)] PS – incomplete sealing.	[8)] Con – contamination inside unit

Таблица D.1 – Результаты контроля термисторов

Номер термистора	Стандартная оценка	Контролер А		Контролер В		Контролер С	
		Проверка № 1	Проверка №2	Проверка № 1	Проверка №2	Проверка № 1	Проверка №2
1	G	G	G	G	G	G	G
2	G	G	G	G	G	G	G
3	PS	PS	PS	PS	PS	Con	PS
4	GC	GC	GC	GC	GC	GC	GC
5	G	G	G	G	G	G	G
6	HT	GC	GC	HT	HT	HT	HT
7	G	G	G	G	G	G	G
8	GC	GC	GC	GC	HT	GC	GC
9	G	G	G	G	G	G	G
10	CT	CT	CT	G	CT	CT	CT
11	G	G	G	G	G	G	G
12	LO	LO	LO	LO	G	LO	LO
13	G	G	G	G	G	G	G
14	Con	Con	Con	Con	G	Con	PS
15	G	G	G	G	G	G	G
16	G	G	G	G	G	G	G
17	GC	GC	GC	G	GC	GC	GC
18	PS	PS	PS	PS	PS	PS	PS
19	G	G	G	G	G	G	G
20	Bub	Bub	Pass	Bub	Bub	Bub	Bub
G – отсутствие дефектов.							

Поскольку стандартная оценка (истинное состояние термистора) в данном случае известна, программа позволяет проверить два дополнительных совпадения:

c) оценок каждого контролера со стандартной оценкой: показано совпадение оценок каждого контролера для каждой выборки со стандартной оценкой;

d) оценок всех контролеров со стандартной оценкой: показано совпадение оценок всех контролеров со стандартной оценкой.

Для каждого типа совпадения оценок приведено значение κ-статистики Флейса.

D.7.1 Совпадение оценок одного контролера

На основе оценок одного контролера может быть проанализировано их совпадение при повторной проверке.

Каждый контролер проверяет 20 термисторов дважды. В таблице D.2 показано совпадение результатов контроля в двух испытаниях для каждого контролера. У контролера № 1 результаты контроля 19 из 20 проверок термисторов совпали (95 %). У контролера № 2 совпало 15 результатов контроля при проверке 20 термисторов (75 %), а у контролера № 3 совпало 17 результатов контроля из 20 (85 %).

Доверительные интервалы с уровнем доверия 95 % для процента совпадений результатов контроля имеют вид: (75,13 %, 99,87 %), (50,90 %, 91,34 %) и (62,11 %, 96,79 %) для контролеров № 1, № 2, и № 3 соответственно.

Таблица D.2 – Процент совпадения результатов контроля для одного контролера

Номер контролера	Количество проверенных термисторов	Количество совпадений результатов контроля при повторной проверке	Процент совпадений результатов контроля	Доверительный интервал с уровнем доверия 95 %
№ 1	20	19	95,00	(75,13, 99,87)
№ 2	20	15	75,00	(50,90, 91,34)
№ 3	20	17	85,00	(62,11, 96,79)

Для анализа совпадений результатов контроля для каждого контролера также может быть

использована κ-статистика.

Существует два основных вида κ-статистики: κ-статистика Кохена и κ-статистика Флейса. При этих двух подходах результаты двух контролеров обрабатывают по-разному, вычисляя вероятность случайного совпадения результатов. При вычислении κ-статистики Кохена предполагают, что контролер(ы) выбраны и фиксированы, тогда как при вычислении κ-статистики Флейса предполагают, что контролеры выбраны случайным образом из группы аналогичных контролеров. Это приводит к двум различным методам определения вероятности. В рассматриваемом случае три контролера были выбраны случайным образом из группы аналогичных контролеров. Таким образом, использование κ статистики Кохена неуместно. Далее рассмотрена только κ-статистика Флейса.

Чем выше значение κ, тем больше совпадений оценок имеется у данного контролера. Значение $\kappa = 1$ указывает на полное совпадение результатов контроля. Если $\kappa = -1$, это указывает на полную противоречивость результатов контроля. Значение $\kappa = 0$ указывает на то, что совпадение результатов контроля носит случайный характер. Значения $\kappa > 0{,}9$ считаются превосходными. Значение $\kappa < 0{,}7$ указывает на то, что система контроля требует улучшения, а значение $\kappa < 0{,}4$ указывает на то, что система контроля, возможно, является несоответствующей. Обычно требуется значение $\kappa > 0{,}70$, но предпочтительно значение κ близкое к 0,90.

В соответствии с таблицей D.3 κ-статистика Флейса для контролера № 1 имеет значение 0,92495, которое является хорошим; для контролера № 2 κ-статистика равна 0,59, что является неприемлемым. Для контролера № 3 κ-статистика принимает значение 0,79, что является приемлемым.

В таблице D.3 приведены значения κ-статистики для каждого контролера. Например, $\kappa = -0{,}02$ для контролера № 1 и дефекта «наличие воздушных пузырей». Это значение указывает на то, что результаты контролера № 1 были противоречивыми в процессе испытаний по этим категориям дефектов. Дальнейший анализ показывает, что контролер № 1 по одному и тому же объекту в двух выборках принимал решение «наличие пузырей» и «отсутствие дефектов». Это означает, что контролер № 1 может путать категории «наличие воздушных пузырей» и «отсутствие дефектов». Такая информация полезна для анализа улучшения системы определения значений контролируемой характеристики.

На основе *p*-значения принимают решение о справедливости нулевой гипотезы ($\kappa = 0$, т.е. совпадение результатов одного контролера является случайным), если она (нулевая гипотеза) верна. Если *p*-значение тестовой статистики меньше заданного уровня значимости α (обычно $\alpha = 0{,}05$), нулевую гипотезу отклоняют. Поскольку *p*-значение для трех общих κ-статистик Флейса меньше 0,05, нулевую гипотезу отклоняют. Таким образом, несовпадения результатов контроля со стандартной оценкой являются чисто случайными. В таблице D.3 показаны *p*-значения, соответствующие категориям дефектов и контролерам.

D.7.2 Совпадение результатов каждого контролера со стандартной оценкой

Необходимо определить насколько результаты каждого контролера в каждой выборке соответствуют стандартной оценке.

В таблице D.4 для каждого контролера указан процент совпадения результатов со стандартной оценкой и соответствующий доверительный интервал с уровнем доверия 95 %.

Контролерам № 1, № 2 и № 3 соответствуют проценты совпадения результатов контроля со стандартной оценкой 90 %, 75 %, и 85 % соответственно.

Таблица D.3 – Значения κ-статистики Флейса для контролеров

Контролер	Решение контролера о виде дефекта	κ	Стандартная ошибка, σ_κ	Z	p-значение
1	Bub	–0,02564	0,223607	–0,11467	0,5456
	Con	1,00000	0,223607	4,47214	0,0000
	CT	1,00000	0,223607	4,47214	0,0000
	GC	1,00000	0,223607	4,47214	0,0000
	HT	–[a]	–[a]	–[a]	–[a]
	LO	1,00000	0,223607	4,47214	0,0000
	G	0,89975	0,223607	4,02380	0,0000
	PS	1,00000	0,223607	4,47214	0,0000
	По всем дефектам	0,92495	0,124203	7,44712	0,0000
2	Bub	1,00000	0,223607	4,47214	0,0000
	Con	–0,02564	0,223607	–0,11467	0,5456
	CT	–0,02564	0,223607	–0,11467	0,5456
	GC	0,44444	0,223607	1,98762	0,0234
	HT	0,63964	0,223607	2,86056	0,0021
	LO	–0,02564	0,223607	–0,11467	0,5456
	G	0,58333	0,223607	2,60875	0,0045
	PS	1,00000	0,223607	4,47214	0,0000
	По всем дефектам	0,59016	0,118732	4,97054	0,0000
3	Bub	1,00000	0,223607	4,47214	0,0000
	Con	–0,05263	0,223607	–0,23538	0,5930
	CT	1,00000	0,223607	4,47214	0,0000
	GC	1,00000	0,223607	4,47214	0,0000
	HT	1,00000	0,223607	4,47214	0,0000
	LO	1,00000	0,223607	4,47214	0,0000
	G	0,89975	0,223607	4,02380	0,0000
	PS	0,31429	0,223607	1,40553	0,0799
	По всем дефектам	0,79275	0,109805	7,21958	0,0000

[a] Значения не могут быть определены.

Таблица D.4 – Процент совпадения результатов контроля со стандартной оценкой для каждого контролера

Номер контролера	Количество проверенных объектов	Количество результатов, совпадающих со стандартной оценкой	Процент результатов, совпадающих со стандартной оценкой	Доверительный интервал с уровнем доверия 95 %
№ 1	20	18	90,00	(68,30, 98,77)
№ 2	20	15	75,00	(50,90, 91,34)
№ 3	20	17	85,00	(62,11, 96,79)

В таблице D.5 приведены данные о соответствии результатов контроля для каждого контролера по категориям дефектов стандартной оценке. Эта информация помогает выявить, при определении какой категории дефекта контролер чаще ошибается. Например, контролеру № 1 соответствует высокое значение общей κ-статистики; однако он испытывает трудности при выявлении дефектной категории «наличие воздушных пузырей в покрытии».

Таблица D.5 – Значения κ-статистики Флейса (сопоставление результатов контроля со стандартной оценкой для каждого контролера)

Контролер	Решение контролера о виде дефекта	κ	Стандартная ошибка, σ_κ	Z	*p*-значение (>0)
1	Bub	0,48718	0,158114	3,0812	0,0010
	Con	1,00000	0,158114	6,23246	0,0000
	CT	1,00000	0,158114	6,23246	0,0000
	GC	0,82684	0,158114	5,2294	0,0000
	HT	–0,02564	0,158114	–0,1622	0,5644
	LO	1,00000	0,158114	6,23246	0,0000
	G	0,94987	0,158114	6,0075	0,0000
	PS	1,00000	0,158114	6,3246	0,0000
	По всем дефектам	0,89015	0,082380	10,8053	0,0000
2	Bub	1,00000	0,158114	6,3246	0,0000
	Con	0,48718	0,158114	3,0812	0,0010
	CT	0,48718	0,158114	3,0812	0,0010
	GC	0,77143	0,158114	4,8789	0,0000
	HT	0,81982	0,158114	5,1850	0,0000
	LO	0,48718	0,158114	3,0812	0,0010
	G	0,79798	0,158114	5,0469	0,0000
	PS	1,00000	0,158114	6,3246	0,0000
	По всем дефектам	0,81075	0,081050	10,0030	0,0000
3	Bub	1,00000	0,158114	6,3246	0,0000
	Con	0,30700	0,158114	1,9416	0,0261
	CT	1,00000	0,158114	6,3246	0,0000
	GC	1,00000	0,158114	6,3246	0,0000
	HT	1,00000	0,158114	6,3246	0,0000
	LO	1,00000	0,158114	6,3246	0,0000
	G	0,94987	0,158114	6,0075	0,0000
	PS	0,60794	0,158114	3,8449	0,0001
	По всем дефектам	0,89550	0,078382	11,4249	0,0000

D.7.3 Совпадение результатов контроля для нескольких контролеров

В таблице D.6 приведены данные об общем количестве совпадения результатов контроля для всех контролеров (55 %).

Таблица D.6 – Процент совпадения результатов контроля для всех контролеров

Количество проверенных объектов	Количество совпадений результатов контроля для контролеров	Процент количества совпадений результатов контроля для контролеров	Доверительный интервал с уровнем доверия 95 %
20	11	55,00	(31,53, 76,94)

В таблице D.7 приведены данные совпадения результатов контроля для всех контролеров по категориям дефектов. Общее значение κ составило 0,74. Результаты показывают, что необходимо улучшение по некоторым категориям дефектов таким как «Con» и «HT».

Таблица D.7 – Значения κ-статистики Флейса для всех контролеров по категориям дефектов

Решение контролера о виде дефекта	κ	Стандартная ошибка, σ_κ	*Z*	*p*-значение (>0)
Bub	0,791304	0,0577350	13,7058	0,0000
Con	0,457391	0,0577350	7,9222	0,0000
CT	0,791304	0,0577350	13,7058	0,0000
GC	0,764706	0,0577350	13,2451	0,0000
HT	0,457391	0,0577350	7,9222	0,0000
LO	0,791304	0,0577350	13,7058	0,0000
G	0,799107	0,0577350	13,8409	0,0000
PS	0,741193	0,0577350	12,8378	0,0000
Для всех видов дефектов	0,742308	0,0298699	24,8513	0,0000

D.7.4 Совпадение результатов контроля для всех контролеров со стандартной оценкой

В таблице D.8 приведены данные совпадения результатов контроля для всех контролеров. Общее значение κ (см. таблицу D.9) показывает, что система определения значений контролируемой характеристики является приемлемой; однако она может быть еще улучшена.

Таблица D.8 – Процент совпадения результатов контроля для всех контролеров со стандартной оценкой

Количество проверенных объектов	Количество совпадений результатов контроля для всех контролеров	Процент количества совпадений результатов контроля для всех контролеров	Доверительный интервал с уровнем доверия 95 %
20	11	55,00	(31,53, 76,94)

Таблица D.9 – Значения κ-статистики Флейса для совпадения результатов контроля для всех контролеров со стандартом

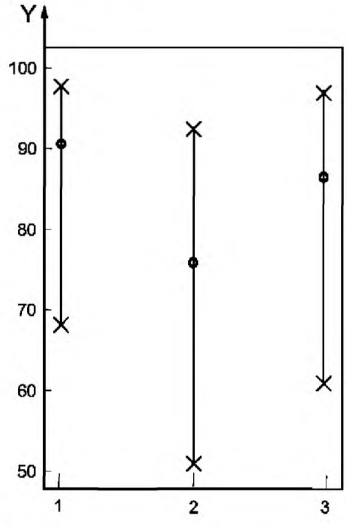
Решение контролера о виде дефекта	κ	Стандартная ошибка, σ_κ	*Z*	*p*-значение (>0)
Bub	0,829060	0,0912871	9,0819	0,0000
Con	0,598060	0,0912871	6,5514	0,0000
CT	0,829060	0,0912871	9,0819	0,0000
GC	0,866089	0,0912871	9,4875	0,0000
HT	0,598060	0,0912871	6,5514	0,0000
LO	0,829060	0,0912871	9,0819	0,0000
G	0,899243	0,0912871	9,8507	0,0000
PS	0,869312	0,0912871	9,5228	0,0000
Для всех видов дефектов	0,865467	0,0465467	18,5935	0,0000

D.7.5 Графики совпадения результатов контроля

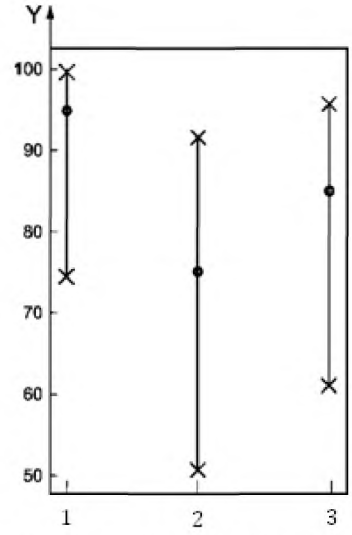
По результатам исследования AAA могут быть построены два графика: процент совпадения результатов контроля и соответствующий доверительный интервал с уровнем доверия 95 % для каждого контролера (слева, на рисунке D.1); процент совпадения результатов контроля и соответствующий доверительный интервал с уровнем доверия 95 % для всех контролеров со стандартной оценкой (справа, на рисунке D.1).

Данные рисунка D.1 a) показывают непротиворечивость результатов контроля для каждого контролера, а рисунка D.1 b) – непротиворечивость и правильность. Черные кружки на графиках указывают соответствующий процент, вертикальные линии, проходящие через них, показывают доверительный интервал с уровнем доверия 95,0 %.

Следует отметить, что процент совпадения результатов контроля изменился с 95 % на 90 % (см. рисунки D.1 a), b).



a) Совпадение результатов контроля одного контролера



b) Совпадение результатов контроля одного контролера со стандартной оценкой

Y – процент совпадения результатов;
1 – контролер № 1;
2 – контролер № 2;
3 – контролер № 3.

Рисунок D.1 – Процент совпадения результатов контроля и доверительный интервал с уровнем доверия 95 %

D.8 Выводы

Существует несколько направлений улучшения системы определения значений контролируемой характеристики:

a) Контролеру № 2 соответствует низкое значение κ при сопоставлении его результатов контроля между собой. Это указывает на то, что контролеру № 2 требуется дополнительное обучение.

b) Все контролеры испытывают затруднения с определением некоторых видов дефектов. Это видно при сопоставлении их результатов контроля между собой или со стандартной оценкой. Рекомендуется проведение тренингов по определению видов дефектов.

c) Вся система (в том числе определение дефектов, процедуры контроля, обучение) требует анализа для дальнейшего улучшения.

Приложение Е
(справочное)
Определение уровня трудоспособности после инсульта[1)]

Е.1 Общие положения

Модифицированная шкала Ранкина (МШР) является методом количественных измерений для определения степени трудоспособности после острого инсульта. Метод основан на интервью (при личной встрече или по телефону), что исключает возможность его использования в ретроспективных исследованиях.

Целью исследования является анализ непротиворечивости и правильности определения значений контролируемой характеристики по модифицированной шкале Ранкина, полученной исключительно на основе данных истории болезни (без интервью), по сравнению с традиционным использованием метода с применением беседы с пациентом.

Е.2 Описание эксперимента

В результате случайного отбора выборки отобрано 46 пациентов из исследуемой совокупности пациентов, выписанных из больницы после острого инсульта. Исследуемая совокупность включает всех амбулаторных больных, выписанных после инсульта, кто согласился дать интервью для определения их трудоспособности по модифицированной шкале Ранкина. Доктора, проводящие интервью, не знали, что их записи будут впоследствии использованы другими докторами, как единственный источник данных для определения состояния больных по МШР.

Два различных эксперта независимо определили оценки с применением МШР по каждой истории болезни в случайной выборке, не зная значения по МШР, полученного первоначально на основе интервью.

Е.3 Наблюдаемая переменная

Наблюдаемой переменной являются оценки трудоспособности по МШР, данные каждым из двух экспертов. Эти данные являются упорядоченными данными (неправильной записи соответствует значение 5).

Е.4 Стандартная оценка

Не существует истинного значения оценки трудоспособности по МШР. «Стандартную оценку по МШР» обычно используют для проверки непротиворечивости и правильности оценки трудоспособности по данным истории болезни. Значения стандартной оценки трудоспособности по МШР основаны на согласованной оценке полученной семью сертифицированными докторами, которые просмотрели видеозаписи, сделанные во время интервью, и согласовали «значение стандартной оценки МШР».

[1)]Этот пример был подготовлен Институтом Гардинера Университета г. Глазго Великобритании.

Е.5 Метод определения значений контролируемой характеристики

Схема процесса определения значений контролируемой характеристики приведена на рисунке Е.1.



Рисунок Е.1 – Схема процесса определения значений контролируемой характеристики

Е.6 Возможные причины ошибочных решений

Значения по МШР охватывают все типы нетрудоспособности, вызванной инсультом, и требующей определения ущерба здоровью, включая физическое, когнитивное, эмоциональное и функциональное состояние. Такие данные всегда могут быть полностью записаны в истории болезни по результатам беседы с пациентом. Эту «недостающую» информацию в данном случае для оценки значения МШР получают на основе записей в истории болезни.

Е.7 План отбора выборки

Начальный план отбора выборки предусматривает отбор случайной выборки из 50 пациентов в возрасте от 30 лет до 92 лет с различными видами инсульта. Впоследствии четыре пациента были удалены из выборки, чтобы избежать возможного смещения результатов. Это привело к тому, что объем выборки сократился до 46 пациентов.

Е.8 Исходные данные

В таблице Е.1 приведены исходные данные, использованные для исследования ААА. Для каждой записи, идентифицированной в столбце «случай», эксперт № 1 и эксперт № 2 указали свои оценки трудоспособности пациента по МШР, полученные исключительно на основе данных истории болезни пациента. Они также указали число (показатель уверенности), характеризующее степень уверенности в правильности оценок по МШР. Это число определяют по пятизначной шкале Лайкерта (5 = «абсолютно уверен», 0 = «совсем не уверен»).

Е.9 Исследование ААА для одного эксперта

Е.9.1 Общие положения

Исследование ААА проведено для анализа непротиворечивости и правильности субъективных оценок каждого эксперта, нескольких экспертов и по отношению к стандарту с применением программного средства SAS.

Е.9.2 Непротиворечивость оценок каждого эксперта

Каждый эксперт дает оценку трудоспособности пациента по каждой записи только один раз. Таким образом, процент совпадения оценок одного эксперта не может быть вычислен и графически изображен.

Таблица Е.1 – Данные истории болезни и стандартная оценка трудоспособности по МШР

Код пациента	Оценка эксперта № 1	Показатель уверенности эксперта № 1	Оценка эксперта № 2	Показатель уверенности эксперта № 2	Стандартная оценка по МШР
50031857m	0	3	0	5	0
50483367l	0	2	0	4	0
50500248r	0	3	0	4	0
50602527e	0	3	0	4	0
50801931m	0	4	0	3	0
50820176H	3	2	1	5	0
50828930x	1	3	0	5	0
64011236B	0	2	0	3	0
50122659h	1	3	1	4	1
50276841R	2	2	0	3	1
50347977m	3	2	1	4	1
50351138w	1	3	1	4	1
50399655k	0	3	1	5	1
50521845a	1	3	0	3	1
50553167M	1	2	1	3	1
50200558v	0	3	0	5	2
50228542H	1	1	0	3	2
50235821v	3	2	2	4	2
50298296a	1	2	0	3	2
50408096r	2	3	1	4	2
50454900m	1	3	1	5	2
50461470b	2	2	1	4	2
50493161e	1	2	1	3	2
50520426r	3	3	2	5	2
50631560h	3	3	1	4	2
50660334r	0	2	1	4	2
50698843V	1	3	2	4	2
50799408a	1	2	0	3	2
50802838b	2	2	1	4	2
51035746x	2	2	1	4	2
64068090w	2	4	1	4	2

Е.9.3 Совпадение оценок каждого эксперта со стандартной оценкой

В таблице Е.2 приведены оценки по МШР, полученные каждым экспертом и стандартные оценки. Оба эксперта подчеркивали случаи, когда стандартная оценка по МШР равна двум или выше и случаи, когда стандартная оценка по МШР равна 0.

Таблица Е.2 – Общая таблица необработанных данных

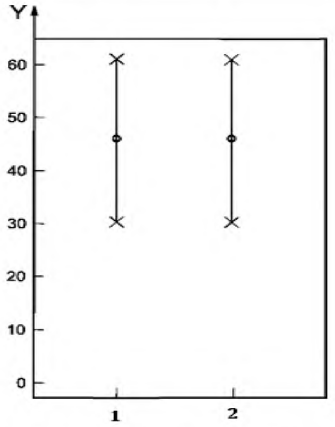
Стандартная оценка по МШР	Объем серии *N*	Среднее значение оценки по МШР (эксперт № 1)	Среднее значение оценки по МШР (эксперт № 2)
0	8	0,50	0,13
1	7	1,29	0,71
2	17	1,65	1,00
3	9	2,11	1,78
4	5	3,40	3,20

Е.9.3.1 Процент совпадения оценок экспертов со стандартной оценкой

В таблице Е.3 и на рисунке Е.2 приведены данные о том, насколько хорошо оценки, полученные экспертом, соответствуют стандартной оценке по МШР. У обоих экспертов процент совпадения составил 45,6 %.

Таблица Е.3 – Совпадение оценок, полученных каждым экспертом со стандартной оценкой

Номер эксперта	Количество проверок	Количество совпадений оценок эксперта со стандартной оценкой	Процент совпадений со стандартной оценкой	Границы доверительного интервала с уровнем доверия 95 %
1	46	21	45,65	(30,90, 60,99)
2	46	21	45,65	(30,90, 60,99)



Y – процент совпадения оценок эксперта со стандартной оценкой;
1 – эксперт № 1;
2 – эксперт № 2.

Рисунок Е.2 – Процент совпадения оценок для каждого эксперта и со стандартной оценкой

Е.9.3.2 Значения κ-статистики Флейса

В таблице Е.4 приведены значения κ-статистики Флейса для оценок, полученных каждым экспертом по каждой истории болезни, а также по отношению к стандартной оценке и общей статистики для каждого эксперта.

Таблица Е.4 – Значения κ-статистики Флейса для одного эксперта

Номер эксперта	Оценка эксперта	κ	Стандартная ошибка, σ_κ	Z	p-значение (>0)
1	0	0,535689	0,147442	3,63322	0,0001
	1	0,270368	0,147442	1,83372	0,0333
	2	0,210784	0,147442	1,42961	0,0764
	3	0,163636	0,147442	1,10984	0,1335
	4	0,452381	0,147442	3,06820	0,0011
	По всем оценкам	0,304926	0,076890	3,96574	0,0000
2	0	0,436275	0,147442	2,95896	0,0015
	1	0,246377	0,147442	1,67101	0,0474
	2	0,130435	0,147442	0,88465	0,1882
	3	0,326007	0,147442	2,21109	0,0135
	4	0,452381	0,147442	3,06820	0,0011
	По всем оценкам	0,299848	0,077733	3,85738	0,0001

Для каждого эксперта средняя оценка по МШР составляет 0,30, (довольно низкое значение, указывающее на низкую возможность определения трудоспособности по МШР на основе записей в истории болезни) по сравнению со стандартной оценкой. Значение 0 указывает на то, что совпадения являются случайными. Значение 0,30 указывает на неслучайное, но достаточно редкое совпадение оценок. Приемлемым значением κ является 0,70 или выше.

Е.9.4 Совпадение оценок для каждого эксперта

Для анализа различий в оценках, полученных по данным истории болезни и стандартных

оценок, в таблице Е.5 приведены коэффициенты Кендалла для каждого эксперта.

Таблица Е.5 – Коэффициент Кендалла для совпадения оценок каждого эксперта со стандартной оценкой

Номер эксперта	Коэффициент Кендалла	Стандартная ошибка Кендалла, σ_K	Z	p-значение (>0)
№ 1	0,506194	0,102046	4,95100	0,0000
№ 2	0,557761	0,102046	5,45633	0,0000

Нулевую гипотезу о том, что совпадение полученных оценок со стандартной оценкой является случайным, отклоняют, поскольку p-значение близко к нулю.

Е.10 Анализ совпадения оценок для нескольких экспертов

Е.10.1 Непротиворечивость оценок

Е. 10.1.1 Процент совпадения оценок для двух экспертов

В соответствии с данными таблицы Е.6 в 23 случаях из 46 оценки экспертов совпадают, т.е. процент совпадения оценок равен 50 %.

Таблица Е.6 – Совпадение оценок эксперта № 1 и эксперта № 2

Количество проверок	Совпадение оценок эксперта со стандартной оценкой	Процент совпадения оценок	95 % CI
46	23	50,00	(34,90, 65,10)

Е.10.1.2 Значения k-статистики Флейса

В таблице Е.7 приведены значения κ-статистики Флейса для двух экспертов и каждого значения МШР. Значение κ = 1 для значения МШР указывает на то, что оценки обоих экспертов полностью совпали для этого значения МШР. Самые большие противоречия в оценках экспертов соответствуют значениям МШР 1 и 2.

Таблица Е.7 – Значения κ-статистики Флейса для непротиворечивости оценок экспертов

Оценка эксперта	κ	Стандартная ошибка, σ_K	Z	p-значение (> 0)
0	0,52821	0,147442	3,58246	0,0002
1	0,07589	0,147442	0,51473	0,3034
2	0,01461	0,147442	0,09906	0,4605
3	0,44745	0,147442	3,03474	0,0012
4	1,00000	0,147442	6,78233	0,0000
По всем значениям	0,34102	0,080439	4,23942	0,0000

Е.10.1.3 Совпадение оценок нескольких экспертов

В таблице Е.8 приведены данные совпадения оценок двух экспертов. Для этих экспертов характерно больше случаев совпадения оценок друг с другом (коэффициент Кендалла 0,90), чем у каждого эксперта со стандартной оценкой (коэффициенты Кендалла 0,5 и 0,56).

Таблица Е.8 – Коэффициент Кендалла

Коэффициент Кендалла	χ^2	Число степеней свободы	p-значение (> 0)
0,901749	81,1574	45	0,0008

Е.10.2 Совпадение оценок всех экспертов по отношению к стандартной оценке

Е.10.2.1 Процент случаев совпадения оценок всех экспертов со стандартной оценкой.

В таблице Е.9 показано, что только в 14 случаях оценки обоих экспертов совпали со стандартной оценкой.

Таблица Е.9 – Процент совпадения оценок всех экспертов со стандартной оценкой

Количество проверок	Количество совпадений оценок всех экспертов со стандартной оценкой	Процент совпадений оценок всех экспертов со стандартной оценкой	Доверительный интервал с уровнем доверия 95 %
46	14	30,43	(17,74, 45,75)

Е.10.2.2 Значения κ-статистика Флейса для оценок всех экспертов по отношению к стандартной оценке.

По данным таблицы Е.10 видно, что наибольшее количество совпадений оценок экспертов соответствует значениям МШР 0 и 4, но даже для этих значений оно достаточно низко по сравнению с рекомендуемым значением 0,7.

Таблица Е.10 – Значения κ-статистики Флейса для оценок всех экспертов

Оценка эксперта	κ	Стандартная ошибка, σ_κ	*Z*	*p*-значение (>0)
0	0,485982	0,104257	4,66137	0,0000
1	0,258372	0,104257	2,47822	0,0066
2	0,170610	0,104257	1,63643	0,0509
3	0,244822	0,104257	2,34825	0,0094
4	0,452381	0,104257	4,33909	0,0000
По всем оценкам	0,302387	0,054668	5,53128	0,0000

Е.10.2.3 Коэффициент Кендалла для совпадения оценок всех экспертов со стандартной оценкой.

В таблице Е.11 приведены значения коэффициента Кендалла для оценок экспертов со стандартной оценкой. В соответствии с данными предыдущих таблиц коэффициент существенно отличается от 0, что указывает на то, что существует некоторая взаимосвязь оценок, значение коэффициента недостаточно высоко вследствие низкой согласованности оценок со стандартной оценкой.

Таблица Е.11 – Коэффициент Кендалла для обоих экспертов по отношению к стандартной оценке

Коэффициент Кендалла	Стандартная ошибка Кендалла, σ_κ	*Z*	*p*-значение (> 0)
0,531978	0,0721572	7,36579	0,0000

Е.11 Выводы

Оценки эксперта № 2 являются более достоверными по сравнению с оценками эксперта № 1 с соответствующими значениями степени уверенности (3,9 и 2,5). Отсутствует взаимосвязь между степенью уверенности и долей корректных оценок.

Оценки, полученные по записям в истории болезни, показывают низкую согласованность со стандартной оценкой с большей согласованностью для случаев, когда стандартная оценка равна 0 или 4.

В результате проведенного исследования рекомендовано отказаться от определения трудоспособности по МШР на основе записей в истории болезни и продолжить определение этой оценки на основе личной беседы с пациентом или беседы по телефону.

Приложение ДА
(справочное)

Сведения о соответствии ссылочных международных стандартов ссылочным национальным стандартам Российской Федерации (и действующим в этом качестве межгосударственным стандартам)

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ИСО 3534-1:2006	–	*
ИСО 3534-2:2006	–	*
* Соответствующий национальный стандарт отсутствует. До его утверждения рекомендуется использовать перевод на русский язык данного международного стандарта. Перевод данного международного стандарта находится в Федеральном информационном фонде технических регламентов и стандартов.		

Библиография

[1] FUTRELL, D. When quality is a matter of taste, use reliability indexes. Qual. Prog. 1995, 28(5),pp. 81-86

[2] MINITAB, INC. Meet Minitab 16. Available (viewed 2010-09-02) at: gost.e-ecolog.ru Files/Shared_Resources/Documents/MeetMinitab/EN16_MeetMinitab.pdf

[3] AUTOMOTIVE INDUSTRY ACTION GROUP (AIAG). Measurement systems analysis: Reference manual, 3rd edition. Troy, MI: Daimler Chrysler, Ford Motor, General Motors Supplier Quality Requirements Task Force, 2002, 225 p.

[4] IWA 1:2005 Quality management systems — Guidelines for process improvements in health service organizations

[5] ISO/TS 16949 Quality management systems — Particular requirements for the application of ISO 9001:2008 for automotive production and relevant service part organizations

[6] QUINN, T.J., RAY, G., ATULA, S., WALTERS, M.R., DAWSON, J., LEES, K.R. Deriving modified Rankin scores from medical case-records. Stroke 2008, 39, pp. 3421-3423

УДК 658.562.012.7:65.012.122:006.352 **ОКС 03.120.30**

Ключевые слова: контроль по альтернативному признаку, система измерений, согласованность оценок, стандартная оценка, выборка, партия, приемлемый уровень качества, единица продукции, несоответствие, несоответствующая единица продукции.

Подписано в печать 01.09.2014. Формат 60x84 1/8.
Усл. печ. л. 5,12. Тираж 51 экз. Зак. 2986.
Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ФГУП «СТАНДАРТИНФОРМ»
123995 Москва, Гранатный пер., 4.

gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru