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

МЕЖГОСУДАРСТВЕННЫЙ СТАНДАРТ

**ГОСТ ISO 11136— 2017**

# Органолептический анализ

# МЕТОДОЛОГИЯ

## Общее руководство по проведению гедонических испытаний потребителями в контролируемой зоне

(ISO 11136:2014,
Sensory analysis — Methodology — General guidance for conducting hedonic tests with consumers in a controlled area,
IDT)

Издание официальное

Москва
Стандартинформ
2017

## Предисловие

Цели, основные принципы и основной порядок проведения работ по межгосударственной стандартизации установлены в ГОСТ 1.0—2015 «Межгосударственная система стандартизации. Основные положения» и ГОСТ 1.2—2015 «Межгосударственная система стандартизации. Стандарты межгосударственные, правила и рекомендации по межгосударственной стандартизации. Правила разработки, принятия, обновления и отмены»

**Сведения о стандарте**

1 ПОДГОТОВЛЕН Акционерным обществом «Всероссийский научно-исследовательский институт сертификации» (АО «ВНИИС») на основе официального перевода на русский язык англоязычной версии указанного в пункте 5 стандарта, который выполнен ФГУП «Стандартинформ»

2 ВНЕСЕН Федеральным агентством по техническому регулированию и метрологии

3 ПРИНЯТ Межгосударственным советом по стандартизации, метрологии и сертификации (протокол от 1 июня 2017 г. № 51)

За принятие проголосовали:

| Краткое наименование страны по МК (ИСО 3166) 004—97 | Код страны по МК (ИСО 3166) 004—97 | Сокращенное наименование национального органа по стандартизации |
|---|---|---|
| Армения | AM | Минэкономики Республики Армения |
| Беларусь | BY | Госстандарт Республики Беларусь |
| Казахстан | KZ | Госстандарт Республики Казахстан |
| Киргизия | KG | Кыргызстандарт |
| Молдова | MD | Молдова-Стандарт |
| Россия | RU | Росстандарт |
| Украина | UA | Минэкономразвития Украины |

4 Приказом Федерального агентства по техническому регулированию и метрологии от 19 сентября 2017 г. № 1159-ст межгосударственный стандарт ГОСТ ISO 11136—2017 введен в действие в качестве национального стандарта Российской Федерации с 1 января 2019 г.

5 Настоящий стандарт идентичен международному стандарту ISO 11136:2014 «Сенсорный анализ. Методология. Общее руководство по проведению гедонических испытаний потребителями в контролируемой зоне» («Sensory analysis — Methodology — General guidance for conducting hedonic tests with consumers in a controlled area», IDT).

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для увязки с наименованиями, принятыми в существующем комплексе межгосударственных стандартов.

Международный стандарт разработан Подкомитетом ISO TC 34/SC 12 «Сенсорный анализ» Технического комитета по стандартизации ISO/TC 34 «Пищевые продукты» Международной организации по стандартизации (ISO).

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

6 ВВЕДЕН ВПЕРВЫЕ

7 Некоторые элементы настоящего стандарта могут являться объектами патентных прав

*Информация об изменениях к настоящему стандарту публикуется в ежегодном информационном указателе «Национальные стандарты», а текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ежемесячном информационном указателе «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 (www.gost.ru)*

© Стандартинформ, 2017

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

## Содержание

1 Область применения ... 1
2 Нормативные ссылки ... 1
3 Термины и определения ... 2
4 Подготовительные мероприятия ... 3
4.1 Предложение о проведении испытания ... 3
4.2 Параметры целевой совокупности ... 4
4.3 Дополнительные параметры ... 4
5 Выборка потребителей ... 4
5.1 Общие положения ... 4
5.2 Набор потребителей ... 5
5.3 Составление выборки потребителей ... 5
5.4 Сегментация выборки ... 6
6 Объем выборки потребителей ... 6
6.1 Решение об измерении ... 6
6.2 Определение объема выборки ... 7
6.3 Работа с подгруппами ... 7
6.4 Количество образцов продуктов и план эксперимента ... 8
6.5 Надежность набора участников ... 8
7 Зона проведения испытания ... 8
8 Продукты ... 8
8.1 Анонимное представление продуктов ... 8
8.2 Подготовка и представление продуктов ... 9
9 Методы ... 9
9.1 Две группы гедонических испытаний ... 9
9.2 Рейтинговый метод (метод шкал) ... 10
9.3 Метод парного сравнения образцов ... 10
9.4 Метод ранжирования ... 11
9.5 Дополнительные вопросы ... 11
9.6 Инструкции для испытателей ... 12
10 Организация испытательных сессий ... 12
10.1 Общие положения ... 12
10.2 Планы ... 13
10.3 Количество продуктов, оцениваемых за одну сессию ... 13
10.4 Характер продуктов, оцениваемых за одну сессию ... 14
10.5 Продолжительность сессии ... 14
10.6 Повторение оценки продукта в одном испытании ... 14
10.7 Предварительное испытание ... 14
11 Анализ результатов ... 15
11.1 Общие положения ... 15
11.2 Анализ выводов ... 15
11.3 Рейтинги (см. 9.2) ... 15
11.4 Парные сравнения (9.3) ... 15
11.5 Ранжирование (9.4) ... 15
12 Отчет об испытании ... 16
12.1 Общие положения ... 16
12.2 Продукты ... 16
12.3 Метод испытания ... 17
12.4 Испытатели (потребители) ... 17
12.5 Результаты ... 17
12.6 Приложения к отчету ... 17
Приложение А (справочное) Пример анкеты для набора участников ... 19
Приложение В (справочное) Методы оценивания подгрупп при сегментации ... 22
Приложение С (справочное) Планы распределения продуктов по испытателям ... 24
Приложение D (справочное) Влияние объема выборки на точность одного среднего значения ... 25
Приложение Е (справочное) Минимальный размер выборки для рейтинговых тестов ... 26

Приложение F (справочное) Минимальный объем выборки для ранжирования ..... 29
Приложение G (справочное) Примеры анализа данных для рейтингов ..... 31
Приложение H (справочное) Примеры анализа данных для ранжирования ..... 34
Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов межгосударственным стандартам ..... 36
Библиография ..... 37

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

**Органолептический анализ**

**МЕТОДОЛОГИЯ**

# Общее руководство по проведению гедонических испытаний потребителями в контролируемой зоне

Sensory analysis. Methodology.
General guidance for conducting hedonic tests with consumers in a controlled area

**Дата введения — 2019—01—01**

## 1 Область применения

Настоящий стандарт описывает подходы к определению, в пределах контролируемой зоны, степени, в которой потребителям нравится или относительно нравится предлагаемая продукция.

В стандарте описаны испытания, основанные на сборе ответов на вопросы, обычно представленных в бумажном виде или посредством клавиатуры, или сенсорной панели. Испытания поведенческого характера (например, параметры записей, использованные потребителями по своему усмотрению) не попадают под настоящий стандарт.

Гедонические испытания, описанные в настоящем стандарте, можно использовать при:

- сравнении предложенного продукта с конкурирующими продуктами;
- оптимизации продукта с целью получения высокой гедонической оценки или с целью понравиться большому количеству потребителей;
- определении ассортимента продуктов, соответствующих конкретной целевой совокупности потребителей;
- определении срока годности (продукта);
- оценке воздействия изменения состава продукта на удовольствие, доставляемое этим продуктом;
- изучении воздействия органолептических характеристик продукта на степень, в которой этот продукт нравится, независимо от косвенных характеристик, таких как бренд, цена или реклама;
- изучении влияния коммерческой переменной или формы преподнесения продукции, например, упаковки.

Описанные методы являются эффективными для определения:

- существуют или нет предпочтения в восприятии (различие в степени предпочтительности), или
- существуют или нет отрицательные предпочтения (тест на проверку подобия методом парного сравнения).

## 2 Нормативные ссылки

Следующие документы, на которые приводится ссылка, являются обязательными для применения настоящего стандарта. В отношении датированных ссылок применимо только указанное издание. В отношении недатированных ссылок действительно только последнее издание ссылочного документа, включая любые изменения к нему

ISO 4121, *Sensory analysis — Guidelines for the use of quantitative response scales* (Органолептический анализ. Руководящие указания по применению шкалы количественных результатов)

ISO 5492, *Sensory analysis — Vocabulary* (*Органолептический анализ. Словарь*)

ISO 5495, *Sensory analysis — Methodology — Paired comparison test* (Сенсорный анализ. Методология. Метод парного сравнения)

**Издание официальное**

ISO 8587, *Sensory analysis — Methodology — Ranking* (Сенсорный анализ. Методология. Ранжирование)

ISO 8589, *Sensory analysis — General guidance for the design of test rooms* (Органолептический анализ. Общее руководство по проектированию помещений для исследований)

ISO 29842, *Sensory analysis — Methodology — Balanced incomplete block designs* (*Органолептический анализ. Методология. Неполностью сбалансированные блочные конструкции*)

## 3 Термины и определения

В данном документе используются термины с соответствующими определениями, приведенные в ISO 5492, а также следующие.

3.1 **группа испытателей** (access panel): Примерная база данных потенциальных респондентов, которые заявили о своем желании сотрудничать в планируемых сборах данных, если будут выбраны.

[ИСТОЧНИК: ISO 20252]

Примечание — Респонденты могут выбираться из базы на более или менее регулярной основе.

3.2 **вопрос, предусматривающий стереотипные альтернативные ответы** (closed-ended question): Вопрос, который требует от респондента выбрать из предварительно составленного перечня возможные ответы.

3.3 **сторона-инициатор** (commissioning party): Лицо, орган или служба, которые уполномочивают третью сторону на проведение испытания.

Примечания

1 Понятие «сторона-инициатор» соответствует концепции «покупатель» или «заказчик» в области контроля качества и противоположно концепции «поставщик», «поставщик услуг» или «лаборатория».

2 Обычно сторона-инициатор и поставщик услуг принадлежат к разным предприятиям. Там, где они принадлежат к двум департаментам одного и того же предприятия, соответствующие обязанности этих двух департаментов должны четко дифференцироваться и разъясняться в предложении о проведении испытания.

3.4 **потребитель (испытатель)** (consumer), **респондент** (respondent). Лицо, использующее продукт.

Примечание — В настоящем стандарте это значение ограничено лицом, которое не является экспертом, специалистом-дегустатором или экспертом по оценке, выбранным в соответствии с ISO 5492.

3.5 **выборка потребителей (испытателей)** (consumer sample): Группа потребителей, набранных на основе критериев, установленных в соответствии с характером целевой совокупности.

3.6 **контролируемая зона** (controlled area): Место, обеспечивающее:

- контролируемую подготовку и представление продуктов,
- комфортабельные условия для потребления продуктов и опроса потребителей,
- отсутствие обмена информацией (вербального и невербального) между потребителями (испытателями), гарантирующее независимость ответов.

3.7 **гедоническое восприятие** (hedonic perception): Количество удовольствия, доставляемого органолептическими свойствами продукта.

3.8 **гедоническое испытание** (hedonic test): Испытание, целью которого является определение общего гедонического восприятия продукта потребителями.

3.9 **вопрос, допускающий неограниченное число ответов, открытый вопрос** (open-ended question): Вопрос, который не предусматривает готовые альтернативные ответы для респондентов, а просит дать ответ собственными словами.

Примечание — Он может касаться каждого из продуктов, представленных потребителям (испытателям), или только некоторых из них (например, когда задается вопрос, почему определенный продукт является наиболее (наименее) привлекательным для потребителя (испытателя).

3.10 **необъективность, связанная с положением** (position bias): Косвенное влияние, связанное с расположением продукта в наборе оцениваемых продуктов.

3.11 **семейство продуктов** (product family): Группа продуктов, которые могут быть взаимозаменяемыми в нормальных условиях потребления.

Примечание — Семейства продуктов могут зависеть от культуры и привычек потребителей.

3.12 **репрезентативность** (representativeness) <выборки>: Степень, в которой качественные признаки выборки согласуются с соответствующими качественными признаками совокупности, из которой эта выборка получена.

3.13 **необъективность, связанная с последовательностью** (sequential bias): Косвенное влияние временного положения или положения в последовательности продукта в наборе оцениваемых продуктов.

3.14 **сессия** (session): Заседание потребителей (испытателей), работающих одновременно по одному и тому же протоколу в течение определенного периода времени.

Примечание — Сессию можно разделить на несколько сеансов, между которыми дается определенное время на отдых. Например, сессию продолжительностью 1 ч 30 мин можно разделить на два сеанса с 10-минутным перерывом между ними.

3.15 **целевая совокупность** (target population): Совокупность потребителей, чей гедонический отзыв на продукт предполагается оценить.

## 4 Подготовительные мероприятия

### 4.1 Предложение о проведении испытания

Предложение о проведении испытания выдвигается предоставляющей услуги лабораторией на основе информации (обычно касающейся цели испытания и целевой совокупности), выданной стороной-инициатором, как письменный документ, который связывает лабораторию и сторону-инициатора договором. Он включает следующие элементы:

- цель испытания;
- подлежащие оценке продукты (и процедуры для их закупки, особенно в тех случаях, когда лаборатории требуется обеспечить один или более продуктов);
- целевая совокупность для испытания;
- размер выборки потребителей (объем выборки) и способ набора участников;
- любые ограничения в отношении потребителей, ранее участвовавших в подобных испытаниях;
- любые ограничения на другие продукты, которые предполагается оценить, за одну и ту же сессию испытаний;
- условия оценивания;
- тип интервью (самостоятельное заполнение форм или интервьюеры);
- вопросник (анкета);
- место проведения оценивания;
- количество продуктов для оценивания на сессию в предлагаемом испытании, наряду с указанием других продуктов для оценивания, там, где сессия охватывает несколько семейств продуктов (там, где эти продукты неизвестны на момент составления предложения, лаборатории следует взять на себя обязательство связаться с испытателями, как только лаборатория узнает о характере продуктов);
- условия подготовки, конкретно для оцениваемых продуктов (если они известны на данном этапе);
- любые инструкции, конкретно для оцениваемых продуктов (если они известны на данном этапе) наряду с конкретной информацией, предоставляемой о продуктах или испытании в начале сессии;
- наименование статистических критериев, которые предполагается использовать для интерпретации результатов; тип использованных рисков и их допустимых значений; величина минимального различия, которое предполагается обнаружить в тестах на проверку различия или максимально допустимого различия в тестах на проверку подобия;
- любое предусмотренное разделение (сегментация) совокупности потребителей на основе характеристик, известных до получения данных оценивания или на основании полученных ответов;
- метод изучения и получения итогов по вопросам, допускающим неограниченное число ответов;
- любые рекомендации, которые сторона-инициатор желает, чтобы лаборатория разработала на основе результатов испытания;
- данные, предусмотренные для тестов и предоставления отчета.

Если какие-либо из указанных выше элементов неизвестны на момент внесения предложения об испытании, это предложение включает для каждого рассматриваемого элемента следующее заявление: «Данный элемент будет установлен стороной-инициатором посредством... [задание этапа испытания]».

Примечание — Если спецификация уже существует, предложение о проведении испытания дает на нее ссылку, не повторяя технические пункты, описанные в рабочем порядке в этой спецификации.

### 4.2 Параметры целевой совокупности

Чтобы установить целевую совокупность, необходимо ответить на следующие вопросы. Эти вопросы должны быть рассмотрены для каждого испытания, которое предполагается осуществить, потому что целевая группа потребителей, участвующая в одном испытании, может отличаться от группы для другого испытания, даже при испытании одного и того же продукта.

- Представлен ли уже подлежащий испытаниям продукт на рынке? Если да, то можно ли отличить реальных потребителей от потенциальных?
- Являются ли представляющие интерес потребители реально использующими представленный продукт потенциальными потребителями, которые еще не использовали этот продукт, или группа будет смешанной?
- Будут ли результаты, полученные от установленных подгрупп, изучаться и сравниваться?
- Будут ли результаты всей выборки потребителей проанализированы с намерением идентификации подгрупп?
- Представляют ли интерес расхождения между результатами отдельных представляющих интерес потребителей?

### 4.3 Дополнительные параметры

После задания параметров целевой совокупности, необходимо установить следующие аспекты:
- место проведения испытания (см. раздел 7);
- ожидаемая прецизионность измерения (см. 6.1);
- метод испытания (см. раздел 9);
- план представления продукта для выбранных методов испытания (см. 10.2);
- спецификации для проверяемой гипотезы (см. 6.1).

Для проверяемой гипотезы затем должны быть установлены следующие аспекты:
- состав выборки потребителей (см. раздел 5);
- объем выборки потребителей (см. раздел 6).

## 5 Выборка потребителей

### 5.1 Общие положения

Целью гедонического испытания является определение приемлемости продуктов и/или определение предпочтений по отношению к двум или более продуктам установленной совокупностью потребителей.

Состав выборки потребителей является решающим для любого гедонического испытания, поскольку рассматривают, сможет или нет результат ответить на интересующие вопросы стороны-инициатора.

Потребители должны быть добровольцами, что означает запрет на получение вознаграждения или оплаты за участие в испытаниях.

Важно получить спонтанную реакцию.

Подготовленные потребители должны быть исключены из таких испытаний, поскольку они будут систематически отличаться от целевой совокупности потребителей.

Обычно не рекомендуется набирать потребителей из персонала компании, производящей испытуемый продукт. Если используют группу внутренних потребителей, то среди факторов, наиболее неблагоприятно сказывающихся на результатах, будет:
- риск узнавания испытуемых продуктов,
- тенденция вынести суждение в пользу узнанных продуктов, и
- хорошая осведомленность в отношении испытуемых продуктов.

Эти факторы несут серьезный риск нарушения репрезентативности результатов для целевой совокупности потребителей.

Все критерии, которым следуют при выборе потребителей (например, можно или нет набирать их из штата компании-изготовителя или ее конкурентов или из конкретного социально-экономического сектора), должны быть согласованы с клиентом, заказавшим проведение испытания, независимо от используемого метода набора.

### 5.2 Набор потребителей

#### 5.2.1 Общие положения

**ПРЕДУПРЕЖДЕНИЕ — Необходимо соблюдать:**

**- конфиденциальность и соответствующие законы при создании и оформлении файлов с персональными данными.**

**- законодательство в отношении возрастных ограничений.**

Потребителей можно набирать либо на основе конкретной задачи (там, где потребителей набирают заново непосредственно под конкретное испытание) или из группы ожидания (перечень потенциальных потребителей с известными характеристиками). Набор под конкретную задачу часто считают средством набора потребителей, которые активно не участвовали в подобных испытаниях, но эта идея может стать ошибочной, когда происходит значительное увеличение мест, где организуют гедонические испытания квазистационарным образом.

#### 5.2.2 Набор под конкретную задачу

Набор под конкретную задачу можно проводить в общественном месте, в торговой точке, по телефону, по обычной или электронной почте или по объявлению в местной газете или на радиостанции, или в социальных сетях, или иными путями через Интернет. В большинстве случаев необходимо достичь предварительного соглашения с корпорацией или организацией, несущей ответственность за участок, на котором ведется набор.

Способ набора и время подхода к потребителю (в частности, когда контакт происходит в общественном месте или по телефону и при распространении информации через средства массовой информации) может в значительной степени повлиять на состав выборки потребителей.

#### 5.2.3 Набор из группы ожидания

Для эффективности набора также можно укомплектовать группу ожидания, участников которой внести в файл описания потребителей. Перед каждым испытанием и после установления целевой совокупности, можно сделать предварительный выбор на основе критериев, записанных в этом файле.

Описание потребителей может включать следующие данные о потребителе:

- идентификационные данные рассматриваемого лица, контактные данные;
- возраст, пол;
- семейное положение, состав семьи;
- социальное положение, профессия, доход;
- возможные проблемы со здоровьем (например, аллергия, диабет, проблемы со зрением, и т. д.) в отношении испытываемых продуктов;
- другая информация, связанная с испытываемыми продуктами (например, привычки в еде и питье, религиозные запреты, образ жизни, бытовое оборудование);
- информация о доступности потребителя для участия в испытаниях;
- информация о случаях предыдущего участия в испытаниях потребителем.

Описание каждого потребителя необходимо регулярно актуализировать.

Описанная база данных позволит эффективно работать в диалоге между стороной-инициатором и поставщиком услуг в такой мере, в какой это позволит последнему обеспечить удовлетворение основных пожеланий стороны-инициатора в отношении целевых потребителей.

#### 5.2.4 Частота привлечения потребителей

По сравнению с набором под конкретную задачу существует большой риск «профессионализации потребителя», при наборе из существующей совокупности потребителей.

Важно избежать эффекта обучения отдельного потребителя в отношении испытуемой группы продуктов, поэтому нельзя привлекать одних и тех же потребителей слишком часто. В качестве руководства рекомендуется проводить отдельные испытания одного и того же продукта не менее чем с трехмесячным интервалом.

Лаборатория должна сохранять записи частоты участия каждого потребителя в испытаниях каждого продукта и каждого семейства продуктов. История участия потребителей должна быть включена в отчет об испытании, возможно наряду с информацией о семействе продуктов по запросу стороны-инициатора.

### 5.3 Составление выборки потребителей

#### 5.3.1 Выбор потребителей с помощью анкеты

Когда потребителей набирают с помощью анкеты, относящейся к критериям, установленным для состава выборки потребителей, потребитель не должен иметь возможность идентифицировать, какие критерии в анкете являются решающими в выборе для участия в испытаниях.

Примеры возможных критериев:

- предыдущее участие в испытаниях рассматриваемой группы продуктов;
- частота использования рассматриваемого продукта;
- обычно используемый бренд, описание продукта, место приобретения и т. д.;
- привычки в использовании или потреблении рассматриваемого продукта.

Могут подойти другие критерии, такие как возраст, пол, социальный слой, группа занятости или географическое местоположение. Из соображений практичности следует установить классы для каждого из этих критериев. Для каждого класса следует установить процент, которым этот класс будет представлен в выборке потребителей.

Для каждого потребителя, выбранного для участия в испытании, анкета должна быть заполнена полностью.

Можно проверить требуемые детали набора.

Пример анкеты для набора участников приведен в приложении А.

Анкету или информацию из этой анкеты необходимо хранить, чтобы обеспечить прослеживаемость параметров набора участников испытания.

**5.3.2 Репрезентативность выборки потребителей**

Репрезентативность в отношении целевой совокупности является главным свойством, требуемым от выборки. Определение такой совокупности обычно вытекает из взаимодействия между стороной-инициатором испытания, которая предлагает целевую совокупность, и лабораторией, которая оценивает осуществимость этого предложения.

Для выбора потребителей, которые будут представительными для целевой совокупности, необходимо использовать следующие критерии:

- частота использования или потребления продукта (этот критерий зачастую считается основным в отношении его предполагаемого влияния на различающую способность испытания);
- возраст, пол, социально-экономический статус, занятие, географическое местоположение;
- бренд продукта, обычно используемого или потребляемого;
- место приобретения этого продукта.

### 5.4 Сегментация выборки

Существует два пути возможной сегментации выборки (потребителей):

a) Сегментация на начальной стадии, чтобы получить ответы на вопросы:

1) Отвечают женщины и мужчины одинаково или по-разному?
2) Одинаковы или нет ответы разных возрастных групп?
3) Зависят ли ответы от привычек потребления?

В этом случае каждую подгруппу анализируют по отдельности и сравнивают результаты.

b) Сегментация может быть выведена на основе собранных данных, если имеется подтверждение (например, мультимодальность) наличия различных подгрупп.

В приложении В приведено руководство в отношении того, как обращаться с данными сегментации.

## 6 Объем выборки потребителей

### 6.1 Решение об измерении

Численная прецизионность любых средних значений или оценок, полученных по результатам, увеличивается с увеличением численности выборки. В то же время повышенная прецизионность измерения сама по себе не обеспечивает релевантность и достоверность полученных заключений. Большое значение имеет также правильный выбор потребителей и методов.

Приложение D содержит некоторые особенности влияния объема выборки потребителей на прецизионность измерения.

Прецизионность также зависит от вариантности ответов потребителей в отношении одного продукта. Чем больше разброс, тем больше должен быть размер выборки, чтобы получить заданную прецизионность. Знание данного параметра вытекает из опыта как лаборатории, так и стороны-инициатора.

Прецизионность результатов зависит также от метода испытания. Например, при прочих равных условиях различие между двумя продуктами устанавливается более прецизионно, если каждый потребитель оценивает оба продукта, а не в случае оценки продуктов разными потребителями.

Требующаяся прецизионность зависит от цели испытания. При прочих равных условиях на прецизионность влияет.

- наименьшее различие между двумя средними оценками удовольствия, которое считается достаточно большим, чтобы использоваться, когда целью испытания является демонстрация различия (чем различие меньше, тем больше надо брать размер выборки);
- наибольшее различие между двумя средними оценками удовольствия, которое считается достаточно малым, чтобы им пренебречь, когда целью испытания является демонстрация равноценности (чем равноценность меньше, тем больше надо брать размер выборки),
- риск $\alpha$ *(альфа)*, выбираемый стороной-инициатором на принятие ошибочного заключения о наличии различия (чем меньше риск α, тем больше надо брать размер выборки);
- мощность *P*, которую сторона-инициатор желает для критерия. ($P = 1 - \beta$, где β представляет собой риск ошибочного заключения, что различие отсутствует); чем больше желательная мощность, тем больше надо брать размер выборки);
- направленность предполагаемого статистического критерия. Если сторона-инициатор заинтересована только в повторном подтверждении, что продукт *B* не уступает продукту *A*, направленный критерий подходит и мощность этого критерия больше, чем, если заключения, что *B* уступает *A* или что *B* превосходит *A*, представляют одинаковый интерес.

### 6.2 Определение объема выборки

Определение объема выборки является критическим этапом в разработке испытания и требует помощи специалиста в области статистики. В приложении E даны формулы (сопровождаемые примерами), которые можно использовать для расчетов в тестах на проверку различия, либо на проверку подобия на основе рейтингов. Приложение F дает то же самое для испытаний, основанных на предпочтениях.

Во всех случаях не допускается объем выборки потребителей, соответствующий настоящему стандарту, меньше 60.

Там, где лаборатория не имеет данных, требующихся для методов, описанных в приложении E или приложении F, настоящий стандарт фиксирует минимальный размер выборки — 100.

Там, где лаборатория использует методы, описанные в приложении E или приложении F, размер выборки задается результатом расчета, который:

- может быть снижен до 100, когда рассчитанное количество потребителей более 100, но лаборатория и/или сторона-инициатор не имеет средств для опроса большего количества потребителей. Если это происходит, лаборатория должна рассчитать оцененную мощность критерия для количества потребителей, равного 100, и получить согласие стороны-инициатора на принятие более низкой мощности, и
- должен быть увеличен до 60, если рассчитанное количество меньше 60.

Значения 100 и 60 предполагают, что сторона-инициатор не надеется на заключения, сегментированные согласно характеристикам потребителей, таким как возраст и пол. Там, где решение о разделении принято до испытания, минимальное число потребителей на сегмент составляет 60.

### 6.3 Работа с подгруппами

#### 6.3.1 Общие положения

Подгруппы выборки потребителей можно сформировать до сбора данных испытания или используя эти данные. Для любых вариантов, перечисленных ниже, необходимо получить рекомендации у статистиков на стадии планирования испытания.

#### 6.3.2 Отдельные испытания с использованием заранее сформированных групп потребителей

Сторона-инициатор может пожелать изучить отдельные подгруппы, сформированные с учетом характеристик потребителей (таких как возраст и пол), которые известны до начала испытаний. Один вариант заключается в независимом анализе полученных результатов. В этом случае каждая подгруппа должна состоять как минимум из 60 испытателей.

#### 6.3.3 Унифицированное испытание с использованием заранее сформированных групп потребителей

Процедура 6.3.2 менее эффективна, чем уравнивающие презентации для всей выборки потребителей с последующим комплексным анализом всех результатов с учетом количества членов в подгруппах. Такая же степень доверия обычно достигается с меньшим, чем 60, количеством потребителей в каждой подгруппе. Такой стиль анализа в высшей степени эффективен, когда все подгруппы одинаковы или очень схожи по размеру.

#### 6.3.4 Сегментация потребителей

Количество членов в подгруппе может быть результатом, выведенным по результатам испытаний на большой выборке потребителей. Подгруппы тогда классифицируют по оценке всех имеющихся данных (результаты измерений и/или социально-демографические данные). Такая сегментация служит для описания выборки потребителей и может использоваться для разработки новых гипотез. Число сформированных групп, их характеристики и их размеры зависят от количества и характера полученных результатов.

### 6.4 Количество образцов продуктов и план эксперимента

Количество образцов и план эксперимента влияют на размер выборки потребителей. Если для презентации продукта используется полный план, то количество потребителей, необходимое для испытания, будет равно числу полученных отзывов на продукт. Если используется неполный план, количество потребителей должно быть больше.

### 6.5 Надежность набора участников

Реально полученное число отзывов обычно меньше числа набранных для испытания потребителей, поскольку некоторые потребители в выборке могут отсутствовать, а некоторые участники неправильно использовать анкету. Чтобы компенсировать это, следует включить резерв при расчете требующегося количества потребителей.

## 7 Зона проведения испытания

Настоящий стандарт касается гедонических испытаний, проводимых в контролируемой зоне согласно 3.6. Это отличается от гедонических испытаний, проводимых, например, в торговой точке или на дому у потребителей, что выходит за рамки области применения настоящего стандарта. Описано три примера соответствующих зон испытания.

a) Стационарная лаборатория органолептической оценки, специально оборудованная в соответствии с ISO 8589 для проведения органолептических испытаний. Она имеет приемную для встречи участников, помещение для подготовки продуктов и помещение, оснащенное боксами для испытаний, температурным контролем и системой вентиляции для периодического обновления воздуха. Такая зона обеспечивает наилучшую подготовку продуктов, представление продуктов и условия протоколирования ответов.

b) Передвижная лаборатория органолептической оценки, установленная на транспортном средстве, которое специально переоборудовано для проведения органолептических испытаний. Обычно такая лаборатория имеет меньше площадей, чем стационарная. Зоны для приема испытателей и подготовки продуктов очень ограничены.

c) Комната или комнаты, оснащаемые на время в соответствии с «конкретной задачей» для оценивания продуктов, которые требуют минимум подготовки. Эту зону разбивают на две отдельные части, одну для проведения испытаний, другую для подготовки и кодирования продуктов. Планировка зоны испытания позволяет потребителям работать в физической изоляции друг от друга. Условия испытания (температура, комнатное освещение, вентиляция, хранение и подготовка образцов) можно отслеживать. В отличие от стационарных и передвижных лабораторий в данном примере подготовку продуктов и систему презентации каждый раз определяют заново, адаптируя помещение под конкретную задачу.

## 8 Продукты

### 8.1 Анонимное представление продуктов

Подлежащие испытанию продукты представляют испытателям после того, как их максимально лишают индивидуальности и единственным средством идентификации остается произвольный код (например, трехзначное случайное число), присвоенный лабораторией. Все ссылки на бренд или знак качества должны быть удалены или закрыты, за исключением двух случаев:

a) когда удаление или маскировка невозможны;

b) когда целью является определение влияния бренда или знака качества.

Примечание — В настоящем стандарте под знаком качества подразумевают любой знак, который потребитель может воспринять как влияние на оценку. Например, тип упаковки может представлять собой знак качества.

### 8.2 Подготовка и представление продуктов

#### 8.2.1 Общие положения

Подлежащие оценке продукты должны быть подготовлены согласно методам, рекомендованным изготовителями, или, если такие методы неизвестны, то согласно методу, официально оформленному по предварительным испытаниям. Допуская надлежащим образом обоснованные ожидания, условия подготовки должны приближаться к обычным условиям использования и потребления.

То же самое правило применяется к представлению продуктов (температура, сопутствующие продукты, посуда и способ подачи).

**ПРЕДУПРЕЖДЕНИЕ — Любой продукт меняется от партии к партии и вследствие хранения. Такая изменчивость может представлять интерес для стороны-инициатора и в некоторых случаях может стать основной задачей испытания. Во всех других случаях изменчивость следует минимизировать, используя продукты из одной и той же партии, которые хранились в одинаковых условиях и в течение одинакового времени.**

#### 8.2.2 Количество представляемых продуктов

Количество продукта, представленного испытателю, должно соответствовать обычно потребляемой порции продукта, установленной стороной-инициатором. Испытателя необходимо информировать о минимальном количестве образца для испытания и, при необходимости, о максимальном количестве образца для испытания или потребления. Необходимо следить за тем, чтобы порция образца не была слишком маленькой (риск получить только мимолетное исчезающее начальное впечатление) или слишком большой (риск добиться пресыщения или отвращения).

Представленное количество может потребовать уменьшения по сравнению с обычно потребляемым количеством, в тех случаях, когда участники пробуют несколько продуктов одного и того же семейства. В то же время, представленного количества всегда должно быть достаточно, чтобы избежать проведения оценки на слишком малом количестве, спровоцировав, таким образом, получение мимолетного исчезающего начального впечатления.

Порция готовой рыбы, предоставленная каждому участнику, должна пропорционально включать все составляющие испытуемого продукта.

#### 8.2.3 Форма представленных продуктов

Иногда существует два варианта представления продукта: продукт в состоянии непосредственно после закупки и продукт в форме, готовой к употреблению.

***Пример 1 — Примеры включают хлеб (целиком или нарезанный), сосиску (целиком или нарезанную), сыр (целиком или нарезанный), пиццу (целиком или порциями), растворимый кофе (в чашке до или после добавления воды).***

Лаборатория и сторона-инициатор могут считать, что обе формы имеют одну и ту же цель: наилучшим образом оценить удовольствие, вызываемое продуктом. Нет необходимости знать, провоцируется или нет эффект ореола той или иной формой. Если да, то две формы представляют одновременно (целиком/кусочками) или последовательно (растворимый кофе). Потребитель дает отдельную оценку обеим формам представления, причем потребители не знают, что один и тот же продукт представлен в двух формах. Такой подход соответствует нормальному поведению потребителя, которому в повседневной жизни не требуется разделять свои восприятия.

***Пример 2 — Слова инструкции и вопроса могут звучать так: Вам дают сосиску под кодом 148. Вам следует осмотреть ее, понюхать, съесть несколько кусочков и затем сообщить свою оценку с помощью указанной ниже шкалы.***

Если лаборатория и сторона-инициатор согласны, что обе формы вносят вклад в удовольствие, получаемое от продукта, они не нуждаются в обработке как разные продукты. Наоборот обе эти формы можно представить вместе или по очереди и получить один рейтинг.

Примечание — Под настоящий стандарт не подпадает случай, когда лаборатория и сторона-инициатор считают необходимым узнать эффект ожидания, создаваемый визуальной информацией. Такая ситуация попадает в рамки области применения специальных планов испытания, включающих конкретные концепции и методы анализа.

## 9 Методы

### 9.1 Две группы гедонических испытаний

Две основные группы гедонических испытаний устанавливают разные задачи для потребителя (испытателя):

a) **проверка приемлемости**, используемая для измерения глубины удовлетворения (удовольствия) при потреблении. В настоящем стандарте единственным описанным типом проверки приемлемости является **рейтинговый метод (метод шкал)**;

b) **методы предпочтения**, используемые для измерения степени получаемого удовольствия от разных продуктов (например, «Какой образец Вам нравится больше?» или «Пожалуйста, расположите представленные образцы в порядке от менее понравившегося до более понравившегося»). Информация, полученная методом оценки моментального предпочтения, имеет относительный характер. Она ничего не говорит о приемлемости продуктов, поскольку можно отдать предпочтение одному продукту перед другим, хотя оба они неприемлемы. Методы предпочтения подразделяются:

- **на методы парного сравнения**, когда сравнивают два образца;
- **методы ранжирования**, когда работают с несколькими образцами.

### 9.2 Рейтинговый метод (метод шкал)

#### 9.2.1 Общие положения

Рейтинговые методы различаются:

- по шкале: шкала может быть структурированной, не структурированной, числовой, семантической или графической в соответствии с ISO 4121.
- методом представления продуктов.

Когда число продуктов два или более, можно использовать три метода представления продуктов:

a) **строгое монадическое представление (отдельная оценка),** когда каждый испытатель оценивает один продукт;

b) **последовательное монадическое представление (неполный или полный план)**, когда несколько продуктов оцениваются испытателем за одну или несколько сессий. Испытатель получает один продукт за один раз, при этом информации об уже оцененных продуктах или полученных при оценке ответах ему не дается; необходимо обеспечить невозможность для испытателя вернуться к оценке ранее испытанного продукта;

c) **сравнительное представление (неполный или полный план),** когда несколько продуктов оцениваются испытателем одновременно, и испытателям позволяется пересмотреть оценки, которые они дали другим продуктам.

Наиболее общей формой представления является последовательная монадическая. Она менее затратная, чем первая, но требует полностью освоить способ распределения продуктов по испытателям (см. приложение С). Сравнительное представление применяют редко, поскольку имеет тенденцию к преувеличению различий между продуктами, а также затрудняет сравнение между результатами испытания, когда условия испытания не строго идентичны.

#### 9.2.2 Планы представления продуктов

Образцы следует представлять испытателям в соответствии с планом представления (см. 10.2).

### 9.3 Метод парного сравнения образцов

#### 9.3.1 Общие положения

Методы парного сравнения образцов должны осуществляться в соответствии с ISO 5495.

Принцип парного сравнения образцов заключается в сравнении продуктов, представленных парами. Испытатель оценивает образцы в установленном порядке и указывает, какой из двух образцов предпочтительнее.

Парное сравнение образцов с указанием предпочтения в отношении одного из двух продуктов можно проводить в два разных дня, а именно с принудительным выбором (вариант А ниже) или с допустимым ответом «предпочтения нет» (вариант В ниже).

a) **Вариант А:** Испытатели отвечают на вопрос, предполагающий только один ответ:

«Попробуйте два представленных Вам продукта, начиная с продукта слева, и затем отметьте крестиком код продукта, который для Вас предпочтительнее».

Продукт xxx Продукт yyy

□ □

Здесь испытателям предстоит решить, какому продукту отдать предпочтение.

b) **Вариант В:** Испытатели отвечают на вопрос, предполагающий только один ответ:

«Попробуйте два представленных Вам продукта, начиная с продукта слева, и затем отметьте крестиком код продукта, который для Вас предпочтительнее. Если продукты Вам нравятся в равной степени, поставьте крестик в клетку «предпочтения нет».

Продукт xxx Предпочтения нет Продукт yyy

Для сравнения одной пары образцов вариант А имеет преимущество большего потенциала для проверки различия, чем вариант В.

В анкетах не должно содержаться вопросов, способных повлиять на решение испытателей в пользу того или иного продукта или оценку продуктов в отношении конкретных качественных характеристик. Например, если испытателей просят высказать замечания в отношении флейвора продуктов, то испытателям следует сконцентрироваться именно на этом аспекте при оценке приемлемости.

Разрешается, в то же время, задать вопрос, допускающий неограниченное количество ответов, в конце испытания, давая испытателям возможность прокомментировать, какие продукты понравились и какие не понравились.

**9.3.2 Планы представления образцов**

Порядок представления образцов А и В должен уравновешиваться парами АВ и ВА одинаково часто, предлагая их испытателям случайным образом.

### 9.4 Метод ранжирования

**9.4.1 Общие положения**

Методы ранжирования проводят в соответствии с ISO 8587.

Принцип метода ранжирования заключается в представлении нескольких образцов одновременно и просьбе отсортировать их в соответствующем порядке.

Метод рангового порядка довольно требователен к испытателю, и его сложность увеличивается с числом образцов для сравнения, поскольку испытатель должен попробовать продукты несколько раз, чтобы отсортировать и расположить их в соответствующем порядке.

Допускается задавать порядок испытания образцов (обычно слева направо) только для первоначальной оценки каждого образца испытателем. После этого испытателю позволяется пробовать образцы в любом порядке. Более того, нельзя задавать метод принятия решения испытателем в отношении порядка распределения.

Метод рангового порядка может быть выполнен либо с принудительным выбором, либо, позволяя испытателю классифицировать два или более продуктов как равные. В то же время анализ второго из этих вариантов не подпадает под данный стандарт.

**9.4.2 Анкетирование**

Испытатели получают задание следующего типа:

«Необходимо попробовать четыре продукта, один за другим, начиная с продукта, расположенного в крайней левой позиции, и заканчивая продуктом, расположенным в крайней правой позиции. Затем можно пробовать любой из продуктов еще раз, сколько требуется, чтобы составить о нем свое мнение. Затем, внесите коды продуктов по порядку от наименее понравившегося до наиболее понравившегося»

| | | | |
|---|---|---|---|
| Понравился меньше всех | | | Понравился больше всех |

### 9.5 Дополнительные вопросы

**9.5.1 Общие положения**

В дополнение к любому из описанных методов стороне-инициатору может понадобиться дополнительная информация, помогающая понять общую оценку продуктов или характеристик испытателей. Например, набирая потребителей из базы данных, лаборатории может потребоваться проверка, что все испытатели выбраны действительно из целевой совокупности. Для этого можно задать вопрос о конкретной привычке потребления.

**9.5.2 Вопросы о потребителях**

Выбор вопросов и этапа, на котором их задавать, требует пристального внимания лаборатории.

Если вопрос задается после сессии, необходимо остерегаться, что на ответы могут повлиять уже проведенные слепые испытания. Например, после испытания сначала томатного соуса, а потом риса, испытатели могут скорее склониться к заявлению, что в действительности они пробовали рис с соусом.

С другой стороны, вопросы, задаваемые до сессии, могут спровоцировать испытателей рассматривать продукты особым образом или уделить особое внимание некоторым аспектам.

#### 9.5.3 Вопросы о продуктах

Для этой цели можно использовать вопросы, предусматривающие один или неограниченное число ответов. Эти вопросы должны задаваться после всех вопросов, касающихся общей оценки и на листе (или экране), отличающемся от других, касающихся общей оценки, чтобы снизить риск, что испытатели ответят таким образом, что их ответы окажутся единообразными.

Количество таких вопросов следует максимально сократить.

Примечания

1 Дополнительные вопросы не предназначены ставить под сомнение результаты основной оценки. Если существует какое-либо несоответствие между общей оценкой и сообщением о намерении потреблять (продукт), предполагается найти причину этого, возможно посредством дополнительного испытания.

2 Дополнительные вопросы часто дают мало сведений. Руководитель лаборатории органолептической оценки должен информировать сторону-инициатора о существовании методов (например, составление карт предпочтений или декомбинационный анализ), разработанных специально для комбинирования гедонических данных и данных органолептического профиля и для определения, почему испытателям нравится или не нравится продукт.

### 9.6 Инструкции для испытателей

Инструкции могут быть даны в письменной или устной форме. Их значение для соответствия полученных результатов нельзя недооценивать. Особое внимание должно уделяться инструкциям, они должны составить предмет конкретного раздела в отчете об испытании. Инструкции должны фокусироваться на следующем:

a) минимальные и максимальные величины, которые каждый испытатель должен оценить до выставления оценки. Эти величины зависят от типа продукта, числа продуктов, оцененных в ходе сессии, и пожеланий стороны-инициатора. Лаборатория должна не только давать инструкции в отношении минимальных и максимальных величин, но также должна определять рабочий протокол, чтобы обеспечить выполнение этих инструкций;

b) что необходимо сделать после заполнения опросного листа. При проведении последовательной монадической оценки лаборатория должна использовать систему, которая не допускает, чтобы испытатели ссылались на ответы ранее оцененного продукта;

c) продолжительность перерывов и любая деятельность во время этих перерывов (очистка рецепторов с помощью хлеба, яблока или воды, выход из помещения, где проводилась оценка, чтение журналов и т. д.);

d) формат и смысл задаваемых вопросов.

Примечание — Формулировка вопросов не бывает нейтральной. Например, чтобы предложить испытателям возможность пояснения своей общей оценки продукта внешним видом, запахом, вкусом, ароматом или причинами, связанными с текстурой, ведет к повышению рациональности ответов, чем наличие простого раздела: *Замечания*.

Лицо, проводящее испытание, обычно дает испытателям в начале сессии информацию о количестве продуктов, количестве, которое необходимо попробовать и вопросах, на которые необходимо ответить. Он/она могут также представить другую информацию, чтобы выполнить оценку менее обезличенную или чтобы выполнить другие задачи испытания. Поскольку такая информация может вызвать ожидания у испытателей, способ ее выражения должен тщательно рассматриваться и включаться дословно в отчет испытания. Там, где испытание включает несколько сессий, в каждой сессии формулировки должны быть строго одинаковыми.

## 10 Организация испытательных сессий

### 10.1 Общие положения

Испытательные сессии организуют с целью достижения оптимального баланса среди ряда задач, требования которых в некоторой степени противоречат друг другу. Они включают обеспечение:

- чтобы сравнение продуктов было действительным, полученным от эквивалентных испытателей;
- чтобы сравнение продуктов было действительным, и необъективность за счет положения и за счет последовательности была минимальной;
- чтобы условия испытаний были максимально репрезентативными для нормального потребления;
- чтобы испытатели не были перегружены, что ведет к более случайным ответам.

Чтобы объединить эти разные задачи, существует три способа, которыми испытателей можно связать с продуктами:

a) **полный план:** каждый испытатель оценивает все продукты;

b) **неполный план:** каждый испытатель оценивает подгруппу продуктов;

c) **отдельная оценка:** каждый испытатель оценивает только один продукт.

### 10.2 Планы

#### 10.2.1 Полный план

Оценивание всех продуктов гарантирует их оценку одними и теми же испытателями (поэтому нет проблем эквивалентности), но несет огромный риск перегрузки испытателей.

Существует также опасность эффектов порядка и переноса, в которых положение в последовательности представления продукта влияет на среднее и/или изменчивость его оценки.

Порядок представления продуктов каждому испытателю играет значительную роль. Один способ заключается в использовании **строго** случайного порядка, распределяя образцы независимо для каждого испытателя. Альтернативно, можно принять систематический план представления, чтобы скомпенсировать эффекты порядка и переноса. В то же время, если эффекты переноса представляют проблему, тогда применение периодов отмывания или других методов снижения вероятности эффектов переноса обычно наиболее важны.

#### 10.2.2 Неполный план

Эффекты порядка и меры против них аналогичны полным планам. Здесь опасность перегрузки испытателей представляет меньшую опасность, но недостатками являются возможные эффекты контекста и дополнительная сложность в планировании и анализе эксперимента. Особые проекты существуют для сбалансированных неполных планов (см. [9]).

#### 10.2.3 Отдельная оценка

Чтобы устранить эффекты порядка, переноса и контекста полностью можно набрать отдельную выборку потребителей на каждый продукт. Этот подход ведет к наиболее простым методикам и наиболее простому анализу; в то же время, необходимо составлять выборки очень внимательно, чтобы обеспечить их эквивалентность и репрезентативность для целевой совокупности.

Более того, поскольку этот план делает невозможным учет в анализе индивидуальных различий, прецизионность оценок ниже, чем в других планах. Равноценно, при прочих равных условиях, выборки потребителей должны быть больше, чтобы достичь той же степени доверия и прецизионности в результатах.

### 10.3 Количество продуктов, оцениваемых за одну сессию

#### 10.3.1 Общие положения

Количество продуктов, которое можно оценить эффективно за одну сессию, зависит от следующего:

- характера представленных продуктов. Не все продукты обеспечивают чувство сытости; некоторые продукты (агрессивные) могут изменить чувствительные рецепторы; другие (такие, которые содержат спирт) могут повлиять на корковую интеграцию; некоторые продукты представляют собой один элемент, в состав других (например, готовые блюда) входят различные элементы;
- продолжительности сессии. Длительную лабораторную сессию можно подразделить на сеансы; но потребители, набранные без определения на должность, обычно имеют очень мало времени в распоряжении;
- количества каждого потребленного продукта;
- чувств, задействованных для оценивания продукта (по внешнему виду или тактильным ощущениям можно оценить больше продуктов, чем по вкусу);
- числа задаваемых вопросов, особенно, если какие-либо из них требуют представления продуктов заново.

Следовательно, настоящий стандарт не устанавливает конкретные пределы. Он требует только, чтобы лаборатория была способна обосновать свой выбор перед стороной-инициатором.

#### 10.3.2 Слишком много продуктов для одной сессии

Там, где количество продукции слишком велико, для всех позиций, которые предполагается оценить в отдельной сессии, возможны два подхода:

a) организовать несколько сессий, при этом каждый испытатель примет участие во всех сессиях (полный план);

b)организовать одну сессию, причем каждый испытатель получит только некоторые продукты (неполный план или отдельная оценка).

**10.3.3 Несколько сессий**

Если есть возможность организовать несколько сессий, все испытатели могут оценить все продукты в процессе всего испытания. В идеале, все продукты представляют на каждой сессии, но не все испытатели в конкретной сессии получают одинаковые продукты.

Если ограничения, связанные с подготовкой этих продуктов, или количество имеющихся продуктов сделают такой вариант невозможным, продукты распределяют случайным образом по сессиям. Все испытатели получают за сессию одинаковый набор продуктов, но в различном порядке. Такой выбор обычно требует более сложного статистического анализа.

При возможности в начале каждой сессии следует представлять идентичный продукт (известный как «продукт для разминки») всем испытателям, так чтобы на каждой сессии все испытатели имели одинаковый «отправной пункт».

В один и тот же день не допускается проведение двух сессий с участием одних и тех же испытателей.

**10.3.4 Отдельная сессия**

Сбалансированный неполноблочный план (*BIB*) позволяет при каждой сессии использовать меньшее число наименований, чем общее число наименований, и сравнивать каждую пару образцов продуктов с равной прецизионностью. При возможности следует также исключить или скомпенсировать влияние порядка и переноса в испытании в целом. Такой выбор обычно требует большего числа испытателей и более сложного анализа (см. [9]). На стадии планирования следует предусмотреть консультации специалистов в области статистики, если такой вариант рассматривается.

Метод применения сбалансированного неполноблочного плана должен выполняться в соответствии с ISO 29842.

**10.4 Характер продуктов, оцениваемых за одну сессию**

В одной и той же сессии можно оценивать продукты из разных семейств, если выполняются следующие условия:

- сочетание продуктов соответствует установившейся практике в области пищевых продуктов;
- целевой потребитель для всех семейств один и тот же.

Потребителей, про которых заранее известно, что они не принадлежат к целевой совокупности в отношении рассматриваемого продукта, обычно не выбирают для оценки этого продукта.

Лаборатория должна обеспечить конфиденциальность оценок, независимо от лиц, присутствующих на оценке. Сторона-инициатор при желании может присутствовать на сессии, но ей не сообщают имен других заказчиков, участвующих в сессии.

**10.5 Продолжительность сессии**

Рекомендуется организовывать сессию во время, соответствующее обычному потреблению продуктов, так чтобы потребители (испытатели) находились в условиях, близких к привычному в реальной жизни.

**10.6 Повторение оценки продукта в одном испытании**

Если оценивание продукта необходимо повторить в том же самом испытании, считается, что число продуктов увеличилось на единицу. Тогда план испытания будет включать вместо продукта *A* продукты *A*1 и *A*2, к которым относятся как к двум разным продуктам, как при определении порядка представления, так и для анализа результатов.

**10.7 Предварительное испытание**

Перед основным испытанием можно провести предварительное испытание только с небольшой группой испытателей, чтобы:

- опробовать анкеты для проверки, например, насколько соответствующие отличительные характеристики продукта понятны и воспринимаются испытателями и насколько пригодны оценочные шкалы,
- опробовать приготовления к управлению испытанием,
- обеспечить соответствие продолжительности испытания, количества вопросов и объема выборки.

Испытатели на предварительные испытания должны выбираться из той же совокупности, что и на основное испытание, но не должны входить в выборку испытателей для основного испытания.

## 11 Анализ результатов

### 11.1 Общие положения

Каждый анализ данных должен начинаться проверкой внутренней согласованности данных и расчетами элементарной описательной статистики. Должен быть как минимум один параметр дисперсии (стандартное отклонение, диапазон изменчивости, и т. д.) для каждого параметра положения (среднее, медиана и т. д.), указанного в отчете.

Очень полезно пользоваться графическими методами. Для каждой выборки продуктов следует составить гистограмму поставленных оценок. Она покажет, является ли распределение одновершинным или нет и велики ли расхождения между выборками.

Чтобы получить визуальное представление о согласованности и противоречиях между испытателями, можно использовать Метод главных компонент [Principal Components Analysis (PCA)] или Восходящую иерархическую классификацию [Ascending Hierarchical Classification (AHC)].

Может потребоваться дополнительная статистика вывода, чтобы ответить на вопросы, такие как: «Ощущают ли испытатели заметное различие между продуктами, которые они оценивают?»

Можно использовать параметрические и непараметрические методы, причем непараметрические методы особенно подходят в случае отсутствия нормальных распределений.

### 11.2 Анализ выводов

Анализ выводов экстраполирует результаты, полученные от выборки испытателей к совокупности, из которой они набраны. Метод, который необходимо использовать, будет зависеть от цели испытания и от количества испытываемых продуктов. Если несколько из испытуемых продуктов имеют особый статус, являясь, например, контрольными образцами, это тоже повлияет на анализ.

Применение плана представления продуктов с целью компенсации систематических погрешностей не устраняет систематические погрешности, но обеспечивает только, чтобы ни одному из продуктов при представлении не были несправедливо предоставлены более благоприятные или более невыгодные условия. В то же время анализ выводов может воспользоваться систематическим планом, чтобы улучшить метод оценки и допуска систематических погрешностей и, таким образом, повысить доверие к основным выводам. Однако такой анализ, в общем, не является простым и прямым (см. [4] и [5]).

Стандартные компьютерные пакеты предоставляют все анализы, которые могут потребоваться, но выбор и применение соответствующего анализа требует консультаций со специалистами в области статистики.

### 11.3 Рейтинги (см. 9.2)

Для рейтинговых методов подойдет дисперсионный анализ (ANOVA).

В приложении G описано несколько различных методов анализа данных.

### 11.4 Парные сравнения (см. 9.3)

Если допускается ответ испытателей «предпочтений нет» в парном сравнении (вариант B), одним способом анализа данных является сообщение ряда ответов «предпочтений нет» и выполнение теста на значимость, который использует только предпочтения, которые были выражены. Для выборки уменьшенного размера можно использовать такие же методы, как в случае принудительного выбора. Эти методы приведены в ISO 5495.

### 11.5 Ранжирование (см. 9.4)

Можно использовать различные анализы, касающиеся различных аспектов данных, которые описаны в ISO 8587 (см. также приложение H).

Можно использовать параметрические и непараметрические методы, при этом непараметрические методы практичнее в тех случаях, когда нормальность ответов не внушает доверия. Можно пользоваться методами, приведенными в таблице 1.

Таблица 1 — Методы анализа данных

| Ситуация | | Метод | |
|---|---|---|---|
| | | Непараметрический | Параметрический |
| Два продукта | Продукты были оценены разными группами испытателей | *U*-тест Манна-Уитни (Mann-Whitney's *U*-test) также известный как критерий Уилкоксона-Манна-Уитни (Wilcoxon-Mann-Whitney test) | *t*-тест для независимых выборок |
| | Каждый из продуктов был оценен всеми испытателями | Знаковый (ранговый) критерий Уилкоксона Wilcoxon signed-rank test | t- тест для парных выборок |
| Более двух продуктов | Продукты были оценены разными группами испытателей | Критерий Краскела-Уоллиса (Kruskal-Wallis test) | Дисперсионный анализ (ANOVA) |
| | Каждый испытатель оценил не менее двух из этих продуктов, а в идеале, все продукты | Критерий Фридмана (Friedmann) (ANOVA по рангам), при необходимости в варианте для неполных блоков | Дисперсионный анализ (ANOVA) с эффектом испытателя и эффектом продукта |

## 12 Отчет об испытании

### 12.1 Общие положения

На первой странице отчета можно разместить краткий обзор результатов и наиболее важные выводы.

Отчет необходимо составить так, чтобы он был понятен без ссылки на план испытания. Кроме того, рекомендуется, чтобы каждая таблица и диаграмма результатов были пронумерованы и имели заголовок и емкое пояснение, чтобы можно было разобраться в нем самостоятельно.

Рекомендации, основанные на испытании и запрашиваемые стороной-инициатором, можно привести в разделе, названном «рекомендации» или в отдельном документе.

В дополнение к результатам отчет может включать следующую информацию:

- название испытания и ссылки;
- даты проведения испытания;
- дата составления отчета;
- идентификация стороны-инициатора;
- полная идентификация лаборатории и ответственного за испытание лица (наряду с идентификацией субподрядчиков, если использовались);
- цель испытания (подведение итогов в соответствии с планом испытания);
- процедуры получения доступа к необработанным данным для стороны-инициатора;
- единая идентификация отчета и каждой страницы, включая общее число страниц;
- ссылка на настоящий стандарт.

### 12.2 Продукты

Информация о продукции должна содержать следующее:

- описание продуктов, наряду с указанием полного или частичного состава, если известно от испытательных лабораторий. (Описание может сопровождаться фотографией каждого продукта);
- вся информация, предоставленная стороной-инициатором в отношении образцов, представленных на испытание;
- дата производства образцов, дата «годен до» или «использовать до», номер партии изготовителя или серийный номер;
- дата получения в лаборатории; температура, продолжительность и условия хранения (особенно в случае свежих или замороженных продуктов);
- метод отбора проб, если отбор проб осуществлялся поставщиком услуг. (Если выбирают контрольные или референтные образцы, метод должен быть описан очень точно);
- метод подготовки продуктов для оценки;

- температура продуктов при представлении испытателям;
- количество, предоставленное каждому испытателю, и инструкции, касающиеся минимального и максимального количества для испытания, включая протокол всех отклонений от инструкций;
- сопутствующие материалы или уничтожение;
- все другие продукты, оцененные во время сессии, и порядок их представления на сессию.

### 12.3 Метод испытания

Описание метода испытания должно включать следующее:
- метод испытания со ссылкой на соответствующий стандарт или, в случае нестандартизованных методов, с полным описанием;
- задаваемые вопросы и использованные шкалы ответов;
- метод сбора данных;
- рабочие процедуры для испытания;
- количество сессий и сеансов;
- даты, время и продолжительность сессий;
- условия окружающей среды, например, помещение, температура или освещение;
- использованная столовая посуда;
- инструкции и информация, предоставленные испытателям, и способ представления: на бумаге, на экране или в устной форме.

### 12.4 Испытатели (потребители)

#### 12.4.1 Общие положения

Информация об испытателях должна включать следующее:
- описание целевой совокупности потребителей;
- описание выборки потребителей, включая объем и количество в каждой из категорий, установленных стороной-инициатором;
- заявление о том, производился набор испытателей специально для поставленной задачи или из совокупности потребителей.

#### 12.4.2 Набор под поставленную задачу

Информация о наборе должна включать следующее:
- место или метод набора,
- выполняемые процедуры;
- метод отбора подходящих потребителей.

#### 12.4.3 Набор из имеющейся совокупности

Если набор выполняется из совокупности потребителей, отчет должен включать следующее:
- описание частоты участия в испытаниях данной выборкой потребителей;
- таблица с указанием всех семейств продуктов, которые были предварительно оценены членами этой совокупности;
- таблица с указанием частоты участия в предварительных испытаниях, включающих оцениваемый продукт или семейство продуктов, к которому рассматриваемый продукт принадлежит.

Вся информация, требующаяся в данном подразделе, относится к предшествующим 12 мес. Если частота участия равна нулю для всех потребителей, это единственный факт, который требуется внести в отчет.

### 12.5 Результаты

Результаты испытаний должны включать следующее:
- указание о предоставлении необработанных данных и описание их предоставления. Краткое описание необработанных данных с помощью таблиц и графиков;
- численные сводки о результатах, включая прецизионность и все оценки или средние значения;
- методы, используемые для анализа, и интерпретации результатов;
- статистические выводы, сделанные на основе результатов;
- заключения по результатам со ссылкой на цели испытания.

### 12.6 Приложения к отчету

Следующая дополнительная информация может быть приложена в форме приложений к отчету:

образец анкеты или форма для внесения ответов, использованная в испытании;

образец анкеты, использованной для набора потребителей;

стенограммы ответов на любые открытые вопросы (допускающие неограниченное количество ответов).

**Приложение А**
**(справочное)**

## Пример анкеты для набора участников

| № | Проект контактного интервью **XY 04/05** | | Фильтр | Код |
|---|---|---|---|---|
| 1 | Здравствуйте. Сегодня мы проводим анализ и просим Вас ответить на несколько вопросов. Принимали Вы когда-нибудь участие в опросе с целью изучения рынка? | | | |
| | Да | 1 | ► 2 | |
| | Нет | 2 | ► 3 | |
| 2 | Когда Вы последний раз принимали участие в испытании? | | ► 3 | |
| | ***Информация для интервьюера: ЕСЛИ ЭТО БЫЛО МЕНЕЕ ПОЛУГОДА НАЗАД*** ► **КОНЕЦ; В ДРУГОМ ВАРИАНТЕ ПРОДОЛЖАЮТ С ВОПРОСА 3** | | | |
| 3 | Работали Вы, Ваши друзья или родственники в одной из следующих отраслей промышленности?<br>***Информация для интервьюера: ПРЕДОСТАВИТЬ ПЕРЕЧЕНЬ «ОТРАСЛЕЙ»*** | | | |
| | Отрасль 1 | 1 | ► 4 | |
| | Отрасль 2 | 2 | ► 4 | |
| | Изучение рынка | 3 | КОНЕЦ | |
| | Отрасль 3 | 4 | ► 4 | |
| | Отрасль 4 | 5 | КОНЕЦ | |
| | Отрасль 5 | 6 | ► 4 | |
| | Реклама | 7 | КОНЕЦ | |
| | Журналистика | 8 | КОНЕЦ | |
| | Маркетинг | 9 | КОНЕЦ | |
| | | | | |
| | Ни в одной из вышеперечисленных | 10 | ► 4 | |
| 4 | Можно узнать Ваш возраст? | | QUOT. | |
| | От 20 до 39 лет | 1 | ►5 | |
| | От 40 до 60 лет | 2 | ►5 | |
| | Ваш пол: | | | |
| | Мужской | 1 | ►5 | |
| | Женский | 2 | ►5 | |
| 5 | У нас есть перечень различных пищевых продуктов ежедневного пользования. Который из этих продуктов Вы покупали недавно?<br>***Информация для интервьюера: ПРЕДОСТАВИТЬ ПЕРЕЧЕНЬ «ПИЩЕВЫЕ ПРОДУКТЫ»*** | | | |
| | Пищевой продукт 1 | 1 | | |
| | Пищевой продукт 2 | 2 | | |
| | Пищевой продукт 3 | 3 | ►6 | |
| | Пищевой продукт 4 | 4 | | |
| | Пищевой продукт 5 | 5 | | |
| | Пищевой продукт 6 | 6 | | |
| | Пищевой продукт 7 | 7 | | |
| | Пищевой продукт 8 | 8 | | |
| | ***Информация для интервьюера: ЕСЛИ «3» НЕ УПОМЯНУТО ►КОНЕЦ*** | | | |

*Продолжение таблицы*

| № | Проект контактного интервью **XY 04/05** | Фильтр | Код |
|---|---|---|---|
| 6 | Вы только что сказали, что используете пищевой продукт № 3. Примерно как часто Вы используете пищевой продукт № 3?<br>***Информация для интервьюера: ПРЕДОСТАВИТЬ ПЕРЕЧЕНЬ «ЧАСТОТА ИСПОЛЬЗОВАНИЯ»***<br>Более одного раза в день 1<br>Ежедневно 2 | ►7 | |
| | Более одного раза в неделю 3<br>Примерно раз в неделю 4<br>Примерно раз в две недели 5 | | |
| | Примерно раз в месяц 6<br>Примерно раз в три месяца 7<br>Реже указанного выше 8 | ►КОНЕЦ | |
| 7 | Каких брендов пищевой продукт № 3 вы регулярно покупаете для дома, т. е. не реже раза в месяц?<br>***Информация для интервьюера: ЗАДАТЬ ВОПРОС ОБ ИЗГОТОВИТЕЛЕ И ПРОДУКТЕ! ЕСЛИ УПОМЯНУТ ТОЛЬКО ИЗГОТОВИТЕЛЬ, ЗАДАТЬ ВОПРОС СНОВА!***<br>Бренд изготовителя Продукт<br>1) _____ ►1) _____<br>2) _____ ►2) _____<br>3) _____ ►3) _____<br>4) _____ ►4) _____<br>5) _____ ►5) _____ | ►8A | |
| 8A | У нас есть различные бренды пищевого продукта № 3 в этом перечне. Какой из этих брендов Вы используете дома, даже если только случайно? | | |
| | ***Информация для интервьюера: ПРЕДОСТАВИТЬ ПЕРЕЧЕНЬ «БРЕНДЫ» И ОБВЕСТИ НУЖНОЕ В 8A!*** | ►8B | |
| 8B | Ответьте, пожалуйста, как часто Вы используете каждый из указанных брендов? | | |
| | ***Информация для интервьюера: ПРЕДОСТАВИТЬ ПЕРЕЧЕНЬ «ЧАСТОТА ИСПОЛЬЗОВАНИЯ» И, ТОЛЬКО В ТОМ СЛУЧАЕ, ЕСЛИ ОБВЕДЕН ВОПРОС 8A, СДЕЛАТЬ ПОМЕТКУ В СХЕМЕ ПОД ВОПРОСОМ 8B!*** | ►8C | |
| 8C | Какие из этих брендов Вы используете наиболее часто? | | |
| | ***Информация для интервьюера: ВОЗМОЖЕН ТОЛЬКО ЕДИНСТВЕННЫЙ ВЫБОР; ОБВЕСТИ НУЖНОЕ В 8C)!*** | | |
| | 8A 8B 8C<br>Бренд 1 1 _____ 1<br>Бренд 2 2 _____ 2<br>Бренд 3 3 _____ 3<br>Бренд 4 4 _____ 4<br>Бренд 5 5 _____ 5<br>Бренд 6 6 _____ 6<br>Бренд 7 7 _____ 7<br>Бренд 8 8 _____ 8<br>Бренд 9 9 _____ 9<br>Бренд 10 10 _____ 10<br>Бренд 11 11 _____ 11<br>Другая информация, а именно<br>_____ 12 _____ 12 | ►9 | |

*Окончание таблицы*

| № | Проект контактного интервью **XY 04/05** | Фильтр | Код |
|---|---|---|---|
| 9 | Есть ли у Вас аллергия на какие-либо пищевые продукты? | | |
| | Да 1 | ►КОНЕЦ | |
| | Нет 2 | ►10 | |
| 10 | Страдаете ли Вы от диабета? | | |
| | Да 1 | ►КОНЕЦ | |
| | Нет 2 | ОК | |

**Приложение В**
**(справочное)**

## Методы оценивания подгрупп при сегментации

### В.1 Общие положения

В зависимости от типа испытания полезно сравнивать подгруппы целевой совокупности. Например, иногда интересно определить, имеются ли различия между подгруппами и, если да, то на каких уровнях приемлемости.

На основе всей выборки потребителей возможны два пути сегментации:

a) **Подгруппы совокупности:** подгруппы представляют собой части выборки, для которых должны проверяться или отвергаться представляющие интерес гипотезы. Эти подгруппы следует установить до набора потребителей. Для каждого установленного критерия и для каждой части выборки необходимо установить размер, так чтобы отдельные подгруппы включали достаточное количество потребителей (не менее 60 человек), чтобы получить желаемую точность измерения при оценивании данных.

Примечание — В зависимости от критериев выбора подгруппы необязательно должны быть одинакового размера.

b) **Сегментация потребителей:** подгруппы должны формироваться по оценке всех имеющихся данных (результаты измерений и/или социально-демографические данные). Такая сегментация служит для описания выборки потребителей и может использоваться для разработки новой гипотезы. Последующая сегментация будет целесообразна только в случае, когда выборка потребителей имела достаточный объем.

В обоих случаях необходимое количество потребителей в испытании зависит от количества сегментов, выбранных или ожидаемых, соответственно, и согласно принимаемой дисперсии (разброса) внутри отдельных сегментов.

### В.2 Подгруппы совокупности

Сравнение подгрупп используется для получения информации в отношении вопросов следующего типа.

- «Имеют ли одинаковые предпочтения мужчины и женщины?»
- «Зависят ли предпочтения от возрастной группы?»
- «Зависят ли предпочтения от страны?»
- «Зависят ли предпочтения от традиций потребления?»

В случае, если используется дисперсионный анализ и каждый из испытателей оценивал каждый образец, применяется следующая модель:

Ответ = группа + испытатель (группа) + образец • группа + погрешность

Подгруппы тестируют по факторному потребителю в группе, тогда как другие факторы тестируются в отношении погрешности [это равно влияющий образец , испытатель (группа)].

В данную модель можно включить эффект порядка, если необходимо, а также эффект переноса. При включении эффекта порядка модель выглядит следующим образом:

Ответ = группа + испытатель (группа) + образец + порядок + образец • группа + порядок • группа + образец • порядок + погрешность

С помощью такого анализа можно определить, располагают ли разные группы образцы по-разному (влияющий образец • группа), если эффект порядка одинаков для всех групп (влияющий порядок • группа) и для всех образцов (влияющий образец • группа).

Если существует заметное различие между группами в отношении образцов (влияющий образец • группа), то каждую из подгрупп оценивают по отдельности.

### В.3 Сегментация потребителей

Последовательная сегментация имеет целью объяснить выявленные типы предпочтений, если имеются.

Это, в частности, применимо, если, при проверке гистограмм были обнаружены бимодальности.

Внутреннее отображение предпочтений дает возможность анализа данных без учета средних по образцам по потребителям. Это, фактически, анализ основных компонентов матрицы градаций, в которой строки представляют продукты, а столбцы потребителей.

Примечание — Анализ основных компонентов: метод для анализа многомерных данных с целью выражения их изменчивости в минимальном количестве основных компонентов или линейных комбинаций исходных или частично скоррелированных переменных.

«Одновременное» представление позволяет совмещать испытателей и продукты.

В то же время информация, предоставленная первыми несколькими компонентами достоверна только в том случае, если эти компоненты объясняют высокий процент исходной вариации. Эта процедура, в общем,

применима только тогда, когда используются полные планы (см. [11]). В то же время, предложен метод для неполных планов испытаний (см. [10]).

Таким образом, при визуальном изучении первых компонентов в анализе основных компонентов, можно различить группы потребителей (кластеры), если они присутствуют. Чтобы идентифицировать кластеры потребителей, можно использовать методы классификации.

**Приложение С**
**(справочное)**

## Планы распределения продуктов по испытателям

Полностью достоверным и даже рекомендуемым предложением по распределению продуктов по испытателям является рандомизированный полноблочный план. Это означало бы, что каждый испытатель испытывает каждый продукт в точности один раз. Порядок представления рандомизирован независимо для каждого испытателя. Преимущество этого предложения заключается в том, что оно подтверждает правильность анализа в простой блок-модели.

Существуют, однако, причины, почему в большинстве случаев экспериментатор может пожелать использовать другой, более сложный план.

- Неполноблочные планы: Если число образцов слишком велико, то иногда невозможно каждому испытателю оценить каждый продукт. Число образцов, которые может испытать один испытатель, зависит от характера продукции. В таком случае, предпочтительно использовать **сбалансированный неполноблочный план,** см. ISO 29842. И снова порядок следования продуктов обычно рандомизируют, независимо для каждого потребителя.

- Эффекты порядка (представления): Зачастую экспериментаторы обеспокоены присутствием эффекта порядка. Хорошо известно, что продукты, которые опробовали раньше, обычно получают лучшую оценку, чем продукты, представленные позже. Одной из причин, по которой полноблочный план и сбалансированный неполноблочный план требуют рандомизации, является этот **эффект порядка:** При рандомизации порядка представления продуктов на испытание можно избежать систематических погрешностей за счет эффектов порядка. Рандомизация гарантирует каждому продукту равные шансы быть представленным раньше.

Если экспериментатор опасается, что эффекты порядка могут оказаться слишком сильными, подходящей альтернативой рандомизированному полноблочному плану будет обобщенный латинский квадрат. Здесь, представление продуктов испытателям организовано таким образом, чтобы каждый продукт одинаково часто попадал на оценку первым, вторым и т. д.

В случае неполноблочного плана, также можно организовать порядок представления продуктов таким образом, чтобы каждый продукт одинаково часто попадал в каждую позицию в очереди. Планы такой структуры называют обобщенными планами Юдена (Youden).

Следует отметить, однако, что применение обобщенного латинского квадрата или обобщенного плана Юдена не подтверждает анализ в простой блок-модели. Вместо этого следует использовать модель строк-столбцов, которая учитывает эффекты порядка.

Эффекты переноса: несмотря на применение периодов отмывки, может оказаться, что опробуемый продукт влияет на вкус продукта, который испытывают следующим. Если испытатель обеспокоен эффектами переноса, имеет смысл использовать так называемый план «сбалансирования ближнего соседа», см., например, [2]. Некоторые возможности анализа планов «сбалансирования ближнего соседа» описаны в [2] и [9].

Организационные вопросы: в некоторых случаях существует организационные вопросы, которые требуют более сложной структуры плана. Например, может случиться, что имеется восемь продуктов для сравнения и только четыре нагревательные плитки, так что в любой сессии можно представить только четыре продукта. Подобные ситуации требуют более сложного, нестандартного плана, и, соответственно, более сложного, нестандартного анализа. В таких случаях испытатель должен обратиться за помощью к специалисту в области статистики.

**Приложение D**
**(справочное)**

## Влияние объема выборки на точность одного среднего значения

Каждая группа испытателей должна иметь определенный минимальный размер, так чтобы различия могли стать очевидными при оценке продуктов и, что более важно, чтобы позволить среднему значению оценок, данных группой испытателей для опробуемого продукта, иметь определенную точность. То же самое применимо к полной выборке потребителей или к классу выборок, или к сегменту, выявленному при изучении индивидуальных различий.

Точность зависит от размера выборки и от изменчивости оценок, выставленных в группе.

При допущении, что оценки, выставленные продукту, распределятся нормально со стандартным отклонением, $\sigma$, минимальный размер выборки испытателей, требующихся для оценки среднего значения с желательной точностью, $A_c$, можно определить по формуле (D.1):

$$A_c = \frac{2\sigma}{\sqrt{n}}, \qquad (D.1)$$

где $A_c$ равно половине доверительного интервала при $P = 0,95$ от среднего.

Например, $A_c = 0,2$ означает, что доверительный интервал от среднего значения составляет от ($m$ – 0,2) до ($m$ + 0,2). В таблице D.1 приведены взаимосвязь между $\sigma$, $n$, и $A_c$.

Таблица D.1 — Размер выборки, требующийся для получения желаемой точности

| Точность $A_c$ | Стандартное отклонение $\sigma$ | | |
|---|---|---|---|
| | 1,5 | 2,0 | 2,5 |
| 0,2 | 225 | 400 | 625 |
| 0,3 | 100 | 178 | 278 |
| 0,4 | 56 | 100 | 156 |
| 0,5 | 36 | 64 | 100 |

$\sigma$ и $A_c$ должны выражаться в одинаковых единицах.

Но чтобы определить число потребителей, которых необходимо набрать, рекомендуется учитывать β-риск (см. приложения E и F).

Примечание — Анализ статистической мощности следует проводить, когда принимают решение о требующемся объеме выборки перед испытанием: $n$ определено не будет, пока не будет зафиксирована величина эффекта (ES) (Cohen[8]).

## Приложение Е
## (справочное)

## Минимальный размер выборки для рейтинговых тестов

### Е.1 Общие положения

Разумное требование к объему выборки в испытаниях с привлечением потребителей предусматривает получение не менее 60 оценок, а предпочтительно порядка 100 оценок, на продукт. В том случае, когда каждый испытатель испытывает только один продукт, потребуется $n = 100 \cdot p$ испытателей, где $p$ — количество продуктов в испытании. Если, однако, каждый испытатель испытает $k$ продуктов, то грубо-эмпирическим методом можно подсчитать, что потребуется $n = 100 \cdot (p/k)$ потребителей.

Однако это правило грубого приближения. Существуют более сложные методы. Если экспериментатор желает ими воспользоваться, ему нужно ответить на ряд вопросов.

- Какова цель испытания? Продемонстрировать, что существуют некоторые различия между одним продуктом и одним опорным значением или между двумя или более продуктами? Или показать, что существует подобие между одним испытуемым продуктом и контрольным продуктом или между двумя и более продуктами?
- Сколько продуктов предложено на испытание? Один? Два? Три и более?
- Сколько продуктов предполагается представить одному испытателю?
- Каков выбранный α-риск?
- Какова желаемая мощность $P$ или β-риск для испытания? [или: Каков приемлемый β-риск? ($P = 1 - \beta$-риск)]
- Каково значение различия $\delta$, которое предполагается выявить в тесте на проверку различия? Чему равно неприемлемое значение разности $\Delta$ в тесте на проверку подобия?
- Каков разброс оценок в рамках испытания продукта (продуктов)?
- Является ли альтернативная гипотеза односторонней или двухсторонней?
- Какова взаимосвязь между рейтингами, присвоенными разным продуктам испытателями, когда испытатели оценивают несколько продуктов?

Данное приложение может не охватить все возможные случаи. Оно рассматривает только испытание на двух продуктах. Для других случаев, более простого (только один продукт) или более сложного (три или более продуктов), читатель (пользователь данного стандарта) должен проконсультироваться со специалистом по статистике или использовать соответствующую компьютерную программу. Читатель также может обратиться к документу [2].

Данное приложение рассматривает четыре случая: два случая проверки различия и два случая проверки подобия.

### Е.2 Случай 1: Цель — продемонстрировать различие при оценке испытателями только продукта А или только продукта В

Количество испытателей на каждый продукт $n$ задается формулой (Е.1):

$$n = 2 \cdot \left[ \frac{\left( u_{\alpha/2} + u_{\beta} \right) \cdot \sigma_{within\ products}}{\delta} \right]^2, \qquad (Е.1)$$

где $u_{\alpha/2}$ — квантиль $(1 - \alpha/2)$ нормального распределения для α-риска и для двухсторонней альтернативной гипотезы; равен 1,960 для α-риска = 5 %. Для односторонней альтернативной гипотезы, $u_{\alpha/2}$ заменяют на $u_{\alpha}$; для α-риска = 5 %; он равен 1,645;

$u_{\beta}$ — квантиль $(1 - \beta)$ нормального распределения для допустимого β (желаемая мощность $P = 1 - \beta$);

$\sigma_{within\ products}$ — стандартное отклонение в рамках испытания продуктов; предполагается однородное стандартное отклонение для рассматриваемых двух продуктов;

$\delta$ — различие, выявляемое между двумя продуктами.

Примечание — $\sigma_{within\ products}$ и $\delta$ измеряются по одной шкале, например, категориальной шкале от 1 до 9. Категории от 1 до 9 категориальной шкалы дают оценки приблизительно согласующиеся с нормальным распределением.

***Пример — Для следующих условий:***

| $\alpha$ | $u_{\alpha/2}$ | $\beta$ | $(P = 1 - \beta)$ | $u_{\beta}$ | $\sigma_{within\ products}$ | $\delta$ |
|---|---|---|---|---|---|---|
| *0,05* | *1,960* | *0,10* | *(0,90)* | *1,282* | *1,50* | *0,50* |

***Формула (Е.1) дает n = 189,2, использование 190 испытателей на продукт и 380 испытателей на два продукта.***

***Примечание — Для односторонней гипотезы при А > В, $u_{\alpha = 1,645}$ (для $\alpha$ = 5 %) формула (Е.1) дает n = 154,2, а именно 155 испытателей.***

**Е.3 Случай 2: Цель: продемонстрировать различие при оценке испытателями обоих продуктов, продукта А и продукта В**

Количество испытателей задается формулой (Е.2):

$$n = 2 \cdot \left[ \frac{\left(u_{\alpha/2} + u_{\beta}\right) \cdot \sigma_{within\ products} \cdot \sqrt{1-\rho}}{\delta} \right]^2, \qquad (Е.2)$$

где $u_{\alpha/2}$ — квантиль $(1 - \alpha/2)$ нормального распределения для α-риска для двухсторонней альтернативной гипотезы; равен 1,960 для α-риска = 5 %. Для односторонней альтернативной гипотезы, $u_{\alpha/2}$ заменяют на $u_{\alpha}$; для α-риска = 5 %, $u_{\alpha}$ равен 1,645;

$u_{\beta}$ — квантиль $(1 - \beta)$ нормального распределения для допустимого β-риска (желаемая мощность, $P = 1 - \beta$);

$\sigma_{within\ products}$ — стандартное отклонение в рамках испытания продуктов; предполагается однородное стандартное отклонение для рассматриваемых двух продуктов;

$\delta$ — различие, выявляемое между двумя продуктами;

$\rho$ — корреляция в совокупности между оценками, присвоенными двум рассматриваемым продуктам испытателями.

Примечание — $\sigma_{within\ products}$ и $\delta$ выражаются в единицах одной шкалы, например по категориальной шкале от 1 до 9 категории.

Поскольку $\sigma_d = 2 \cdot \sigma_{within\ products} \cdot \sqrt{(1-\rho)}$, где $\sigma_d$ является стандартным отклонением различий между рейтингами, присвоенными двум рассматриваемым продуктам каждым испытателем совокупности, формула (Е.2) принимает вид:

$$n = \left[ \frac{\left(u_{\alpha/2} + u_{\beta}\right) \cdot \sigma_d}{\delta} \right]^2. \qquad (Е.3)$$

Возможно формула (Е.3) является наиболее часто используемым выражением.

***Пример — Для следующих условий***

| $A$ | $u_{\alpha/2}$ | $\beta$ | $(P = 1 - \beta)$ | $u_{\beta}$ | $\sigma_d$ | $\delta$ |
|---|---|---|---|---|---|---|
| *0,05* | *1,960* | *0,10* | *(0,90)* | *1,282* | *1,70* | *0,50* |

***Формула (Е.3) дает n = 121,5; используют 122 испытателя.***

***Примечание — Для односторонней гипотезы при A > B, $u_{\alpha}$ = 1,645 (для α = 5 %) формула (Е.3) дает n = 99,0.***

**Е.4 Случай 3: Цель: продемонстрировать, что продукт *А* не уступает продукту *В* (не хуже), причем каждый испытатель оценивает только продукт *А* или продукт *В***

Количество испытателей для каждого продукта задается формулой (Е.4):

$$n = 2 \cdot \left[ \frac{\left(u_{\alpha} + u_{\beta}\right) \cdot \sigma_{within\ products}}{\Delta} \right]^2, \qquad (Е.4)$$

где $u_{\alpha}$ — квантиль $(1 - \alpha)$ нормального распределения для $\alpha$-риска для односторонней альтернативной гипотезы (испытание типа «не хуже»);

$u_{\beta}$ — квантиль $(1 - \beta)$ нормального распределения для допустимого $\beta$-риска (желаемая мощность $P = 1 - \beta$);

$\sigma_{within\ product}$ — стандартное отклонение в рамках испытания продукта; предполагается однородное стандартное отклонение для рассматриваемых двух продуктов;

$\Delta$ — различие, не приемлемое для того, чтобы считать два продукта подобными (взаимозаменяемыми).

Примечание — $\sigma_{within\ products}$ и $\Delta$ выражают в единицах одной шкалы, например категориальной шкалы от 1 до 9 категории.

***Пример — Для следующих условий***

| $\alpha$ | $u_{\alpha}$ | $\beta$ | $(P = 1 - \beta)$ | $u_{\beta}$ | $\sigma_{within\ products}$ | $\Delta$ |
|---|---|---|---|---|---|---|
| *0,05* | *1,645* | *0,10* | *(0,90)* | *1,282* | *1,40* | *0,30* |

***Формула (Е.4) даст n = 373,2; используют 374 испытателя на продукт и 748 испытателей для двух продуктов.***

**Е.5 Случай 4: Цель: продемонстрировать, что продукт *A* не уступает продукту *B* (не хуже), причем каждый испытатель оценивает оба продукта — и продукт *A*, и продукт *B*.**

Количество испытателей задается формулой (Е.5):

$$n = 2 \cdot \left[ \frac{\left(u_{\alpha} + u_{\beta}\right) \cdot \sigma_{within\ products}}{\Delta} \right]^2, \qquad (E.5)$$

где $u_{\alpha}$ — квантиль $(1 - \alpha)$ нормального распределения для $\alpha$-риска для односторонней гипотезы (испытание типа «не хуже»);

$u_{\beta}$ — квантиль $(1 - \beta)$ нормального распределения для $\beta$-риска (желаемая мощность $P = 1 - \beta$);

$\sigma_d$ — стандартное отклонение различий между рейтингами, присвоенными двум рассматриваемым продуктам каждым испытателем совокупности;

$\Delta$ — различие, не приемлемое для того, чтобы считать два продукта подобными (взаимозаменяемыми).

Примечание — $\sigma_d$ и $\Delta$ выражают в единицах одной шкалы, например, категориальной шкалы от 1 до 9.

***Пример — Для следующих условий***

| $\alpha$ | $u_{\alpha}$ | $\beta$ | $(P = 1 - \beta)$ | $u_{\beta}$ | $\sigma_d$ | $\Delta$ |
|---|---|---|---|---|---|---|
| *0,05* | *1,645* | *0,10* | *(0,90)* | *1,282* | *1,70* | *0,30* |

***Формула (Е.5) дает n = 275,1; используют 276 испытателей.***

Примечание — Случаи 3 и 4 являются испытаниями типа «не хуже». Проверки на строгое подобие, в которых лаборатория должна подтвердить, что продукт *A* не хуже или лучше продукта *B*, не рассматриваются. Расчет количества испытателей для таких испытаний является сложным (не существует выражения, которое непосредственно даст это количество) и дает высокие значения для числа испытателей. Например, число испытателей равно 480 на продукт для условий, например, случая 3 (960 испытателей для двух продуктов) и 698 для примера в случае 4, соответственно. Для испытаний на строгое подобие предполагают, что лаборатория использует соответствующие компьютерные программы. Необходимо отметить, что [3] рассматривает проверку на строгое подобие, только если они выполняются методом парного сравнения, когда каждый испытатель получает оба продукта, A и B.

**Приложение F**
**(справочное)**

## Минимальный объем выборки для ранжирования

### F.1 Общие положения

Метод ранжирования может включать два или более продуктов. По аналогии с приложением E, рассматривающим рейтинговые тесты, приложение F ограничено представлением выражений для расчетов количества испытателей в случае ранжирования относительно двух продуктов. В отношении ранжирования для трех и более продуктов лаборатория может проконсультироваться со специалистами в области статистики.

В сфере гедонических измерений метод ранжирования для двух продуктов называют методом предпочтения. Каждый испытатель получает два продукта, *A* и *B*; он должен определить тот продукт, который ему нравится больше: *A* или *B*.

### F.2 Случай 1: Продемонстрировать предпочтение в отношении двух продуктов — *A* и *B*

Количество испытателей $n$ задается формулой (F.1):

$$n = \left[ \frac{\left( u_{\alpha/2} \cdot \sqrt{p_0 \cdot (1-p_0)} + u_\beta \cdot \sqrt{p_a \cdot (1-p_a)} \right)}{(\delta / 2)^2} \right]^2, \quad \text{(F.1)}$$

где $u_{\alpha/2}$ — квантиль $(1 - \alpha/2)$ нормального распределения, соответствующий риску α, выбранному в случае двусторонней альтернативной гипотезы. Для $\alpha = 5\ \%$, $u_{\alpha/2} = 1{,}960$. В случае односторонней гипотезы, $u_{\alpha/2}$ заменяют на $u_\alpha$ (где: $u_\alpha = 1{,}645$ для $\alpha = 5\ \%$);

$u_\beta$ — квантиль $(1 - \beta)$ нормального распределения, соответствующий требуемой мощности $P$, где $P = 1 - \beta$;

$p_0$ — вероятность предпочтения, соответствующая нулевой гипотезе $H_0$; всегда равна 0,50 в тесте на предпочтение;

$p_a$ — вероятность предпочтения более понравившегося продукта, соответствующая альтернативной гипотезе $H_a$; равна $p_0 + \delta/2$;

$\delta$ — различие в предпочтении между двумя продуктами, считающееся заметным.

Примечания

1 Формула (F.1) основана на аппроксимации биномиального распределения нормальным распределением; такая аппроксимация действительна постольку, поскольку две величины $n \cdot p_a$ и $n \cdot (1 - p_a)$ как минимум равны 5.

2 В настоящем стандарте значение $\delta$ равно разности между вероятностями предпочтения более понравившегося продукта и менее понравившегося продукта. Некоторые авторы выражают значение $\delta$ в отношении к среднему от этих двух вероятностей, т. е. в отношении $p_0$. Читатель должен обратить внимание, чтобы не спутать эти два способа выражения.

***Пример — Для следующих условий***

| $p_0$ | δ | $p_a$ | *A* | $u_\alpha$ | β | *P* | $u_\beta$ |
|---|---|---|---|---|---|---|---|
| *0,50* | *0,20* | *1,14* | *0,05* | *1,960* | *0,10* | *0,90* | *1,282* |

***Формула (F.1) дает n = 261,8, т. е. 262 испытателя.***

Если лаборатория анкетирует 262 потребителя и существует расхождение в частоте предпочтения, равное 0,20 между двумя продуктами *A* и *B*, можно с определенностью сделать вывод, что различие в предпочтении происходит в среднем в 9 из 10 испытаний.

Примечание — В случае односторонней гипотезы при $A > B$, формула (F.1) дает $n = 235{,}6$. Если лаборатория опрашивает 236 испытателей и, если истинное предпочтение для *A* равно 0,60, можно с определенностью сделать заключение о значимом предпочтении в среднем в 9 из 10 испытаний.

### F.3 Случай 2: Продемонстрировать, что частота предпочтения для продукта *A* не меньше частоты предпочтения для продукта *B*

Количество испытателей задается формулой (F.2):

$$n = \left[ \frac{\left( u_\alpha \cdot \sqrt{p_a \cdot (1-p_a)} + u_\beta \cdot \sqrt{p_0 \cdot (1-p_0)} \right)}{(\Delta / 2)^2} \right]^2, \quad \text{(F.2)}$$

где $u_\alpha$ — квантиль $(1 - \alpha)$ центрированного и приведенного нормального распределения, соответствующий выбранному риску, α. В испытании по типу «не хуже» гипотеза односторонняя, поэтому $\alpha = 5\ \%$, $u_\alpha = 1{,}645$;

$u_\beta$ — квантиль $(1 - \beta)$ нормального распределения, соответствующий мощности $P$, требуемой для испытания, где $\beta = 1 - P$;

$\Delta$ — разность вероятностей предпочтения, которая не позволяет считать продукт $A$ «не хуже», чем продукт $B$;

$p_a$ — вероятность предпочтения продукта $A$ в случае нулевой гипотезы в испытании по типу «не хуже»; равная $p_0 - \Delta/2$;

$p_0$ — вероятность предпочтения для обоих продуктов, когда не имеется разницы в предпочтении между ними; она равна 0,50.

Примечания

1 Формула (F.2) основана на аппроксимации биномиального распределения нормальным распределением.

2 В настоящем стандарте значение $\Delta$ равно разности двух вероятностей $p_a$ и $(1 - p_a)$. Некоторые авторы выражают значение $\Delta$ по отношению к среднему от двух этих величин, т. е. по отношению к вероятности $p_0$. Читатель должен быть внимательным, чтобы не спутать два эти способа выражения.

***Пример — Для следующих условий***

| $\Delta = \lvert p_a - (1 - p_a) \rvert$ | $p_a$ | $p_0$ | $A$ | $u_\alpha$ | $\beta$ | $P$ | $u_\beta$ |
|---|---|---|---|---|---|---|---|
| *0,10* | *0,45* | *0,50* | *0,05* | *1,645* | *0,10* | *0,90* | *1,282* |

***Формула (F.2) дает n = 851,9, т. е. 852 испытателя.***

**Приложение G**
**(справочное)**

## Примеры анализа данных для рейтингов

### G.1 Общие положения

Данное приложение использует примеры, описанные в приложении Е. Первые два примера касаются проверки различия, где нулевую гипотезу $H_0$, можно определить как: два продукта *A* и *B* извлекают из двух множеств, которые имеют одинаковое среднее.

Последние два примера касаются испытания по типу «не хуже», где нулевая гипотеза $H_0$, может быть определена как: продукт *A* хуже продукта *B* на величину, по крайней мере, равную Δ; следовательно, продукт *A* не равноценен продукту *B*.

Для проверки различия данные примеры рассматриваются в двухсторонней гипотезе. В случае односторонней гипотезы необходимо консультироваться со специалистом в области статистики.

### G.2 Пример 1 (случай 1 приложения Е): подтвердить, что два продукта *A* и *B* воспринимаются по-разному; каждый испытатель дает оценку только одному продукту

- После обсуждения с заказчиком лаборатория опросила 230 потребителей: 110 потребителей в отношении продукта *A* ($n_A$ = 110) и 120 потребителей в отношении продукта *B* ($n_B$ = 120). Для двух продуктов лаборатория получила средние значения и стандартные отклонения, а именно: $m_A$ = 7,23, $s_A$ = 1,85, $m_B$ = 6,87, and $s_B$ = 1,65.

- Значение $t_{cal}$ задается формулой (G.1):

$$t_{cal} = \frac{m_A - m_B}{s_{within\ products} \cdot \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}, \tag{G.1}$$

где $s_{within\ products}$ — стандартное отклонение в пределах испытанных продуктов.

$s_{within\ products}$ рассчитывают по формуле (G.2):

$$s_{within\ products}{}^2 = \frac{s_A^2 \cdot (n_A - 1) + s_B^2 \cdot (n_B - 1)}{n_A - n_B - 2}, \tag{G.2}$$

Следовательно, в данном примере:

$$s_{within\ products}{}^2 = \frac{1{,}85^2 \cdot (110 - 1) + 1{,}65^2 \cdot (120 - 1)}{110 - 120 - 2},$$

$$s_{within\ products} = \sqrt{s_{within\ products}{}^2} + \sqrt{3{,}057} = 1{,}748.$$

Значение $t_{cal}$: $t_{cal} = \frac{7{,}23 - 6{,}87}{1{,}748 \cdot \sqrt{\left(\frac{1}{100} + \frac{1}{120}\right)}} = \frac{0{,}36}{0{,}231} = 1{,}558.$

Это значение сравнивают с заданным значением $t_{the}$ для выбранного α-риска. Это значение $t_{the}$ задается квантилем $(1 - \alpha/2)$ *t*-распределения с $n_A + n_B - 2$ степенями свободы. Для α-риска = 5 %, в данном примере оно равно 1,970.

Чтобы отвергнуть $H_0$, $t_{cal}$ должно быть > $t_{the}$. Поскольку 1,558 < 1,970, это условие не подтверждается в данном примере; лаборатория не может сделать заключение о том, что существует заметное различие между двумя средними значениями продуктов *A* и *B*.

- Поэтому лаборатория рассчитывает эффективную мощность критерия по формуле (G.3):

$$P = 1 - \beta, \tag{G.3}$$

Значение $t_\beta$ рассчитывают по формуле (G.3):

$$t_\beta = t_{the} - \left(\frac{\delta}{t_{calden}}\right), \tag{G.4}$$

где $t_{the}$ — значение, ведущее к отвержению $H_0$;

$\delta$ — различие, выявляемое между двумя продуктами;

$t_{calden}$ — знаменатель $t_{cal}$ [формула (G.1)].

$$t_\beta = 1{,}970 - (0{,}50/0{,}231) = -0{,}195.$$

Если $t_\beta < 0$, значение $\beta$ задается функцией распределения $t$-распределения при $n_A + n_B - 2$ степенях свободы при $t_\beta$. Функция распределения $t$-распределения с 110 + 120 − 2 степенями свободы при −0,195 равна 0,423.Таким образом, мощность критерия будет равна 1 − β = 1 – 0,423 ≈ 0,58. Это неудовлетворительное значение гораздо меньше, чем требуемое значение: 90 %.

Примечание — Если $t_\beta > 0$, то значение мощности напрямую задается функцией распределения $t$-распределения при $(-1)\cdot t_\beta$. Например, для $\delta$ = 0,30, $t_\beta$ = 1,970 – (0,30/0,231) = + 0,671; расчет функции $t$-распределения дает $P$ = 0,251 ≈ 0,25.

### G.3 Пример 2 (случай 2 приложения Е): подтвердить, что два продукта *A* и *B* воспринимаются по-разному; каждый испытатель дает оценку обоим продуктам

Лаборатория опросила 120 потребителей ($n$ = 120). Получено $m_A$ = 6,33, $m_B$ = 6,66 и $s_d$ = 1,92, где $s_d$ — стандартное отклонение разностей между рейтингами, присвоенными двум продуктам каждым потребителем.

- Значение $t_{cal}$ задается формулой (G.5):

$$t_{cal} = \frac{m_A - m_B}{s_d \cdot \sqrt{\left(\frac{1}{n}\right)}}, \tag{G.5}$$

$$t_{cal} = \frac{m_A - m_B}{s_d \cdot \sqrt{\left(\frac{1}{n}\right)}} = \frac{6,33 - 6,66}{1,92 \cdot \sqrt{\left(\frac{1}{120}\right)}} = \frac{0,33}{0,175} = -1,886,$$

- Это значение сравнивают со значением $t_{the}$ для выбранного $\alpha$-риска. Это значение $t_{the}$ задается квантилем $(1 - \alpha/2)$ $t$-распределения с $n - 1$ степенями свободы. В этом примере для $\alpha$-риска = 0,05 оно равно 1,980.

Чтобы отвергнуть $H_0$ и сделать вывод, что существует заметное различие между *A* и *B*, необходимо, чтобы абсолютное значение $t_{cal}$ было > $t_{the}$. Поскольку это условие не подтверждается в данном примере [abs(−1,886) < < 1,980], лаборатория не может прийти к заключению о заметном различии между двумя рейтингами продуктов *A* и *B*.

- Поэтому лаборатория рассчитывает эффективную мощность критерия, который она применила, по формуле

$$t_\beta = t_{the} - \frac{\delta}{t_{calden}}, \tag{G.6}$$

где $t_{the}$ — значение, ведущее к отвержению $H_0$;

$\delta$ — различие, выявляемое между двумя продуктами;

$t_{calden}$ — знаменатель $t_{cal}$ [формула (G.5)].

$t_\beta$ = 1,980 – (0,50/0,175) = –0,873.

Поскольку правила, касающиеся расчета β, одинаковы для формул (G.4) и (G.6), значение β задается функцией $t$-распределения с 120 − 1 степенями свободы при –0,873; оно равно 0,192. Мощность критерия поэтому близка к 1 – 0,192 = 0,808, т. е. 81 %, значение обычно считается удовлетворительным, поскольку оно выше 80 %.

### G.4 Пример 3 (случай 3 из приложения Е): подтвердить, что продукт *A* не хуже продукта *B*; каждый потребитель оценивает только один продукт

Расчет, выполненный в приложении Е, дает $n$ = 374 потребителя на продукт. Лаборатория опросила только 300 потребителей для *A* ($n_A$ = 300) и 316 потребителей для *B* ($n_B$ = 316). Получилось: $m_A$ = 7,01, $s_A$ = 1,95, $m_B$ = 6,87 и $s_B$ = 1,80.

- В испытании типа «не хуже» значение $t_{cal}$ задается формулой (G.7):

$$u_{cal} = \frac{m_A - m_B + \Delta}{s_{within\ products} \cdot \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}, \tag{G.7}$$

где $s_{within\ products}{}^2 = \dfrac{s_A{}^2 \cdot (n_A - 1) + s_B{}^2 \cdot (n_B - 1)}{n_A + n_B - 2}$,

В нашем примере, $s_{within\ products}{}^2 = \dfrac{1,95^2 \cdot (300 - 1) + 1,80^2 \cdot (316 + -1)}{300 + 316 - 2} = 3,514$,

и $s_d = \sqrt{s_{within\ products}} = \sqrt{3,514} = 1,875$.

Следовательно: $t_{cal} = \dfrac{7,01 - 6,87 + 0,30}{1,875 \cdot \sqrt{\left(\frac{1}{300} + \frac{1}{316}\right)}} = \dfrac{0,44}{0,151} = 2,914$.

- Это значение сравнивают со значением $t_{the}$ для выбранного $\alpha$-риска. Это значение $t_{the}$ задается квантилем $(1-\alpha)$ $t$-распределения с $n_A + n_B - 2$ степенями свободы. Для $\alpha$-риска = 5 % в данном примере оно равно 1,647.

Чтобы отвергнуть $H_0$ и сделать вывод о том, что продукт $A$ не хуже продукта $B$, $t_{cal}$ должен быть > $t_{the}$. Поскольку это условие в данном примере верно (2,914 > 1,647), лаборатория может отвергнуть $H_0$ и сделать вывод, что продукт $A$ подобен продукту $B$ в отношении качества.

- Если лаборатория не может сделать заключение о заметном подобии в отношении критерия «не хуже», она должна рассчитать мощность этого критерия. Он задается формулой (G.8):

$$t_\beta = t_{the} - \left(\frac{\Delta}{t_{calden}}\right), \tag{G.8}$$

где $t_{the}$ — значение, ведущее к отвержению $H_0$;

$\Delta$ — разность, неприемлемая для того, чтобы считать два рассматриваемых продукта подобными (т. е. взаимозаменяемыми);

$t_{calden}$ — знаменатель $t_{cal}$ (формула G.7).

Например, при $m_A = 6,87$, $s_A = 1,95$, $m_B = 7,01$, и $s_B = 1,80$:

$$t_{cal} = \frac{6,87 - 7,01 + 0,30}{1,875\sqrt{\left(\frac{1}{300} + \frac{1}{316}\right)}} = \frac{0,16}{0,151} = 1,060.$$

Это значение не ведет к отвержению $H_0$, поскольку 1,060 < 1,647. Значение $t_\beta$: $t_\beta$ = 1,647 – (0,30/0,151) = = 1,647 – 1,987 = – 0,340.

Правила, касающиеся расчета $\beta$, одинаковы для формул (G.4), (G.6) и (G.8). Поскольку $t_\beta < 0$, значение $\beta$ задается функцией распределения $t$-распределения с $n_A + n_B - 2$ степенями свободы при $t_\beta$. Функция распределения $t$-распределения с 300+316–2 степенями свободы при −0,340 равна 0,367. Таким образом, мощность критерия равна 1 – $\beta$ = 1 – 0,367 = 0,633 ≈ 0,63. Это неудовлетворительное значение гораздо ниже, чем требуемое значение 90 %.

### G.5 Пример 4 (случай 4 из приложения E): подтвердить, что продукт *A* не хуже продукта *B*; каждый потребитель оценивает оба продукта

- Расчет, сделанный в приложении E, дает $n$ = 374 потребителя. Но лаборатория опросила только 200 потребителей ($n$ = 200). Получила: $m_A = 7,26$, $m_B = 7,31$, и $s_d = 2,08$, где $s_d$ — стандартное отклонение совокупности разностей между двумя рейтингами, присвоенными каждым потребителем двум рассматриваемым продуктам.

- В испытании типа «не хуже», значение $t_{cal}$ задается формулой (G.9)

$$t_{cal} = \frac{m_A - m_B + \Delta}{s_d \cdot \sqrt{\left(\frac{2}{n}\right)}}, \tag{G.9}$$

В нашем примере это $t_{cal} = \dfrac{7,26 - 7,31 + 0,30}{2,08 \cdot \sqrt{\left(\frac{2}{200}\right)}} = \dfrac{0,25}{0,208} = -1,202.$

- Это значение сравнивают со значением $t_{the}$ для выбранного $\alpha$-риска. Это значение $t_{the}$ задается квантилем $(1-\alpha)$ $t$-распределения с $n - 1$ степенями свободы. В данном примере и для $\alpha$-риска = 0,05, оно равно 1,653.

Чтобы отвергнуть $H_0$ и сделать вывод о заметном подобии между $A$ и $B$, $t_{cal}$ должно быть > $t_{the}$. Поскольку это условие несправедливо в данном примере (1,202 < 1,653), лаборатория не может отвергнуть $H_0$ и сделать вывод, что продукт $A$ подобен продукту $B$ в испытании по типу «не хуже».

- Тогда необходимо рассчитать мощность критерия по формуле (G.10)

$$t_\beta = t_{the} - \left(\frac{\Delta}{t_{calden}}\right), \tag{G.10}$$

где $t_{the}$ — значение, ведущее к отвержению $H_0$;

$\Delta$ — разность, неприемлемая для того, чтобы считать два рассматриваемых продукта подобными (т. е. взаимозаменяемыми);

$t_{calden}$ — знаменатель $t_{cal}$ [формула (G.9)].

В нашем примере $t_\beta$ = 1,653 – (0,30/0,208) = 0,211.

Правила расчета $\beta$ одинаковы для формул (G.4), (G.6), (G.8) и (G.10). Поскольку $t_\beta > 0$, мощность непосредственно задается функцией распределения $t$-распределения с 200 — 2 степенями свободы при $(-1)\cdot t_\beta$ = – 0,211. Следовательно, она равна 0,417 ≈ 0,42. Это значение является исключительно неудовлетворительным.

## Приложение Н
## (справочное)

## Примеры анализа данных для ранжирования

### Н.1 Общие положения

Данное приложение использует примеры приложения F; поэтому оно имеет дело только с двумя продуктами.

Для полноблочных планов (*CB*) с тремя и более продуктами лаборатория может пользоваться ISO 29842. В ISO 29842 используют критерий Фридмана (Friedman), чтобы определить, существуют ли заметные различия в ранжировании между продуктами и в случае положительного ответа предлагает множественное сравнение для каждой пары. Для сбалансированных неполноблочных планов используют критерий Дарбина (Durbin) для определения, заметны или не заметны различия между продуктами.

### Н.2 Пример 1 (случай 1 из приложения F): продемонстрировать, что существует разное предпочтение в отношении двух продуктов, *A* и *B*

- Лаборатория хочет продемонстрировать, что существует предпочтение между двумя продуктами, *A* и *B*. Нулевая гипотеза $H_0$, следующая:

Два продукта *A* и *B* выбирают из двух множеств, которые имеют одинаковое среднее.

Если экспериментальные значения ведут к отвержению $H_0$, лаборатория должна сделать вывод о заметном различии между средними обоих продуктов *A* и *B*.

- Количество потребителей, подлежащих опросу, вычисленное в приложении F, было равно 262 потребителя. После обсуждения с заказчиком лаборатория опросила только 200 потребителей. Она получила 82 ответа в пользу *A* и 118 ответов в пользу *B*. Относительная частота предпочтения поэтому равна $f_A = 0,41$ для *A* и $f_A = 0,59$ для *B*.

- Значение $u_{cal}$ задается формулой (Н.1)

$$u_{cal} = \frac{X - 0,5 \cdot n}{(n \cdot 0,5 \cdot 0,5)^{0,5}}, \qquad (Н.1)$$

где $X$ — наивысшее число ответов;

$n$ — общее число ответов.

Следовательно, для данного примера

$$u_{cal} = \frac{118 - 0,5 \cdot 200}{(0,50 \cdot 0,5 \cdot 200)^{1/2}} = \frac{18}{7,07} = 2,55,$$

Это значение сравнивают со значением $u_{the}$ для выбранного $\alpha$-риска. Для $\alpha$-риска = 5 % и двусторонней альтернативной гипотезы, оно равно 1,960.

Чтобы сделать вывод о заметном предпочтении для одного из двух продуктов, $u_{cal}$ должно быть > $u_{the}$. Поскольку это условие удовлетворяется в данном примере, лаборатория может сделать вывод о заметном различии в предпочтении в пользу *A*.

- Если бы это условие удовлетворено не было, все равно гипотезу $H_0$ нельзя было бы отвергнуть. Тогда лаборатории следовало бы рассчитать мощность критерия, поскольку она опросила меньше, чем 264 потребителя, установленных по формуле (F.1).

### Н.3 Пример 2 (случай 2 из приложения F): продемонстрировать, что частота предпочтения для продукта *A* не хуже, чем для продукта *B*

- Нулевую гипотезу $H_0$ можно сформулировать следующим образом:

Продукт *A* хуже продукта *B* на величину, равную не менее $\Delta$; следовательно, продукт *A* не равноценен продукту *B*.

Если экспериментальные значения приводят к отвержению $H_0$, необходимо сделать заключение о подобии продуктов, т. е. продукт *A* не хуже продукта *B*.

- Число потребителей, которых предполагается опросить, данное в приложении F равно 853. Это число очень высоко. Это объясняется низким значением $\Delta$ (равным 0,10), на основании которого заказчик больше не считает, что продукт *A* не хуже *B*.

После обсуждения с заказчиком лаборатория опросила 300 потребителей и получила 147 ответов в пользу *A* и 153 ответа в пользу *B*. Следовательно, $f_A = 0,49$ и $f_B = 0,51$. Следовательно, $f_A - f_B = 0,49 - 0,51 = -0,02$.

- Значение $u_{cal}$ задается формулой (Н.2):

$$u_{cal} = \frac{f_A - f_B + \Delta}{s_d \cdot \sqrt{\left(\frac{1}{n}\right)}}, \qquad (Н.2)$$

где $\Delta$ — разность частот, считающаяся неприемлемой;

$s_d$ — стандартное отклонение разности при $H_0$; равно $2 \cdot \sqrt{(f_A \cdot f_B)}$;

$n$ — число потребителей.

В данном примере:

$$s_d = 2 \cdot \sqrt{(f_A \cdot f_B)} = 2 \cdot \sqrt{(0,55 \cdot 0,45)} = 2 \cdot 0,4975 = 0,9950 \approx 1.$$

Тогда:

$$u_{cal} = \frac{0,49 - 0,51 + 0,10}{1 \cdot \sqrt{\left(\frac{1}{300}\right)}} = \frac{0,08}{0,0577} = 1,386.$$

Это значение сравнивают со значением $u_{the}$ для $\alpha$ = 5 % и односторонней гипотезы, т. е. 1,645. Поскольку $u_{cal} < u_{the}$, лаборатория не может отвергнуть $H_0$. Следовательно, не существует заметного подобия в том, что продукт *A* не хуже продукта *B*.

- Поскольку значение $u_{cal}$ не позволяет отвергнуть $H_0$, лаборатории следует рассчитать мощность критерия по формуле (F.2).

**Приложение ДА**
**(справочное)**

## Сведения о соответствии ссылочных международных стандартов межгосударственным стандартам

Таблица ДА.1

| Обозначение ссылочного международного стандарта | Степень соответствия | Обозначение и наименование соответствующего межгосударственного стандарта |
|---|---|---|
| ISO 4121 | IDT | ГОСТ ISO 4121—2016 «Органолептический анализ. Руководящие указания по применению шкал количественных характеристик» |
| ISO 5492 | IDT | ГОСТ ISO 5492—2014 «Органолептический анализ. Словарь» |
| ISO 5495 | — | *,[1)] |
| ISO 8587 | IDT | ГОСТ ISO 8587—2015 «Органолептический анализ. Методология. Ранжирование» |
| ISO 8589 | IDT | ГОСТ ISO 8589—2014 «Органолептический анализ. Общее руководство по проектированию лабораторных помещений» |
| ISO 29842 | — | * |

* Соответствующий межгосударственный стандарт отсутствует. До его принятия рекомендуется использовать перевод на русский язык международного стандарта. Официальный перевод данного международного стандарта находится в Федеральном информационном фонде стандартов.

Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандартов:

- IDT — идентичные стандарты.

[1)] В Российской Федерации действует ГОСТ Р 53161—2008 (ИСО 5495:2005) «Органолептический анализ. Методология. Метод парного сравнения».

## Библиография


[1] ABDI H., EDELMAN B., VALENTIN D., DOWLING W.G. *Experimental Design and Analysis for Psychology*. Oxford University Press, 2009, pp. 1—538.

[2] AFNOR NF V09-500, *Directives Générales pour la réalisation d'épreuves hédoniques effectuées avec des consommateurs dans un espace contrôlé. 2012*

[3] ASTM E1958-12, *Guide for Sensory Claim Substantiation. 2012*

[4] BALL R. Incomplete block designs for the minimization of order and carry-over effects in sensory analysis. *Food Qual. Prefer.* 1997, 18 pp. 111—118

[5] BI J. Similarity testing using paired comparison method. *Food Qual. Prefer.* 2007, 18 pp. 500—507

[6] BLACKWELDER W.C. Proving the null hypothesis» in clinical trials. *Control. Clin. Trials*. 1982, 3 pp. 345—353

[7] COCHRAN W.G., COX G.M. *Experimental Designs*. Wiley, New York, Second Edition, 1957

[8] COHEN J. *Statistical Power Analysis for the Behavioural Sciences*. Elsevier-Academic Press, New York, Second Edition, 1988

[9] GACULA M.C. *Singh J., Bi J., Altan S., Statistical methods in food and consumer research*. Academic Press, New York, Second Edition, 2009, pp. 1—853

[10] GREEHOUSE S.W., GEISSER S. On methods in the analysis of profile data. *Psychometrika*. 1959, 24 pp. 98—112

[11] HOWELL D.C. *Statistical Methods for Psychology*. Duxbury, Pacific Grove, Fifth Edition, 2002, pp. 1— 802

[12] KUNERT J., SAILER O. On Nearly Balanced Designs for Sensory Trials. *Food Qual. Prefer.* 2006, 17 pp. 219—227

[13] KUNERT J. Sensory Experiments as Crossover Studies. *Food Qual. Prefer.* 1998, 9 pp. 243—253

[14] LAWLESS H.T., HEYMANN H. *Sensory Evaluation of Food — Principles and Practices*. Springer, New York, Second Edition, 2010, pp. 1—596

[15] MAC FIE H.J., BRATCHELL N., GREENHOFF K., VALLIS L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *J. Sens. Stud.* 1989, 4 pp. 129—148

[16] MEILGAARD M.C., CIVILLE G.V., CARR B.T. *Sensory evaluation techniques*. CRC Press, Boca Raton, Fourth Edition, 2007, pp. 1—448

[17] MEYNERS M. Least equivalent allowable differences in equivalence testing. *Food Qual. Prefer.* 2007, 18 pp. 541—547

[18] MEYNERS M. Equivalence tests — A review. *Food Qual. Prefer.* 2012, 26 pp. 231—245

[19] NAES T., BROCKHOFF P.B., TOMIC B. *Statistics for Sensory and Consumer Science*. Wiley, Chichester, Second Edition, 2010, pp. 1—287

[20] SCHUIRMANN D. A comparison of the Two One Sided Tests procedure and the Power Approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* 1987, 15 pp. 657—680

[21] SIEGEL S., CASTELLAN N. *Nonparametric systems for the behavioural sciences*. 2nd ed., 1988, Singapour: Mc Graw-Hill International Editions

[22] TEN KLEIJF F., MUSTERS P.A.D., *Text analysis of open-ended survey responses: a complementary method to preference mapping*. Food Qual. Prefer., Vol. 14, 2003, pp. 43—52

[23] WAKELING I.N., MACFIE H.J.H. Designing consumer trials balanced for first and higher orders of carryover effect when only a subset of k samples from t may be tested. *Food Qual. Prefer.* 1995, 6 pp. 299—308

[24] WAKELING I.N., HASTED A., BUCK D. Cyclic presentation order designs for consumer research. *Food Qual. Prefer.* 2001, 12 pp. 39—46

[25] ISO 20252, Market, opinion and social research — Vocabulary and service requirements (*Исследование рынка, общественного мнения и социальных проблем. Словарь и сервисные требования)*

УДК 543.92:006.35 МКС 67.240 IDT

Ключевые слова: органолептический анализ, методология, пищевая продукция, общее руководство, гедоническое восприятие, гедоническое испытание, контролируемая зона, целевая совокупность, потребители (испытатели)

БЗ 8—2017/173

Редактор *Л.В. Коретникова*
Технический редактор *В.Н. Прусакова*
Корректор *С.И. Фирсова*
Компьютерная верстка *Е.О. Асташина*

Сдано в набор 21.09.2017. Подписано в печать 09.10.2017. Формат 60×84 1/8. Гарнитура Ариал.
Усл. печ. л. 5,12. Уч.-изд. л. 4,64. Тираж 25 экз. Зак. 1916.
Подготовлено на основе электронной версии предоставленной разработчиком стандарта

Издано и отпечатано во ФГУП «СТАНДАРТИНФОРМ», 123001 Москва, Гранатный пер., 4
www.gostinfo.ru info@gostinfo.ru

ISO 11136-2017