ICS 67.240

X 04

中华人民共和国国家标准

 GB/T XXXX—20XX/ISO 11136:2014

感官分析 方法学 受控区域

消费者喜好测试一般导则

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

（ISO11136:2014和ISO11136:2014/AMD 1:2020, NEQ）

（征求意见稿）

202X- - 发布 202X- 实施

 **国家市场监督管理总局**

**国 家 标 准 化 管 理 委 员 会 发布**

**目 录**

[前 言 Ⅲ](#_Toc49175445)

[1 范 围 1](#_Toc49175446)

[2 规范性引用文件 1](#_Toc49175447)

[3 术语和定义 2](#_Toc49175448)

[4 前期准备工作 4](#_Toc49175479)

[4.1 实验方案 4](#_Toc49175480)

[4.2 目标群体确定依据 5](#_Toc49175481)

[4.3 其他规范 5](#_Toc49175482)

[5 消费者样本 5](#_Toc49175483)

[5.1 总则 5](#_Toc49175484)

[5.2 消费者招募 6](#_Toc49175485)

[5.3 消费者样本构建 7](#_Toc49175490)

[5.4 消费者样本细分 8](#_Toc49175493)

[6 消费者样本量 8](#_Toc49175494)

[6.1 测量结果 8](#_Toc49175495)

[6.2 样本量确定 9](#_Toc49175496)

[6.3 消费者子群使用 9](#_Toc49175497)

[6.4 产品数量和实验设计 10](#_Toc49175502)

[6.5 招募数量安全界限 10](#_Toc49175503)

[7 测试区域 10](#_Toc49175504)

[8 产品 11](#_Toc49175505)

[8.1 产品匿名提供 11](#_Toc49175506)

[8.2 产品制备和提供 11](#_Toc49175507)

[9 程序 12](#_Toc49175511)

[9.1 喜好测试分类 12](#_Toc49175512)

[9.2 评分法 12](#_Toc49175513)

[9.3 成对偏爱测试 13](#_Toc49175516)

[9.4 排序法 14](#_Toc49175519)

[9.5 其他问题 14](#_Toc49175522)

[9.6 给消费者的提示语 15](#_Toc49175526)

[10 组织测试 15](#_Toc49175527)

[10.1 总则 15](#_Toc49175528)

[10.2 实验设计 16](#_Toc49175529)

[10.3 单次评价的产品数量 16](#_Toc49175533)

[10.4 同次评价中产品性质 18](#_Toc49175538)

[10.5 评价时间 18](#_Toc49175539)

[10.6 测试中产品的重复次数 18](#_Toc49175540)

[10.7 预实验 18](#_Toc49175541)

[11 结果分析 18](#_Toc49175542)

[11.1 总则 18](#_Toc49175543)

[11.2 推断分析 19](#_Toc49175544)

[11.3 评分法（见9.2） 19](#_Toc49175545)

[11.4 成对比较法（见9.3） 19](#_Toc49175546)

[11.5 排序法（见9.4） 19](#_Toc49175547)

[12 测试报告 20](#_Toc49175548)

[12.1 总则 20](#_Toc49175549)

[12.2 产品 20](#_Toc49175550)

[12.3 测试程序 21](#_Toc49175551)

[12.4 消费者 21](#_Toc49175552)

[12.5 测试结果 22](#_Toc49175556)

[12.6 附件 22](#_Toc49175557)

[附录A（资料性附录）招募问卷示例 23](#_Toc49175558)

[附录B（资料性附录）消费者细分数据评价方法 25](#_Toc49175561)

[附录C（资料性附录）产品呈送方案 27](#_Toc49175567)

[附录D（资料性附录）样本量对均值精度的影响 28](#_Toc49175570)

[附录E（资料性附录）评分法的最小样本量 29](#_Toc49175573)

[附录F（资料性附录）评分法数据分析的示例 32](#_Toc49175587)

[附录G（资料性附录）适合中国消费者的喜好标度（评分法） 37](#_Toc49175595)

[参 考 文 献 38](#_Toc49175601)

前 言

本标准非等效采用国际标准ISO 11136:2014《感官分析—方法学—受控区域消费者喜好测试一般导则》和ISO 11136:2014/修改单1: 2020《感官分析—方法学—受控区域消费者喜好测试一般导则—修改单1》。

本标准与ISO 11136和ISO 11136:2014/修改单1的主要差异为：

——修改了“术语和定义”中顺序，并增加术语“偏好测试”、“接受性”和“评分”（见3）

——修改了原标准11.4“成对比较法”中的内容（见11.4）；

——删除了原标准中附录F“排序法的最小样本量”和附录H“排序法数据分析的示例”

——修改了原标准中附录G“评分法数据分析的示例”为附录F“评分法数据分析的示例”

——增加了附录G“适合中国消费者常用喜好标度”（见附录F）。

——将6.1中第七段中的“差异越小，样本量越大”更改为“差异越大，样本量越大”，纠正原文错误；

——将原标准附录E.2表格中的“B”更改为“β”，E.3表格中的 “A”更改为“α”，纠正原文错误；

——将原标准附录E.3中的“$σ\_{d}=2×σ\_{产品内}×\sqrt{(1-ρ)}$”更改为“$σ\_{d}=σ\_{产品内}×\sqrt{2(1-ρ)}$”，纠正原文错误；

——将原标准修改单1附录G公式(G.7)中的“$s\_{产品内}^{2}$”更改为“$s\_{产品内}$”，纠正原文错误；

——将原标准修改单1附录G.4注中的“T.INV(0.05,614)”更改为“T.INV(0.95,614)”，纠正原文错误；

本标准按照GB/T 1.1—2020给出的规则起草。

与本标准中规范性引用的国际文件有一致性对应关系的我国文件如下：

——GB/T 10221—2012 感官分析 术语（ISO 5492: 2008，MOD）

——GB/T 12310—2012 感官分析方法 成对比较检验 (ISO 5495:2005，MOD)

——GB/T 12315—2008 感官分析 方法学 排序法（ISO 8587:2006, NEQ）

——GB/T 13868—2009 感官分析 建立感官分析实验室的一般导则 (ISO 8589:2007，NEQ)

本标准由全国感官分析标准化技术委员会（SAC/TC566）提出并归口。

本标准主要起草单位：。

本标准主要起草人：。

感官分析 方法学 在受控区域对消费者

开展喜好测试的一般导则

# 1 范 围

本标准给出了受控区域内测量消费者对产品喜好或相对喜好程度的方法。

本标准中给出的方法是基于收集消费者对测试问题的回答，包括纸质问卷或电子问卷。

本标准不适用于有关消费行为的测试，例如记录消费者自由消费的产品数量等。

本标准中的喜好测试方法（包括接受性测试与偏爱测试），具体应用如下列所示：

——比较产品与竞争产品的喜好差异；

——提升产品，以获得更高的喜好评分或被更多消费者喜爱；

——确定适合某个特定目标消费人群的产品系列；

——确定产品的最佳食用期；

——评估配方变化对产品喜好的影响；

——研究产品感官特性及产品包装对其喜好程度的影响。

本标准中的方法对确定下列情况有效：

——是否存在可感知的偏爱差异（喜好程度差异检验）；

——是否不存在可感知的偏爱差异（成对相似检验）。

# 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

ISO 4121 感官分析 定量响应标度使用导则（Sensory analysis—Guidelines for the use of quantitative response scales）

ISO 5492 感官分析 术语（Sensory analysis—Vocabulary）

ISO 5495 感官分析方法 成对比较检验（Sensory analysis—Methodology—Paired comparison test）

ISO 8587 感官分析 方法学 排序法（Sensory analysis—Methodology— Ranking）

ISO 8589 感官分析 建立感官分析实验室的一般导则（Sensory analysis— General guidance for the design of test rooms）

ISO 29842 感官分析 方法学 平衡不完全区组设计（Sensory analysis—Methodology —Balanced incomplete block designs）

# 3 术语和定义

ISO 5492界定的以及下列术语和定义适用于本文件。

## 3.1

## 受控区域 controlled area

区域应提供：

——准备和提供产品的受控区域；

——消费者评价产品和接受提问的舒适环境；

——避免消费者之间的口头和非口头交流，确保每位消费者独立反馈意见。

## 3.2

## 消费者/受访者 consumer/respondent

产品的使用者。

注：本标准中定义的消费者，不包括ISO 5492中定义的专家、专家评价员及优选评价员。

## 3.3

## 消费者样本 consumer sample

以目标群体特性作为筛选标准招募的消费者群体。

## 3.4

## 喜好感知 hedonic perception

对产品感官特征的喜好程度。

## 3.5

## 喜好测试 hedonic test

测量消费者对产品整体喜好感知的测试。

## 3.6

## 偏爱测试 preference test

在2个或者多个产品中评估更偏好或喜欢样品的测试。

## 3.7

## 接受性 acceptability

对产品整体或特定感官特性喜欢或不喜欢的程度。

## 3.8

## 评分 scoring

用对产品或产品特性具有数学意义的数值评价产品或产品特性。

## 3.9

## 委托方 commissioning party

委托第三方进行研究的个人、团体或服务机构。

注1：质量控制中，委托方相当于采购方或客户，与供应商、服务提供商或实验室相对应。

注2：通常，委托方和服务提供商应分属不同企业。若二者隶属同一企业的不同部门时，应在研究方案中对双方职责进行详细区分和明确说明。

## 3.10

## 目标群体 target population

用于产品喜好响应评价的人群。

## 3.11

## 轮次 session

规定时间内消费者按照同一方案开展相同的感官评价活动。

注：每个轮次应分成多个环节，中间插有规定的休息时间。例如1个90 min的轮次应分为2个环节，中间有10 min的休息时间。

## 3.12

## 开放式问题 open-ended question

不给受访者提供现成选项，要求其自由作答的问题。

注：问题应涉及提供给消费者的所有或部分产品（例如询问消费者为何某一特定产品对其具有最大或最小吸引力的问题）。

## 3.13

## 封闭式问题 closed-ended question

要求受访者从一组预先确定的可能答案中进行选择的问题。

## 3.14

## 产品集 product family

正常消费过程中可以相互替代的产品组。

注：产品集可能会因文化和消费习惯差异而有所不同。

## 3.15

## 代表性 representativeness

待测样的特性与抽样群体相关特性的符合程度。

## 3.16

## 可访问样本库 access panel

同意后期被选中后会配合完成数据采集的潜在消费者样本数据库。

[来源：ISO 20252]

注：受访者筛选应遵照一定规则。

## 3.17

## 位置偏差 position bias

由一个产品在待测产品组中所处位置引起的偏差。

## 3.18

## 顺序偏差 sequential bias

由一个产品在待测产品组中所呈现时间或顺序引起的偏差。

# 4 前期准备工作

## 4.1 实验方案

测试实验室依据委托方提供的信息（尤其是研究目标和目标群体）制定研究方案，并以书面文件形式交付于委托方，实验方案中应包括下列内容：

——研究目的；

——待测产品（若实验室需采购1种或多种产品，应提供采购程序）；

——测试的目标群体；

——消费者样本大小（样本量）及招募方式；

——消费者前期参与过的消费者测试项目的限制条件；

——同一轮测试中其他待测产品的限制条件；

——评价条件；

——调查类型（自填式问卷或访问式问卷）；

——调查问卷；

——评价场所；

——每轮测试中预计评价的产品数量，若同一轮次中需评价不同产品系列时，其他待测产品的情况说明（若制定研究方案时，这些产品情况是未知的，当明确产品性质后，实验室应及时与委托方沟通）；

——待测产品的准备条件（若已知）；

——待测产品的提示语（若已知），产品评价前应提供有关产品或测试的具体信息；

——用于结果解释的统计分析方法，方法的风险类型及可接受水平，喜好差异性测试中可测出的最小差异值或喜好相似性测试中允许的最大差异值；——基于获取判断数据前已知的消费者人群特征或可能响应，对消费者群体细分的设想方案；

——开放式问题的核查和汇总方法；

——委托方希望实验室提供的与测试结果相关的内容；

——预定的测试时间和报告提交日期。

若制定实验方案时上述任何一项内容尚未知晓，应在方案中列出下列陈述：“该项内容将由委托方指定…（指定的研究阶段）。”

注：若已有相关测试标准，实验方案中可直接参考该标准，无需重复给出标准中操作步骤和技术要点。

## 4.2 目标群体确定依据

下列问题的回答对确定目标群体非常重要。每次测试均应重复询问这些问题，因为即使是同一产品的测试，目标消费者也会存在差异。

——待测产品是否已上市？若已上市，是否有可能区分实际消费者和潜在消费者？

——对产品感兴趣的消费者是已使用过产品的实际消费者，还是未使用过产品的潜在消费者？或者两者兼有？

——特定消费消费子群的测试结果是否需要核查和比较？

——消费者总样本结果分析时是否需要考虑消费子群的影响？

——对产品感兴趣的个体之间是否存在差异？

## 4.3 其他规范

确定目标群体后，应进一步确定下列内容：

——测试地点（条款7）；

——预期的测量精度（6.1）和测试方法（条款9）；

——所选测试方法的产品提供方案（10.2）；

——假设检验中参数的确定（6.1）；

假设检验时，应明确消费者样本的构成（条款5）和数量（条款6）。

# 5 消费者样本

## 5.1 总则

喜好测试旨在确定特定消费群体对产品的可接受性和/或对不同产品（2种或多种）的偏爱性。

消费者样本构成对喜好测试结果是否可以回答委托方关注的问题具有决定性作用。

消费者应自愿参与测试，并允许接受测试相关的费用或奖励。

获得自然响应对于测试非常重要，应删除接受过训练的消费者，因为他们与目标消费群体存在显著性系统差异。

不宜从待测产品生产企业的工作人员中招募消费者。使用生产商内部人员构成消费者测试小组时，下列因素均可能导致结果产生偏差：

——待测产品存在被识别的风险；

——倾向于选择被识别的产品；

——对产品过度熟悉；

上述因素会带来测试结果无法代表目标消费群体的严重风险。无论采用何种方式招募消费者，测试人员的筛选要求（例如是否可以来自生产企业、竞争对手企业或特定的社会经济部门等）均应与委托方协商一致。

## 5.2 消费者招募

### 5.2.1 总则

**注意：招募消费者应遵循以下2点：**

**­——创建和记录包含个人数据文件时，应遵守保密规定及相关法律；**

**——遵守有关年龄限制的法规。**

消费者招募可基于特定任务（为某特定研究任务重新招募的消费者）或从消费者可访问样本库中筛选（具有已知特征的潜在消费者名单）。基于特定任务的招募通常被认为是一种筛选很少参与消费者测试的消费者招募方式，但随着组织准永久性喜好测试的开展地的成倍增加，，这种观念可能并不正确。

### 5.2.2 基于特定任务的招募

对于特定任务的消费者招募，通常在公共场所（例如销售点），通过电话、书面邮件或电子邮件、当地报纸或电台的广告、社交网络或其他互联网等多种方式进行。多数情况下，事先与消费者招募区域负责的公司或组织达成一致非常重要。

应注意，招募方式及接触消费者的时间（特别是公共场所或通过电话接触以及通过媒体发布信息）均对消费者样本构成具有强烈影响。

### 5.2.3 基于消费者可访问样本库的招募

为提高招募效率，可从已有的消费者样本库中筛选消费者构建评价小组。每次测试开展前及随后的特定目标群体细分时，均可根据样本库中消费者的记录信息进行预筛选。

消费者信息应包括下列内容：

——身份信息和联系方式；

——年龄和性别；

——婚姻状况和家庭组成；

——社会状况、职业和收入；

——与待测产品相关的健康问题（例如过敏症、糖尿病、视力障碍等）；

——与待测产品相关的其他信息（例如饮食习惯、宗教禁忌、生活方式和家用装备等）；

——能参加消费者测试的时间信息；

——前期参加过消费者测试的相关信息。

每位消费者的情况说明应定期更新。该数据库能确保服务商预先满足委托方关于目标消费者的初始需求，实现委托方和服务商之间的有效沟通。

### 5.2.4 消费者招募频次

相比基于特定任务的招募，从现有消费者群体中进行招募的“消费者专业化”风险更大。

避免个别消费者形成与产品组相关联的训练效应非常重要。因此，应避免频繁招募同一消费者。通常，同一消费者参与相同产品消费者测试的间隔时间宜为三个月及以上。

实验室应记录每位消费者参与的每个产品及产品集的测试频次。实验方案中应包括消费者参加测试的记录，若委托方要求，还需列明其参加测试中产品的相关信息。

## 5.3 消费者样本构建

### 5.3.1 采用招募问卷筛选消费者

采用含筛选标准的招募问卷筛选和构建消费者样本时，应确保消费者并不知道问卷中哪些是用于筛选的关键问题。

下列是部分常用筛选标准的示例：

——前期参加过相关产品组的测试情况；

——使用产品的频次；

——通常使用产品的品牌、产品说明和购买地点等；

——使用或消费产品的习惯。

其他相关的条件还包括年龄、性别、社会阶层、职业群体或地理位置等。出于实用性目的，应对符合上述标准的人群进行分类，并指出消费者样本中每类人群的构成比例。

应要求参加测试的每位消费者完整填写招募问卷，并对招募中的部分细节进行核实。

附录A给出了一个招募问卷示例。

应保留招募问卷或保存招募的相关信息，确保参加消费者测试人员招募信息的可追溯性。

### 5.3.2 消费者样本的代表性

目标群体的代表性是构建消费者样本最重要的质量要求。测试人群通常由提出目标群体的委托方与评估方案可行性的实验室共同确定。

筛选消费者目标群体时，应采用下列条件：

——产品的使用或消费频次（该条标准是基本要求，因其会影响测试的辨别力）；

——年龄、性别、社会经济地位、职业和地理位置；

——通常使用或消费产品的品牌；

——产品的购买地。

## 5.4 消费者样本细分

消费者样本细分通常采用下列2种方式：

a）测试开始前进行细分，以获取下列问题的回答：

1）女性和男性的响应方式相同吗？

2）不同年龄组响应有所不同吗？

3）响应取决于消费习惯吗？

该情况下，对每个子群单独分析并比较结果。

b）有证据表明存在多个子群（例如数据呈现多峰形分布）时，应根据获取的数据进一步对消费者细分。

附录B给出了如何处理消费者细分数据的指导。

# 6 消费者样本量

## 6.1 测量结果

测量结果平均值或估计值的数值精度随样本量增加而增大，但测量精度提高并不能确保所得结论的相关性和有效性。选择适宜的消费者和测量程序同样重要。

附录D给出了消费者样本量对测量精度的影响研究。

精度也取决于消费者对单个产品响应值的变化范围。数据越分散，达到既定精度所需的样本量则越大。该参数的相关信息来源于实验室和委托方的经验。

结果精度同样受到测试程序的影响。其他条件相同时，同一消费者同时评价2个产品比不同消费者分别评价2个产品的结果具有更高的精度。

精度也取决于测试目的。其他条件相同时，精度受到下列因素的影响：

——若测试目的是确定产品存在喜好差异，2个喜好评分均值间的最小差异可被认为大到足够被有效（差异越小，样本量越大）。

——若测试目的是确定产品具有喜好相似，2个喜好评分均值间的最大差异可被认为小到足以被忽略（差异越大，样本量越大）；

——委托方愿意接受的误判风险*α*（*α*风险越小，样本量越大）；

——委托方期望的检验效力*P*（*P* = 1- *β*，*β*是漏判风险；预期效力越高，样本量越大）；

——预期的测试统计结果。若委托方仅关注确保产品B不劣于产品A时，宜采用定向检验，得到产品B劣于或优于产品A的结论比得到产品B和A具有相似喜好的结论具有更高的检验效力**。**

## 6.2 样本量确定

消费者样本量确定是测试开展的关键步骤，应由统计人员协助完成。附录E给出了采用评分法进行差异性或相似性测试所需样本量的计算方法（附有示例）。

任何情况下，本标准中消费者样本量不应低于60人。

若实验室未采用或无法实现附录E中的所述方法，本标准中的最小样本量应调整为100人。

若实验室采用附录E中的所述方法，样本量应由计算结果得出。但最终样本量应依据下列情况具体而定：

——若计算所需的消费者样本量超过100人，但实验室和/或委托方无法提供这么多数量消费者时，样本量可减少至100人。该情况下，实验室应计算100位消费者的检验效力，并获得委托方同意接受此降低的效力；

——若计算所需的消费者样本量低于60人时，测试人数应增至60人。

样本量定为100和60均是假设委托方未依据消费者特征（例如年龄或性别）分组的基础上计算出的所需样本量。若测试前已对消费者分组，则每组最小样本量应为60人。

## 6.3 消费者子群使用

### 6.3.1 总则

收集或使用测试数据前，宜先建立消费者子群。下列测试方法使用时，应在实验设计阶段加入统计的意见。

### 6.3.2 采用预先分好组的消费者群独立测试

通常委托方希望在测试前先根据已知的消费者特征（例如性别或年龄）进行消费者分组，再对形成的消费子群单独测试，并独立分析每组结果。该种情况下，每个消费子群的样本量应不低于60人。

### 6.3.3 采用预先分好组的消费者群统一测试

相比6.3.2中方法，对消费者总样本进行平衡供样后再综合分析所有子群结果的效率更高。即使每组消费者样本量低于60人，也会获得相同的置信度。若所有子群具有相同或高度相似的样本量时，该种方法的效率最高。

### 6.3.4 消费者细分

大样本量的消费者测试结果宜进行消费者细分，包括测量结果和/或社会人口学的统计结果。消费者细分有利于描述消费者样本的特征，并形成新的假设。最终的分组数量、特征和规模均取决于所得结果的数量和性质。

## 6.4 产品数量和实验设计

消费者样本量受到产品数量和实验设计的影响。采用完全供样时，测试所需消费者数量为每个产品所需的响应数量。采用不完全供样时，所需消费者数量更大。

## 6.5 招募数量安全界限

由于部分招募的消费者可能未能参加招募测试，或参加了测试却未通过考核，因此实际获得的有效消费者数量通常少于招募的数量。为弥补这一点，所需消费者数量计算时应包括安全界限。

# 7 测试区域

本标准中的喜好测试均在受控区域（定义3.1）内开展，不包括在销售点或消费者家中进行的喜好测试。下列给出了3种适宜作为测试区域的示例。

a）固定式感官评价实验室，应符合ISO 8589中的规定，包括1个消费者接待区、1个样品制备区以及配备有独立评价小间、温度控制装置和定期通风换气装置的评价区。该区域能提供最佳的样品制备、样品提供和消费者响应记录的条件。

b）移动式感官评价实验室，安装在为感官测试而专门改造的车辆上。其空间小于固定式感官实验室，用于消费者接待和样品制备的空间也有限。

c）临时按“特定任务”进行布局的房间，通常用于准备工作较少的产品评价。该区域应包括2个独立区域，1个用于感官测试，另1个用于产品准备和编码。测试区域布局应实现将消费者彼此隔开。测试条件应进行监控，包括测试温度、室内照明、通风和样品的存放及制备等。与上述固定式或移动式实验室不同，当需要开展新的特定任务测试时，测试区域的产品准备和提供系统均应重新配备。

# 8 产品

## 8.1 产品匿名提供

测试产品应以匿名形式提供给消费者，唯一的识别手段是实验室指定的随机编码（例如随机三位数编码）。应删除或遮蔽所有可能会带来品牌或质量标识的信息，除非：

——无法删除或遮蔽产品信息；

——测试的目的就是确定品牌或质量标识的影响。

注：本标准中，质量标识是指任何可能影响消费者期望的标记，例如包装类型也是一种质量标识。

## 8.2 产品制备和提供

### 8.2.1 总则

应依据生产商推荐的步骤制备待测产品。若无法获取制备步骤时，应依据预实验建立的规范步骤制备产品。若如无特殊理由，产品的制备条件应接近通常的使用和消费条件。

产品的提供方式（例如温度、相关的产品、餐具和呈送方式）均应遵循上述规则。

**注意：产品会受到批次和存放条件的影响产生差异。委托方可能会关注到这种变化，某些情况下也作为调查的重点。应使用同一批次、具有相同贮存时间和条件的产品进行测试，使得差异最小化。**

### 8.2.2 产品提供量

提供给消费者的产品量应与通常消费的产品量或委托方指定的产品量一致。应告知消费者测试或消费样品的最少用量，必要时也应告知最大用量。应确保测试量不会出现太少（存在感官印象停留在最初、转瞬即逝的风险）或太多（存在吃饱或厌恶的风险）的情况。

提供量宜少于消费者通常消费同系列几种产品时的用量。但应确保有足够的提供量，避免提供量太小导致评价时存在感官印象仅停留在最初、稍纵即逝的风险。

提供成品菜肴时，给每位消费者提供的样品应按比例包含原产品的全部成分。

### 8.2.3 产品提供形式

产品提供形式通常包括产品购买时形式和待消费时形式。

示例1 面包（整个或切片），香肠（整根或切片），干酪（整块或切片），比萨（整份或部分），速溶咖啡（放咖啡后再加水或加水后再放咖啡）。

实验室和委托方应确保上述2种产品提供形式具有相同的目标，即最能接近真实产品带来的喜好感知。无需考虑2种形式是否会产生光环效应。若存在光环效应，可同时（整体/切片）或相继（速溶咖啡，放咖啡后再加水或加水后再放咖啡）提供产品，且无需向消费者特别指出产品是以2种不同的形式提供。消费者在产品提供后对产品评分，该方法类似于消费者日常生活中无需细分的、普遍的感知行为。

示例2 提示语和问题的陈述应如下所示：您面前有一个香肠，编码为148。看一看，闻一闻，吃几块，然后用下列的响应标度记录您对它的评分值。

若实验室和委托方均认为2种产品提供形式对产品喜好具有相似的贡献，则无需将不同提供形式的产品视为不同产品。否则，2种形式应同时或逐一提供，并获得单独的评分值。

注：本标准不涉及实验室和委托方认为有必要了解的视觉信息产生的期望效应情况。该情况涉及特定的概念和分析方法，属于特殊实验设计范畴。

# 9 程序

## 9.1 喜好测试分类

根据给消费者设定的任务，喜好测试通常分为2类：

a）接受性测试，用于衡量消费时的喜好程度。本标准仅给出了一种接受性测试方法，即评分法。

b）偏爱测试，用于衡量不同产品的喜好排序（例如“您最喜欢哪种样品？”或“请将提供的产品由最不喜欢到最喜欢进行排序。”）。偏爱测试中获得的信息是相对的，并不表明任何产品的接受情况。1种产品也许比另1种产品更受偏爱，但2种产品可能均不被接受。偏爱测试分为成对偏爱测试（比较2个样品）和偏爱排序法（比较2个以上样品）。

## 9.2 评分法

### 9.2.1 总则

评分法有多种分类方式：

a）根据采用的响应标度分类。响应标度可以是结构化的或非结构化的、数字的、语义的或图片形式的，具体见ISO 4121。

b）根据产品提供方式分类。测试2个或更多产品时，有下列3种产品提供方式：——严格的单一提供（单个评价），每位消费者只评价1个产品；

——连续逐一提供（完全或不完全实验设计），消费者通过一轮或多轮测试完成多个产品评价。消费者每次只评价1个产品，且不提供已评价过的产品或相关的评价信息。应确保消费者不能回看之前的产品评分；

——比较提供（完全或不完全实验设计），将几种产品同时提供给消费者，并允许消费者回看对其他产品的评分。

连续逐一提供是最常见的样品提供方式。相比严格的单一提供，该方式更加经济，但需掌握如何分配产品给消费者（附录C）。比较提供的方式较少使用，因其会夸大产品间差异，且测试条件达不到严格一致时，难以实现不同测试结果之间的比较。

### 9.2.2 样品提供

按照样品提供方案将样品提供给消费者（见10.2）。

## 9.3 成对偏爱测试

### 9.3.1 总则

成对比较测试应按照ISO 5495中的规定进行。

成对比较测试的原理是产品以成对形式呈现，消费者按照指定的顺序进行评价，指出更喜欢哪一个产品。

2种产品的成对偏爱检验有2种方式，即强迫选择（选项A）或允许有“无偏爱”响应（选项B）。

1. 选项A：消费者回答下列封闭问题：

“从左到右品尝提供给您的2种产品，然后在您更喜欢产品编码下标记̔̕ʻ×ʼ”。

 产品xxx 产品yyy

* □

 此处，消费者应选择更喜欢其中的哪一个产品。

b）选项B：消费者回答下列封闭问题：

“从左到右品尝提供给您的2种产品，然后在您更喜欢产品编码下标记ʻ×ʼ。如果您对2种产品的喜好相同，则在ʻ无偏爱ʼ下标记 ”。

产品xxx 无偏爱 产品yyy

□ □ □

当仅测试一对样品时，选项A比选项B具有更大的潜在区分性。

问卷中不应包含任何会影响消费者偏爱选择或有关待测产品特性的相关问题。例如若要求消费者评价产品风味，他们可能会在评价产品接受性时偏向关注风味的接受性。

测试结束时可以提出一个开放式问题，让消费者对喜欢或不喜欢产品的具体方面进行评价。

### 9.3.2 样品提供

样品A和B应按照AB和BA的顺序平衡提供，并以随机形式成对提供给消费者。

## 9.4 排序法

### 9.4.1 总则

排序法应按照ISO 8587中的规定进行。

排序法的原理是同时提供多个样品，要求消费者按照接受性顺序进行排序。

排序法对消费者的要求更高，其复杂性也随样品数量增加而增大。应要求消费者多次评价产品，使其能正确排序。

消费者对样品进行初评时，应指定待测样品的评价顺序（通常从左到右）。随后，消费者可自由选择产品的评价顺序，以自己的方式确定产品的喜好排序。

排序法可以是强迫选择检验，也允许消费者将2种或多种产品排为同一秩次。但本标准不包括后者的测试分析。

### 9.4.2 问卷

消费者通常收到下列提示语：

“请从左到右依次品尝4种产品。产品可多次品尝，直至您做出决定为止。最后，请按照从最不喜欢到最喜欢的顺序，依次写下产品编码。”

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 最不喜欢 |  |  | 最喜欢 |

## 9.5 其他问题

### 9.5.1 总则

除上述测试程序，委托方通常希望获得更多信息帮助理解产品的总体评价或消费者特征。例如采用消费者样本库的招募方式时，若实验室希望核查是否所有消费者来自于目标群体时，可设置一个特定的消费习惯问题。

### 9.5.2 消费者相关的问题

实验室应关注所选的问题及问题的询问时间。测试后的提问时应注意刚结束的测试是否会影响消费者的回答。例如番茄酱搭配米饭测试结束后，消费者可能倾向于认为番茄酱配米饭是常吃食物。测试前的提问易导致消费者倾向于以某种特定的方式评价产品，或重点关注产品的部分特性。

### 9.5.3 产品相关的问题

产品相关的问题可以是开放式问题或封闭式问题。问题应设置在与整体评价相关的所有问题之后，但应单独列出，不宜和整体评价的相关问题列在同一张纸上或屏幕上，减少消费者尝试给出一致性回答的风险。此外，此类问题数量应尽量少。

注1：附加问题并非是质疑评价结果。若产品的整体评价和报告的消费意向之间存在任何差异，宜开展研究寻求原因。

注2：附加问题通常提供的信息很少。感官评价实验室的负责人应告知委托方，可采用喜好数据和感官数据的联合分析方法（例如偏爱映射图或联合分析），来确定消费者喜欢或不喜欢产品的原因。

## 9.6 给消费者的提示语

提示语通常以书面或口头形式提供给参加测试的消费者。不应低估提示语对测试结果的重要性，需重点关注提示语，并在实验方案中专门提及。提示语中应重点关注下列内容：

a）评价前告知每位消费者待测产品的最小和最大使用量。使用量取决于产品类型、待测产品的数量和委托方的意愿。除产品的最小和最大使用量外，实验室还应明确给出标准的操作规范并确保其执行；

b）填写完问卷后的安排事项。采用连续逐一方式提供样品时，实验室应使用能有效避免消费者参考之前评价结果的操作系统；

c）休息时长及休息期间宜进行的活动（例如用面包、苹果或水进行感官清零、离开评价间或阅读杂志等）；

d）所提问题的陈述方式和含义。

注：陈述问题时应避免使用中性语气。例如提示消费者从产品的外观、气味、滋味或质地等方面描述他们对产品的整体评价，使回答更为合理，而并非简单如“有任何评价请填写”的陈述。

测试开始时提供给消费者关于产品数量、需消耗数量和需回答问题等相关信息。也可提供其他信息，使评价过程更加人性化或满足测试的其他目标。由于提供的信息会让消费者产生一定预期，因此应仔细考虑信息的陈述形式，并在报告中逐一说明。若测试由数轮评价组成，应确保每轮测试中的提示语完全相同。

# 10 测试开展

## 10.1 总则

宜合理组织安排测试，确保在多个研究目标存在时取得最佳平衡，应包括下列内容：

——采用相等的消费者评价产品，确保产品间比较的有效性；

——最小化位置偏差和顺序偏差的影响，确保产品间比较的有效性；

——确保测试条件应尽可能地代表正常的消费行为；

——确保消费者不会因超负荷评价导致随意回答数量增多。

为协调不同的研究目标，有下列3种测试方式：

a) 完全测试：每位消费者评价所有产品；

b) 不完全测试：每位消费者仅评价1组产品；

c) 单个评价：每位消费者仅评价1个产品。

## 10.2 实验设计

### 10.2.1 完全测试

每位消费者评价所有产品，能有效避免测试条件非等同的问题，但最大风险是消费者负荷过重，且存在顺序效应和转移效应的风险。产品在一组测试产品中的提供位置会影响评价结果的均值和/或变化性。

消费者的产品提供顺序至关重要。一种方法是采用严格的随机顺序，每位消费者的产品提供顺序遵循完全独立和随机原则。另一种方法是设计系统的样品提供方案来平衡顺序效应和转移效应。转移效应的影响过大时，应采用足够长的时间间隔确保感觉清零，或采用其他方法减少影响。

### 10.2.2 不完全测试

不完全测试中，顺序效应和转移效应的处理措施与完全测试类似。消费者超负荷评价的风险较小，但缺点是可能存在情境效应，且试验设计和分析较复杂。平衡不完全实验设计见参考文献[9]。

### 10.2.3 单个测试

为完全消除顺序效应、转移效应和情境效应，可为每个产品生成独立的消费者样本。该方法开展和结果分析简单方便，但应确保消费者样本的等效性和目标群体的代表性。此外，由于单个评价无法顾及个体差异，结果的精度低于完全测试和不完全测试。其他条件相同的情况下，为确保测试结果达到相同的置信度和精度，单个测试的消费者样本量应更大。

## 10.3 单次评价的产品数量

### 10.3.1 总则

单次评价中，有效评价产品数量取决于下列因素：

——待测产品的性质，例如不同产品具有不同的饱腹感；有些产品（例如具有侵略性）能改变感觉受体；含酒精的产品可能影响皮质层的信息整合；产品分单一组分和多组分产品（例如成品菜肴）；

——评价时长，评价时间较长的评价宜分为多个环节，但未预约消费者的可用时间通常较短；

——每种产品的消费量；

——评价产品的感觉器官（主要通过使用视觉或触觉而非味觉进行评价，或其他感觉器官）；

——询问问题的数量，尤其需新的产品提供时。

因此，本标准中并未设置评价产品数量的具体限制，但要求实验室向委托方合理论证其测量能力。

### 10.3.2 单次评价的产品数量过多

产品数量过多、1次评价无法完成时，可采取下列方式：

a）组织多次评价，每位消费者参加所有的评价（完全测试）；

b）组织单次评价，每位消费者只评价部分产品（不完全测试或单个评价）。

### 10.3.3 多次评价

若能安排多次评价时，每位消费者应在测试期间对所有产品展开评价。理想情况是每次评价提供所有产品，但并非每次评价时消费者均能拿到相同产品。

若受到产品准备或可用数量的限制，无法保证每次评价均能提供所有产品时，应将产品随机分配于多次评价中。所有消费者在评价期间均收到相同的产品子集，但顺序应不同。这种方式通常需要更复杂的统计分析。

每次评价开始时，宜向所有消费者提供1个相同产品（常称“热身产品”），以确保消费者每次评价时使用相同的参比样。

若同一批消费者需开展2次评价，不应在同一天进行。

### 10.3.4 单次评价

平衡不完全区组（BIB）设计允许每次评价的产品数量少于产品总数量，但应确保是在相同精度下进行产品对比较。整个测试过程中，样品提供应遵循平衡原则，避免产生顺序效应和转移效应。这种评价方式通常需更多数量的消费者和更复杂的分析方法（参考文献[9]），因此开展实验设计时应寻求相关统计建议。

BIB的设计方法应按照ISO 29842中的规定进行。

## 10.4 同次评价中产品性质

同一轮次中包含不同系列产品时，应满足下列条件：

——产品组合符合惯常的消费习惯；

——所有产品系列的目标消费者一致。

通常不选用非产品目标群体的消费者进行产品评价。

无论参加评价的人员有谁，实验室都应确保评价的机密性。若委托方要求，应允许其参加评价，但无需向其指出参加评价的其他委托方信息。

## 10.5 评价时间

宜选择与该产品的惯常消费时段相同的时间段开展评价，便于消费者在接近真实的情形下进行评价。

## 10.6 测试中产品的重复次数

同一测试中需重复评价某产品时，应视为产品数量的增加。实验方案中应提到的是产品A1和A2，而并非产品A。在确定产品提供顺序和进行结果分析时，都应将产品A1和A2视为2种不同的产品。

## 10.7 预实验

正式测试前，应先采用少量消费者开展预实验。旨在：

——通过回答问卷，核查产品的相关特性能否被理解和感知，确定响应标度是否适宜；

——预演如何安排和管理测试工作；

——确定适宜的评价时长、问题数量和样品数量。

参加预实验的消费者应来自与参加正式测试相同的消费群体，但不应是参加正式测试的消费者样本中的一部分。

# 11 结果分析

## 11.1 总则

首先检查结果数据的内部一致性，计算基本的描述性统计量。报告中应给出数据（例如平均值和中位数等）的至少1个离散参数（例如标准偏差和偏差范围等）。

绘图法非常有效，宜创建每个产品的直方图，通过数据分布（例如单峰）表明样品之间是否存在较大差异。

主成分分析（PCA）或凝聚层次聚类（AHC）可直观展示消费者间的一致性和差异性。

回答“消费者群体是否察觉出评价产品之间的显著差异？”等问题时，需更多的数据统计量。

结果分析可采用参数统计分析或非参数统计分析，后者尤其适用于数据未呈现正态分布的情况。

## 11.2 推断分析

推断分析是利用获取的消费者样本结果外推消费者总体特征的方法。采用的方法取决于测试目标和待测产品的数量。若某些产品被赋予了特殊身份，如作为参比样，将直接影响结果分析。

尽管采用平衡设计的产品提供方案无法消除检测偏差，但能确保产品未受到过度偏爱或提供顺序的不利影响。同时，系统设计的推断分析能减小估计偏差和可接受偏差，提高主要结论的置信度。但推断分析并不容易，也难以直接开展（参考文献[4]和[5]）。

标准的数据统计分析软件包中提供了多种统计方法，应由统计分析师选择适宜的分析方法。

## 11.3 评分法（见9.2）

评分结果可采用ANOVA进行分析。

附录F给出了几种常见的数据分析方法。附录G给出了适用中国消费者的常用喜好标度，包括9点喜好、15cm线性喜好标度和标记情感（LAM）标度。

## 11.4 成对比较法（见9.3）

对于消费者在成对偏爱测试中报告“无偏爱”（选项B）的情形，应先统计“无偏爱”的回答数，将“无偏爱”回答数平均分配到2个产品中（若为基数，减1后再平均分配），再使用“有偏爱”的数据进行显著性分析。

## 11.5 排序法（见9.4）

ISO 8587 给出了针对不同情形所采用的分析方法。

参数统计和非参数统计分析方法均可采用，非参数统计适用于数据不符合正态分布的情况。表1中列出了具体的数据分析方法。

表1 数据分析方法

|  |  |
| --- | --- |
| **情 形** | **方 法** |
| **非参数统计分析** | **参数统计分析** |
| 2个产品 | 分组，每种产品被不同组消费者评价。 | Mann-Whitney’s U检验（也称为Wilcoxon-Mann-Whitney检验） | 独立样本*t*检验 |
| 不分组，所有消费者均评价2种产品。 | Wilcoxon秩和检验 | 成对样本*t*检验 |
| 2个以上产品 | 分组，每种产品被不同组的消费者评价。 | Kruskal-Wallis检验 | 方差分析（ANOVA） |
| 每位消费者至少评价其中的2种产品，理想情况是评价所有产品。 | Friedmann检验（秩方差分析），必要时可采用不完全区组变量 | 消费者效应和产品效应的双因素方差分析（ANOVA） |

# 12 测试报告

## 12.1 总则

报告首页应简要介绍测试结果和重要结论。

报告撰写应简洁明了，确保在无研究计划参照的情况下仍清晰易懂。此外，建议对报告中所有图表进行编号，并附上标题和题注帮助理解。

若委托方要求，可根据测试结果给出相应的建议，并在报告中指出此部分为建议内容或以单独文件形式附上。

除测试结果外，报告还应给出下列详细信息：

——研究题目和参考文献；

——开展研究的日期；

——报告的日期；

——委托方信息；

——实验室和研究负责人的详细身份信息（如果有分包商，其信息应一并提供）；

——测试目标（研究方案总结）；

——委托方访问原始数据的程序；

——报告和每一页的统一标识，包括总页数；

——测试所参考的标准，例如ISO 11136:2014。

## 12.2 产品

产品信息中应给出下列内容：

——产品描述，包括产品全部或部分的组分，若测试实验室了解的话。（宜附上每种产品的照片）

——委托方提供的待测样品的所有信息；

——样品的生产日期、最佳使用期或使用期限，生产批号或批次号；

——实验室收到产品的日期、储存温度、储存时间和储存条件（特别是新鲜和冷冻产品）；

——服务提供商的抽样程序（选择对照样或参比样时，应精确描述抽样方法）；

——待测产品的准备方法；

——提供给消费者时的产品温度；

——提供给消费者的样品数量以及最小和最大消费量的说明，若未遵守该要求，应予以说明；

——相关的产品或废弃处置方式；

——测试期间评价的所有其他产品及提供顺序。

## 12.3 测试程序

测试程序中应给出下列内容：

——采用的测试程序和参照的标准。若采用的是非标准中的程序，应给出完整的步骤描述；

——询问的问题和使用的响应标度；

——数据采集方法；

——测试操作程序；

——评价轮次或环节的总次数；

——评价的日期、次数和持续时长；

——环境条件，例如环境、温度和照明等；

——使用的餐具；

——给消费者的提示语、信息以及提供方式，例如纸质的、电子屏幕的或口头的。

## 12.4 消费者

### 12.4.1 总则

消费者信息应包括下列内容：

——目标消费者群体的描述；

——消费者样本描述，包括委托方指定的每个样本类别的规模和数量；

——消费者来源说明，是特定任务招募还是来自消费者样本库。

### 12.4.2 基于特定任务招募

招募信息中应包括下列内容：

——招募的地点或方式；

——招募程序；

­——筛选适宜消费者的方法。

### 12.4.3 基于消费者样本库招募

若从消费者样本库中招募消费者，报告中应包括下列内容：

——消费者参加消费者测试的频次说明；

——消费者评价过的所有产品的列表；

——消费者参与过的待测产品或产品集的频次。

这些信息通常仅包括测试前12个月内的相关情况。若所有消费者的测试频次为0，也应报告。

## 12.5 测试结果

测试结果中应包括下列内容：

——原始数据提供形式及其描述说明，数据应以图形或表格形式汇总；

——结果数值，包括估计值或均值精度；

——分析和解释结果的方法；

——结果的统计推断；

——根据研究目标，通过测试结果得到的结论。

## 12.6 附件

下列补充资料宜作为研究报告的附件：

——测试中使用的调查问卷或回答表；

——消费者的招募问卷；

——开放式问题回答的逐字记录。

# 附录 A

# （资料性附录）

# 招募问卷示例

| **编号** | **接触式访谈项目××××× 日期** | **跳至问题** | **编码** |
| --- | --- | --- | --- |
| 1 | 您好。我们正在做一个调查，想请您回答几个问题。您之前有参加过有关市场调研的访谈吗？ |  |  |
| 参加过 1 | 2 |  |
| 未参加过 2 | 3 |  |
| 2 | 您上一次参加调查是什么时候？ | 3 |  |
| ***提示信息：受试者的答案小于半年时，结束访谈；反之，继续提问问题3。*** |  |  |
| 3 | 您本人或有无朋友或亲戚在下列行业工作？***提示信息：提供“行业”名单。*** |  |  |
| 行业1 1 | 4 |  |
| 行业2 2 | 4 |  |
| 市场调研 3 | 结束 |  |
| 行业3 4 | 4 |  |
| 行业4 5 | 结束 |  |
| 行业5 6 | 4 |  |
| 广告业 7  | 结束 |  |
| 新闻业 8 | 结束 |  |
| 市场营销 9  | 结束 |  |
|  |  |  |
| 以上皆无 10 | 4 |  |
| 4 | 您的年龄 |  |  |
| 20~39岁 1 | 5 |  |
| 40~60岁 2 | 5 |  |
| 您的性别 |  |  |
| 男性 1 | 5 |  |
| 女性 2 | 5 |  |
| 5 | 这里有一个日常食品清单，请丛中选择您最近购买过的食品。***提示信息：提供“食品”列项。***食品1 1食品2 2食品3 3食品4 4食品5 5食品6 6食品7 7 食品8 8 | 6 |  |
| ***提示信息：如果没有选择“3”，则结束面试。*** |  |  |
| 6 | 您刚说购买了食品3。您大概多久食用1次食品3？***提示信息：提供“使用频次”列项。***每天大于一次 1每天一次 2每周大于一次 3大约每周一次 4大约每两周一次 5 | 7 |  |
| 大约每月一次 6大约每三个月一次 7低于以上频次 8 |  结束 |  |
| 7 | 您定期（如每月至少一次）为家庭采购哪种品牌的食品3？***提示信息：询问生产商和产品！若只提到生产商，请再问一遍！***品牌/生产商 产品1) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 1) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_3) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 3) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_4) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 4) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_5) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 5) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | 8A |  |
| 8A | 该清单中列有不同品牌的食品3。您家使用（即使只是偶尔使用）哪个品牌？ |  |  |
| ***提示信息：提供“品牌”列项并在问卷中选出8A！*** | 8B |  |
| 8B | 您使用这些品牌的频次是多久1次？ |  |  |
| ***提示信息：提供“使用频次”列项，只有当问题8A被圈出时，才回答8B！***  |  |  |
| 8C | 这些品牌中，您最常用的是哪一种? | 8C |  |
| ***提示信息：只有1种选择；根据问题8C，在问卷中圈出您的选择！*** |  |  |
|  | 8A 8B 8C品牌 1 1 \_\_\_\_\_\_ 1品牌 2 2 \_\_\_\_\_\_ 2品牌 3 3 \_\_\_\_\_\_ 3品牌 4 4 \_\_\_\_\_\_ 4品牌 5 5 \_\_\_\_\_\_ 5品牌 6 6 \_\_\_\_\_\_ 6品牌 7 7 \_\_\_\_\_\_ 7品牌 8 8 \_\_\_\_\_\_ 8品牌 9 9 \_\_\_\_\_\_ 9品牌 10 10 \_\_\_\_\_\_ 10品牌 11 11 \_\_\_\_\_\_ 11 | 9 |  |
| 其他信息， \_\_\_\_\_\_\_\_\_\_12 \_\_\_\_\_\_ 12 |  |  |
| 9 | 您有食物不耐受症吗？是 1 | 结束 |  |
| 否 2 | 10 |  |
| 10 | 您有糖尿病吗？是 1 | 结束 |  |
| 否 2 | OK |  |

# 附录B

# （资料性附录）

# 消费者子群细分数据评价方法

## B.1 总则

目标群体子群比较通常依据研究目的开展。例如明确不同组别间是否存在喜好差异，若有差异，可接受的水平为多少。

消费者总样本分组通常采用下列2种方式：

a）消费者子群：依据需要进行证实或证伪的假设类别对消费者展开分组。消费者招募前应先明确分组依据，规定每组样本量大小，确保评价时每组样本量不少于60位消费者，以获得期望的测量精度。

注：根据选择标准，不同类别的样本量无需完全一致。

b）消费者细分：对现有的消费者样本数据（例如测量结果和/或社会人口学的统计数据）分析后再分组。该分组方式能描述消费者样本的特征，并提出新的假设。样本量应足够大，确保后续消费者细分的合理性。

上述2种情况中，研究中所需的消费者数量会依据选用或预期的分组数量以及每组的具体情况而有所差异。

## B.2 消费子群

子群的结果可回答下列问题：

——女性和男性具有相同偏爱吗？

——年龄对偏爱有影响吗？

——国别对偏爱有影响吗？

——消费习惯对偏爱有影响吗？

若每位消费者均评价所有产品，方差分析时应采用下列模型：

响应=组别+消费者（组）+样本×组别+误差

组别根据嵌套在组内的消费者因素进行检验，其他因素则针对误差[等同于样本×消费者（组）的交互效应]进行检验。

该模型中可能还包括顺序效应，有必要时还包括转移效应。若含有顺序效应，则模型如下：

响应=组别+消费者（组）+样本+顺序+样本×组别+顺序×组别+样本×顺序+误差

采用该模型时，若所有组别（顺序×组别的交互效应）和所有样本（样本×顺序的交互效应）的顺序效应相同，即可确定组别对样本的影响（样本×组别的交互效应）。

若各组别在样本上(样本×组别的交互效应)存在显著性差异，则各组别应单独评价。

## B.3 消费者细分

后续进一步进行消费者细分旨在解释结果的偏爱性（若存在该种情形）。

若直方图检测到双峰分布时，开展消费者细分非常有效。

不考虑消费者样本时可采用内部偏爱映射图对数据进行分析。通常，可采用主成分分析的等级矩阵，横列代表产品，竖列代表消费者。

注：主成分分析是一种多变量数据分析方法，旨在用最小数量的主成分或原始变量和部分相关联变量的线性组合表达结果的变异性。

“同时的”表示产品可以叠加在消费者之前。

只有排列靠前的少数主成分能够解释很高比例的原始变化时，这些主成分提供的信息才有效。通常，只有完全测试才能使用主成分分析（参考文献[11]）。不完全测试的结果分析见参考文献[10]。

因此，第一主成分中的变量可以区分消费者组（簇）。若需进一步辨别消费者组，应采用聚类分析。

# 附录C

# （资料性附录）

# 产品呈送方案

最有效和最常采用的消费者产品提供方案是随机完全区组设计。对每位消费者随机独立提供样品，且每个样品只评价1次。该方案的优点是使用简单区组模型便可进行结果验证。但多数情况下，由于诸多因素的影响，实验者需采用更复杂的产品提供方案。

——不完全区组设计：样品量过大时，每位消费者可能无法评价完所有产品。消费者能有效评价的样品量取决于产品性质。该情况下，通常采用平衡不完全区组设计，具体见ISO 29842。每位消费者的产品提供顺序应遵循随机和独立原则。

——顺序效应：实验者经常担心存在顺序效应，先品尝产品的评分通常高于后品尝产品。因此，完全区组和平衡不完全区组实验设计时应考虑随机因素，确保每个产品提供的先后顺序具有同等机会。

若实验者担心顺序效应影响过强时，应采用广义拉丁方有效替代随机完全区组。广义拉丁方是产品提供顺序中，每个产品出现在每个位置（例如第1位、第2位等）的频次相同。

不完全区组设计中，按照所有产品在每个位置上以相同几率出现的方式安排产品提供顺序称为广义尧敦设计（Youden design）。

应注意，由于广义拉丁方或广义尧敦设计无法采用简单区组模型验证，应采用考虑了顺序效应的行列模型。

——延滞效应：尽管评价期间设置了感觉清除期，但前1种产品仍有可能影响下一个产品品尝的味道。若实验者担心延滞效应，可使用相邻平衡设计（参考文献[2]）。参考文献[2]和[9]报道了使用相邻平衡设计的相关情况。

——组织事项：某些情况下，测试需要更复杂的实验设计方案。例如需要对8种产品进行比较，由于只有4个加热点，因此每次只能提供4种产品进行评价。这时应采用更复杂、更具针对性的实验设计和分析，并在统计分析师指导下进行方案设计。

# 附录D

# （资料性附录）

# 样本量对均值精度的影响

应明确每个消费者组的最小样本量，确保产品小组评分值具有明显差异，更重要是确保评分均值具有一定的精度。上述同样适用于消费者总样本或一类样本或显示存在个体差异的消费者子群。

精度取决于样本量和评分值在组内的变化性。

假定产品评分呈现正态分布，*σ*为标准偏差，采用公式（D.1）可以获得预测精度*Ac*下估算平均值所需的最小消费者样本量：

$A\_{C}=\frac{2σ}{\sqrt{n}}$ （D.1）

式中：

$A\_{C}$——平均值95%置信区间的一半宽度。

例如*AC*＝0.2时，平均值的置信区间为（*m*-0.2，*m*+0.2）。

表D.1给出了*σ*，$n$和$A\_{C}$之间的关联。

表D.1 获得预测精度所需的消费者样本量

|  |  |
| --- | --- |
| **精 度**$$A\_{C}$$ | **标准偏差** *σ* |
| **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 |

*σ*和$A\_{C}$应采用相同单位。

确定消费者招募数量时，*β*风险宜予以考虑（附录E）。

注：若要求测试前计算所需样本量时，应先进行统计效力分析，固定效应量（ES）（参考文献[8]）后再确定样本量$n$值。

# 附录E

# （资料性附录）

# 评分法的最小样本量

## E.1 总则

消费者测试中，每个产品适宜评价数量应不少于60次，100次左右最佳。每位消费者仅检验1种产品时，需要消费者数量$n=100×p$。其中，$p$是待测产品数量。若每位消费者检验$k$个产品，按经验所需消费者数量为$n=100×(p/k)$位。

该经验法并不够完善，应采用其他更精确的方法计算样本量。计算前应先回答下列问题：

——项目的研究目的？是证明产品和参比样之间或2个及更多产品之间存在差异或具有相似性？

——待测产品的数量？1个、2个、3个或更多？

——每位消费者应评价多少个产品？

——选择的*α*风险是多少？

——期望的检验效力*P*或*β*风险是多少？（或可接受的*β*风险是多少，$P=1-β$）

——差别检验的差别度δ是多少？对于相似性检验，不接受的差异值*Δ*是多少？

——产品评分值的分散度是多少？

——采用的是单边还是双边的备择假设？

——消费者同时评价多个产品时，不同产品评分值是否存在关联？

本附录仅针对2种产品的检验。对于更简单（1种产品）或更复杂（3种或更多产品）的情况，应咨询统计分析师或使用适宜的统计分析软件，也可参考AFNOR NF V09-500：2012 [2]。

本附录包括下列4种情况：2种差异性检验和2种相似性检验。

## E.2 示例1：确定2种产品A和B之间存在差异，每位消费者仅对1种产品评分

每个产品所需的消费者数量$n$由公式（E.1）计算得到：

 $n=2×[\frac{(u\_{α/2}+u\_{β})×σ\_{产品内}}{δ}]^{2}$ （E.1）

式中：

$u\_{α/2}$*——*双边备择假设中标准正态分布的（1 - *α*/2）分位数，*α*=0.05时，*uα/*2=1.960；当单边备择假设时，*uα*替代*uα*/2，*α*=0.05时，*uα*=1.645；

$u\_{β}$*——*可接受*β*风险下，标准正态分布的(1-*β*)分位数（期望效力*P* = 1-*β*）；

*σ*产品内*——*产品内的标准偏差，假定2种产品具有相同的标准偏差；

*δ——*2种产品间的检验差异。

量表是用来根据正态分布进行近似评分的

**注**：*σ*产品内和*δ*采用相同标度表示，例如9点标度。9点标度应根据正态分布进行近似评分。

示例：如下列条件所示

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *α* | *uα*/2 | β | (*P* = 1-*β*) | *uβ* | *σ*产品内 | *δ* |
| 0.05 | 1.960 | 0.10 | (0.90) | 1.282 | 1.50 | 0.50 |

由公式（E.1）得到$n$= 189.2，即1种产品采用190位消费者，2种产品采用380位消费者。

注：对于A> B的单边假设，由*uα*= 1.645（*α*= 0.05）和公式（E.1）得到$n$= 154.2，即155位消费者。

## E.3 示例2：确定2种产品A和B之间存在差异，每位消费者对2种产品均评分

消费者数量由公式（E.2）计算得到：

$n=2×[\frac{(u\_{α/2}+u\_{β})×σ\_{产品内}×\sqrt{1-ρ}}{δ}]^{2}$ （E.2）

式中：

$u\_{α/2}$*——*双边备择假设中标准正态分布的（1 - *α*/2）分位数，*α*=0.05时，*uα/*2=1.960；当单边备择假设时，*uα*替代*uα*/2，*α*=0.05时，*uα*=1.645；

$u\_{β}$*——*可接受*β*风险下，标准正态分布的(1-*β*)分位数（期望效力*P* = 1-*β*）；

*σ*产品内——产品内的标准偏差，假定2种产品具有相同的标准偏差；

*δ*——2种产品之间的检验差异。

$ρ$——消费者对2种产品评分值在群体中的相关性。

 注：*σ*产品内和*δ*采用相同标度表示，例如9点标度。

由于

$σ\_{d}=σ\_{产品内}×\sqrt{2(1-ρ)}$

式中：

$σ\_{d}$——每位消费者对2种产品评分差值的标准偏差；

因此，公式（E.2）表示为

$n=[\frac{(u\_{α/2}+u\_{β})×σ\_{d}}{δ}]^{2}$ （E.3）

公式（E.3）是目前最常用的表达式。

示例：如下列条件所示

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| α | *uα*/2 | *β* | (*P* = 1 − *β*) | *uβ* | *σ*d | *δ* |
| 0.05 | 1.960 | 0.10 | (0.90) | 1.282 | 1.70 | 0.50 |

由公式（E.3）得到$n$= 121.5，即采用122位消费者。

注：对于A> B的单边假设，由*uα*= 1.645（*α*=0.05）和公式（E.3）得到$n$= 99.0。

## E.4 示例3：确定产品A不劣于产品B，每位消费者仅对1种产品评分

每种产品所需消费者的数量由公式（E.4）得到：

 $n=2×[\frac{(u\_{α}+u\_{β})×σ\_{产品内}}{∆}]^{2}$ （E.4）

式中：

*uα*——单边备择假设（非劣效性检验）中标准正态分布的（1 - *α*）分位数；

*uβ*——可接受*β*风险下，标准正态分布的(1 *– β*)分位数（期望效力*P* = 1-*β*）；

*σ*产品内——产品内的标准偏差，假定2种产品具有相同的标准偏差；

*Δ*——认为2种产品相似（可互换）时所不能接受的差异值。

注：*σ*产品内和*δ*采用相同标度表示，例如9点标度。

示例：如下列条件所示

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *α* | *uα* | *β* | (*P* = 1 − *β*) | *uβ* | *σ*产品内 | *Δ* |
| 0.05 | 1.645 | 0.10 | (0.90) | 1.282 | 1.40 | 0.30 |

由公式(E.4)得到$n$= 373.2，即1种产品采用374位消费者，2种产品共采用748位消费者。

## E.5 示例4：确定产品A不劣于产品B，每位消费者对2种产品均评分

消费者的数量由公式（E.5）计算得到：

$n=[\frac{(u\_{α}+u\_{β})×σ\_{d}}{∆}]^{2}$ （E.5）

式中：

*uα*——单边备择假设（非劣效性检验）中标准正态分布的（1 - *α*）分位数；

*uβ*——可接受*β*风险下，标准正态分布的(1 *– β*)分位数（期望效力*P* = 1-*β*）；

*σd*——每位消费者对2种产品评分差值的标准偏差；

*Δ*——认为2种产品相似（可互换）时不能接受的差异值。

注：*σ*产品内和*δ*采用相同标度表示，例如9点标度。

示例：如下列条件所示

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *α* | *uα* | *β* | (*P* = 1 − *β*) | *uβ* | *σ*d | *Δ* |
| 0.05 | 1.645 | 0.10 | (0.90) | 1.282 | 1.70 | 0.30 |

由公式（E.5）得到*n* = 275.1，即采用276位消费者。

注：示例3和4是非劣效性检验。若实验室要确定产品A不劣于或优于产品B，应采用严格相似性检验。该检验所需的消费者数量很难计算（没有直接计算该数值的公式），通常需要更多的样本量。如示例3中（2种产品共需960位消费者），每个产品需480位消费者，示例4中消费者数量为698。使用严格相似性检验时，实验室应在统计分析师指导下、并使用适宜的计算机统计分析软件完成。应明确ASTM E1958-07[3]中的方法仅适用于成对比较、且每位消费者同时收到产品A和B的严格相似性检验。

# 附录F

# （资料性附录）

# 评分法数据分析的示例

## F.1 总则

本附录采用附录E中的示例。前2个示例是差异检验，其中，原假设*H*0设置为：2种产品A和B的群体评分均值相同；后2个示例是非劣效性检验，其中原假设*H*0设置为：产品A劣于产品B的量至少等于*Δ*。

差异检验均采用双边假设，单边假设的情况应咨询统计分析师。非劣效性检验通常采用单边假设。

## F.2 示例1（附录E示例1）：证实2种产品A和B存在喜好差异，每位消费者仅对1种产品评分

——与委托方讨论后，实验室采用230位消费者进行测试：A产品110位消费者（*n*A=110）和B产品120位消费者（*n*B=120）。2种产品的评分均值和标准偏差如下：$m\_{A}$= 7.23，$s\_{A}$= 1.85，$m\_{B}$ = 6.87，$s\_{B}$=1.65。

—— *t*cal值由公式（F.1）计算得到：

$t\_{cal}=\frac{m\_{A}-m\_{B}}{S\_{产品内}×\sqrt{（\frac{1}{n\_{A}}+\frac{1}{n\_{B}}）}} $ （F.1）

式中：

*s*产品内——产品内的标准偏差；

*s*产品内采用公式（F.2）计算得到：

$s\_{产品内}^{2}=\frac{s\_{A}^{2}×\left(n\_{A}-1\right)+s\_{B}^{2}×\left(n\_{B}-1\right)}{n\_{A}+n\_{B}-2} $ （F.2）

因此，本示例中：

$$ s\_{产品内}^{2}=\frac{1.85^{2}×\left(110-1\right)+1.65^{2}×\left(120-1\right)}{110+120-2} $$

和

$$ \sqrt{s\_{产品内}^{2}}=\sqrt{3.057}=1.748$$

*t*cal值为

$$t\_{cal}=\frac{7.23-6.87}{1.748×\sqrt{（\frac{1}{110}+\frac{1}{120}）}}=\frac{0.36}{0.231}=1.558 $$

将*t*cal值与所选*α*风险下的*t*the值进行比较，*t*the值对应于自由度为（*n*A + *n*B - 2）时*t*分布的（1 -*α*/ 2）分位数。本例中，*α*＝0.05时，*t*the=1.970。

注1：*t*the值应采用Excel[[1]](#footnote-1)）中的函数计算：*α*＝0.05时，T.INV.2T (*α*;d*f*) = T.INV.2T (0.05;110 + 120–2)=1.970。

*t*cal>*t*the时，拒绝*H*0。本例中，1.558 <1.970，该条件不满足，无法得到产品A和B均值之间存在显著性差异的结论。

——因此，实验室使用公式（F.3）计算检验效力：

*P* = 1 – *β* （F.3）

$t\_{β}$值采用公式（F.4）计算得到：

$t\_{β}=t\_{the}-\frac{δ}{t\_{calden}}$ （F.4）

式中：

*t*the——拒绝*H*0的值；

*δ*——2种产品之间的偏爱概率差异；

*t*calden——*t*cal的分母[公式（F.1）]。

因此，

$$t\_{β}=1.970-\left(\frac{0.50}{0.231}\right)=-0.195$$

$t\_{β}$<0时，*β*值由$t\_{β}$和自由度为（*n*A + *n*B - 2）的t分布函数得到。$t\_{β}= $-0.195、*df* =110 +120-2=298时，*β=*0.423。因此，检验效力$P=1-β=1-0.423≈0.58$，该值远低于要求值0.9。

**注2**：*β*值应采用Excel中的函数计算：T.DIST(x;deg\_freedom;TRUE) = T.DIST(−0.195;228;TRUE) = 0.423。

**注3**：*t*β> 0时，*P*值由t分布函数在（-1）×*t*β计算得到。例如δ=0.30时，*t*β= 1.970 -（0.30 / 0.231）=0.671；通过t分布函数得到*P* =0.251≈0.25。

**注4**：*t*the值应采用Excel中的函数计算：T.INV.2T(0.05;119) = 1.980；其中，*df* =（120 − 1）= 119。

## G.3 示例2（附录E示例2）：证实2种产品A和B存在喜好差异，每位消费者对2种产品均评分

实验室采用120位消费者（*n* = 120）进行测试，得到$m\_{A}$= 6.33，$ m\_{B}$= 6.66，$s\_{d}$=1.92。

其中，

$s\_{d}$*——*每位消费者对2种产品评分差值的标准偏差；

——*t*cal值由公式(F.5)计算得到：

 $t\_{cal}=\frac{m\_{A}-m\_{B}}{s\_{d}×\sqrt{1/n}}=\frac{6.33-6.66}{1.92×\sqrt{1/120}}=\frac{-0.33}{0.175}=-1.886$ （F.5）

——将*t*cal值与所选*α*风险下的*t*the值进行比较。*t*the值由分位数为（1-*α*/ 2）和自由度为（*n*-1）时的t分布函数得到。本例中，当*α*= 0.05，*t*the=1.980。

当∣*t*cal∣>*t*the时，拒绝*H*0，产品A和B之间存在显著性差异。本例中并不满足该条件（∣-1.886∣<1.980）不能得到产品A和B的评分有显著性差异的结论。

——因此，实验室采用公式(F.6)计算测试的检验效力：

$ t\_{β}=t\_{the}-(\frac{δ}{t\_{calden}})$ （F.6）

式中：

*the*——拒绝*H*0的值；

*δ*——2种产品之间的偏爱差异；

*t*calden——*t*cal的分母[公式(F.5)]；

*tβ* = 1.980 – (0.50/0.175) = –0.877

由于公式(G.4)和(G.6)中*β的*计算规则相同，*β*值由分位数为（-0.877）和自由度为（120-1）时t分布函数得到，*β=*0.192。因此，检验效力*P=*1-0.192 = 0.808，约81%。通常*P*值高于80%认为有效。

注：*β*值应采用Excel中函数计算：T.DIST(−0.873;119;TRUE) = 0.192, *P* = 1− 0.192 = 0.81。

## G.4 示例3（附录E示例3）：通过$∆$值证实产品A不劣于产品B，每位消费者仅对1种产品评分

附录E计算得出每个产品应有*n* = 374位消费者。实验室采用300位消费者对产品A（*n*A = 300）和316位消费者对产品B（*n*B = 316）进行测试。得到：*m*A = 7.01，*s*A = 1.95，*m*B = 6.87，*s*B = 1.80。

——非劣效性测试中，*t*cal值由公式（F.7）得到：

$ t\_{cal}=\frac{m\_{A}-m\_{B}+∆}{s\_{产品内}^{}×\sqrt{(\frac{1}{n\_{A}}+\frac{1}{n\_{B}})}} $ （F.7）

式中：

$$s\_{产品内}^{2}=\frac{s\_{A}^{2}×\left(n\_{A}-1\right)+s\_{B}^{2}×\left(n\_{B}-1\right)}{n\_{A}+n\_{B}-2}$$

本例中，

$$s\_{产品内}^{2}=\frac{1.95^{2}×\left(300-1\right)+1.80^{2}×\left(316-1\right)}{300+316-2}=3.514$$

$$s\_{d}=\sqrt{s\_{产品内}^{2}}=\sqrt{3.514}=1.875$$

*t*cal值为

$$t\_{cal}=\frac{7.01-6.87+0.30}{1.875×\sqrt{（\frac{1}{300}+\frac{1}{316}）}}=\frac{0.44}{0.151}=2.914$$

——将*t*cal值与所选*α*风险的*t*the值进行比较。 *t*the值对应于自由度为（*n*A + *n*B - 2）时*t*分布的（1 -*α*）分位数。本例中，*α*= 0.05，*t*the=1.647。

*t*cal>*t*the时，拒绝*H*0。本例中，2.914> 1.647，满足该条件，实验室可拒绝*H*0并得到结论：非劣效性检验产品A与B具有显著相似性。

注：*tthe*值应采用Excel1）中的函数计算：T.INV(1−*α;n*1+*n*2-2) = T.INV(0.95,614) = 1.647。此处，自由度= (300 + 316 − 2) = 614。

——当实验室不能得到非劣效性检验具有显著性相似的结论时，应通过公式（F.8）计算检验效力。

 $t\_{β}=t\_{the}-(\frac{∆}{t\_{calden}})$ （F.8）

式中：

*t*the——拒绝*H*0的值；

*Δ*——认为2种产品相似（即可互换）的不可接受的差异；

*t*calden——*t*cal的分母[公式（F.7）]。

示例：若*m*A = 6.87，*s*A = 1.95，*m*B = 7.01，*s*B = 1.80；

$$t\_{cal}=\frac{6.87-7.01+0.30}{1.875×\sqrt{(\frac{1}{300}+\frac{1}{316})}}=\frac{0.16}{0.151}=1.060$$

由于1.060＜1.647，不能拒绝*H*0。*tβ*值为：

*tβ* = 1.647 – (0.30/0.151) = 1.647 – 1.987 = -0.340

公式（F.4）、（F.6）和（F.8）中，*β*计算规则均相同。由于*tβ*<0，*β*值由在*tβ*处自由度为（nA + nB-2）时*t*分布的分布函数得到。*tβ*=-0.340，自由度=300 + 316-2时，*t*分布函数等于0.367。因此，检验效力*P*=1-*β*= 1-0.367 =0.633≈0.63。这个值远低于要求值90%。

## G.5 示例4（附录E示例4）：证实产品A不劣于产品B，每位消费者对2种产品均评分

——附录E中得到*n*=374位消费者，但实验室仅采用200位消费者进行测试（*n* = 200），得到：*m*A = 7.26，*m*B = 7.31，*s*d = 2.08。

式中：

*s*d——每位消费者对2种产品评分差值的标准偏差；

——非劣效性测试中，*t*cal值由公式（G.9）得到：

 $t\_{cal}=\frac{m\_{A}-m\_{B}+∆}{s\_{d}×\sqrt{1/n}}$ （F.9）

本例中，

$$t\_{cal}=\frac{7.26-7.31+0.30}{2.08×\sqrt{1/200}}=\frac{0.25}{0.208}=1.202$$

 注1：公式（F.9）中，*tthe*值应采用Excel中的函数计算：T.INV(0.95*;199*) = 1.653。

——将该值与所选*α*风险的*t*the值进行比较。*t*the值由自由度为（*n*-1）时t分布的（1– *α*）分位数得到。本例中，*α*= 0.05，*t*the=1.653。

*t*cal>*t*the时，拒绝*H*0。本例中，1.202 <1.653，该条件不满足，实验室不能拒绝*H*0，无法得到非劣效性检验产品A与B具有显著相似性的结论。

——应采用公式（G.10）计算其检验效力：

 $t\_{β}=t\_{the}-(\frac{∆}{t\_{calden}})$ （F.10）

式中：

*t*the*——*拒绝*H*0的值；

*Δ——*认为2种产品相似（即可互换）的不可接受的差异；

*t*calden*——t*cal的分母，见公式（F.9）；

本例中，

*t*β= 1.653 − (0.30/0.208) = 0.211

公式（F.4）、（F.6）、（F.8）和（F.10）中*β*计算规则相同。由于*t*β> 0，所以效力由-1×*t*β= - 0.211、自由度（200 - 1）的*t*分布函数直接得到，等于0.417≈0.42。这个值并不令人满意。

注2：*P*值应采用Excel中的函数计算：T.DIST(−0.211;199;1) = 0.417计算。

# 附 录 G

# （规范性附录）

# 适合中国消费者的喜好标度（评分法）

## G.1 9点喜好标度

表 1 9点喜好标度

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 语义 | 极不喜欢 | 很不喜欢 | 不喜欢 | 有点不喜欢 | 谈不上喜欢不喜欢 | 有点喜欢 | 喜欢 | 很喜欢 | 极喜欢 |
| 数值 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

## G.2 线性喜好标度



图2 15厘米喜好标度

## G.3 标记情感（LAM）标度



图3 LAMS

# 参 考 文 献

[1] Abdi H., Edelman B., Valentin D., Dowling W.G. 2009. Experimental design and analysis for psychology. Oxford University Press, *pp*. 1–538.

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

[3] ASTM E1958–12 2012. Guide for Sensory Claim Substantiation.

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

[5] Bi J. Similarity testing using paired comparison method. 2007. Food Qual. Prefer., 18, 500–507.

[6] Blackwelder W.C. 1982. “Proving the null hypothesis” in clinical trials. Control. Clin. Trials., 3, 345–353.

[7] Cochran W.G., Cox G.M. 1957. Experimental Designs. Wiley, New York, Second Edition.

[8] Cohen J. Statistical Power analysis for the behavioural sciences. 1988. Elsevier-Academic Press, New York, Second Edition.

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

[10] Greehouse S.W., Geisser S. 1959. On methods in the analysis of profile data. Psychometrika, 24, *pp.* 98–112.

[11] Howell D.C. 2002. Statistical methods for psychology. Duxbury, Pacific Grove, Fifth Edition, *pp.* 1–802.

[12] Kunert J., Sailer O. 2006. On nearly balanced designs for sensory trials. Food Qual. Prefer., 17**,** *pp*. 219–227.

[13] Kunert J. 1998. Sensory experiments as crossover studies. Food Qual. Prefer., 9,*pp.* 243–253.

[14] Lawless H.T., Heymann H. 2010. Sensory evaluation of food – principles and practices. Springer, New York, Second Edition, *pp*. 1–596.

[15] Mac Fie H.J., Bratchell N., Greenhoff K., Vallis L.V. 1989. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. J. Sens. Stud., 4**,** *pp.* 129–148.

[16] Meilgaard M.C., Civille G.V., Carr B.T. 2007. Sensory evaluation techniques. CRC Press, Boca Raton, Fourth Edition, *pp.* 1–448.

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

[18] Meyners M. 2012. Equivalence tests – a review. Food Qual. Prefer., 26,*pp.* 231–245.

[19] Naes T., Brockhoff P.B., Tomic B. 2010. Statistics for sensory and consumer science. Wiley, Chichester, Second Edition, *pp.* 1–287.

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

[21] Siegel S., Castellan N. 1988. Nonparametric systems for the behavioural sciences. Singapour: Mc Graw-Hill International Editions, Second Edition.

[22] Ten Kleijf F., Musters P.A.D. 2003. Text analysis of open-ended survey responses: a complementary method to preference mapping. Food Qual. Prefer., 14, *pp.* 43–52.

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

[24] Wakeling I.N., Hasted A., Buck D.2001. Cyclic presentation order designs for consumer research. Food Qual. Prefer., 12, *pp.* 39–46.

[25] ISO 20252. Market, opinion and social research — Vocabulary and service requirements.



1. ） Excel是Microsoft旗下的一个产品。提供此信息仅为方便本文件的用户，并不构成标准对所提到产品的认可。如果其他同类产品可以获得相同的结果，亦可采用。 [↑](#footnote-ref-1)