



中华人民共和国国家标准

GB/T ×××××—202×

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

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

××××-××-×× 发布

××××-××-×× 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 测试前准备工作	3
4.1 实验方案	3
4.2 目标群体确定依据	4
4.3 其他	4
5 消费者样本	5
5.1 总则	5
5.2 消费者招募	5
5.3 消费者样本构建	6
5.4 消费者样本细分	6
6 消费者样本量	7
6.1 测量精度	7
6.2 样本量确定	7
6.3 消费者子群使用	8
6.4 待测产品数量和实验设计	8
6.5 招募数量的安全限	8
7 测试区域	8
8 产品	9
8.1 产品盲评	9
8.2 产品制备和提供	9
9 程序	10
9.1 喜好测试分类	10
9.2 评级法	10
9.3 成对偏爱法	10
9.4 排序法	11
9.5 其他	11
9.6 给消费者的提示语	12
10 测试	12
10.1 总则	12
10.2 实验设计	13
10.3 单一轮次测试的产品数量	13
10.4 同轮次测试中产品的性质	14

10.5	测试的时间	14
10.6	产品重复测试的次数	14
10.7	预实验	14
11	结果分析	15
11.1	总则	15
11.2	推断分析	15
11.3	评级法(见 9.2)	15
11.4	成对偏爱法(见 9.3)	15
11.5	排序法(见 9.4)	15
12	测试报告	16
12.1	总则	16
12.2	产品	16
12.3	测试程序	17
12.4	消费者	17
12.5	测试结果	17
12.6	附件	18
附录 A (资料性)	招募问卷示例	19
附录 B (资料性)	消费者子群细分数据评价方法	22
附录 C (资料性)	样本量对结果精度的影响	24
附录 D (资料性)	评级法的最小样本量	25
附录 E (资料性)	产品提供方案	28
附录 F (资料性)	评级法数据分析的示例	29
附录 G (资料性)	中文语义喜好标度	33
参考文献	34

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

本文件参考 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》起草，一致性程度为非等效。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国感官分析标准化技术委员会(SAC/TC 566)提出并归口。

本文件起草单位：中国标准化研究院、泸州老窖股份有限公司、四川郎酒股份有限公司、重庆德庄农产品开发有限公司、中国绿色食品有限公司、华测检测认证集团股份有限公司、内蒙古伊利实业集团股份有限公司、浙江工商大学、北京林业大学、四川源坤教育科技有限公司。

本文件主要起草人：钟葵、赵镭、史波林、汪厚银、王明、蒋英丽、李德建、唐飞、李丰勇、苏玉芳、田师一、张璐璐、朱保庆、杨宇、刘龙云、杜娟、沈毅、张丽、魏阳吉、林兆盛。

感官分析 方法学

受控区域消费者喜好测试一般导则

1 范围

本文件给出了在受控区域内测量消费者对产品喜好程度或相对喜好程度的方法。该方法是基于问卷调查(包括纸质问卷和电子问卷)的消费者数据收集和分析方法。

本文件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。

本文件中的喜好测试方法(包括接受性测试与偏爱测试)可应用于:

- 比较产品与竞争产品的喜好差异;
- 优化产品,以期获得更高的喜好级别或更多消费者的喜爱;
- 确定适合某个特定目标消费人群的产品系列;
- 确定产品的最佳消费期;
- 评估配方变化对产品喜好的影响;
- 研究产品的感官特性(非外在特性,如品牌,价格,广告等)对产品喜好程度的影响;
- 研究商业行为或展示形式(如包装)等产品喜好程度的影响;
- 确定不同产品间存在可感知的偏爱差异(喜好程度差异检验)或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

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;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。

- GB/T 10221—2021 感官分析 术语(ISO 5492: 2008, IDT)
- GB/T 12315—2008 感官分析 方法学 排序法(ISO 8587:2006, IDT)
- 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, IDT)
- GB/T 39501—2020 感官分析 定量响应标度使用导则(ISO 4121: 2003, IDT)
- GB/T 39992—2021 感官分析 方法学 平衡不完全区组设计(ISO 29842: 2011, IDT)

3 术语和定义

GB/T 10221—2021 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

提供以下功能的区域:

- 准备和提供产品的可控区域;
- 消费者测试产品和接受提问的舒适环境;
- 避免消费者之间口头和非口头交流,确保每位消费者独立反馈意见的区域。

[来源:GB/T 10221—2021,6.56,有修改]

3.2

消费者 consumer

受访者 respondent

产品使用者。

注：本文件中定义的消费者，不包括 GB/T 10221 中定义的专家、专家感官评价员和优选评价员。

[来源：GB/T 10221—2021, 3.12, 有修改]

3.3

消费者样本 consumer sample

以目标群体特性作为招募标准的消费者群体。

3.4

喜好感知 hedonic perception

产品感官特性引起的愉悦程度。

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相对更喜欢产品的测试。

[来源：GB/T 10221—2021, 6.28]

3.7

[可]接受性 acceptability

对刺激整体或特定感官特性喜欢或不喜欢的程度。

[来源：GB/T 10221—2021, 3.23]

3.8

评级[法] rating

采用顺序标度测量各种感知的量级，并将其表示为多个可能类别之一的方法。

注：常用标度包括类项标度、线性标度以及 CATA 法的列表。

[来源：GB/T 10221—2021, 6.6]

3.9

评分[法] scoring

用对产品或产品特性具有数学意义的数字来评价产品或产品特性。

[来源：GB/T 10221—2021, 6.7]

3.10

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分属不同企业或机构。二者隶属同一企业或机构的不同部门时，在研究方案中对双方职责进行详细划分和明确说明。

3.11

目标群体 target population

参与产品喜好响应测试的人群。

3.12

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：通过设定的休息时间可将1个轮次分成多个环节，例如1个90 min的轮次可分为2个环节，中间休息10 min。

[来源：GB/T 10221—2021,6.63,有修改]

3.13

开放式问题 open-ended question

不给受访者提供现成的选项，要求其自由作答的问题。

注：问题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）。

[来源：GB/T 10221—2021,6.61,有修改]

3.14

封闭式问题 closed-ended question

要求受访者从预先确定的可能选项中选择答案的问题。

[来源：GB/T 10221—2021,6.53,有修改]

3.15

产品集 product family

正常消费过程中可以相互替代的产品群。

注：产品集会因文化和消费习惯差异而有所不同。

3.16

代表性 representativeness

样本的个体特性与其所在群体的相关特性的一致程度。

3.17

可访问样本库 access panel

承诺被选中后会配合完成后期数据采集的潜在受访者样本数据库。

注：不包括连续调查的受访者固定样本组，如收视率调查固定样本组和消费者固定样本组。

[来源：GB/T 26315—2010,5.2.11]

3.18

位置偏差 position bias

由一个产品在待测产品组中所处的空间位置引起的偏差。

[来源：GB/T 10221—2021,3.44]

3.19

顺序偏差 sequential bias

由一个产品在待测产品组中所处的时间或顺序引起的偏差。

[来源：GB/T 10221—2021,3.47]

4 测试前准备工作

4.1 实验方案

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

——研究目的；

- 待测产品(若实验室需采购 1 个或多个待测产品时,应提供采购程序);
- 测试目标人群;
- 消费者样本量及招募方式;
- 对消费者前期参与过的消费者测试项目的限制条件;
- 对同轮次测试中其他待测产品的限制条件;
- 测试条件;
- 调查类型(如自填式问卷或访问式问卷);
- 调查问卷;
- 测试场所;
- 每轮次中待测产品数量,若同一轮次中需测试多个产品系列,应说明其他待测产品的情况(若实验室在制定实验方案时尚不清楚产品的相关情况,可在明确产品性质后再及时与委托方沟通);
- 待测产品的准备条件(若已知);
- 待测产品的提示语(若已知),产品测试前提供有关产品或测试的具体信息;
- 用于结果分析和解释的统计方法,方法的风险类型以及可接受水平,如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;
- 基于测试前已知晓的消费者的群体特征或可能响应预判,提出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;
- 预期的测试时间和报告提交日期。

若制定实验方案时上述任何 1 项内容尚未知晓,可在方案中列出下列陈述:“该项内容将由委托方在…(特定研究阶段)指定。”

注:若已有相关测试标准,实验方案中可直接列出参考标准的名称,无需重复给出标准中的操作步骤和技术要点。

4.2 目标群体确定依据

回答下列问题对目标群体确定非常重要,应在每次测试时重复询问这些问题,因为即使是相同产品的测试,目标消费者也可能存在差异。

- 待测产品是否已上市?若已上市,是否有可能区分现实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现实消费者,还是未使用过产品的潜在消费者?或者两者兼有?
- 是否需要对特定消费子群的测试结果进行检验和比较?
- 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

目标群体确定后,还应进一步确定下列内容:

- 测试场所(见第 7 章);
 - 预期的测量精度(见 6.1)和测试方法(见第 9 章);
 - 选用测试方法中所对应的产品提供方案(见 10.2);
 - 假设检验中参数确定(见 6.1)。
- 假设检验时,应明确消费者样本组成(见第 5 章)和数量(见第 6 章)。

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接受性和/或对不同产品(2个或多个)的偏爱。

消费者样本组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,且可以接受测试相关的酬谢(如金钱或奖励)。

消费者的本能回应当对于测试结果非常重要,应剔除接受过培训的消费者,因为他们与目标消费群体存在显著系统性差异。

通常不建议使用生产企业内部人员构建消费者测试小组,下列因素有可能导致结果产生偏差:

- 待测产品存在被识别的风险;
- 消费者倾向于选择能识别出的产品;
- 消费者对产品过于熟悉。

上述因素存在测试结果无法代表目标消费群体的风险。无论采用哪种招募方式,测试人员的招募标准(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

消费者招募应遵循以下2点:

- 创建和记录个人数据文件时,应遵守相关法律及保密规定;
- 遵守有关年龄限制的法规。

消费者招募通常有2种方式,一种是基于特定任务(为某特定研究任务新招募的消费者)招募者,另一种是从可访问样本库中招募(具有已知特征的潜在消费者样本库)。通常认为基于特定任务的招募能得到较少参与消费者测试活动的消费者,但随着喜好测试越来越多地在多地重复展开,这种认知可能并不正确。

5.2.2 基于特定任务的招募

特定任务的消费者招募,通常在公共场所(例如商场、超市等)通过电话、邮件(书面或电子邮件)、当地报纸、电台广告、社交网络或其他互联网等多种形式开展。事先与消费者招募区域的负责公司或组织沟通并达成一致意见对开展招募非常重要。

招募方式及与消费者接触的时间(特别是公共场所采用电话沟通以及通过媒体发布信息)均对消费者样本的组成具有重要影响。

5.2.3 基于可访问样本库的招募

从可访问样本库中抽取消费者构建消费者群体是一种高效的招募方式。在测试开展及目标群体确定前,可先根据样本库中消费者的基本信息进行预筛选。

消费者信息包括下列内容:

- 身份信息和联系方式;
- 年龄和性别;
- 婚姻状况和家庭组成;
- 社会群体、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段；
- 前期参与过的消费者测试的相关信息。

应定期更新每位消费者的信息。样本库能协助服务商提前满足委托方对目标群体的初始要求,实现委托方和服务商之间的有效沟通。

5.2.4 消费者招募的频次

与特定任务招募相比,从可访问样本库中招募存在更高的“职业消费者”风险。应重点预防消费者个体与产品集之间产生关联效应。避免频繁招募同一个消费者,每个消费者参与相同类型产品测试的时间间隔宜为3个月及以上。

实验室应记录每位消费者参与过的所有产品及产品集的测试频次。实验方案中应列出消费者参与过测试的记录。若委托方要求,还应列明其测试过的产品的相关信息。

5.3 消费者样本构建

5.3.1 消费者招募问卷

采用问卷调查方式构建消费者样本时,应确保消费者并不知晓招募问卷中哪些问题是用于招募的标准。以下是常用的招募问题:

- 前期是否参与过相关产品的测试；
- 产品的使用频次；
- 常用产品的品牌、说明和购买场所等；
- 使用或消费产品的习惯。

其他问题还包括年龄、性别、社会群体、职业或地理位置等。出于实用性和可行性考虑,可将符合上述标准的人群分类,并指出消费者样本中每类人群的构成比例。

要求参加测试的每位消费者完整填写招募问卷,并且可对招募中的部分细节进行核实。

附录A给出了1个招募问卷示例。此外,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消费者样本构建中最重要的质量要求,通常由提出目标群体要求的委托方与评估方案可行性的实验室共同确定测试群体。

选择目标群体时,宜包括下列问题:

- 产品的使用或消费频次(这是基本要求,会影响测试的辨别力)；
- 消费者的年龄、性别、社会地位、职业和地理位置；
- 常用或消费产品的品牌；
- 产品的购买场所。

5.4 消费者样本细分

消费者样本细分可采用下列2种方式。

a) 测试前细分,可使用下列问题:

- 1) 女性和男性的响应方式相同吗?

2) 不同年龄组的响应方式相同吗?

3) 消费习惯对响应方式有影响吗?

上述情况存在时,应单独分析每个消费者子群的结果并对结果进行比较。

b) 有证据表明存在多个消费者子群(例如数据呈现多峰分布)时,可根据获得的数据进一步再对消费者细分。

附录 B 给出了如何对消费者细分数据进行处理和指导。

6 消费者样本量

6.1 测量精度

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,但提高了测量精度也并不保证结论的相关性和有效性。选择适当的消费者和测量程序同样非常重要。

附录 C 给出了消费者样本量对测量精度的影响研究。

测量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。

测试程序同样影响测量精度。在其他条件相同的情况下,1 个消费者同时测试 2 个产品比不同消费者分别测试 2 个产品的结果具有更高的精度。

此外,测量精度也取决于测试目的。其他条件相同时,测量精度还受到下列因素的影响:

- 若测试目的是确定产品间存在喜好差异,喜好评分均值间的最小差异可被认为大到足够有效(差异越小时,则样本量越大);
- 若测试目的是确定产品间存在喜好相似,喜好评分均值间的最大差异可被认为小到足以被忽略(差异越大时,则样本量越大);
- 委托方愿意接受的误判风险 α (α 风险越小时,则样本量越大);
- 委托方期望的检验效力 P ($P = 1 - \beta$, β 是漏判风险;预期效力越高时,则样本量越大);
- 统计结果的预期方向。若委托方仅关注“确保产品 B 不劣于产品 A”,宜采用定向检验(单边检验)。因为与委托方同时关注“B 不劣于 A”和“B 优于 A”相比,采用定向检验的效力会更高。

6.2 样本量确定

确定消费者样本量是开展测试的关键步骤,需在统计人员协助下完成。附录 D 给出了采用评级法进行差异性 or 相似性测试时所需样本量的计算方法(附有示例)。

任何情况下,本文件中的消费者样本量都不应低于 60 人。

若实验室未采用或无法实现附录 D 中的所述方法,本文件中的最小样本量调整为 100 人。

若实验室采用附录 D 中的所述方法,可通过结果计算得到样本量,但最终样本量可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/或委托方无法提供该数量的消费者时,样本量可减少至 100 人。实验室应重新计算 100 位消费者时的检验效力,并取得委托方同意接受降低的检验效力值;
- 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。

上述的 100 人和 60 人是建立在委托方并未根据消费者特征(例如年龄或性别)分组的假设上的样本量。若测试前已对消费者分组,则每组最小样本量为 60 人。

6.3 消费者子群使用

6.3.1 总则

测试数据收集或使用前宜先构建好消费者子群。使用下列测试方法时,实验设计阶段宜加入统计的考量。

6.3.2 消费者子群独立测试

委托方可能希望根据测试前已知的消费者特征(例如性别或年龄)构建消费者子群,再对消费者子群的结果独立分析。该种情况下,每个消费者子群的样本量应不低于 60 人。

6.3.3 消费者子群统一测试

相比 6.3.2 中对消费者子群结果独立分析的方法,先对消费者总样本均衡供样开展测试,再将子群信息(如将其作为 1 个变量)代入到总样本结果里进行综合分析的方法会更有效率。即使每个子群样本量低于 60 人,也会获得相同的置信度。当所有子群具有相同或高度相似的样本量时,该方法的效率最高。

6.3.4 消费者细分

大样本量的消费者测试结果宜进一步进行细分,测试结果和/或社会人口学的统计结果均作为消费者细分的依据。消费者细分可用于描述消费者样本的特征,并提出新的假设。最终细分的子群数量、特征和规模取决于测试结果的数量和性质。

6.4 待测产品数量和实验设计

消费者样本量会受到待测产品数量和实验设计的影响。采用完全供样(每个消费者测试所有待测样品)时,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(每个消费者测试部分待测样品)时,所需的消费者数量更大。

6.5 招募数量的安全限

由于部分招募的消费者可能无法参加测试,或参加了测试却未能正确进行测试,因此实际获得的有效答案数量通常少于最初招募数量。为弥补这一点,计算消费者招募数量时宜设置 1 个安全余量,即安全限。

7 测试区域

本文件中的喜好测试均在受控区域(见 3.1)内开展,不包括在公共场所或消费者家中进行的喜好测试。下列给出了 3 种适用于测试的区域示例。

- a) 固定式感官评价实验室:应符合 GB/T 13868—2009 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。其空间通常小于固定式感官评价实验室,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”布置的评价室:通常适用于准备工作较少的产品评价。包括 2 个独立区域,

1 个用于感官评价,另 1 个用于产品准备和编码。评价区能将消费者彼此隔开,测试环境可被监控,如房间温度、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的是,当需要开展新的特定任务测试时,应重新设置测试区域的产品准备和提供系统。

8 产品

8.1 产品盲评

测试产品应以盲样形式提供给消费者,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)。删除或遮蔽所有与品牌或质量标识相关的信息,除非无法删除或遮蔽产品信息或者测试的目的就是确定品牌或质量标识的影响。

注:本文件中,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

按照生产商推荐步骤制备待测产品。若无法获取制备步骤时,按照预实验建立的规范步骤制备产品。若无特殊理由,产品的制备条件应接近惯常的使用和消费条件。

产品的提供方式(例如温度、相关产品、餐具和呈送方式)均遵循上述规定。

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,某些情况下也会将这种变化列为调查重点。除上述情况外,宜使用同一批次、具有相同贮存时间和条件的产品进行测试,使产品间的差异最小化。

8.2.2 提供量

提供给消费者的产品量应与该产品的日常消费量或委托方指定的消费量保持一致。告知消费者产品测试时的最小消费用量,必要时也应告知最大消费用量。应确保测试样品量既不会太少(存在感官印象停留在最初、转瞬即逝的风险)也不会太多(存在吃饱或厌恶的风险)。

当消费者需要测试同系列的多个产品时,产品提供量宜少于产品的日常消费用量。但仍应确保样品有足够的提供量,避免提供量太小导致测试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,提供给每位消费者的样品应成比例地包含原产品的全部组分。

8.2.3 提供形式

产品的提供形式包括产品购买时的形式和待消费时的形式 2 种。

示例 1:面包以整个或切片;香肠以整根或切片;干酪以整块或切片;比萨以整份或切块;速溶咖啡以在杯中未冲泡或加水后再放咖啡。

实验室和委托方可认为上述 2 种产品提供形式都能达到相同目的,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,在杯中未冲泡或加水后再放咖啡)提供产品,且无需向消费者特别指出这是 2 种不同提供形式的相同产品。消费者在产品提供后对产品评分,这种方法类似于消费者日常生活中的常见消费行为,无需对所有感知再细分。

示例 2:提示语和问题的陈述可以如下所示:您面前有一香肠,编码为 148。请看一看,闻一闻,吃几块,然后用下列的响应标度记录您对它的感觉。

若实验室和委托方均认为 2 种提供形式对产品喜好影响相似,则无需将不同提供形式的 2 个样品

视为不同产品。否则,2种提供形式可以同时或逐一提供,并独立获得喜好值。

注:本文件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。该情况涉及特定的定义和分析方法,属于特殊实验设计范畴。

9 程序

9.1 喜好测试分类

根据对消费者设定的不同任务,喜好测试主要有2类。

- a) 接受性测试,用于测量对产品的喜好程度。本文件中仅给出了一种接受性测试的方法,即评级法。
- b) 偏爱测试,用于对不同产品的喜好程度进行排序(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”)。偏爱测试获得的是相对喜好信息,并不表明对产品的喜好程度。某产品也许比另一个产品更被偏爱,但2个产品可能均不被喜爱。偏爱测试分为成对偏爱测试(2个样品比较)和偏爱排序测试(2个以上样品比较)。

9.2 评级法

9.2.1 总则

评级法通常分为以下2类。

- a) 根据使用的响应标度分类。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。响应标度的使用应按照 GB/T 39501—2020 中的规定进行。
- 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。
 - 严格的单样品提供(单样品测试),每位消费者只测试1个产品。
 - 连续单样品提供(完全或不完全实验设计),消费者进行1轮或多轮测试完成多个产品的测试。消费者每次只测试1个产品,且不提供已测试过产品及相关的测试信息。确保消费者不能回看之前的产品评价结果。
 - 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价结果。

连续单样品提供是最常见的样品提供方式。相比严格的单样品提供,该方式更加经济,但需掌握如何将产品分配给消费者(见附录 E)。比较提供方式较少使用,因其会放大产品间差异,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 样品提供方案

按照样品提供方案将样品提供给消费者(见 10.2)。

9.3 成对偏爱法

9.3.1 总则

成对偏爱测试的原理是产品以成对形式呈现,消费者按照指定的序列进行测试,指出更偏爱哪一个产品。

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

- a) 选项 A:消费者回答下列封闭式问题:

“从左到右品尝提供给您 的 2 个产品,然后在您更喜欢产品编码下标记‘√’”。

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式问题:

“从左到右品尝提供给您 的 2 个产品,然后在您更喜欢产品编码下标记‘√’。如果您对 2 个产品的喜好相同,则在‘无偏爱’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。例如若要求消费者对产品风味发表意见,可能会使他们在对产品整体评价时倾向于关注对产品风味的评价。

测试结束时可以提出一个开放式问题,让消费者阐述喜欢或不喜欢产品的具体方面。

9.3.2 样品提供方案

样品 A 和样品 B 应按照 AB 和 BA 的顺序均衡提供,并随机成对提供给消费者。

9.4 排序法

9.4.1 总则

排序法应按照 GB/T 12315—2008 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,消费者为了能正确排序需要多次评价产品,随着品尝样品数量增加方法的复杂性也随之增大。消费者对样品进行初评时,应按照指定顺序评价待测样品(通常从左到右)。随后,消费者可自由选择产品的评价顺序,按照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,也可以允许消费者将 2 个或多个产品排为同一秩次。但本文件不包括后者的测试分析。

9.4.2 问卷

排序法问卷中给消费者的提示语通常如下:

“请从左到右依次品尝 4 种产品,每个产品均可品尝多次,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

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了解测试产品的整体评价或消费者特征。例如从可访问样本库中招募消费者时,若实验室希望核查是否所有消费者来自目标群体,可设置一个关于特定消费习惯的问题。

9.5.2 与消费者相关的问题

实验室要密切关注所提的问题以及问题提出的时间段。需注意放在测试后的提问是否会受到刚结束测试的影响。例如在 1 项番茄酱配米饭测试结束后,消费者可能倾向于认为番茄酱配米饭是一种常吃食物。此外,放在测试开始前的提问易使消费者在测试期间倾向于以某种特定方式评价产品,或重点关注产品的某部分特性。

9.5.3 与产品相关的问题

与产品相关的问题可以是开放式问题,也可以是封闭式问题。问题应设置在与整体评价相关的所有问题之后,并单独列出。不宜和整体评价的问题列在同一页面或屏幕上,以减少消费者尝试给出一致性回答的风险。此外,此类问题数量宜尽量少。

注 1: 附加问题并非质疑评价结果。产品的整体评价和消费意向之间存在差异时,采用附加问题研究差异原因。

注 2: 附加问题能提供的信息很少。感官评价实验室的负责人和委托方常采用喜好数据和感官特性数据的关联分析(例如偏爱映射法或联合分析法)进一步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,需要重点关注并在实验方案中专门提及。

提示语重点关注的内容如下列所示。

- a) 测试前告知每位消费者待测产品的最小和最大消费量。消费量取决于产品类型、待测产品数量和委托方意愿。此外,实验室还应明确给出标准操作规范并确保消费者按照其执行。
- b) 问卷填写完后的相关安排。采用连续单样品提供方式时,实验室应使用能有效避免消费者参考之前评价结果的系统。
- c) 休息时长及休息期间的建议活动(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涵义。问题的陈述不应中立,例如提示消费者可从产品的外观、气味、滋味或质地等方面描述他们对产品的整体评价,使得消费者回答更为理性,而并非简单如“有任何评价请填写”的陈述。

通常会在测试开始时提供给消费者关于产品的数量、消费量和需要回答的问题等相关信息。也可提供其他信息,使测试过程更加人性化或满足测试的其他目的。为避免提供信息让消费者产生一定预期,应仔细考虑信息的陈述方式,并在报告中呈现信息的完整内容。若测试包括多个轮次,应确保每个轮次测试中的提示语完全相同。

10 测试

10.1 总则

宜合理组织测试,确保存在多个研究目的时能实现最佳平衡,通常包括下列内容:

- 采用具有等效性的消费者对产品进行测试,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;
- 确保测试条件能尽可能地代表正常的消费行为;
- 确保消费者不会因超负荷测试导致随意回答数量增多。

为协调不同的研究目的,通常采用3种测试方法:

- 完全测试:每位消费者测试所有产品;
- 不完全测试:每位消费者测试部分产品;
- 单样品测试:每位消费者仅测试1个产品。

10.2 实验设计

10.2.1 完全测试

每位消费者测试所有产品,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,且存在序列效应和残留效应的风险。产品在测试产品群中的位置会影响测试结果的均值和/或离散度。

此外,给消费者的产品提供序列至关重要。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,减少序列效应和残留效应的影响。残留效应的影响过大时,应采用足够长的时间间隔确保感官清零,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,对于序列效应和残留效应的处理措施与完全测试类似。消费者超负荷测试的风险较小,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[10]。

10.2.3 单样品测试

为完全消除序列效应、残留效应和情境效应,可为每个产品生成独立的消费者样本。该方法操作和结果分析简单方便,但要确保消费者样本的等效性和目标群体的代表性。此外,由于单样品测试无法考虑个体差异,被估计参数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,要达到相同的置信度和精度,单样品测试的消费者样本量需更大。

10.3 单一轮次测试的产品数量

10.3.1 总则

单一轮次测试中,可进行有效测试的产品数量取决于下列因素:

- 待测产品的性质:例如不同产品产生不同的饱腹感;有些产品(例如具有侵略性)能影响感觉受体;含酒精的产品可能影响大脑皮质层的信息整合;产品可以是单一组分或多组分产品(例如成品菜肴);
- 测试时长:需较长时间的测试宜分为多个环节,但未经预约的消费者可用测试时间通常较短;
- 每种产品的消费量;
- 测试产品所使用的感觉器官(使用视觉或触觉能比味觉测试更多数量的产品);
- 询问问题的数量,尤其是提供一种新产品时。

本文件中并未对1个轮次中测试产品的数量进行限制,但实验室应向委托方证明其单一轮次测试中产品数量设置的合理性。

10.3.2 单一轮次测试的产品数量过多

当测试产品数量过多、单一轮次测试无法完成时,可采取下列方式:

- a) 组织多次测试,每位消费者参加所有的测试(完全测试);
- b) 组织单次测试,每位消费者只测试部分产品(不完全测试或单样品测试)。

10.3.3 多轮次测试

若能安排多轮次测试时,每位消费者应在整个测试期间对所有产品进行测试。理想情况是每轮次测试中提供所有产品,但实际上并非每轮次测试时消费者评估相同产品。

若受到产品准备或可用数量的限制,无法保证每轮次测试均能提供所有产品时,将产品随机分配于多个轮次测试中。所有消费者在测试期间均收到相同的产品子集,但顺序不同。这种方式通常需要更复杂的统计分析。

每轮测试开始时,宜向所有消费者提供 1 个相同产品,确保消费者每次测试时使用相同参比样。

若同一批消费者需要开展 2 次测试,不应在同天进行。

10.3.4 单轮次测试

采用平衡不完全区组(BIB)设计可以实现每个轮次测试的产品数量少于产品总数,但确保是在相同精度下进行所有的产品对比较。整个测试过程中,样品提供宜遵循平衡原则,避免产生序列效应和残留效应。这种测试方式通常需更多数量的消费者和更复杂的分析方法(参见参考文献[10]),且在实验设计阶段时应寻求相关统计的建议。

BIB 的设计方法应按照 GB/T 39992—2021 中的规定进行。

10.4 同轮次测试中产品的性质

同轮次中包含不同性质的产品时,应满足下列条件:

- 产品组合符合日常的消费习惯;
- 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测试。

无论参加测试的人员是谁,实验室都应确保测试的机密性。若委托方要求,可以参加测试,但无需让其知晓参加测试的其他委托方信息。

10.5 测试的时间

宜选择与该产品的日常消费相同的时间段开展测试,便于消费者在接近真实的情形下进行测试。

10.6 产品重复测试的次数

同一测试中,如果需要重复测试某产品,应视为测试产品数量的增加。实验方案中应设置为产品 A1 和 A2,而并非产品 A。在确定产品提供序列和进行结果分析时,都应将产品 A1 和 A2 视为 2 个不同的产品。

10.7 预实验

正式测试前,先选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,并确定采用的响应标度是否合适;
- 预演测试过程;
- 确定适宜的测试时长、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,但不应再参加正式测试。

11 结果分析

11.1 总则

结果数据分析时应先检查数据的内部一致性,计算基本的描述性统计量。报告中应对每个数据(例如平均值和中位数等)至少给出1个对应的离散参数(例如标准偏差和偏差范围等)。

绘图法能有效呈现数据分布。宜创建每个产品的直方图,通过数据分布(例如单峰)呈现样品之间是否存在较大差异。主成分分析(Principal Component Analysis, PCA)或凝聚层次聚类(Agglomerative hierarchical clustering, AHC)能直观展现消费者之间的一致性和差异性。

回答“消费者群体是否能察觉出测试产品之间的显著差异”等问题时,需要更多的数据统计量。结果分析可采用参数或非参数统计分析,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,再外推样本所代表的消费者群体特征的分析方法。采用方法取决于测试目的和待测产品数量。若某些产品被赋予了特殊身份,如参比样,会直接影响结果分析。

尽管采用平衡设计过的产品提供方案仍然无法消除偏差,但能确保产品不会只受到产品提供顺序带来的有利影响或是不利影响。同时,系统设计的推断分析能改进估计偏差,提升主要结论的可信度。但推断分析既不简单也并不直接(参见参考文献[5]和[6])。

标准的数据统计分析软件包中提供了多种可用的统计方法,统计分析师可选择适宜的分析方法进行分析。

11.3 评级法(见 9.2)

评级法数据分析中,方差分析(Analysis of Variance, ANOVA)是一种有效的分析方法。

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中文语义喜好标度,包括 9 点喜好标度、15 cm 线性喜好标度和标记情感(LAM)标度。

11.4 成对偏爱法(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(若为奇数,减 1 后再平均分配),再使用新的数据进行统计分析。

11.5 排序法(见 9.4)

GB/T 12315—2008 给出了针对不同情况时采用的分析方法。

可采用参数统计或非参数统计分析方法,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	消费者分组, 每组消费者只测试 1 个产品	Mann-Whitney's U 检验 (也称为 Wilcoxon-Mann-Whitney 检验)	独立样本 t 检验
	每位消费者均测试 2 个产品	Wilcoxon 秩和检验	成对样本 t 检验
2 个以上产品	消费者分组, 每组消费者只测试 1 个产品	Kruskal-Wallis 检验	方差分析(ANOVA)
	每位消费者至少测试其中的 2 个产品, 理想情况是测试所有产品	Friedmann 检验(秩方差分析), 必要时可采用不完全区组设计	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

报告首页可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, 报告中所有图表宜进行编号, 并附上标题和注帮助理解。

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, 报告还可给出下列详细信息:

- 研究题目和参考文献;
- 开展研究的日期(时间段);
- 报告的日期;
- 委托方身份信息;
- 实验室和研究负责人的完整身份信息(如果有分包商, 其身份信息一并提供);
- 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, 例如本文件。

12.2 产品

产品信息中应给出下列内容:

- 产品描述, 测试方宜列出产品全部或主要的成分(宜附上每个产品的照片);
- 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；
- 测试期间测试的所有其他待测产品及提供序列。

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，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。
- 测试轮次或环节的总次数。
- 测试的日期、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：

- 目标消费群体描述；
- 消费者样本描述，包括委托方要求的每类消费者子群的规模和数量；
- 消费者来源说明，是基于特定任务的招募还是来自可访问样本库。

12.4.2 基于特定任务的招募

招募信息中应包括下列内容：

- 招募的场所或方式；
- 招募程序；
- 招募合适消费者的方法。

12.4.3 基于消费者样本库的招募

若从消费者样本库中招募消费者，报告中可包括下列内容：

- 消费者参加测试的频次说明；
- 消费者测试过的所有产品的产品列表；
- 消费者参与过的本次待测产品或产品集测试的频次说明。

上述信息通常只包括测试前 12 个月内的相关情况。若所有消费者均未参与过消费者测试，同样在报告中列出。

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，数据以绘图或表格形式汇总；
- 结果的数值汇总，包括估计值精度或均值；

- 分析和解释结果的方法；
- 结果的统计推断；
- 根据研究目的,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)
招募问卷示例

表 A.1 给出了招募问卷的示例。

表 A.1 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,请从中选择出您最近购买过的食品。 提示信息:提供“食品”列项。			
	食品 1	1	▶6	
	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		
提示信息:如果没有选择“3”,则结束面试。				

表 A.1 招募问卷示例 (续)

编号	接触式访谈项目×××××	日期	跳至问题	编码																																																
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																																																	
	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? 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!		▶8C																																																	
	<table style="width:100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th style="text-align: center;">8A</th> <th style="text-align: center;">8B</th> <th style="text-align: center;">8C</th> </tr> </thead> <tbody> <tr><td>品牌 1</td><td style="text-align: center;">1</td><td style="text-align: center;">_____</td><td style="text-align: center;">1</td></tr> <tr><td>品牌 2</td><td style="text-align: center;">2</td><td style="text-align: center;">_____</td><td style="text-align: center;">2</td></tr> <tr><td>品牌 3</td><td style="text-align: center;">3</td><td style="text-align: center;">_____</td><td style="text-align: center;">3</td></tr> <tr><td>品牌 4</td><td style="text-align: center;">4</td><td style="text-align: center;">_____</td><td style="text-align: center;">4</td></tr> <tr><td>品牌 5</td><td style="text-align: center;">5</td><td style="text-align: center;">_____</td><td style="text-align: center;">5</td></tr> <tr><td>品牌 6</td><td style="text-align: center;">6</td><td style="text-align: center;">_____</td><td style="text-align: center;">6</td></tr> <tr><td>品牌 7</td><td style="text-align: center;">7</td><td style="text-align: center;">_____</td><td style="text-align: center;">7</td></tr> <tr><td>品牌 8</td><td style="text-align: center;">8</td><td style="text-align: center;">_____</td><td style="text-align: center;">8</td></tr> <tr><td>品牌 9</td><td style="text-align: center;">9</td><td style="text-align: center;">_____</td><td style="text-align: center;">9</td></tr> <tr><td>品牌 10</td><td style="text-align: center;">10</td><td style="text-align: center;">_____</td><td style="text-align: center;">10</td></tr> <tr><td>品牌 11</td><td style="text-align: center;">11</td><td style="text-align: center;">_____</td><td style="text-align: center;">11</td></tr> </tbody> </table>		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	
	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																																																	
	其他信息, _____ 12 _____ 12																																																			

表 A.1 招募问卷示例（续）

编号	接触式访谈项目×××××	日期	跳至问题	编码
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)

消费者子群细分数据评价方法

B.1 总则

依据研究目的,开展目标群体的子群比较。例如检测不同子群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进行分组。消费者招募前宜先明确分组依据,确定每组样本量大小,确保测试时每组消费者样本量不少于 60 人,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)分析后再对消费者进行细分。这种细分方式可用于描述消费者样本的特征,并提出新的假设。样本量应足够大,才能确保后续消费者细分的合理性。

上述 2 种情况中,所需的消费者数量会依据已确定或预期的子群数量以及每个子群的具体情况而有所变化。

B.2 消费子群

消费者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(或文化)对偏爱有影响吗?

——消费习惯对偏爱有影响吗?

若每位消费者均测试所有产品,并采用方差分析时,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序列效应,有时还包括残留效应。若含有序列效应,则模型如下:

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

采用该模型时,若所有组别(序列 × 组别的交互效应)和所有样本(样本 × 序列的交互效应)的序列效应相同,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,则各组别单独测试。

B.3 消费者细分

根据测试结果再对消费者细分可用于结果偏爱性的解释(若存在该种情形)。

当直方图检测到双峰分布时,开展消费者细分非常有效。

不考虑消费者总样本时,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，即原始和部分相关联变量的线性组合，表达结果的变异性(离散性)。

该方法可以“同时”展示产品和消费者分布，两者叠加同时展示了两者的关联。

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[12])。对于不完全测试，结果分析见参考文献[11]。

因此，第一主成分中的变量可以用来区分消费者组(簇)。如需进一步分辨消费者组别，可采用分类分析。

附录 C

(资料性)

样本量对结果精度的影响

应明确每个消费者组的最小样本量,确保产品差异可以被测量出,更重要的是确保每组产品均值估计具有一定的精度。上述原则同样适用于消费者总样本或某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离散性。

假定产品的评分呈现正态分布, σ 为标准偏差,采用公式(C.1)可以获得预测精度 A_c 来估算平均值时所需的最小消费者样本量:

$$A_c = \frac{2\sigma}{\sqrt{n}} \dots\dots\dots (C.1)$$

式中:

A_c ——平均值 95%置信区间的一半宽度。

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

表 C.1 给出了 σ 、 n 和 A_c 之间的关联。

表 C.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,宜予以考虑 β 风险(附录 D)。

注:若测试前计算所需样本量,宜先进行统计效力分析,即固定效应量(ES)(参考文献[9])后再确定样本量 n 值。

附录 D

(资料性)

评级法的最小样本量

D.1 总则

消费者测试中,每个产品的测试数量宜不少于 60 次,100 次左右最佳。若每位消费者仅测试 1 个产品时,需要的消费者数量 $n=100 \times p$ 。其中, p 是待测产品数量。若每位消费者测试 k 个产品,按经验所需消费者数量为 $n=100 \times (p/k)$ 位。

这种基于经验的方法并不够完善,应采用其他更复杂的方法计算样本量。计算前应先回答下列问题。

- 项目的研究目的是什么? 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?
- 待测产品的数量有多少? 1 个、2 个、3 个或更多?
- 每位消费者需要测试多少个产品?
- 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, $P=1-\beta$)
- 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?
- 产品评价结果的离散度是多少?
- 采用的是单边还是双边的备择假设?
- 消费者同时测试多个产品时,不同产品评价结果是否存在关联?

本附录仅涉及 2 个产品的测试。对于更简单(1 个产品)或更复杂(3 个或更多产品)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^[3]。

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。

D.2 示例 1:确定 2 个产品 A 和 B 之间存在差异,每位消费者仅评价 1 个产品

每个产品所需的消费者数量 n 由公式(D.1)计算得到:

$$n = 2 \times \left[\frac{(u_{\alpha/2} + u_{\beta}) \times \sigma_{\text{产品内}}}{\delta} \right]^2 \dots\dots\dots (D.1)$$

式中:

- $u_{\alpha/2}$ ——双边备择假设时标准正态分布中对应 $(1-\alpha/2)$ 分位数的值, $\alpha=0.05$ 时, $u_{\alpha/2}=1.960$; 当单边备择假设时, u_{α} 替代 $u_{\alpha/2}$, $\alpha=0.05$ 时, $u_{\alpha}=1.645$;
- u_{β} ——可接受 β 风险下,标准正态分布中对应 $(1-\beta)$ 分位数的值(期望检验效力 $P=1-\beta$);
- $\sigma_{\text{产品内}}$ ——产品内的标准偏差,前提是假定 2 个产品具有相同的标准偏差;
- δ ——2 个产品间可被检验出的差异。

注: $\sigma_{\text{产品内}}$ 和 δ 采用相同标度表示,例如 9 点标度法。9 点标度法得到的数据可认为近似正态分布。

示例:如下列条件所示:

α	$u_{\alpha/2}$	β	$(P=1-\beta)$	u_{β}	$\sigma_{\text{产品内}}$	δ
0.05	1.960	0.10	(0.90)	1.282	1.50	0.50

由公式(D.1)得到 $n=189.2$,即 1 个产品需要 190 位消费者参与测试,2 个产品需要 380 位消费者参与测试。

注：对于 A>B 的单边假设，由 $u_\alpha = 1.645 (\alpha = 0.05)$ 和公式 (D.1) 得到 $n = 154.2$ ，即 155 位消费者。

D.3 示例 2: 确定 2 个产品 A 和 B 之间存在差异，每位消费者均评价 2 个产品

消费者数量由公式 (D.2) 计算得到：

$$n = 2 \times \left[\frac{(u_{\alpha/2} + u_\beta) \times \sigma_{\text{产品内}} \times \sqrt{1 - \rho}}{\delta} \right]^2 \dots\dots\dots (D.2)$$

式中：

$u_{\alpha/2}$ —— 双边备择假设时标准正态分布中对应 $(1 - \alpha/2)$ 分位数的值， $\alpha = 0.05$ 时， $u_{\alpha/2} = 1.960$ ；当单边备择假设时， u_α 替代 $u_{\alpha/2}$ ， $\alpha = 0.05$ 时， $u_\alpha = 1.645$ ；

u_β —— 可接受 β 风险下，标准正态分布中对应 $(1 - \beta)$ 分位数的值（期望检验效力 $P = 1 - \beta$ ）；

$\sigma_{\text{产品内}}$ —— 产品内的标准偏差，前提是假定 2 个产品具有相同的标准偏差；

δ —— 2 个产品间可被检验出的差异；

ρ —— 消费者群体对 2 个产品评分值的线性相关系数。

注： $\sigma_{\text{产品内}}$ 和 δ 采用相同标度表示，例如 9 点标度法。

由于

$$\sigma_d = \sigma_{\text{产品内}} \times \sqrt{2(1 - \rho)}$$

式中：

σ_d —— 所有消费者对 2 个产品评分差值的标准偏差；

因此，公式 (D.2) 表示为：

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

公式 (D.3) 是目前最常用的表达式。

示例：如下列条件所示：

α	$u_{\alpha/2}$	β	$(P = 1 - \beta)$	u_β	σ_d	δ
0.05	1.960	0.10	(0.90)	1.282	1.70	0.50

由公式 (D.3) 得到 $n = 121.5$ ，即采用 122 位消费者参与测试。

注：对于 A>B 的单边假设，由 $u_\alpha = 1.645 (\alpha = 0.05)$ 和公式 (D.3) 得到 $n = 99.0$ 。

D.4 示例 3: 确定产品 A 不劣于产品 B，每位消费者仅评价 1 个产品

每种产品所需消费者的数量由公式 (D.4) 得到：

$$n = 2 \times \left[\frac{(u_\alpha + u_\beta) \times \sigma_{\text{产品内}}}{\Delta} \right]^2 \dots\dots\dots (D.4)$$

式中：

u_α —— 单边备择假设（非劣效性检验）时标准正态分布中对应 $(1 - \alpha)$ 分位数的值；

u_β —— 可接受 β 风险下，标准正态分布中对应 $(1 - \beta)$ 分位数的值（期望检验效力 $P = 1 - \beta$ ）；

$\sigma_{\text{产品内}}$ —— 产品内的标准偏差，前提是假定 2 个产品具有相同的标准偏差；

Δ —— 认为 2 个产品相似（可互换）时所不能接受的差异值。

注： $\sigma_{\text{产品内}}$ 和 Δ 采用相同标度表示，例如 9 点标度法。

示例：如下列条件所示：

α	u_α	β	$(P = 1 - \beta)$	u_β	$\sigma_{\text{产品内}}$	Δ
0.05	1.645	0.10	(0.90)	1.282	1.40	0.30

由公式(D.4)得到 $n=373.2$, 即 1 个产品采用 374 位消费者参与测试, 2 个产品共采用 748 位消费者参与测试。

D.5 示例 4: 确定产品 A 不劣于产品 B, 每位消费者均评价 2 个产品

消费者的数量由公式(D.5)计算得到:

$$n = \left[\frac{(u_\alpha + u_\beta) \times \sigma_d}{\Delta} \right]^2 \dots\dots\dots (D.5)$$

式中:

- u_α —— 单边备择假设(非劣效性检验)时标准正态分布中 $(1-\alpha)$ 分位数的值;
- u_β —— 可接受 β 风险下, 标准正态分布中 $(1-\beta)$ 分位数的值(期望检验效力 $P=1-\beta$);
- σ_d —— 所有消费者对 2 个产品评分差值的标准偏差;
- Δ —— 认为 2 个产品相似(可互换)时所不能接受的差异值。

注: $\sigma_{\text{产品内}}$ 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	u_α	β	$(P=1-\beta)$	u_β	σ_d	Δ
0.05	1.645	0.10	(0.90)	1.282	1.70	0.30

由公式(D.5)得到 $n=275.1$, 即采用 276 位消费者参与测试。

注: 示例 3 和示例 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, 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 个产品共需 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。明确 ASTM E1958-20^[4] 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 E
(资料性)
产品提供方案

随机完全区组设计是最有效和最常采用的消费者产品提供方案。对每位消费者随机独立提供样品,消费者要测试所有产品,且每个产品只测试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,实验者需要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测试每个产品。消费者能有效测试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 GB/T 39992—2021。每位消费者的产品提供序列遵循独立随机原则。

——序列效应:实验者经常担心存在序列效应,先品尝的产品喜好值通常高于后品尝的产品。因此,完全区组和平衡不完全区组实验设计时需考虑随机顺序提供产品,确保每个产品在先或后的提供序列上具有同等机会。

若实验者担心序列效应影响过强时,一个有效替代随机完全区组设计的方法是广义拉丁方设计。广义拉丁方是在产品提供序列设计中,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序列称为广义尧敦设计(generalized Youden design)。

但是,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,应采用考虑了序列效应的行列模型。

——延滞效应:尽管测试期间设置了感官清零期,但前一个产品的品尝仍有可能影响下一个产品的味道。若实验者担心延滞效应,可使用相邻平衡设计(参考文献[3])。参考文献[3]和[10]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种产品进行测试。这时宜采用更复杂、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 F

(资料性)

评级法数据分析的示例

F.1 总则

本附录采用附录 D 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 个产品 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 的评分至少等于 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 D 示例 1):证实 2 个产品 A 和 B 存在喜好差异,每位消费者仅评价 1 个产品

——与委托方讨论后,实验室使用 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 个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:

$$t_{\text{cal}} = \frac{m_A - m_B}{s_{\text{产品内}} \times \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \quad \dots\dots\dots(\text{F.1})$$

式中:

$s_{\text{产品内}}$ ——产品内的标准偏差。

$s_{\text{产品内}}$ 采用公式(F.2)计算得到:

$$s_{\text{产品内}}^2 = \frac{s_A^2 \times (n_A - 1) + s_B^2 \times (n_B - 1)}{n_A + n_B - 2} \quad \dots\dots\dots(\text{F.2})$$

因此,本例中:

$$s_{\text{产品内}}^2 = \frac{1.85^2 \times (110 - 1) + 1.65^2 \times (120 - 1)}{110 + 120 - 2}$$

和

$$\sqrt{s_{\text{产品内}}^2} = \sqrt{3.057} = 1.748$$

t_{cal} 值为

$$t_{\text{cal}} = \frac{7.23 - 6.87}{1.748 \times \sqrt{\left(\frac{1}{110} + \frac{1}{120}\right)}} = \frac{0.36}{0.231} = 1.558$$

将 t_{cal} 值与所选 α 风险下的 t_{the} 值进行比较, t_{the} 值为对应于自由度为($n_A + n_B - 2$)时 t 分布的($1 - \alpha/2$)分位数的值。本例中, $\alpha=0.05$ 时, $t_{\text{the}}=1.970$ 。

注 1: t_{the} 值应采用 Excel¹⁾ 中的函数计算: $\alpha=0.05$ 时, $T.INV.2T(\alpha; df) = T.INV.2T(0.05; 110+120-2) = 1.970$ 。

$t_{\text{cal}} > t_{\text{the}}$ 时,在 α 水平上拒绝 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:

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。

$$P = 1 - \beta \dots\dots\dots (F.3)$$

t_β 值采用公式(F.4)计算得到:

$$t_\beta = t_{the} - \frac{\delta}{t_{calden}} \dots\dots\dots (F.4)$$

式中:

- t_{the} ——拒绝 H_0 的值;
- δ ——2 个产品之间的偏爱概率差异;
- t_{calden} —— t_{cal} 的分母[公式(F.1)].

因此,

$$t_\beta = 1.970 - \left(\frac{0.50}{0.231} \right) = -0.195$$

$t_\beta < 0$ 时, β 值由 t_β 和自由度为 $(n_A + n_B - 2)$ 的 t 分布函数得到。 $t_\beta = -0.195$ 、 $df = 110 + 120 - 2 = 228$ 时, $\beta = 0.423$ 。因此, 检验效力 $P = 1 - \beta = 1 - 0.423 \approx 0.58$, 该值远低于要求值 0.9。

注 2: β 值采用 Excel 中的函数计算: $T.DIST(x; deg_freedom; TRUE) = T.DIST(-0.195; 228; TRUE) = 0.423$ 。

注 3: $t_\beta > 0$ 时, P 值由 t 分布函数在 $(-1) \times t_\beta$ 计算得到。例如 $\delta = 0.30$ 时, $t_\beta = 1.970 - (0.30/0.231) = 0.671$; 通过 t 分布函数得到 $P = 0.251 \approx 0.25$ 。

注 4: t_{the} 值应采用 Excel 中的函数计算: $T.INV.2T(0.05; 228) = 1.970$; 其中, $df = (230 - 2) = 228$ 。

F.3 示例 2(附录 D 示例 2): 证实 2 个产品 A 和 B 存在喜好差异, 每位消费者均评价 2 个产品

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 个产品评分差值的标准偏差;
- t_{cal} 值由公式(F.5)计算得到:

$$t_{cal} = \frac{m_A - m_B}{s_d \times \sqrt{1/n}} = \frac{6.33 - 6.66}{1.92 \times \sqrt{1/120}} = \frac{-0.33}{0.175} = -1.886 \dots\dots\dots (F.5)$$

——将 t_{cal} 值与所选 α 风险下的 t_{the} 值进行比较。 t_{the} 值对应于分位数为 $(1 - \alpha/2)$ 和自由度为 $(n - 1)$ 时的 t 分布函数的值。本例中, 当 $\alpha = 0.05$, $t_{the} = 1.980$ 。

当 $|t_{cal}| > t_{the}$ 时, 拒绝 H_0 , 在 α 水平上产品 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:

$$t_\beta = t_{the} - \left(\frac{\delta}{t_{calden}} \right) \dots\dots\dots (F.6)$$

式中:

- t_{the} —— α 水平上拒绝 H_0 的值;
- δ ——2 个产品之间的偏爱差异;
- t_{calden} —— t_{cal} 的分母[公式(F.5)];

$$t_\beta = 1.980 - (0.50/0.175) = -0.877。$$

由于公式(F.4)和公式(F.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, $\beta = 0.192$ 。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 值采用 Excel 中函数计算: $T.DIST(-0.873; 119; TRUE) = 0.192$, $P = 1 - 0.192 = 0.81$ 。

F.4 示例 3(附录 D 示例 3): 通过 Δ 值证实产品 A 不劣于产品 B, 每位消费者仅评价 1 个产品

附录 D 计算得出每个产品应有 $n = 374$ 位消费者进行测试。实验室使用 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:

$$t_{\text{cal}} = \frac{m_A - m_B + \Delta}{s_{\text{产品内}} \times \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \quad \dots\dots\dots(\text{F.7})$$

式中:

$$s_{\text{产品内}}^2 = \frac{s_A^2 \times (n_A - 1) + s_B^2 \times (n_B - 1)}{n_A + n_B - 2}$$

本例中,

$$s_{\text{产品内}}^2 = \frac{1.95^2 \times (300 - 1) + 1.80^2 \times (316 - 1)}{300 + 316 - 2} = 3.514$$

$$s_d = \sqrt{s_{\text{产品内}}^2} = \sqrt{3.514} = 1.875$$

t_{cal} 值为

$$t_{\text{cal}} = \frac{7.01 - 6.87 + 0.30}{1.875 \times \sqrt{\left(\frac{1}{300} + \frac{1}{316}\right)}} = \frac{0.44}{0.151} = 2.914$$

——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 $(1 - \alpha)$ 分位数的值。本例中, $\alpha = 0.05, t_{\text{the}} = 1.647$ 。

$t_{\text{cal}} > t_{\text{the}}$ 时, 拒绝 H_0 。本例中, $2.914 > 1.647$, 满足该条件, 实验室可在 α 水平上拒绝 H_0 并得到结论: 非劣效性检验产品 A 在统计上不劣于 B。

注: t_{the} 值应采用 Excel¹⁾ 中的函数计算: $T.INV(1 - \alpha; n_1 + n_2 - 2) = T.INV(0.95, 614) = 1.647$ 。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。

$$t_{\beta} = t_{\text{the}} - \left(\frac{\Delta}{t_{\text{calden}}}\right) \quad \dots\dots\dots(\text{F.8})$$

式中:

t_{the} ——拒绝 H_0 的值;

Δ ——认为 2 个产品相似(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 $m_A=6.87, s_A=1.95, m_B=7.01, s_B=1.80$;

$$t_{\text{cal}} = \frac{6.87 - 7.01 + 0.30}{1.875 \times \sqrt{\left(\frac{1}{300} + \frac{1}{316}\right)}} = \frac{0.16}{0.151} = 1.060$$

由于 $1.060 < 1.647$, 不能拒绝 H_0 。 t_{β} 值为:

$$t_{\beta} = 1.647 - (0.30/0.151) = 1.647 - 1.987 = -0.340$$

公式(F.4)、公式(F.6)和公式(F.8)中, β 计算规则均相同。由于 $t_{\beta} < 0$, β 值由在 t_{β} 处自由度为 $(n_A + n_B - 2)$ 时的分布函数的分位数。 $t_{\beta} = -0.340$, 自由度 = $300 + 316 - 2$ 时, t 分布函数等于 0.367。因此, 检验效力 $P = 1 - \beta = 1 - 0.367 = 0.633 \approx 0.63$ 。这个值远低于要求值 90%。

F.5 示例 4(附录 D 示例 4): 证实产品 A 不劣于产品 B, 每位消费者均评价 2 个产品

——附录 D 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 个产品评分差值的标准偏差;

——非劣效性测试中, t_{cal} 值由公式(F.9)得到:

$$t_{cal} = \frac{m_A - m_B + \Delta}{s_d \times \sqrt{1/n}} \dots\dots\dots (F.9)$$

本例中,

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

注 1: 公式(F.9)中, t_{the} 值采用 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 $(1-\alpha)$ 分位数得到。本例中, $\alpha=0.05, t_{the}=1.653$ 。

$t_{cal} > t_{the}$ 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 , 非劣效性检验不能得到产品 A 与 B 具有相似性。

——应采用公式(F.10)计算其检验效力:

$$t_{\beta} = t_{the} - \left(\frac{\Delta}{t_{calden}} \right) \dots\dots\dots (F.10)$$

式中:

t_{the} ——拒绝 H_0 的值;

Δ ——认为 2 个产品相似(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,

$$t_{\beta} = 1.653 - (0.30/0.208) = 0.211$$

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

注 2: P 值采用 Excel 中的函数计算: T.DIST(-0.211;199;1)=0.417 计算。

附录 G

(资料性)

中文语义喜好标度

表 G.1、图 G.1 和图 G.2 给出了适合中国消费者的 9 点喜好中文语义标度、15 cm 线性喜好中文语义标度和标记情感量值(labeled affective magnitude, LAM)中文语义标度。

表 G.1 9 点喜好中文语义标度

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

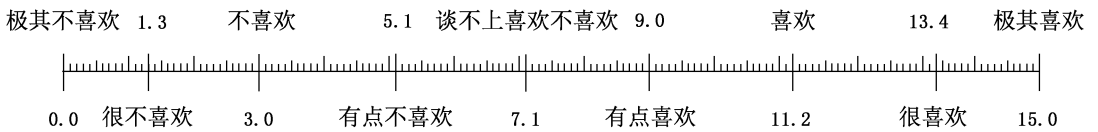


图 G.1 15 cm 线性喜好中文语义标度

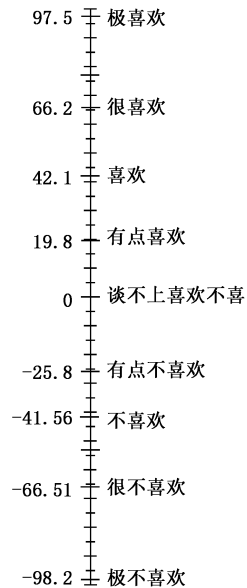


图 G.2 LAM 中文语义标度

参 考 文 献

- [1] GB/T 26315—2010 市场、民意和社会调查 术语
- [2] Abdi H., Edelman B., Valentin D., Dowling W.G. 2009. Experimental design and analysis for psychology. Oxford University Press, *pp.* 1-538.
- [3] AFNOR NF V09-500, 2012. Directives Générales pour la réalisation d'épreuves hédoniques effectuées avec des consommateurs dans un espace contrôlé.
- [4] ASTM E1958-20. Guide for Sensory Claim Substantiation.
- [5] Ball R. 1997. Incomplete block designs for the minimization of order and carry-over effects in sensory analysis. *Food Qual. Prefer.*, 18, 111-118.
- [6] Bi J. Similarity testing using paired comparison method. 2007. *Food Qual. Prefer.*, 18, 500-507.
- [7] Blackwelder W.C. 1982. "Proving the null hypothesis" in clinical trials. *Control. Clin. Trials.*, 3, 345-353.
- [8] Cochran W.G., Cox G.M. 1957. *Experimental Designs*. Wiley, New York, Second Edition.
- [9] Cohen J. *Statistical Power analysis for the behavioural sciences*. 1988. Elsevier-Academic Press, New York, Second Edition.
- [10] 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.
- [11] Greehouse S.W., Geisser S. 1959. On methods in the analysis of profile data. *Psychometrika*, 24, *pp.* 98-112.
- [12] Howell D.C. 2002. *Statistical methods for psychology*. Duxbury, Pacific Grove, Fifth Edition, *pp.* 1-802.
- [13] Kunert J., Sailer O. 2006. On nearly balanced designs for sensory trials. *Food Qual. Prefer.*, 17, *pp.* 219-227.
- [14] Kunert J. 1998. Sensory experiments as crossover studies. *Food Qual. Prefer.*, 9, *pp.* 243-253.
- [15] Lawless H.T., Heymann H. 2010. *Sensory evaluation of food - principles and practices*. Springer, New York, Second Edition, *pp.* 1-596.
- [16] 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.
- [17] Meilgaard M.C., Civille G.V., Carr B.T. 2007. *Sensory evaluation techniques*. CRC Press, Boca Raton, Fourth Edition, *pp.* 1-448.
- [18] Meyners M. 2007. Least equivalent allowable differences in equivalence testing. *Food Qual. Prefer.*, 18, *pp.* 541-547.
- [19] Meyners M. 2012. Equivalence tests - a review. *Food Qual. Prefer.*, 26, *pp.* 231-245.
- [20] Naes T., Brockhoff P.B., Tomic B. 2010. *Statistics for sensory and consumer science*. Wiley, Chichester, Second Edition, *pp.* 1-287.
- [21] 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.
- [22] Siegel S., Castellan N. 1988. *Nonparametric systems for the behavioural sciences*. Sin-

gapour: Mc Graw-Hill International Editions, Second Edition.

[23] 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.

[24] 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.

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

