



# 中华人民共和国国家标准

GB/T ××××—××××/ISO 20784:2021

---

## 感官分析 产品感官宣称证实导则

Sensory analysis — Guidance on substantiation for sensory and consumer  
product claims

(ISO 20784:2021, IDT)

××××-××-×× 发布

××××-××-×× 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 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 注意事项 .....	3
5 指导原则 .....	4
5.1 总则 .....	4
5.2 查询政府发布的法律法规 .....	4
5.3 明确宣称主要内容并设计测试方案 .....	4
5.4 确定宣称类型:单一产品测试或比较性测试 .....	4
5.5 确定决策标准 .....	4
5.6 确定相关产品集 .....	4
5.7 确定相关消费者群体或评价员 .....	5
5.8 确定证据强有力的程度 .....	5
5.9 保证公正性 .....	5
5.10 保证可靠性 .....	5
6 感官宣称分类 .....	5
6.1 分类 .....	5
6.2 用语 .....	6
6.3 非比较性感官宣称 .....	6
6.4 比较性感官宣称 .....	6
7 方法 .....	7
附录 A (资料性) I 类错误率与统计检验次数的函数 .....	9
附录 B (资料性) 感官宣称示例 .....	10
附录 NA (资料性) 比较性感官宣称示例 .....	15
参考文献 .....	18

## 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。

本文件等同采用了 ISO 20784:2021《感官分析 感官宣称证实导则》。

本文件做了下列最小限度的编辑性改动：

- a) 将 3.1 定义中的举例放入了“示例”中，以方便理解和使用；
- b) 增加了 3.6 的注，以方便理解和使用；
- c) 更改了 6.4 示例 1 和示例 3 的表述，以方便理解和使用；
- d) 将附录 B.4 中的“46%”改为“44%”，纠正原文错误；
- e) 增加了附录 B 中 B.5~B.7 的注，以方便理解和使用；
- f) 将 B.8 标题中的“非比较性”更改为“比较性”，纠正原文错误；
- g) 增加了附录 NA，补充了比较性感官宣称示例，以方便理解和使用；
- h) 增加了参考文献，以方便理解和使用。

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、内蒙古蒙牛乳业(集团)股份有限公司、内蒙古伊利实业集团股份有限公司、四川丁点儿食品开发股份有限公司、青岛啤酒股份有限公司、黑龙江飞鹤乳业有限公司、广东美的生活电器制造有限公司、河北凤来仪酒业有限公司、中国绿色食品有限公司、重庆德庄农产品开发有限公司、元气森林(湖北)饮料有限公司、广东海明晖电子科技有限公司、上海测谱检测技术有限公司、武汉市华测检测技术有限公司、星智豫美(上海)生物科技有限公司、钟薛高食品(上海)有限公司、通标标准技术服务(青岛)有限公司、中标能效科技(北京)有限公司。

本文件主要起草人：钟葵、赵镭、李洪亮、史波林、苏玉芳、汪厚银、任康、董建军、李懿霖、苏莹、马辉峰、唐飞、李德建、杨莹、鲍熹珺、朱立宁、高飞、杨传全、鲁楠、邵世伟、陈剑、温焯、尹花、安志丛、罗强祖、张东方、张文娟、李媛、王思思、廖欣、白雪、何晓宁、黄翠玲、李婷婷、吴博。

# 感官分析 产品感官宣称证实导则

## 1 范围

本文件提供了对食品和非食品类产品感官宣称以及产品包装上的宣传语进行证实的导则,给出了感官宣称与其他类型宣称的区别、感官宣称分类、不同类型感官宣称示例以及感官宣称证实示例。

本文件适用于产品感官宣称证实。

本文件不适用于下列情况:

- 支持感官宣称所用测试方法中的具体或详细要求;
- 产品原产地、组分、加工和营养成分的事实宣称;
- 产品技术特征的事实宣称;
- 产品消费或使用,产品的健康特性、医学及治疗功效、生理功效、结构或功能优势的宣称;
- 使用仪器评价产品特性或性能的宣称(仪器评价是指不使用评价员或被试,而使用仪器对产品的特性或性能进行评价);
- 服务类的宣称(例如房屋清洁服务、航空服务、汽车服务等);
- 大件/耐用消费品的宣称(例如汽车、冰箱、灶具等)。

## 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。

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

注: GB/T 10221—2021 感官分析 术语(ISO 5492:2021, IDT)

## 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。

ISO 和 IEC 维护的用于标准化的术语数据库网址如下:

- ISO 在线浏览平台: <https://www.iso.org/obp>
- IEC 电子百科: <http://www.electropedia.org/>

### 3.1

#### 感官宣称 sensory claim

向消费者宣传一个产品的感官特性、功能特性/性能特性以及消费者对该产品使用时(包括使用前、使用期间或使用后)的情感响应和感知响应的表述。

示例:感官特性例如“具有焙烤风味”,功能特性例如“去除油渍”,性能特性例如“口气清新更持久”,情感响应例如“消费者偏爱的”,感知响应例如“让你的肌肤看起来更年轻”。

注:感官宣称包括在公共场所发布的所有形式的广告信息。广告信息包括产品包装、印刷品或数字媒体(如电子、电视或视频)等多种形式。目的是将产品的感官特性传达给产品的购买者或潜在用户,或者强调消费者在使用产品(包括使用前、使用期间或使用后)时的感知。这种广告信息是让购买者或潜在用户了解产品特性,从而影响其购买、消费或使用决定。

3.2

**情感宣称 affective claim**

用户或潜在用户在产品使用(包括使用前、使用期间或使用后)时与其喜好和/或情绪响应相关的表述。

注:该响应包括消费者在产品使用(包括产品使用前、使用期间或使用后)时,由产品引发的消费者喜好、态度、认知或情绪响应。最常用的喜好响应测试是喜好测试和偏好测试。态度响应是指消费者愿意在未来购买该产品,或者是消费者认同该产品具有某种突出特性或能提供特定情感体验的声称。

3.3

**感知/性能宣称 perception/performance claim**

描述产品引起的某种感知特性或者预期效果的表述。

示例:×××产品稠厚(感知)且使用后无残留(性能)。

3.4

**夸大 puffery**

一种极其宽泛、模糊和主观的表述,由于过度夸大导致可信度低,在测量操作或实践角度上不具有可测性。

3.5

**等效性宣称 equivalence claim**

声称两个或多个产品在某个或多个产品特性上相同时所提供的表述。

3.6

**卓越性宣称 unsurpassed claim**

声称产品在一个或多个特性上不会被其他产品超越时所提供的表述。

注:这里的其他产品,是宣称证实测试时的对比产品。对比产品可以是竞品,也可以是同公司的其他产品,或者是标准样品等。在中国地区,与竞品进行比较的宣称内容及用语要符合中国相关的法律法规、规章和有关规定,例如《中华人民共和国广告法》和《中华人民共和国反不正当竞争法》。

3.7

**优越性宣称 superiority claim**

声称产品与另一个或多个产品比较时,产品的性能/特性水平更高或者更被喜爱/偏爱的表述。

3.8

**风险 risk**

伤害发生的可能性和严重性的组合。

注:感官研究人员和利益相关方是需要考虑基于感官测试获得宣称的相关风险。风险是指提出宣称后出现负面后果的可能性。这些负面后果的形式并不明确,包括了消费者在社交媒体上发布负面帖子或在感官、营销或法律社区内发表负面评论,竞争对手的挑战以及自我监督机构、监管机构或政府机构采取的相关行动。宣称公开前要预先识别、讨论和理解相关风险。

[来源:ISO/IEC 导则 51:2014,3.9,有修改]

3.9

**感官分析方法 sensory analysis methods**

感官分析中被广泛使用且有科学依据的一套方法,包括差别检验、描述性分析和性能评价。

注:内部有效性和实验室类型控制是感官测试质量的保证,尤其在测试目标是产品特性时。对于关注产品和关注参与者的感官测试,屏蔽产品的品牌标识并进行独立评价均是最适宜的方法。测试方法的灵敏度、效应量以及评价员的数量和类型都是感官分析方法的考虑因素(参见 ISO 6658)。

3.10

**消费者测试方法 consumer methods**

大多数从事产品感官和消费者测试的专业人员所使用的消费者定量测试系列方法,包括情感测试

和感知/性能测试。

### 3.11

#### 代表性消费者样本 representative sample of consumers

测试中使用的消费者群体,通常是从更大的总群中抽取的一个较小群体。该群体的测试结果涵盖了采用总群体测试时可获得的响应变化范围。

注:消费者情感响应测试中,代表性消费者样本特征包括:

- a) 足够多的消费者样本量,保证测试结果涵盖了总群体的情感响应变化范围,消费者样本量选择参见 ISO 11136;
- b) 使用符合条件的消费者,即选择产品的真实用户/消费者、购买者或者接受产品概念的消费者;
- c) 消费者样本抽取方案的变量包括人口统计、地理、行为或心理等影响群体差异的相关因素。

### 3.12

#### 代表性产品样本 representative sample of products

从市场可获取产品中抽取的,用于产品感官特性、性能特性或产品引起的喜好响应等测试的系列产品。

注:通常,研究人员是等产品从工厂生产端流通到常用销售端后,再通过零售渠道获取产品进行测试。样品的选择原则是测试产品能够代表消费者在市场上可购买得到的产品。选择足够多数量、批次和不同厂商的产品,保证涵盖此类产品的所有变化范围。新品推广、销售和上市前的样板产品能够用于支持宣称。用新样品测试结果证实宣称时,要提供新样品与市场产品是相同的相关信息。此外,产品选择时,测试产品数量和变量数量取决于宣传想要表达的程度,即是否适用于消费者可能会使用到的所有产品。

### 3.13

#### 客观测试结果 objective test result

采用科学领域中被广泛认可的实验方法开展测试所获得的结果,测试结果不受实验者的期望或干预的影响(具有可验证性)。

注:客观研究中,数据收集不会受到研究者的干预,试验设计也允许得到一种以上的可能结果。同样客观研究中,受访者并不会知晓任何可能影响其对潜在研究目标响应的信息。因此,客观测试结果不会受到测试者或测试管理者意见或期望的影响,是包含了相关变量并遵循最佳操作实践得到的研究结果,并非预期的结论。通常,客观研究结果能通过不同的研究方式获得,并能采用其他相关研究或组合测试结果进行验证。

## 4 注意事项

证实感官宣称时,宜考虑下列事项。

- a) 感官宣称是基于受试者对产品直接体验的响应记录。
- b) 产品宣称旨在向潜在购买者介绍产品特性,阐明产品与对比产品的差异并影响其购买决定。
- c) 我国对于儿童或儿童产品相关的宣称有着严格的监管和限制。
- d) 支持宣称的证据是否恰当使用了科学的方法以及支持数据的力度和相关性。

注1:研究人员对自家公司产品通常进行过多次测量,频繁测试中可能会出现某次测试结果与其他测试结果不一致的情况。该情况下,如果大多数测试结果的分布非常集中,对宣称的支持力度会更强。如果单次测试结果是在前期的测试结果范围内,可以作为宣称依据使用。如果没有前期测试结果,与单次结果一致的其他技术信息或证据也能加强对宣称的支持。

注2:供应商或测试机构开展完测试后,如果没有前期测试结果记录导致无法进行比较时,提出宣称的公司对基于单次测试结果支持的宣称风险负责。

- e) 感官宣称是基于下述内容的标准化和科学化测量:
  - 采用感官分析方法确定的产品特性/性能;
  - 用户对产品特性/性能的喜好、偏爱、感受、态度或感知。
- f) 赫尔辛基宣言中给出了临床研究的定义,即临床试验是“对人体受试者(病人和健康志愿者)所

使用的药物或医疗设备进行的系统评价,以证实或揭示产品的功效、不良反应以及试验药物的吸收、分布、代谢和排泄,目的是确定产品的疗效和安全性”<sup>[6]</sup>。因此,需要临床研究的宣称并不属于感官宣称,不在本文件讨论范围内。

- g) 与关注产品疗效或者产品对人体结构或操作产生的可能潜在影响的研究不同,感官研究主要涉及感官效果的评价,例如特定的口感、香气或外观。

示例:如护肤产品,如果研究目的是证明皮肤外观发生变化或改善,裸眼检测和评价员评价是属于感官研究的范围。不使用仪器、采用外部评价员对皮肤触觉特性变化的检测也属于感官研究范围。如果护肤产品宣称是改变了皮肤的深层组织/功能或者是整个皮肤表层、真皮层的状况,这属于临床研究而并非感官研究。

注 3:上述 2 种研究类型存在细微差别。对于化妆品和个人护理用品,不同国家和公司对于“感官”和“临床”类型的区分存在差异。本文件中,宣称“改善人体的深层组织/功能”是临床研究,不在本文件讨论范围内。

- h) 夸大宣称或虚假宣传中的表述是空洞不明确的(例如“这款香水会带你飞”)或者过于夸张(例如“世界上最舒适的鞋子”),没有人会当真。由于表述含糊或者考虑实用性,夸大宣称不能用科学测试数据进行支持。
- i) 计划开展感官宣称测试的研究人员应清楚了解感官宣称相关的监管机构、管理法规、媒体或潜在竞争对手对感官宣称的可能反应。研究人员宜在研究方案设计和分析时就要考虑相关法规和竞争对手的可能反应。

## 5 指导原则

### 5.1 总则

开展感官或消费者测试对感官宣称进行证实时,调查人员宜考虑 5.2~5.10 中给出的 9 项指导原则,以获得足够强有力的测试结果来支持宣称内容。

### 5.2 查询政府发布的法律法规

宣称测试前,应先查询政府、监管机构、广告业和媒体对感官宣称的要求和标准。

### 5.3 明确宣称主要内容并设计测试方案

根据宣称内容设计研究方案。测试前先确定好宣称用语,便于合理设计试验。依据预期的宣称内容,选择研究的主要目标终点(例如特性描述或者消费者响应)。理想情况下,目标数量不宜过多,避免产生多重效应<sup>[3]</sup>。评价项目越多,出现虚假或矛盾结果的可能性越大。检验统计次数增加与 I 类错误概率变化的函数见附录 A。

注:必要时,研究人员先确定好感官宣称的类型,然后再开展宣称的相关研究。

### 5.4 确定宣称类型:单一产品测试或比较性测试

无论是单一产品、成对产品还是多个产品测试,宜先确定宣称类型。非比较性宣称时,宜采用单一测试;比较性宣称时,产品类别宣称宜采用成对产品或一个产品与多个产品比较。

注:比较性宣称时,研究人员先确定对比产品的数量和类型。对比产品选择的具体指导参见 ASTM E1958。

### 5.5 确定决策标准

宜提前清楚且明确地界定决策标准。

### 5.6 确定相关产品集

对于比较性宣称,宜提前确定好测试相关的产品集。

### 5.7 确定相关消费者群体或评价员

对于消费者测试,宜在开展研究前先确定好消费者群体,包括相关用户、当前用户、购买者或潜在消费者。对于关注产品的测试,提前确定好评价员的训练水平和评价资质。

### 5.8 确定证据强有力的程度

宜使用强有力的证据支持宣称,以应对预期中的挑战。

注:研究人员需知晓宣称证据的支持力度,以及宣称受到政府、监管部门或者竞争对手质疑风险的可能性。确定证据支持力度时的考虑因素包括:

- a) 相同测试结果出现多少次时,宣称方才能确定结论是“可靠的”;
- b) 宣称证据的一致性程度,例如确认该结果与其他测试和测量的信息(结果)是否一致;
- c) 受访者样本和产品样本的相关性和代表性程度;
- d) 研究中试验变量和混淆变量的管理程度;
- e) 测试时遵循上述最佳实践原则的程度,同时测试前先选择好主要目标终点并坚持不变(见 5.3)。

### 5.9 保证公正性

提出宣称的组织方宜保证测试方法、测试条件和测试执行的公正性。例如,使用同期生产的产品,采用相同方式制备和提供样品,呈送顺序遵循平衡原则,测试程序相同。

### 5.10 保证可靠性

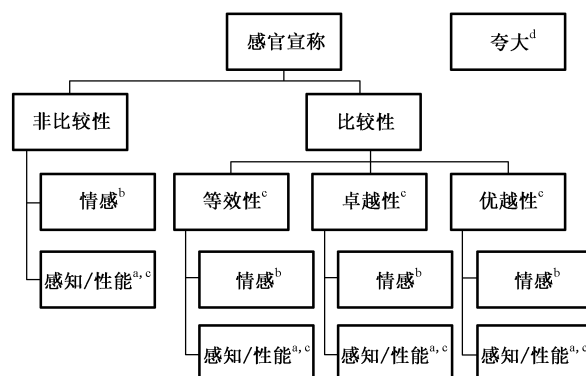
宜遵循最佳实践原则以保证研究的客观性和公正性,研究结果可靠性高且重复性好。

## 6 感官宣称分类

### 6.1 分类

感官宣称分类有利于获得清晰合理的证据以及选择适宜的测试方法。感官宣称的分类如图 1 所示,不同宣称类别的主要差异如下所示:

- a) 是单个产品宣称还是与对比产品的比较宣称;
- b) 是消费者对产品的情感响应还是对感知/性能响应的宣称;
- c) 受访者是消费者还是经过训练的评价员。



<sup>a</sup> 采用经过训练的评价员或相关消费者开展测试。

<sup>b</sup> 采用产品当前购买者或潜在用户作为情感宣称测试中的消费者。

<sup>c</sup> 等效性、卓越性和优越性宣称中使用的受访者类型取决于情感宣称还是感知/性能宣称。

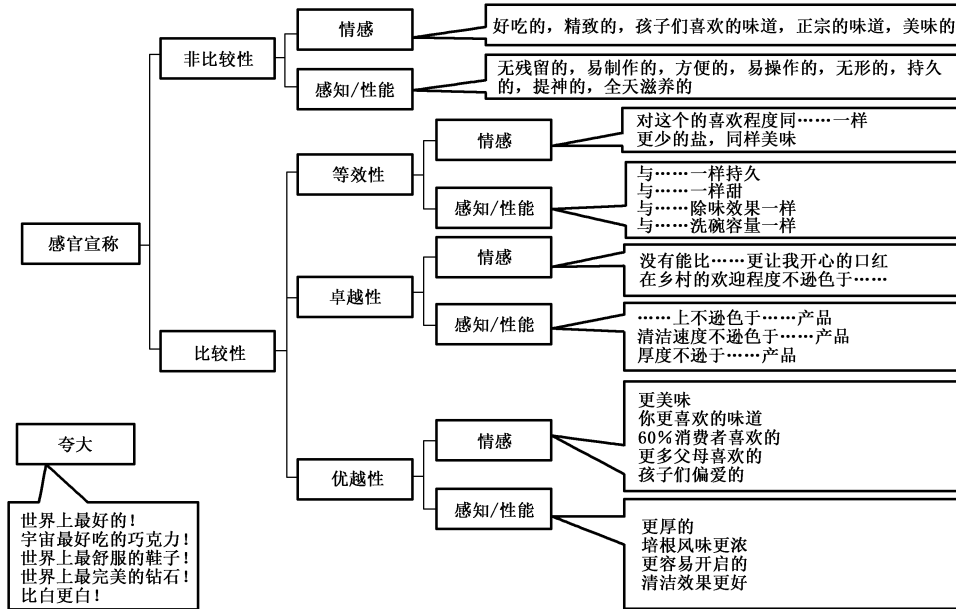
<sup>d</sup> 夸大是一种特殊类型宣称,极度夸张或含糊不清,且无法通过感官方法测试获取。

图 1 感官宣称的分类



## 6.2 用语

图 2 给出了不同类型感官宣称的一小部分用语示例。宣称用语会因产品类别、预期目标消费者和感官宣称实施地区而有所差异。宣称用语的确定取决于感官测试结果是否直接明确地支持宣称。



注：感官宣称类型有很多种，表述方式多样。图 2 给出了一部分感官宣称用语示例，不同类型感官宣称证实的具体示例见附录 B 和附录 NA。

图 2 不同类型感官宣称用语的示例

## 6.3 非比较性感官宣称

非比较性感官宣称测试包括单一测试或单个产品测试，有明确的测试目标，依据决策标准具体展开测试，并对定量数据进行统计分析。情感类的非比较性宣称采用消费者定量测试方法，感知/性能类的非比较性宣称采用感官描述小组的评价结果或者消费者对产品引发感知的自我评价报告支持宣称。此类数据同样要进行统计学处理。

## 6.4 比较性感官宣称

比较性宣称是指将 2 个或多个产品相互比较后得出的宣称。这种比较可以是广告商产品与一个或多个竞品的比较，也可以是同一产品不同版本间的比较。比较内容包括优越性、等效性和非劣效性。比较性宣称能突出不同产品在感官特性、情感响应以及产品性能上的差异或者突出产品的某些新特性。

当广告商想对公司内的 2 款产品进行比较宣称时，例如产品的改进版和当前版，这种挑战的风险通常很低。公司有时候仍然希望获得一些感官数据来支持这种低风险宣称，因此需要进行测试，支持新产品与当前产品存在感官特性差异的宣称。这种情况下，可使用产品的历史测试信息作为决策标准。由于挑战风险很低，选择决策标准时不用像高风险宣称那么严格。

与其他产品公开比较的宣传方式风险更高，其他人员可能会对比较性宣称的内容提出质疑，因此广告商要开展严格的测试来支持所提出的宣称。

产品优越性宣称是声称广告商产品更受欢迎或更被偏爱，具有更多让人满意的产品特性或优于宣称中对比的产品。

示例 1：“比……更被消费者偏爱的”“比……产品清洁效果更好”。

等效性宣称是声称广告商的产品与宣称中的对比产品一样好或是等效的。

示例 2:“味道和……一样好”“清洁效果和……一样”。

非劣效性(或卓越性)宣称是声称广告商的产品至少与宣称中对比产品一样好。

示例 3:“欢迎程度不逊色于……”,“水果味不弱于……”,“美味程度不比……差”,“清洁效果不比……差”。

等效性宣称是指 2 个或多个产品是相同的。等效性宣称的统计假设包括上限和下限(即双边备择假设),优越性和非劣效性宣称的统计假设只涉及下限(即单边备择假设)。每种比较性宣称都有相应的试验设计和分析方法。需要注意一点,测试结果不具有显著性时,不能推断出等效性和非劣效性。与产品优越性或非劣效性的宣称测试相比,等效性宣称需要更多的样本量。

每种比较性宣称(优越性、等效性或非劣效性)的证实都需要使用经过统计学处理的定量测试数据。

情感类比较性宣称时收集的数据通常是喜好数据或偏好数据。感知/性能特性类比较性宣称的数据包括消费者数据和经过训练的感官评价小组数据。

## 7 方法

7.1 用于证实感官宣称的感官分析方法有很多。依据宣称类型和测试目的来选择支持感官宣称的测试方法,具体参见 ISO 6658。

差别检验和描述性分析类的感官分析方法通常以产品为中心,检验目的是确定产品中存在哪些感官特性以及感官特性上是否存在差异。差别检验和描述性分析使用经过训练的或有经验的感官评价人员开展测试。对于希望建立产品相关宣称的感官研究人员,了解与感官特性宣称相关的具体法规非常重要。如果目标群体是终端消费者/产品用户,宣称特性必须是消费者能感知到的感官特性。

开展消费者喜好或偏好类情感测试时,宜选择适宜且具有代表性的用户样本,如果是尚未上市的新产品则选择潜在用户。情感宣称是基于人的声称(例如消费者偏爱的产品),是关于测试中消费者响应的表述,宜选取代表性消费者样本进行情感测试。宜使用足够多数量的代表性消费者,并设置相应的招募条件。

情感测试的消费者包括:

- a) 产品的使用者或购买者,或潜在的使用者/购买者(如果产品尚未上市);
- b) 代表了更广大的潜在消费者、购买者或用户群体的测试人员;
- c) 产品主观评价的提供者,例如提供对产品的喜好或偏好,或采用适合项全选法(check all that apply, CATA)对感知到的产品特性评级或进行自我报告。

与以产品为中心的感官分析方法一样,盲测和独立判定也是消费者测试方法的重要特征。外部效度,即测试结果在多大程度上能被推广到更大的用户或潜在用户群体,以及测试产品在多大程度上等同消费者的日常使用,都是消费者测试方法中最重要的一环。

宜使用市场上可获取的产品作为分析型感官检验和消费者测试中的测试产品。从零售店购买是最直接的方式,获取的产品最能代表是消费者会购买和消费(使用)的产品。如果产品还未上市,研究者应提供证据,证明宣称中使用的测试产品与市面产品的一致性。

当使用一种创新的方法(例如用于新产品研发)证实感官宣称时,测试宜遵循多数感官专业人员所认可的科学性原则。

7.2 不宜采用探索性、假设性或发现性的研究方法作为感官宣称的主要支持证据。没有其他证据支持时,不宜采用定性测试支持宣称。

7.3 下列方法均可用于感官宣称证实,包括:

- a) 整体差别检验(例如三点检验、二-三点检验);
- b) 特性差异测试(例如定向差别检验、特性评级法);
- c) 描述性分析;

d) 消费者情感测试或性能测试。

7.4 证实感官宣称的感官测试指导原则参见第 5 章,具体内容如下:

- a) 明确我国对宣称的相关要求;
- b) 预定义的主要宣称表述;
- c) 数据收集的公平性和测量方法的公正性;
- d) 与宣称相关的参与者和产品的要求。

7.5 宜选用产品的当前用户或潜在用户参与产品的情感宣称测试或感知功能测试。选用经过培训的感官评价员或优选评价员对产品特性或性能宣称进行评价。

## 附录 A

(资料性)

## I 类错误率与统计检验次数的函数

表 A.1 给出了同一项研究中  $\alpha$ -风险 (I 类错误) 随产品数量或特性数量增加而增大的变化。 $\alpha$ -风险随检验统计次数增加而增大的现象表明至少观察到 1 个假阳性 (显著) 结果的概率会随检验次数增加而显著增大。因此, 当实际产品特性并无差异时, 得出“存在差异”结论的概率增大。

支持宣称的感官描述性分析人员宜意识到这种影响。表 A.1 是“多重性”的一个具体示例<sup>[8]</sup>, 例如 5% 显著性水平下, 统计检验次数为 10 个产品或特性时, 至少发生 1 次  $\alpha$ -风险偶然事件的概率 (即使检测的产品特性都是相同的) 约为 40%。

消除多重性的最好方法是在研究开始前先确定产品或特性差异的再现性或可靠性, 并严格设定主要终点, 或同时提供其他证据来支持产品或特性差异的结论。需要注意一点, 任何时候执行多重检验均会出现这个问题, 包括多重成对产品比较、子群 (例如年龄、性别) 多重比较或相同产品多个特性间的比较。

表 A.1 是使用了 R3.6.3 中的下列代码行所得到的结果: `cbind[x<-1:30,round(1-0.95^x,2),round(1-0.99^x,2)]`。

表 A.1 I 类错误率随统计检验 (产品或特性) 次数增加的变化

统计检验次数 (产品或特性)	比较的显著性水平		统计检验次数 (产品或特性)	比较的显著性水平	
	0.05	0.01		0.05	0.01
1	0.05	0.01	16	0.56	0.15
2	0.10	0.02	17	0.58	0.16
3	0.14	0.03	18	0.60	0.17
4	0.19	0.04	19	0.62	0.17
5	0.23	0.05	20	0.64	0.18
6	0.26	0.06	21	0.66	0.19
7	0.30	0.07	22	0.68	0.20
8	0.34	0.08	23	0.69	0.21
9	0.37	0.09	24	0.71	0.21
10	0.40	0.10	25	0.72	0.22
11	0.43	0.10	26	0.74	0.23
12	0.46	0.11	27	0.75	0.24
13	0.49	0.12	28	0.76	0.25
14	0.51	0.13	29	0.77	0.25
15	0.54	0.14	30	0.79	0.26

**附 录 B**  
(资料性)  
感官宣称示例

### B.1 非比较性——情感宣称：“美味”

一家沙拉酱生产公司希望在产品标签正面印上“美味”。感官、营销和法规团队同意进行产品测试,如果大多数消费者品尝产品后均认同产品“美味”,公司就会将在产品包装标签上印上“美味”。研究人员查阅了消费者评价为“美味”的类似产品的历史数据,获取相关背景信息来确定决策标准。最终研究人员提出的决策标准是:70%或更多的消费者在品尝沙拉酱后认同“美味”声称。

共计 120 名消费者采用盲测方式品尝了产品,并对产品声称“美味”的认同程度进行评价。采用 7 点李克特标度进行评价,“7”是“完全同意”,“1”是“绝完全不同意”,“4”是“既不同意也不反对”。研究人员推荐使用最高的 2 个评分,即以选择了“完全同意(7)”和“非常同意(6)”的消费者数量之和作为认同沙拉酱“美味”的消费者数量。这保证了只有强烈认同的消费者才符合宣称决策标准中的消费者要求。

测试结果以列表形式呈现,120 名消费者中有 96 人(80%)选择了“完全同意”或“非常同意”沙拉酱“美味”的声称。基于二项式分布,95%置信区间下 80%的分布范围为(0.717,0.867),即认同沙拉酱“美味”的消费者比例为 72%~87%,95%置信区间的下限高于 70%的决策标准。因此,测试结果支持在包装上印上“美味”表述。

受控区域内消费者喜好测试方法,参见 ISO 11136。

注:可采用程序完成计算,例如使用 R 中代码行 `binom.test(96,120)`,或使用 Microsoft Excel 2010 完成计算。<sup>1)</sup>

### B.2 非比较性——性能宣称：“无残留”

产品研发人员发现了一种成分,添加到护肤霜中不会产生残留感。该公司希望将这种成分添加到公司目前最畅销的护肤霜中,并宣传“无残留”。现在需要感官数据支持新配方护肤霜的“无残留”宣称。感官研究人员建议采用经过培训且考核通过的描述性分析小组对添加和未添加新成分的护肤霜进行评价。测试目标是客观确认这种新添加成分是否具有“无残留”效果。描述性评价小组经过培训后对护肤霜产品展开定期评价。研究团队提出的决策标准是认为护肤霜“无残留”的人数更多,即“无残留”回答数比例应显著高于 0.5(此处对照数据,不是指未添加新成分产品的数据,而是指被添加了新成分、但被判定“有残留”产品的数据)。

描述性评价小组由 12 名经过训练的评价员组成。评价员使用标准涂抹程序将护肤霜涂抹在前臂上,同时对涂抹护肤霜后的前臂表面进行评价。评价员使用标准用量和标准涂抹方法涂抹 2 种护肤霜,2 个前臂各涂 1 种。护肤霜采用三位数字编码识别,其中一种护肤霜含有“无残留”成分,另一种则不含。2 种护肤霜在香气、外观、质地和感觉上均没有差异。

3 min 后,评价员对前臂的“可见残留物”展开评价。

接下来的一周内再次重复相同程序,以确保“无残留”宣称得到强有力的支持。2 项试验数据见表 B.1。

---

1) 文中所提及的软件是目前比较适合的商业化产品。该信息只是为了给本文件使用者提供一些便利,并不涉及推荐认可产品。

表 B.1 2 次试验数据结果

判定	测试 1	测试 2	测试 1+测试 2 的总和
判定为“无残留”的总人数	10	8	18
判定为“无残留”的人数比例	0.83	0.67	0.75

合并 2 次试验数据后进行分析。添加了新成分的护肤霜,在 24 个受试者中总共获得 18 个“无残留”的判定。根据二项式定律,获得“无残留”答案的概率为 0.5,24 次中偶然获得 18 次或更多次“无残留”结果的概率小于 0.023。因此,该公司决定在新配方护肤霜包装上添加“无残留”字样,从而向消费者强调产品的优势。

### B.3 非比较性——性能宣称:“易烹煮”

一家公司研发了一种“易烹煮”的芝士意面,与公司现有销售产品相比,新配方芝士意面的烹煮步骤更少。公司跨职能团队就决策标准达成一致意见,即当前产品的大多数消费者认同新配方芝士意面“易烹煮”。其中,“大多数消费者”被定义为测试中不少于 80% 的消费者。

公司招募了 90 位经常购买公司芝士意面的消费者来进行非比较性的性能宣称测试。向每位消费者提供 1 盒无品牌标识的新研发产品,让消费者在家中按照包装盒上给出的步骤自行烹煮。要求消费者评价新研发芝士意面是否“易烹煮”,并独立完成 1 份关于产品烹煮难易程度的线上问卷。采用 5 点标度评价烹煮的难易程度,“5”表示“非常容易烹煮”,“1”表示“一点也不容易烹煮”,“3”表示“既不容易也不难烹煮”。

90 名消费者中有 80 名(89%)消费者认为新配方产品“非常容易烹煮”或“容易烹煮”。根据二项式定律,95% 置信区间下 89% 的分布范围为(0.81,0.95),即 81%~95%,显著高于决策标准中 80% 的要求( $p=0.035$ )。测试结果满足决策标准,市场部可以开始准备印有“易烹煮”宣称的包装和销售材料,为后期营销做准备。

### B.4 比较性——情感宣称:“减盐 30%,同样美味”

一家薄脆饼干公司希望研发更健康的饼干产品,以满足消费者追求更健康零食的需求。公司决定将 1 款主打的薄脆饼干产品中的盐分减少 30%,同时尽可能保持饼干的原有味道。与当前使用的食盐相比,减盐饼干中使用了一种食盐新产品,这种盐改变了颗粒表面积,可以提供更强的咸感。现在需要感官数据确定减盐饼干和当前饼干的咸度是否相同。为了实现这个要求,感官研究人员决定采用定向比较检验。

公司跨职能团队同时希望获得消费者数据来支持“同样美味”的宣称。团队制定的决策标准是不少于 70% 的消费者在品尝减盐饼干后认同“美味”的声称,从而支持“盐分减少了,但同样美味”的宣称。经过几轮配方改进和内部测试小组测试后,产品研发团队研发了 1 款新配方饼干。与现有饼干相比,新配方饼干具有相近的咸味感知。新配方中盐含量降低了 30%,但咸度感知差异很小,具有与当前产品类似的滋味剖面。

招募了 68 名经验丰富的评价员进行相似性定向成对比较,检验减盐饼干咸度是否低于当前饼干产品。测试目标是认为减盐饼干咸度低于当前饼干的评价员数量(如果有的话)尽可能的“少”。团队制定的决策标准是不超过 20% ( $p_d=0.2$ ) 的评价员能够识别出咸度差异,例如正确回答数比例低于  $p=0.6$ 。

68 名评价员中有 30 名(44%)选择了当前饼干比 30% 减盐饼干“更咸”。这个比例显著低于可接受的比例 0.6(单边  $p=0.006$ ,单边置信区间上限是 0.55)。

团队继续测试,确定是否可以使用“同样美味”的声称。

90名当前饼干产品的消费者采用盲评方式品尝了低盐饼干和当前饼干,并使用5点标度分别评价2种饼干是否“美味”。选择“4”和“5”才被认为是认同“美味”,决策标准是至少有 $2/3$ 的消费者认同“美味”声称。测试采用单产品评价法,即只使用低盐饼干数据,不使用当前饼干的数据。

90名消费者中有65名(72%)给低盐饼干打分不低于“4”。二项式单边检验表明该比率并没有显著高于 $2/3=0.667$ ( $p=0.157$ ,置信区间下限为0.63)。该团队得出结论:测试结果不支持“减盐30%,相同美味”的宣称。因此,新配方低盐饼干不使用“减盐30%,相同美味”的宣称。

### B.5 比较性——卓越性——性能宣称:“清洁效果与领先品牌相媲美”

国外一家洗涤剂生产商旗下的A品牌洗涤剂产品在全国市场上销售。A品牌产品的竞争威胁是B公司一个更昂贵的洗涤剂,其宣称具有“B品牌产品的清洁效果优于其他领先品牌”的优越性能。品牌A和品牌B的销量占全国洗涤剂销售的91%。

A公司研究人员认为A品牌洗涤剂和竞品B具有相同的清洁效果,该公司希望将该宣称印在产品宣传材料上展示给零售商,以保持A品牌的零售市场份额。同时,公司还希望在电视广告上宣传“A品牌的清洁效果与领先品牌相媲美”,以对抗B品牌的优越性宣称。

A品牌的达成标准如下。

- a) 在实验室中使用标准污垢测试,专家评价员使用10点清洁度标尺对清洁效果进行评价,得出2种洗涤剂具有相同清洁性能的标准是评分差异不能超过0.5。
- b) 消费者在家中连续2周使用A品牌和B品牌洗涤剂后,对洗涤剂清洁效果进行比较。采用10点清洁度标尺时,评分差异不能超过0.5。

在实验室准备相同污垢的标准衣物样品,对A品牌与B品牌的清洁性能进行评价。先使用标准用量的A品牌洗涤剂清洁一批衣物上的污垢,再使用相同标准量的B品牌洗涤剂处理具有相同污垢的另一批衣物。公司的10名专家评价员在不知道使用了哪种洗涤剂的前提下检查清洗后的衣服。B品牌洗涤剂的清洁度均值为8.1,A品牌清洁度均值为7.9,成对评分的标准误差为0.4。等效性检验中,双向单侧检验(two one-sided tests, TOST)和等效界值 $\delta=0.5$ 时, $p$ 值为0.0201( $<0.001$ )。基于TOST方法的2个 $p$ 值均低于5%的显著性水平,感官团队得出结论是2种洗涤剂的性能非常相似。

150名消费者居家使用了A品牌和B品牌的洗涤剂。以2个品牌洗涤剂的真实用户作为测试的代表性用户。用户首先连续2周使用其惯常使用的1种洗涤剂,完成洗涤剂性能评价的简短问卷。在使用了惯用洗涤剂一周后,消费者会收到第2种洗涤剂,继续居家使用2周,并完成第2份洗涤剂性能评价问卷。产品使用顺序遵循平衡原则,并采用三位数随机编码。消费者对使用A品牌洗涤剂清洗的衣物整体清洁度评分均值为7.4(满分为10),B品牌的评分均值为7.7,差异的标准误差为1.4。如前所述,使用TOST法成对 $t$ 检验,等效界值 $\delta=0.5$ 时,得到 $p=0.041$ 和 $p<0.001$ 。由于2个 $p$ 值都低于5%的显著性水平,基于TOST方法(不需要多重校正)和预设的非劣效性团队得出结论:A品牌具有与B品牌相同的性能。

A公司研究人员最后得出结论是B品牌并没有表现出更优。同时,A公司决定通过销售文案、路演和电视媒体等多种形式向零售商传达“清洁效果不逊色于其他品牌的衣物洗涤剂”的信息。

注:《中华人民共和国广告法》第十三条提及,广告不得贬低其他生产经营者的商品或者服务。本示例可能不适用中国地区的广告宣称。

### B.6 比较性——优越性——情感宣称:“消费者搭配汉堡时更偏爱的番茄酱味道”

国外一家调味品公司开发并推出了1款口味更丰富、比市场上其他品牌番茄酱更百搭的番茄酱。几家大型快餐连锁店都选用这款番茄酱产品作为汉堡的标配调味酱。调味品公司在这款番茄酱研发期间进行了大量消费者感官测试,认为这款番茄酱的味道比市场上另一家领先品牌番茄酱的味道更胜一筹。这2个品牌番茄酱占全国销售的87%。

调味品公司决定开展这款番茄酱和另一家领先品牌番茄酱的消费者偏好测试。团队制定的决策标准是：使用具有地域代表性的消费者群体对 2 个品牌番茄酱进行成对偏好盲测，确定该款番茄酱的偏好。招募了 213 名消费者，在全国 4 个不同地点使用汉堡包和番茄酱进行测试，设置  $\alpha=0.05$  和  $\beta=0.1$  (90%) 的检验效力来检测单边偏爱测试中  $p_d=0.2$  ( $p=0.6$ ) 的差异。4 个测试地点均为主要人口分布的代表性地区，参与测试的都是经常吃搭配番茄酱汉堡的消费者。给每位消费者 2 份样品，每份样品包括半个汉堡，并涂上不同品牌的番茄酱。2 种番茄酱的品尝顺序遵循平衡原则，询问消费者更偏爱哪种番茄酱的味道。

测试结果表明，共有 123 人更偏爱该款番茄酱，89 人更偏好领先品牌。单边二项式检验  $p$  值为 0.010，置信下限为 0.52。结果符合决策标准，可以得出“消费者更偏爱新研发番茄酱”的结论。这家调味品公司决定在印刷品和电视广告中宣称“消费者搭配汉堡时更偏爱的番茄酱味道”。

注：《中华人民共和国广告法》第十三条提及，广告不得贬低其他生产经营者的商品或者服务。本示例可能不适用中国地区的广告宣称。

### B.7 比较性——优越性——性能宣称：“比其他同等价位的睫毛膏更少结块”

某国家一家化妆品公司新推出 1 款睫毛膏，希望宣称“比其他同等价位的品牌更少结块”。这款睫毛膏含有 1 种特殊成分，能粘附在睫毛上但几乎不结块。因此，公司希望这个宣称能获得强有力的技术支持。品类销售数据显示，该睫毛膏与其他 3 个领先品牌的睫毛膏价格相当。

该公司制定的达成标准是，一个由 12 名经过训练的评价员组成的评价小组对该公司睫毛膏产品的结块强度进行评价，结块强度应显著低于其他 3 款价格相当的睫毛膏产品。

该公司招募了 16 名有经验的消费者为被试，每位被试每天按照标准程序涂抹 1 种睫毛膏。评价员对每位被试的睫毛进行目视评价，采用 10 点标度对结块强度评分，1 分为“无结块”，10 分为“极其结块”。被试者和评价员都不知道使用和评价的睫毛膏品牌，不同品牌的涂抹顺序遵循平衡原则。测试期为 3 天，评价员在涂抹睫毛膏的 2 h 内对每位被试进行评价。结果如表 B.2 所示。

表 B.2 评价结果

参数	新睫毛膏	品牌 B	品牌 C	品牌 D
结块强度均值	1.8	5.4	3.8	6.2
$p$ 值(与新产品的比例比较)	—	0.004	0.022	<0.001

对结块强度数据进行三因素方差分析，包括产品、受试者和评价员 ( $F_{\text{产品}}=3.4$ ,  $df=3$  和 738,  $p=0.017$ )。进一步，采用单边  $t$  检验及事后检验对新产品的结块强度均值与其他 3 种产品进行两两比较，结果表明该公司新睫毛膏的结块强度显著低于其他 3 种品牌产品。

根据研究结果，该公司决定在其所在国针对 3 个竞品提出宣称。

注 1：此处不需要对多重性进行修正，因为只有所有的成对比较结果都具有统计显著性时，才会进行类别范围的宣称。该示例是公司针对 3 个竞品进行测试，并根据测试结果决定具体针对哪些竞品提出宣称，与上述情况不同。上述情况是根据显著性结果作出宣称，因此需要多重校正，例如 Dunnett's 的多重比较。

注 2：《中华人民共和国广告法》第十三条提及，广告不得贬低其他生产经营者的商品或者服务。本示例可能不适用中国地区的广告宣称。

### B.8 比较性——感官特性宣称：“苦味更少”或“烘焙味更浓郁”

一家咖啡公司改变了加工工艺，生产出一种苦味和酸味更少、烘焙风味更浓郁的咖啡。该公司希望有证据支持其在包装上标识“苦味更少”或“烘焙味更浓郁”的声称，从而向消费者强调这些变化。感官研究人员向团队建议，描述性分析小组精通评价不同咖啡的感官特性。支持该宣称的达成标准是，训练



过的描述性感官分析小组能证实采用当前工艺和新工艺生产的咖啡在苦味或烘焙味强度上存在差异(符合宣称内容)。使用 15 cm 线性标度测定时,强度差异均值不少于 1 cm,前期研究认为 1 cm 代表消费者响应存在明显差异。整体显著性水平设为 5%。

感官研究中采用了 11 位经过训练的评价员评价 2 种咖啡(当前工艺和新工艺)的 5 个关键感官特性:咖啡的整体印象、烘焙味、酸味、甜味和苦味,支持宣称的测试重点是烘焙味和苦味。早期咖啡喜好和感官评价研究表明,这 5 个感官特性与消费者对咖啡的喜好最为相关。以标准方式冲泡 2 种咖啡,评价员使用 15 cm 线性标度评价每种咖啡 5 种感官特性的强度,从“无”到“极其”,采用盲评方式,重复 2 次。表 B.3 中列出了强度均值及其单边  $p$  值。

表 B.3 强度评分均值和单边  $p$  值

参数	新工艺制咖啡	当前工艺制咖啡	$p$ 值	$p$ 值(Bonferroni 校正)
咖啡整体印象	8.8	8.1	0.092	—
烘焙味	7.9	6.4	0.016	0.032
酸味	4.7	6.1	0.006	—
甜味	5.6	5.3	0.538	—
苦味	7.1	7.9	0.037	0.074

将产品、评价员以及产品和评价者的交互作为自变量,对数据进行方差分析,5%显著性水平下对产品进行单边检验。由于宣称是由检验结果决定的,因此需要对多重性进行校正。采用 Bonferroni 校正对烘焙味和苦味这 2 个关键特性进行校正。表 B.3 提供了校正后的  $p$  值(乘以 2,测试宣称中的 2 个特性),或者将主要关注的 2 个  $p$  值(未校正)与减半的显著性水平 0.025 进行比较。

结果表明,新工艺咖啡的烘焙味强度显著增强(强度差均值为 1.5 cm)。但多重校正后新工艺咖啡的苦味并未显著低于当前工艺咖啡(校正后  $p$  值=0.074>0.05,未校正的  $p$  值=0.037>0.025),强度差均值仅为 0.8 cm。因此,测试结果不支持“苦味更少”的宣称,但支持“烘焙味更浓郁”的表述。营销部门开始准备带有“烘焙味更浓郁”的新包装。

**附录 NA**  
(资料性)  
**比较性感官宣称示例**

**NA.1 比较性——卓越性——烹饪性能宣称：“嫩度不逊色于中国烹饪大师”**

一家电公司新推出了一款电压力烹饪器具,基于产品的技术创新,产品的烹饪性能大幅提升。公司希望在产品包装上标识“嫩度不逊色于中国烹饪大师”,以向消费者强调新产品的烹饪性能。

制定达成该宣称的标准是:

- a) 在感官评价测试中,12名感官评价员使用10点线性标度对电压力烹饪器具和中国烹饪大师烹饪的黄焖鸡嫩度进行评价,得出两种烹饪方式下嫩度评分均值差异不超过0.5。
- b) 120名以上的消费者,使用10点标度对电压力烹饪器具烹饪的黄焖鸡与厨师自主烹饪的黄焖鸡嫩度进行评价,嫩度差异均值不超过0.5。

准备2份相同的黄焖鸡食材,其中1份直接放置电压力烹饪器具中,采用智能烹饪程序进行烹饪,另外1份由中国烹饪大师自主烹饪。协调好2种烹饪方式的时间,保证2份黄焖鸡均在出品后5 min内完成感官评价。

12名经过培训的专业感官评价员组成的评价小组,对上述2种烹饪方式的黄焖鸡进行嫩度评价,品评顺序遵循平衡原则。电压力烹饪器具烹饪的黄焖鸡嫩度均值为8.9,中国烹饪大师烹饪的黄焖鸡嫩度均值为9.0,成对检验的标准误差为0.4。等效性检验中,等效界值 $\delta=0.5$ 时,双单边检验 $p$ 值为0.016,低于5%的显著性水平。因此,该家电公司感官研究团队得出结论:电压力烹饪器具烹饪的黄焖鸡与中国烹饪大师烹饪的黄焖鸡具有相似嫩度。

120名消费者对2份黄焖鸡进行盲测,2种烹饪方式黄焖鸡的品尝顺序遵循平衡原则。品尝后2 min内记录黄焖鸡嫩度的10点标度,其中电压力烹饪器具嫩度均值为9.1,中国烹饪大师自主烹饪嫩度均值为9.3。2种不同烹饪方式得到的黄焖鸡嫩度均值差值为0.2( $<0.5$ ),成对检验的标准误差为0.6。如前所述,等效界值 $\delta=0.5$ 时, $p$ 值为0.02( $p<0.05$ )。故该家电公司得出结论:电压力烹饪器具的烹饪嫩度不逊于中国烹饪大师水平。

该家电公司最后得出结论:电压力烹饪器具符合宣称的达成标准,可以在销售文案、路演和电视媒体等媒介上向消费者宣传该电压力烹饪器具“嫩度不逊色于中国烹饪大师”的信息。

**NA.2 比较性——优越性——性能特性宣称：“持久留香,香气更强”**

一家洗护公司有1款香氛洗发露产品A,这款产品“持久留香”,一直深受消费者好评。公司近期推出了1款新香氛洗发露产品B,由于采用了新配方,产品B不仅香气持久,且香气比A更强。因此,公司希望宣称与A相比,“持久留香,香气更强”。

公司感官部门和研发部门协商后,制定了决策标准:感官评价小组使用5点类项标度,对2款产品洗过的头发香气强度进行评价。测试时间点为刚洗完头(洗完0 h)、洗完24 h和洗完48 h。如果48 h后评价员还能闻到香气,则认为产品香气持久;3个测试时间点的香气强度值均大于0.5,且B产品洗过的头发香气强度均显著高于A产品,则认为产品B香气更强。

感官评价小组由7名优选感官评价员组成。评价员均进行过同类型洗护产品香气强度评价的培训、考核和测试工作,具备良好的感官评价能力。在专业感官实验室准备了相同的标准头发样品,采用标准洗发程序,用A和B产品对头发样品分别进行清洗。同时,选择了1款非香氛型洗发露产品作为阴性对照,并采用相同头发样品和洗发程序进行清洗。清洗后的头发样品均用吹风机吹干,进行香气强度评价,每个产品重复2次。在24 h和48 h后重复进行香气强度测试,产品评价顺序遵循随机原则。

测试结果如表 NA.1 所示,产品 A 和 B 在洗发后 0 h、24 h 和 48 h 后评价员均能感受到香气,香气强度值均在 0.5 以上,表明 2 款香氛洗发露产品均具有香气持久的特性。成对比较结果表明,3 个测试时间点,产品 B 的香气强度均显著高于产品 A。阴性对照样品 48 h 后已经闻不到香气。因此,团队可以得出结论,B 产品与 A 产品一样持久留香,且 B 产品的香气更强。

公司决定采用“持久留香,香气更强”的宣称,向消费者和经销商进行 B 产品的宣传。

表 NA.1 评价结果

参数	香气强度		F 值
	产品 A	产品 B	
0 h	1.5±0.26	2.7±0.07	99.249*
24 h	1.1±0.09	2.1±0.11	196.244*
48 h	0.6±0.09	0.9±0.14	40.345*

\* 0.05 水平下产品 A 和 B 存在显著差异。

### NA.3 比较性——等效性——感官特性宣称：“麻味同样强劲”

一家调味品公司有 1 款畅销花椒油产品,一直以“麻味强劲”受到经销商和消费者的喜爱。该公司新研发了一种花椒油加工工艺,降低了产品苦味,同时公司希望新工艺没有降低产品麻度,即宣称“麻味同样强劲”,从而向经销商表明新工艺红花椒油的麻度与原配方产品相同。

公司感官团队和研发团队决定采用单边相似性检验,达成该宣称的标准是:设置 $\beta=0.05$ , $\alpha=0.5$ ,同时能够检验出差异的感官评价员比例不超过 20% ( $P_d=0.2$ )。

确定标准后,公司开展红花椒麻度的感官测试工作。根据 GB/T 12310—2012,确定了本测试需要至少 78 个评价员。招募了 78 位感官评价员开展测试。测试方法是将原工艺和新工艺加工的 2 种红花椒油与白灼鸡丝拌匀,评价员品尝加了红花椒油的鸡丝,评价哪个样品麻度更强。样品品尝顺序遵循了随机原则。

测试结果表明,78 位评价员中,共有 25 位评价员指出拌有新工艺红花椒油的鸡丝“更麻”。根据 GB/T 12310—2012 中的表 A.3,当 $n=78$ , $P_d=0.2$ , $\beta=0.05$ 时,推断样品相似的最大值为 39。检测中得到的正确答案数小于该最大值,表明新工艺红花椒油产品与原工艺花椒油产品具有相似的麻度。

因此,公司可以向经销商宣称,新工艺红花椒油产品“麻味同样强劲”。

### NA.4 比较性——优越性——情感宣称：“消费者更喜爱的味道”

一家乳制品公司想知道旗下新推出的同品牌下 3 款不同类型的纯牛奶产品是否存在消费者喜好差异,从而向经销商宣称其中产品 A 是 3 款产品中“消费者更喜爱的味道”。公司的研发、感官和市场团队一起,开展了这 3 款产品的消费者喜好测试。公司联合团队制定的达成标准是:选用具有地域代表性的消费者群体,采用 9 点喜好标度对 3 款产品进行喜好测试。产品 A 的喜好均值显著高于产品 B 和 C ( $p<0.05$ ),且喜好均值差异 $>1.0$ 时,表明产品 A 是 3 款产品中消费者更喜欢的。

公司选择了 4 个城市,分别是西部、北部、南部和中部的代表性地区和主要人口的分布城市。每个城市被试数量为 90 人~95 人,均为日常喝纯牛奶的消费者(喝奶频次不少于 2 次/周)。测试条件符合 GB/T 41408 中的要求,消费者品尝 3 个样品,样品采用盲评方式,样品编号采用 3 位随机编码,递送顺序遵循随机原则。消费者品尝样品后,采用 9 点喜好标度进行评分。

3 个产品的喜好测试结果如表 NA.2 所示。采用成对  $t$  检验比较 3 个产品之间的喜好均值差异。测试结果表明,4 个城市的产品 A 的喜好均值显著高于产品 B ( $p<0.05$ ),但与产品 C 没有显著差异

( $p > 0.05$ )。进一步分析发现,除了城市 2 以外,其余 3 个城市测试结果中,纯牛奶 A 与 B 的喜好均值差异均在 1 以内。因此,测试结果不符合达成标准,公司决定不进行相关宣称。

表 NA.2 评价结果

地区	纯牛奶 A	纯牛奶 B	纯牛奶 C
城市 1	7.09±1.38	6.22±1.51	7.00±1.06
城市 2	6.57±1.32	5.20±1.55	6.24±1.24
城市 3	6.81±1.42	6.12±1.36	6.62±1.47
城市 4	6.39±1.32	5.43±1.47	6.25±1.30

参 考 文 献

- [1] GB/T 12310—2012 感官分析方法 成对比较检验
  - [2] ISO/IEC Guide 51 Safety aspects—Guidelines for their inclusion in standards
  - [3] ISO 6658 Sensory analysis—Methodology—General guidance
  - [4] ISO 11136 Sensory analysis—Methodology—General guidance for conducting hedonic tests with consumers in a controlled area
  - [5] ASTM E1958 Standard guide for sensory claim substantiation
  - [6] Declaration of Helsinki. Bulletin of the World Health Organization, 2001,79(4),p. 373.
  - [7] Corbin R M. et al. A Practical Guide to Comparative Advertising; Dare to Compare. Elsevier, 2018.
  - [8] Martin G. Munchausen's statistical grid, which makes all trials significant. The Lancet, 1984,324(8417-8418):p.1457
  - [9] 中华人民共和国广告法
  - [10] ABC Television Network. Advertising Standards and Guidelines; <https://abcaccess.com/app/uploads/2016/01/2014-Advertising-Guidelines-.pdf>
  - [11] Advertising Standards of Canada. Guidelines for the Use of Comparative Advertising; Guidelines for the Use of Research and Survey Data to Support Comparative Advertising Claims,2010
  - [12] ASAI. Code of Standards for Advertising and Marketing Communications in Ireland. Advertising Standards Authority for Ireland (ASAI),Dublin,Ireland
  - [13] Committee of Advertising Practice (CAP):[www.cap.org.uk](http://www.cap.org.uk)
  - [14] NAD;<http://www.asrcreviews.org/asrc-procedures>
  - [15] NBC Universal Advertising Standards. Advertising Guidelines; [https://nbcuadstandards.com/files/NBC\\_Network\\_Advertising\\_Guidelines.pdf](https://nbcuadstandards.com/files/NBC_Network_Advertising_Guidelines.pdf)
  - [16] Schneider-häder B.,Hamacher E.,Beeren C. Sensory Claims—Methodological approach to development and substantiation. DLG-Expert report 15/2015. DLG e.V.,Frankfurt,2015. Available from: [https://www.dlg.org/fileadmin/downloads/lebensmittel/themen/publikationen/expertenwissen/lebensmittelsensorik/e\\_2015\\_15\\_Expertenwissen\\_SensoryClaims.pdf](https://www.dlg.org/fileadmin/downloads/lebensmittel/themen/publikationen/expertenwissen/lebensmittelsensorik/e_2015_15_Expertenwissen_SensoryClaims.pdf)
  - [17] UK. Advertising Standards Authority (ASA) and Committees of Advertising Practice (CAP) guidelines; <https://www.asa.org.uk/news/a-quick-guide-to-comparative-advertising.html#.VVDGXvCy7-I> with specific codes at;<https://www.asa.org.uk/codes-and-rulings/advertising-codes.html>
-