



中华人民共和国国家标准

GB/T 19547—2025/ISO 11056:2021

代替 GB/T 19547—2004

感官分析 方法学 量值估计法

Sensory analysis—Methodology—Magnitude estimation method

(ISO 11056:2021, IDT)

2025-10-05 发布

2026-05-01 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

引言 IV

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 原理 2

5 基本测试条件 2

6 感官评价员的选拔和培训 3

 6.1 选拔和培训的基本条件 3

 6.2 量值估计法的专项内容培训 3

7 感官评价员的数量 4

 7.1 一般原则 4

 7.2 感官评价小组 4

 7.3 消费者评价小组 4

8 程序 5

 8.1 样品提供 5

 8.2 外部参比样 5

 8.3 样品提供的顺序 5

 8.4 量值估计 5

9 数据分析 6

 9.1 数据分析方法的选择 6

 9.2 原始数据的提供 6

 9.3 产品差异的确定 6

 9.4 回归分析 6

 9.5 重新标度的方法 7

10 测试报告 7

附录 A (资料性) 问卷样式 9

 A.1 无固定模数参比样的问卷样式 9

 A.2 有固定模数参比样的问卷样式 9

附录 B (资料性) 数据分析示例 10

 B.1 无重复且未经重新标度数据的数据分析——所有评价员一次评价完该系列的所有产品 10

 B.2 无重复且有内部重新标度数据的数据分析——所有评价员一次评价完该系列的所有产品 13

 B.3 对子组样品进行内部重新标度数据的数据分析 15

 B.4 外部重新标度数据的数据分析 18

B.5 有重复数据的数据分析 21

B.6 回归曲线斜率的确定(史蒂文斯幂函数,见 9.4)..... 24

参考文献 28

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。

本文件代替 GB/T 19547—2004《感官分析 方法学 量值估计法》，与 GB/T 19547—2004 相比，除结构调整和编辑性改动外，主要技术变化如下：

- a) 增加了本文件的适用范围(见第1章)；
- b) 更改了术语“外部参比样”和“内部参比样”的定义(见 3.2、3.3, 2004 年版的 3.2、3.5)；
- c) 将“量值估计法的专项内容培训”中“注”更改为正文内容(见 6.2.4, 2004 年版的 6.2.4)；
- d) 更改了消费者评价小组的人数(见 7.3, 2004 年版的 7.3)；
- e) 更改了“数据分析方法的选择”中的内容(见 9.1, 2004 年版的 9.1)；
- f) 将“原始数据的提供”中“注”更改为正文内容(见 9.2, 2004 年版的 9.2)；
- g) 将“经参比样的重新标度法”中“注”更改为正文内容(见 9.5.2, 2004 年版的 9.5.2)；
- h) 增加了“测试的日期”(见第10章)。

本文件等同采用 ISO 11056:2021《感官分析 方法学 量值估计法》。

本文件做了下列最小限度的编辑性改动：

- a) 增加了术语“外部参比样”和“内部参比样”的注。

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、安利(上海)科技发展有限公司、黑龙江飞鹤乳业有限公司、厦门元之道生物科技有限公司、百特威(上海)化妆品有限公司、广东健力宝股份有限公司、上海大学、青岛圣桐营养食品有限公司、深圳市新荣阳食品科技有限公司、江苏创健医疗科技股份有限公司、北京科技大学、中国烟草总公司郑州烟草研究院、内蒙古伊利实业集团股份有限公司、河北中烟工业有限责任公司、四川丁点儿食品开发股份有限公司、中国热带农业科学院香料饮料研究所、水羊集团股份有限公司、上海康识食品科技有限公司、北京安德普泰医疗科技有限公司、灏图科技(上海)有限公司、一苇堂检测科技(上海)有限公司、石家庄市惠康食品有限公司、山东固本堂健康产业开发集团股份有限公司、四川源坤教育科技有限公司、光明乳业股份有限公司、泸州老窖股份有限公司、杭州浪澄生物科技有限公司、浙江丰岛食品股份有限公司、四川米老头食品工业集团股份有限公司、茶姬(上海)品牌管理有限公司、舍得酒业股份有限公司、陇南利和萃取科技有限公司、中国农业科学院茶叶研究所。

本文件主要起草人：钟葵、赵镭、李洪亮、史波林、侯姣靓、刘桂荣、温焱、汪厚银、安志丛、项雅科、费雅君、李懿霖、赵菲菲、黄君阳、庄宝霞、郑群、唐素婷、高海燕、秦臻、徐庆利、孙延琳、石叶飞、支瑞聪、乔学义、朱丹晔、苏玉芳、冯文宁、罗强祖、廖欣、董文江、张廷志、张顶武、赵春江、伊宇锋、李明远、蒋莉、李永新、李聚强、裴庆华、杨宇、张锋华、熊燕飞、徐行行、唐莉明、苗宇、李一凡、李青松、赵楠、许勇泉。

本文件及其所代替文件的历次版本发布情况为：

——2004 年首次发布为 GB/T 19547—2004；

——本次为第一次修订。

引言

量值估计法是一种心理物理学标度技术,感官评价员对样品某一感官特性强度进行量值估计,赋以相应的数字。该方法要求评价员给出的估计值要符合比例原则,即若样品 B 某一特性强度是样品 A 的 2 倍,则样品 B 的赋值是样品 A 赋值的 2 倍。样品强度、愉悦度和可接受性等特性都可使用量值估计法进行评价。

相比由测试人员定义的连续或非连续响应标度法,量值估计法更不易受到“末端效应”的影响。当评价员对产品所激发感觉的强度范围不熟悉时,容易发生“末端效应”。于是,评价员可能会将某个初始测试样品的赋值归类到靠近标度端点的类别中,结果发现,由于端点附近的类别缺乏足够的细分刻度,致使感知上存在差别的样品会被归到同一类别。使用量值估计法可避免这种情况,因为其理论上,各个类别都明确对应不同的数值。

每位评价员自行选择任一数字(即评价员使用自己的标度)进行赋值,这会引发一个极为重要的“评价员”效应。然而,针对这一问题,有多种解决方法可供选择:

- 方差分析(ANOVA)可分析评价员效应以及其与其他因子的交互效应;
- 使用一个已赋值的参比样,使评价员使用一个共同的标度范围;
- 重新标度,将每个评价员给出的数据转化为一个共同的标度范围。

由测试负责人依据实际情况来选择最合适的方法。

量值估计法不仅是确定史蒂文斯心理物理学幂函数的首选方法,还能用于解决具体的问题。

注:在区别差异微小的刺激或者在觉察阈附近进行评价时,量值估计法并非是最有效的方法。

示例 1:某公司生产一款市场畅销的饮料产品,但近期其竞争对手推出了一款甜度更高的产品,抢占了该公司产品的一部分市场份额。为了夺回部分市场损失,该公司决定将产品甜度提高三分之一。在产品新配方的研发过程中,若已知甜味剂的幂函数,将有助于估算将甜度提高三分之一时所需的甜味剂量。

示例 2:低糖饮料新配方的研发过程中,已确定饮料所需的甜味强度,但尚未决定使用阿斯巴甜还是蔗糖作为甜味剂。在已知两种甜味剂幂函数的基础上,就可绘制两者的等甜度曲线,从而确定达到所需甜度水平时两种甜味剂分别所需的浓度。结合这些信息与成本/产量信息进行综合分析,即可帮助决定哪种甜味剂更具成本效益。

附录 B 使用 R 软件进行数据计算。R 软件包是免费访问,这些信息是为方便本文件用户使用而提供,并不构成本文件对 R 软件包独家使用的认可或推荐。用户也能使用其他软件执行本文件所需的计算。由于使用软件不同,数据有时因四舍五入而导致结果有细微差异。

感官分析 方法学 量值估计法

1 范围

本文件描述了使用量值估计进行感官特性评价的方法,包括感官评价员的培训、估计值的获取和统计数据解释。

本文件适用于产品的感官特性强度、愉悦度和可接受性等特性的评价。

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 3534-1 统计学 术语和符号 第1部分:一般统计术语与用于概率的术语(Statistics—Vocabulary and symbols—Part 1: General statistical terms and terms used in probability)

注: GB/T 3358.1—2009 统计学词汇及符号 第1部分:一般统计术语与用于概率的术语(ISO 3534-1:2006, IDT)

ISO 3534-3 统计学 术语和符号 第3部分:实验设计(Statistics—Vocabulary and symbols—Part 3: Design of experiments)

注: GB/T 3358.3—2009 统计学词汇及符号 第3部分:实验设计(ISO 3534-3:1999, IDT)

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)

注: GB/T 39501—2020 感官分析 定量响应标度使用导则(ISO 4121:2003, IDT)

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

注: GB/T 10221—2021 感官分析 术语(ISO 5492:2008, IDT)

ISO 6658 感官分析 方法学 总论(Sensory analysis—Methodology—General guidance)

注: GB/T 10220—2012 感官分析 方法学 总论(ISO 6658:2005, IDT)

ISO 8586 感官分析 感官评价员的选拔和培训(Sensory analysis—Selection and training of sensory assessors)

注: GB/T 46555—2025 感官分析 感官评价员的选拔和培训(ISO 8586:2023, IDT)

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)

注: 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)

3 术语和定义

ISO 3534-1、ISO 3534-3、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 电工百科:<http://www.electropedia.org/>。

3.1

量值估计 magnitude estimation

对产品某一特性强度或被喜好程度进行赋值的过程,不同赋值间的比例与感官评价员对感知特性间的比例保持一致。

3.2

外部参比样 external reference

样品系列中的首个样品作为参比,后续所有测试样品均以该样品为参比进行评价。

注:外部参比样并非测试样品组中的样品。

3.3

内部参比样 internal reference

样品系列中首个样品作为测试样品提供给评价员,后续测试样品均以该样品为参比进行评价。

注:内部参比样是测试样品组中的一个样品。

3.4

模数 modulus

由测试负责人给定的外部参比样(3.2)的赋值(固定模数),或由评价员自行选择的外部参比样(3.2)的赋值(非固定模数)。

3.5

重新标度 rescaling

将每个评价员给出的原始数据分别乘以一个校正因子,从而将所有评价员的数据纳入到同一标度范围的过程。

注:先将校正因子与原始数据分别取对数值,再进行加和,其效果与上述乘法过程等效。

3.6

史蒂文斯心理物理幂函数 Stevens's equation psychophysical power function

其关系如下所示:

$$R = KS^n$$

式中:

R ——评价员的响应(如感知的强度);

K ——常数,用于表征 R 和 S 之间的关联关系;

S ——刺激量(物理量或化学物质的浓度);

n ——幂函数指数,为 $\log R$ 与 $\log S$ 回归曲线的斜率。

注:实际应用中,史蒂文斯心理物理幂函数一般转化成自然对数表示; $\ln R = \ln K + n \ln S$ 。

4 原理

将样品连续提供给感官评价员,评价员遵循比例原则对每个样品的某一特性强度进行赋值并记录。评价员依据组内首个样品(参比样)的数值,对后续样品进行评价。参比样的赋值由每个评价员自由赋值或由测试负责人指定一个数值,后一种情况被称为“固定模数”。

5 基本测试条件

对于基本测试条件,如设施、样品的准备、提供和编码等方面,应遵循 ISO 6658、ISO 8589 等感官分析通用方法学的相关标准以及 ISO 4121 中所描述的标度和分类方法。

6 感官评价员的选拔和培训

6.1 选拔和培训的基本条件

感官评价员的选拔和培训应遵循 ISO 8586 中的要求。和其他感官分析方法一样,由感官评价小组组长负责确定评价员所需的资质水平。制定培训计划时应依据测试的目的、评价员的可用性以及召集和培训新评价员所需的经费。通常情况下,评价员在完成 3 次~4 次测试培训后,即可熟练运用量值估计法开展感官评价工作。

6.2 量值估计法的专项内容培训

6.2.1 几何图形面积估计,已被证明是引导感官评价员掌握量值估计法基本概念的有效方法。表 1 给出的 18 个几何图形已被成功应用于感官评价员的培训中,包括 6 个圆形、6 个等边三角形和 6 个正方形,其表面积范围为 $2\text{ cm}^2\sim 200\text{ cm}^2$ 。

对于消费者评价小组,可采用简化版培训内容,如将培训内容范畴限定在图形面积估计方面。

表 1 培训图形的边长和面积

圆形		等边三角形		正方形	
半径/cm	表面积/ cm^2	边长/cm	表面积/ cm^2	边长/cm	表面积/ cm^2
1.4	6.2	2.2	2.1	3.2	10.2
2.5	19.6	4.1	7.3	4.2	17.6
3.7	43.0	7.6	25.0	8.5	72.3
5.4	91.6	12.2	64.4	11.1 ^a	123.2
6.8	145.3	15.5	104.0	11.1 ^a	123.2
8.3	216.4	19.2	159.6	14.2	201.6
^a 设置两个边长均为 11.1 cm 的正方形主要用于评价感官评价员的再现性。					

6.2.2 将这些图形提供给评价员之前,先向评价员介绍量值估计法的基本原理。原理中包括但不限于下列三点:

- 赋值应依据比例原则:若感知到样品某一特性强度是参比样的 2 倍,则其强度赋值也是参比样的 2 倍;
- 使用的标度无上限值;
- 当某一特性感知不到时,赋值为 0。

需提醒评价员,尽管在培训过程中通常使用整数(如 5、10、20、25 等)进行赋值,但在实际使用量值估计法时,所有的数字均被允许使用。由于评价员也会受到培训中所提及比例的影响,因此建议他们在实际操作时就使用不同的比例,如 3/1、1/3、7/5、5/6 等,以避免仅局限于使用 2/1 或 1/2 这类单一的比例形式。

6.2.3 将所有图形进行编号,并分别放置在一张白色 A4 纸(尺寸约为 $21\text{ cm}\times 29.7\text{ cm}$)的正中央位置。从边长为 8.5 cm 的正方形(外部参比样)开始,指导每位评价员依次对这一系列图形的面积响应进行赋值并记录。依据测试阶段所使用的程序,对评价员开展有固定模数或无固定模数的相关培训。在进行固定模数培训时,由测试负责人对边长为 8.5 cm 的正方形面积赋予一个介于 30 到 100 之间的数值。在进行非固定模数培训时,由评价员对首个图形进行自主赋值,但建议其避免选择过低的数值。

每次测试时,以随机顺序提供这些图形,以确保图形的形状和尺寸不会形成某种特定的顺序模式。

6.2.4 在评价员完成一组图形的估值后,让其将各自的赋值结果与评价小组均值进行比较。若实际操作中难以实现与小组均值的比较,则将赋值结果与前一个评价小组的赋值结果进行比较。这些操作旨在提供具有积极意义的反馈信息,确保每位评价员清晰理解培训工作的目的。宜避免让评价员产生存在所谓“正确”答案的印象。除非评价员的赋值结果存在很大差异,否则宜将评价员结果与评价小组结果的偏离解释为顺序效应,即评价员的响应值会受到评价样品递送顺序的影响。同时,宜使评价员相信,尽管个别评价员会受到顺序效应的影响,但整个评价小组的结果仍然是准确的。

如果部分评价员的赋值结果差异很大,应重新向这些评价员详细解释量值估计法的基本原则。

6.2.5 在评价员顺利完成面积估计值的培训后,宜根据实际测试的产品和物质类型,开展进一步的培训工作,这有助于评价员积累更多关于测试样品特性量值估计的经验。评价小组组长宜再设计一些练习活动,使评价员能正确识别待测特性。这部分培训工作可根据 ISO 8586 中给出的一般导则进行设计。

7 感官评价员的数量

7.1 一般原则

与其他标度技术方法一样,感官评价员的数量取决于下列方面:

- 不同测试产品特性之间的相似程度;
- 评价员所接受的培训状况;
- 该评价结果所得结论的重要性程度(见 ISO 8586);
- 确定目标时所依据的统计功效。

若尚未有明确目标时,参照 7.2 和 7.3 中的相关内容。

7.2 感官评价小组

感官评价小组的组成应依据表 2 中的要求进行安排。从个体评价的方差结果和需检测出的差异大小两个方面,来解决统计功效的问题。

表 2 感官评价小组的组成

评价员的类型	评价员最少数量	推荐数量
有经验的评价员,接受过待测产品及特性的严格培训	5	10
有经验的评价员,接受过待测产品及特性的相关培训	15	20~25
新培训的评价员	20	20 及以上

7.3 消费者评价小组

量值估计法也能使用消费者评价小组,如用于市场调查研究。消费者的数量应根据具体测试类型所对应的消费者抽样要求予以确定。在所需评价员数量方面,量值估计法并无明显优势。其消费者数量应与典型消费者测试的要求保持一致,即通常不少于 60 人,在某些情况下可能还需要更多。

8 程序

8.1 样品提供

所有样品应使用相同的方式提供(如使用统一的盛装容器以及保持一致的样品量等)。盛装样品的容器应进行编码,采用随机三位编码方式为最佳。

8.2 外部参比样

对于待测的特性,参比样中该特性的强度宜接近所有待测产品特性强度的几何均值。

注:采用具有某特性极值的参比样,有时会导致评价结果的失真。

在评价员未知的前提下,可在一组待测样品中加入一个或多个随机编码的参比样,以此来评估同一轮次中评价员的重复性。

8.3 样品提供的顺序

样品应一次全部或依次逐一提供给评价员。评价员应按照指定的顺序对样品进行评价。和所有感官测试一样,每个评价员的评价顺序都不相同,理想的样品提供顺序是实现样品间的平衡。

评价小组组长可参照参考文献中的拉丁方表,综合考量平衡顺序设计及其滞后效应^[3]。如果在实际测试中无法采用拉丁方设计,则使用随机顺序提供样品。

8.4 量值估计

8.4.1 一般要求

采用 8.4.2~8.4.4 中所描述的任一方法进行评价。有参比样的问卷样式见附录 A。

8.4.2 无固定模数的外部参比样

评价员先评价外部参比样,对其进行赋值。建议评价员给出的数值不要太低。随后,评价员逐一将其他的编码样品与参比样进行比较,根据其对应参比样的赋值,对待测样品进行赋值。

8.4.3 有固定模数的外部参比样

由评价小组组长指定外部参比样的数值,如 30、50、100,或任何一个评价小组组长认为适宜的数值。随后,评价小组组长指导评价员根据外部参比样的数值,对其他样品进行赋值(固定模数)。

8.4.4 无外部参比样

量值估计法在无外部参比样的情况下也可使用。由于感觉系统(如记忆)的限制,评价员很难始终如一地根据第一个样品对所有样品进行赋值。该情况下,有以下两种解决方法。

- a) 评价员在对后续每个样品赋值时,不必每次重新评价第一个样品。鼓励评价员记住参比样的特性强度,仅在必要时再重新评价参比样。因此,可采用:
 - 测试前:设计一种样品提供顺序,使每个评价员评价的第一个样品都不相同。理想的设计是每个样品都有相同数量的评价员将其作为参比样,从而使样品间平均差异的方差相等;
 - 数据分析时:使用一个具有高权重(理论上趋近于无穷大)的任意常数来评价每个评价员对第一个样品的赋值,以便更正确地估计差异的方差。
- b) 评价员通过将每个样品与其相邻的前一个样品进行比较来开展评价。

注:该方法的问题是导致误差存在自相关性,同时相邻两个样品间差异的方差小于不相邻样品间差异的

方差。

因此,可采用:

- 测试前:设计一种样品提供顺序,使样品所有可能的排列组合都能提供给相同数量的评价员。若实际操作中难以实现,设计一种接近理想模式的样品提供顺序,使样品间平均差的方差相等或相当接近;
- 数据分析时:使用自相关误差模型进行分析,但这种方法较为复杂。

需要说明一点,即使采用 a) 方法(通过与第一个样品即参比样进行比较的方式来系统评价样品),仍需保留与相邻的前一个样品比较评价的自相关误差项,即便该值可能非常小(如 8.4.2 和 8.4.3 中所述的有参比样时的评价结果也证实了这一点)。因此,如前所述,关于均衡样品提供顺序的建议,在所有评价场景下均适用。

9 数据分析

9.1 数据分析方法的选择

分析方法的选择取决于下列因素(见附录 B 数据分析示例):

- 试验设计采用完全设计还是不完全设计;
- 是否进行了重复测试;
- 评价员因子的设定状态(固定因子或随机因子)。样品因子通常设定为固定因子。

9.2 原始数据的提供

测试结果可通过一个两因素表格进行呈现,列代表不同的样品,行代表经对数转化后的评价员的估计值。当所有评价员对每个样品的评价次数都相同时,该设计为完全平衡设计,此时包含评价员效应的模型为正交模型。如果有部分样品并未被所有评价员评价,此时为不完全设计,相应的包含评价员效应的模型为非正交模型。

由于无法对 0 取对数,所以响应数值为 0 时会出现问题。有多种方法处理 0 值的情况。如宜用一个很小的数值来替代 0 值,该数值的选择宜考虑每个评价员所使用的标度(如该评价员给出的最小估计值的一半)。

9.3 产品差异的确定

方差分析可清楚地解释所有的区组因素(包括不平衡设计或非正交因素)。使用经对数转换的数据进行方差分析是最准确的方法。在实际操作中,无法保证每次评价时所有关键因素都是实现了平衡和正交的完全设计。如当一个试验被分成多个轮次开展时,可能无法保证每个测试轮次都由同一组评价员完成。通常建议咨询统计学专家,以制定一个可能的最优实验设计方案。

当方差分析结果显示产品间存在显著差异时,进一步采用均值多重比较法进行分析。采用完全设计(且未进行重新标度的数据)的产品特性差异比较的示例,见附录 B 的 B.1。

9.4 回归分析

当已知相关变量 S 值(如化学浓度、物理量等)与感官响应 R 值存在关联时,可假定遵循史蒂文斯定律,并按公式(1)对该物理变量或化学变量与相应的感官响应进行线性回归分析,并计算模型的参数值:

$$\ln R = \ln K + n \ln S \quad \dots\dots\dots (1)$$

回归分析中,最重要的参数是史蒂文斯定律中 n 值,即回归曲线的斜率。

也可分析不同评价员间回归曲线的斜率是否相等。

9.5 重新标度的方法

9.5.1 全面重新标度法

全面重新标度法的依据如下：由于每个评价员都对同一组样品进行了评价，因此该组样品的总响应量值宜相等。基于此，对于同一组样品，每个评价员所使用的标度总量值也要相等。具体操作流程如下。

对于被全部评价员评价过的所有样品：

- 计算每个评价员的均值，即每个评价员对样品估计值取对数后的均值；
- 计算评价小组的均值，即全部评价员对样品估计值取对数后的均值。

对于每位评价员：

- 计算每位评价员的校正值，即评价小组均值与该评价员均值之差；
- 计算每位评价员的重新标度估计值，即原估计值加上该评价员的校正值。

全面重新标度法的示例见 B.2。

9.5.2 经参比样的重新标度法

当一个或多个随机编码的参比样被插入到测试样品组中，首先计算每个评价员对参比样（首个样品以及可能存在的隐藏参比样）估计值的均值，然后将参比样均值作为一个固定值计算校正值。随后，将评价员对每个样品的估计值乘以由参比样计算得到的校正值，以重新标度测试样品的数据。

需注意进行方差分析时，全面重新标度法得到的均方误差要小于经参比样的重新标度法。如 8.2 中所述，参比样强度估计值宜尽量接近评价小组对所有样品估计值的几何均值。当参比样的强度均值与评价小组估计值的几何均值相等时，此时误差较小^[6]。参比样的强度估计值与实际样品估计值的几何均值越接近越好。

9.5.3 外部重新标度法

不同形式的外部重新标度法见参考文献[5]。测试样品评价完后，提供给评价员一个语义响应标度，标度范围通常为 4 级～11 级，语义描述示例如下：

- 极强；
- 很强；
- 中等强度；
- 稍微有点强，等。

评价小组组长要求评价员对语义描述进行量值估计，不同语义赋值间的比例与评价员对测试样品估计值间的比例相同。随后，再使用全面重新标度法计算每个评价员对语义响应标度估计值的校正值，并使用该校正值对每位评价员的估计值进行重新标度。

外部重新标度法的示例见 B.4。

10 测试报告

测试报告应包括下列内容：

- 研究目的；
- 测试结果；
- 样品数量和样品描述；
- 参比样（若有）和该样品的性质（若测试中有使用）；

- 测试重复次数(若测试中有设定重复);
- 评价员的数量和资质水平;
- 测试的一般条件,如环境、日期和时间;
- 有助于评价测试总体有效性的任何信息;
- 本文件标准编号,若对标准方法有过修改,需对修改内容进行相关说明;
- 测试负责人的姓名;
- 测试的日期。

附 录 A
(资料性)
问卷样式

A.1 无固定模数参比样的问卷样式

无固定模数参比样的问卷样式如图 A.1 所示。

姓名: 日期:

1. 您面前有个编号为“R”的橙汁参比样, 请品尝该样品, 并对样品的酸味强度进行赋值:
.... 参比样“R”酸度值:
.... 记住该样品的强度赋值。

2. 您面前有6个橙汁样品, 请按下列顺序依次评价这6个橙汁样品。参照参比样“R”的酸度值, 按照等比例原则依次对6个样品的酸度进行赋值。每个样品品尝之前, 需要重新品尝一次参比样。

样品561酸度值:
样品274酸度值:
样品935酸度值:
样品803酸度值:
样品417酸度值:
样品127酸度值:

感谢您的参与!

图 A.1 无固定模数参比样的问卷样式

A.2 有固定模数参比样的问卷样式

有固定模数参比样的问卷样式如图 A.2 所示。

姓名: 日期:

1. 您面前有个编号为“R”的橙汁参比样, 该参比样的酸度值为50, 请品尝该样品, 并记住样品的酸度。

2. 您面前有6个橙汁样品, 请按下列顺序依次评价这6个橙汁样品。参照参比样“R”的酸度值(50), 按照等比例原则依次对6个样品的酸度进行赋值。每个样品品尝前, 需要重新品尝一次参比样。

样品561酸度值:
样品274酸度值:
样品935酸度值:
样品803酸度值:
样品417酸度值:
样品127酸度值:

感谢您的参与!

图 A.2 有固定模数参比样的问卷样式

附录 B

(资料性)

数据分析示例

B.1 无重复且未经重新标度数据的数据分析——所有评价员一次评价完该系列的所有产品

B.1.1 示例

7 个经验丰富的评价员对 6 种不同咖啡因浓度饮料的苦味强度进行评价,结果如表 B.1 所示。对于评价员 1 到评价员 3,样品 274 被指定为外部参比样,其强度值为 20;对于评价员 5~评价员 7,样品 803 被指定为外部参比样,强度值为 40;对于评价员 4,样品 935 被指定为外部参比样,强度值为 32。每位评价员均先评价外部参比样,随后再评价其余 5 个样品。5 个样品均以随机顺序提供,每个评价员的样品提供顺序都不相同。评价员并未对外部参比样进行赋值(表 B.1 中外部参比样的评分均由评价小组组长给出),但都品尝了参比样品 5 次以上(每个测试样品评价前,都会先品尝一次参比样),并将其与小组组长给出的强度值进行对应。

表中括号内为 6 个样品赋值的自然对数值,均精确至小数点后三位。

表 B.1 6 个样品的评分数据

评价员	样品编码(咖啡因浓度)					
	561 (9 mg/100 mL)	274 (18 mg/100 mL)	935 (36 mg/100 mL)	803 (40 mg/100 mL)	417 (72 mg/100 mL)	127 (144 mg/100 mL)
	评分值(自然对数值)					
1	10(2.303)	20(2.996)	35(3.555)	40(3.689)	70(4.248)	140(4.942)
2	8(2.079)	20(2.996)	38(3.638)	44(3.784)	85(4.443)	160(5.075)
3	8(2.079)	20(2.996)	36(3.584)	40(3.689)	75(4.317)	150(5.011)
4	7(1.946)	15(2.708)	32(3.466)	37(3.611)	70(4.248)	135(4.905)
5	12(2.485)	25(3.219)	38(3.638)	40(3.689)	75(4.317)	145(4.977)
6	12(2.485)	22(3.091)	35(3.555)	40(3.689)	80(4.382)	160(5.075)
7	9(2.197)	18(2.890)	35(3.555)	40(3.689)	74(4.304)	145(4.977)
自然对数均值	2.225	2.985	3.570	3.691	4.323	4.995

将该示例的数据文件(如位于 F 盘中“ME”文件夹中的文件)通过下列 3 个命令导入 R 中:

```
tableb1<-read.table("F:/ME/Table Annex B1.txt",header = T,sep="\t",dec=".")
attach(tableb1)
names(tableb1)
```

文件 tableb1 包含 7 列(评价员、样品、评分、评分自然对数值、重新标度的评分自然对数值、浓度、浓度自然对数值)和 42 行(6 个样品×7 个评价员)数据。输入以下命令:

```
round(with(tableb1,tapply(LogScore, list(Treatment),mean)),digits=3)
```

得到如表 B.1 中各样品的“自然对数均值”：

1T	2T	3T	4T	5T	6T
2.225	2.985	3.570	3.691	4.323	4.995

B.1.2 显著性差异分析

使用 R 软件对表 B.1 中的评分自然对数值进行双因素方差分析,输入以下命令：

```
summary(aov(LogScore~Assessor+Treatment,data=tableb1))
```

执行命令后,得到结果如表 B.2 所示,样品因子效应极显著。这一数据结果很合理,因为各样品间的咖啡因浓度差异很大(如 9 mg/100 mL、18 mg/100 mL 和 36 mg/100 mL 等)。

表 B.2 表 B.1 数据的双因素方差分析结果

方差来源	自由度	平方和	均方	F 值	$P > F$
评价员	6	0.24	0.040	4.55	0.002
样品	5	33.18	6.635	752.97	$< 2 \times 10^{-16}$
残差	30	0.26	0.009	—	—

进一步进行多重比较检验,采用图基(Tukey's)检验研究任意两个样品间是否存在显著性差异。本示例中,最小显著性差异的计算如公式(B.1)所示：

$$C \times \sqrt{\frac{1}{2} s^2 \times \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad \dots\dots\dots (B.1)$$

式中：

C ——因子,根据残差的自由度、样品总数量和所选 α -风险的函数获得,图基检验的学生化极差的临界值见参考文献[1];

s^2 ——表 B.2 中残差项的均方;

n_i ——第一组样品均值计算时观测值数量;

n_j ——第二组样品均值计算时观测值数量。

本示例中(6个样品,残差自由度为 30),当选择 α -风险=0.05 时, $C=4.30$,最小显著性差异的计算如下：

$$4.30 \times \sqrt{0.5 \times 0.009 \times \left(\frac{1}{7} + \frac{1}{7} \right)} = 0.154$$

显著性分析结果表明,仅有样品组 803 和 935 之间不存在显著性差异,2 个样品的评分自然对数均值差异仅为 0.121。这一结论很合理,因为这两个样品的咖啡因浓度很接近,浓度对数值的差异仅为 0.046,而其他相邻样品组的咖啡因浓度对数值差异较大,分别为 0.255(样品组 417 和 803)、0.301(样品组 274 和 561、样品组 935 和 274、样品组 127 和 417)。

上述计算也可通过 R 软件的 TukeyHSD 命令获得：

```
TukeyHSD(aov(LogScore~Assessor+Treatment,data=tableb1),"Treatment")
```

执行命令后,得到结果如下：

样品	评分差值	置信下限	置信上限	p_{adj}
2T-1T	0.7602857	0.60766673	0.9129047	0.0000000
3T-1T	1.3452857	1.19266673	1.4979047	0.0000000
4T-1T	1.4665714	1.31395244	1.6191904	0.0000000
5T-1T	2.0978571	1.94523816	2.2504761	0.0000000
6T-1T	2.7697143	2.61709530	2.9223333	0.0000000
3T-2T	0.5850000	0.43238101	0.7376190	0.0000000
4T-2T	0.7062857	0.55366673	0.8589047	0.0000000
5T-2T	1.3375714	1.18495244	1.4901904	0.0000000
6T-2T	2.0094286	1.85680958	2.1620476	0.0000000
4T-3T	0.1212857	-0.03133327	0.2739047	0.1825395
5T-3T	0.7525714	0.59995244	0.9051904	0.0000000
6T-3T	1.4244286	1.27180958	1.5770476	0.0000000
5T-4T	0.6312857	0.47866673	0.7839047	0.0000000
6T-4T	1.3031429	1.15052387	1.4557618	0.0000000
6T-5T	0.6718571	0.51923816	0.8244761	0.0000000

对于每个样品组,该命令得出:

- a) 两个样品间评分自然对数的差值;
- b) 置信水平为 95%的置信下限(lwr)和置信上限(upr);
- c) p 值(“adj”这里可忽略)。

在显著性水平 $\alpha=0.05$ 时,样品组 4T-3T(803 和 935)间不存在显著性差异。

注 1: 当评价员因子设定为随机因子时,需要使用 lmer()命令(来自 lmerTest R 包)。输入下列命令:

```
library(lmerTest)
resb1<-lmer(LogScore~Treatment+(1|Assessor),data=tableb1)
anova(resb1)
ranova(resb1)
```

本示例中,采用了平衡试验设计,样品因子的 p 值与表 B.2 中样品的 p 值相同。但评价员因子的 p 值(0.006)高于表 B.2 中评价员的 p 值(0.002)。通常情况下,当将评价员因子设定为随机因子而非固定因子时,“评价员”效应往往偏低(即 p 值偏高)。

注 2: 对表 B.1 中每个评价员的 6 个评分数据进行重新标度,使用方法如 B.2 中所述。重新标度后的评分值对应 R 文件 tableb1 中的“重新标度的评分自然对数值”列。使用 R 软件对重新标度后的数据进行双因素方差分析,输入以下命令:

```
summary(aov(LogScoreresc~Assessor+Treatment,data=tableb1))
```

执行命令后,得到结果如表 B.3 所示。7 个评价员重新标度的评分自然对数均值均相等(3.631),且评价员因子的平方和为 0。

表 B.3 表 B.1 重新标度数据的双因素方差分析结果

方差来源	自由度	平方和	均方	F 值	$Pr(>F)$
评价员	6	0.00	0.000	0.0	1
样品	5	33.17	6.634	754.5	$< 2 \times 10^{-16}$
残差	30	0.26	0.009	—	—

B.2 无重复且有内部重新标度数据的数据分析——所有评价员一次评价完该系列的所有产品

表 B.4 展示了一项研究项目的结果。该项研究中,感官评价小组成员给一系列蔗糖溶液的感知甜度进行评分。使用 2%蔗糖溶液作为内部参比样,所有评价员对该内部参比样的甜度评分为 10。参比样的评分不参与后期的数据分析,同时 2%蔗糖参比样以隐藏样品的形式进行评价。

表 B.4 6 个蔗糖样品的数据

蔗糖浓度/% (质量/体积)	评价员					
	L.M.	B.W.	M.M.	H.H.	J.D.	J.F.
0.5	5	1	1	1	2	0.5
1	1	4	2.5	1	2	2
2	10	10	8	5	10	5
4	20	15	20	30	20	25
8	50	30	40	100	25	50
16	100	50	80	300	40	200
几何均值	13.08	9.83	10.42	12.85	9.63	10.38

注:每位评价员的几何均值计算如下: $GM = \sqrt[n]{(x_1) \times (x_2) \cdots (x_n)}$
如小组成员 L.M.,几何均值为 $GM = \sqrt[6]{(5 \times 1 \times 10 \times 20 \times 50 \times 100)} = 13.08$
表 B.4 中的数据经过自然对数转化后,生成了表 B.5。

表 B.5 表 B.4 经过自然对数转化后的数据

蔗糖浓度/% (自然对数值)	L.M.	B.W.	M.M.	H.H.	J.D.	J.F.	小组均值
0.5(−0.693)	1.609	0.000	0.000	0.000	0.693	−0.693	—
1(0.000)	0.000	1.386	0.916	0.000	0.693	0.693	—
2(0.693)	2.303	2.303	2.079	1.609	2.303	1.609	—
4(1.386)	2.996	2.708	2.996	3.401	2.996	3.219	—
8(2.079)	3.912	3.401	3.689	4.605	3.219	3.912	—
16(2.773)	4.605	3.912	4.382	5.704	3.689	5.298	—
对数均值	2.571	2.285	2.344	2.553	2.266	2.340	2.393
校正因子	−0.178	0.108	0.049	−0.16	0.127	0.053	—

表 B.5 中,评价小组的算数均值为 2.393。

对于每个评价员,使用校正因子来重新标度评分值。校正因子等于评价小组均值与评价员均值的差值。如对于评价员 L.M.,校正因子的值等于 $2.393 - 2.571 = -0.178$ 。应用校正因子后,0.5%浓度蔗糖样品的甜度评分变为 $1.609 - 0.178 = 1.431$,1%浓度蔗糖样品的甜度评分变为 $0 - 0.178 = -0.178$,而 2%浓度蔗糖样品的甜度评分变为 $2.303 - 0.178 = 2.125$ 。所有评价员校正后的评分计算结果见表 B.6。

表 B.6 表 B.4 进行全面重新标度后的数据

蔗糖浓度的对数值	评价员						
	L.M.	B.W.	M.M.	H.H.	J.D.	J.F.	算术均值
−0.693	1.431	0.108	0.049	−0.160	0.820	−0.640	0.268
0.000	−0.178	1.494	0.965	−0.160	0.820	0.746	0.615
0.693	2.125	2.411	2.128	1.449	2.430	1.662	2.034
1.386	2.818	2.816	3.045	3.241	3.123	3.272	3.053
2.079	3.734	3.509	3.738	4.445	3.346	3.965	3.790
2.773	4.427	4.020	4.431	5.544	3.816	5.351	4.598
对数均值	2.393	2.393	2.393	2.393	2.393	2.393	—

将该示例的数据文件(如位于 F 盘中“ME”文件夹中的文件)通过下列 3 个命令导入 R 中：

```
tableb6<-read.table("F:/ME/Table Annex B2.txt",header = T,sep="\t",dec=".")
attach(tableb6)
names(tableb6)
```

文件 tableb6 包含 7 列(评价员、样品、蔗糖浓度、蔗糖浓度自然对数值、评分、评分自然对数值和重新标度的评分自然对数值)和 36 行(6 名评价员×6 个浓度)数据。输入以下命令：

```
round (with(tableb6,tapply(LogScoreresc, list(Treatment),mean)),digits=3)
```

得到如表 B.6 中 6 个浓度蔗糖样品的“算术均值”：

1c	2c	3c	4c	5c	6c
0.268	0.615	2.034	3.053	3.790	4.598

输入以下命令：

```
round (with(tableb6,tapply(LogScoreresc, list(Assessor),mean)),digits=3)
```

得到表 B.6 中 6 个评价员的算数均值,为 2.393。

对表 B.6 的数据进行双因素方差分析,输入以下命令：

```
summary(aov(LogScoreresc~Assessor+Treatment,data=tableb6))
```

结果如表 B.7 所示。由于已对数据进行了全面重新标度,因此评价员因子的平方和等于 0。

表 B.7 表 B.6 的双因素方差分析结果

方差来源	自由度	平方和	均方	F 值	Pr>F
评价员	5	0.00	0.000	0.0	1
样品	5	90.33	18.066	49.5	3.62×10^{-12}
残差	25	9.12	0.365	—	—

浓度因子非常显著,输入以下命令：

```
TukeyHSD(aov(LogScoreresc~Assessor+Treatment,data=tableb6),"Conc")
```

得到结果表明,相邻浓度样品组 3c-2c 存在显著差异($p<0.05$),但样品组 2c-1c、4c-3c、5c-4c 和6-5c 之间不存在显著性差异(p 值分别为 0.916、0.071、0.313 和 0.224)。

样品	评分差值	置信下限	置信上限	p_{adj}
2c-1c	0.3465000	-0.72844156	1.421442	0.9157442
3c-1c	1.7661667	0.69122510	2.841108	0.0004102
4c-1c	2.7845000	1.70955844	3.859442	0.0000003
5c-1c	3.5215000	2.44655844	4.596442	0.0000000
6c-1c	4.3301667	3.25522510	5.405108	0.0000000
3c-2c	1.4196667	0.34472510	2.494608	0.0049459
4c-2c	2.4380000	1.36305844	3.512942	0.0000035
5c-2c	3.1750000	2.10005844	4.249942	0.0000000
6c-2c	3.9836667	2.90872510	5.058608	0.0000000
4c-3c	1.0183333	-0.05660823	2.093275	0.0707337
5c-3c	1.7553333	0.68039177	2.830275	0.0004437
6c-3c	2.5640000	1.48905844	3.638942	0.0000015
5c-4c	0.7370000	-0.33794156	1.811942	0.3129969
6c-4c	1.5456667	0.47072510	2.620608	0.0020171
6c-5c	0.8086667	-0.26627490	1.883608	0.2240927

B.3 对子组样品进行内部重新标度数据的数据分析

在表 B.1 的示例中,假定并非所有评价员都对全部样品进行了评价。评价员 2~评价员 7 仅对 5 个样品进行了评价,因此试验采用的是非平衡设计。数据结果如表 B.8 所示。与表 B.1 相比,表 B.8 中的数据基本相同,只是去掉了评价员 2、评价员 4 和评价员 6 对样品 274 的评分,以及评价员 3、评价员 5 和评价员 7 对样品 417 的评分。

所有评价员都对样品 561、935、803 和 127 进行了评分,因此对该子组样品进行重新标度,具体步骤如下:

- a) 首先,计算每个评价员对子组中 4 个样品评分的自然对数均值,如评价员 1 对子组 4 个样品的评分自然对数均值为 $(2.303+3.555+3.689+4.942)/4=3.622$ (见表 B.8);
- b) 其次,计算整个评价小组对子组样品评分的自然对数均值,结果为 3.620;
- c) 再次,将整个评价小组的均值(3.620)减去每个评价员的均值,得到每个评价员的校正因子,如评价员 1 的校正因子为 $3.620-3.622=-0.002$;
- d) 最后,使用校正因子重新标度每个评价员的评分,得到表 B.9。

表 B.8 评价员评分的自然对数值和校正因子

评价员	样品编号						共同子组的均值	校正因子
	561	274	935	803	417	127		
	估计值的自然对数值							
1	2.303	2.996	3.555	3.689	4.248	4.942	3.622	—0.002
2	2.079	—	3.638	3.784	4.443	5.075	3.644	—0.024
3	2.079	2.996	3.584	3.689	—	5.011	3.591	+0.029
4	1.946	—	3.466	3.611	4.248	4.905	3.482	+0.138

表 B.8 评价员评分的自然对数值和校正因子（续）

评价员	样品编号						共同子组的均值	校正因子
	561	274	935	803	417	127		
	估计值的自然对数值							
5	2.485	3.219	3.638	3.689	—	4.977	3.697	—0.077
6	2.485	—	3.555	3.689	4.382	5.075	3.701	—0.081
7	2.197	2.890	3.555	3.689	—	4.977	3.605	+0.015
全组均值	—	—	—	—	—	—	3.620	—

表 B.9 重新标度的评分自然对数值

评价员	样品编号					
	561	274	935	803	417	127
	估计值的自然对数值					
1	2.301	2.994	3.553	3.687	4.246	4.940
2	2.055	—	3.614	3.760	4.419	5.051
3	2.108	3.025	3.613	3.718	—	5.040
4	2.084	—	3.604	3.749	4.386	5.043
5	2.408	3.142	3.561	3.612	—	4.900
6	2.404	—	3.474	3.608	4.301	4.994
7	2.212	2.905	3.570	3.704	—	4.992
全组均值	2.225	3.017	3.570	3.691	4.338	4.994

将该示例的数据文件(如位于 F 盘中“ME”文件夹中的文件)通过下列 3 个命令导入 R 中：

```
tableb9<-read.table("F:/ME/Table Annex B3.txt",header = T,sep="\t",dec=".")
attach(tableb9)
names(tableb9)
```

文件 tableb9 包含 5 列(评价员、样品、样品编码、评分的自然对数值、重新标度的评分自然对数值)和 36 行(评价员 1 有 6 行,评价员 2 到评价员 7 各 5 行,共计 30 行)数据。输入以下命令：

```
round (with(tableb9,tapply(LogScoreresc, list(Treatment),mean)),digits=3)
```

得到表 B.9 中的 6 个样品的“全组均值”：

1T	2T	3T	4T	5T	6T
2.225	3.017	3.570	3.691	4.338	4.994

采用 R 软件,对表 B.9 的数据进行方差分析,输入以下命令：

```
summary(aov(LogScoreresc~Treatment+Assessor,data= tableb9))
```

结果见表 B.10。

表 B.10 表 B.9 数据的方差分析

方差来源	自由度	平方和	均方	F 值	$Pr > F$
样品	5	30.417	6.083	641.577	$< 2 \times 10^{-16}$
评价员	6	0.010	0.002	0.177	0.981
残差	24	0.228	0.009	—	—

由表 B.10 可见,样品因子极显著。由于本示例是对 4 个样品的子组进行重新标度,因此评价员因子的平方和不等于 0。

输入下列命令,计算每个评价员的均值:

```
round (with(tableb9, tapply(LogScoreresc, list(Assessor), mean)), digits=3)
```

得到结果如下:

As1	As2	As3	As4	As5	As6	As7
3.620	3.780	3.501	3.773	3.525	3.756	3.477

计算最小显著性差异时,需考虑不同样品的均值并非是基于相同数量的观测值计算得出(4 个评价员评价了样品 274、417 和 127,7 个评价员评价了样品 561、935 和 803)。采用图基检验计算各样品对之间是否有显著差异^[2]。本示例中,共有 6 个样品,残差自由度为 24,当 α -风险为 0.05 时, C 等于 4.37。因此,对于包含 4 个观测值的样品组(274、417),其最小显著性差异计算如下:

$$4.37 \times \sqrt{\frac{1}{2} \times 0.009 \times \left(\frac{1}{4} + \frac{1}{4} \right)} = 0.207$$

样品组(561、274)的最小显著性差异计算如下:

$$4.37 \times \sqrt{\frac{1}{2} \times 0.009 \times \left(\frac{1}{7} + \frac{1}{4} \right)} = 0.184$$

样品组(561、935)的最小显著性差异计算如下:

$$4.37 \times \sqrt{\frac{1}{2} \times 0.009 \times \left(\frac{1}{7} + \frac{1}{7} \right)} = 0.157$$

最终结果表明,仅样品组(803、935)间没有显著性差异。这一结论与 B.1 中的结论相同。

上述计算可通过下列命令得到:

```
TukeyHSD(aov(LogScoreresc~Treatment+ Assessor, data=tableb9), "Treatment")
```

命令输出结果如下:

样品	评分差值	置信下限	置信上限	p_{adj}
2T-1T	0.7919286	0.60321700	0.9806401	0.0000000
3T-1T	1.3452857	1.18435195	1.5062195	0.0000000
4T-1T	1.4665714	1.30563766	1.6275052	0.0000000
5T-1T	2.1134286	1.92471700	2.3021401	0.0000000
6T-1T	2.7697143	2.60878052	2.9306481	0.0000000

样品	评分差值	置信下限	置信上限	p_{adj}
3T-2T	0.5533571	0.36464557	0.7420687	0.0000000
4T-2T	0.6746429	0.48593129	0.8633544	0.0000000
5T-2T	1.3215000	1.10860463	1.5343954	0.0000000
6T-2T	1.9777857	1.78907414	2.1664973	0.0000000
4T-3T	0.1212857	-0.03964805	0.2822195	0.2210445
5T-3T	0.7681429	0.57943129	0.9568544	0.0000000
6T-3T	1.4244286	1.26349480	1.5853623	0.0000000
5T-4T	0.6468571	0.45814557	0.8355687	0.0000000
6T-4T	1.3031429	1.14220909	1.4640766	0.0000000
6T-5T	0.6562857	0.46757414	0.84499730	0.0000000

注：当评价员因子设定为随机因子时，使用 lmer() 命令（来自 lmerTest R 包）进行计算。输入下列命令：

```
library(lmerTest)
resb9<-lmer(LogScoreresc~Treatment+(1|Assessor),data=tableb9)
anova(resb9)
ranova(resb9)
```

得到结果中，样品因子的 F 值等于 768.05， p 值等于 2.2×10^{-16} ；评价员因子的 p 值等于 1，评价员间的差值为 0。

B.4 外部重新标度数据的数据分析

完成测试样品评价后，评价员对语义类项标度进行评分。具体步骤如下：首先，建立一个 5 点标度，范围从“极苦”到“轻微苦”；随后，评价小组组长要求每个评价员采用其在测试样品评价时所使用的标度，对 5 个语义描述进行评分。该练习的假设性结果如表 B.11 所示。

表 B.11 语义标度评分的假设性结果

评价员	轻微苦	苦	中等苦	很苦	极苦	校正因子 ^a
1	5	25	50	100	150	-0.002 5
2	5	30	60	100	160	-0.088 0
3	5	25	50	100	150	-0.002 5
4	5	20	45	90	140	+0.098 1
5	5	25	50	100	150	-0.002 5
6	3	30	55	110	170	0.000 1
7	5	25	50	100	150	-0.002 5
^a 校正因子的计算方法与 B.2 中所用方法一致。						

本示例中的测试样品数据如表 B.12 所示。

表 B.12 评价员评分的自然对数值

评价员	样品编号					
	561	274	935	803	417	127
1	2.303	2.996	3.555	3.689	4.248	4.942
2	—	2.996	3.638	3.784	4.443	5.075
3	2.079	—	3.584	3.689	4.317	5.011
4	1.946	2.708	—	3.611	4.248	4.905
5	2.485	3.219	3.638	—	4.317	4.977
6	2.485	3.091	3.555	3.689	—	5.075
7	2.197	2.890	3.555	3.689	4.304	—

使用表 B.11 中的校正因子对表 B.12 中的数据进行重新标度,结果如表 B.13 所示,具体步骤如下:

a) 首先,使用全面重新标度法,计算语义标度评分的校正因子;

b) 其次,将每个评价员的评分值加上该评价员的校正值,对表 B.12 中的数据进行重新标度。

需注意在表 B.12 中,仅有评价员 1 评价了全部的 6 个测试产品,其他评价员均只评价了 5 个产品(如评价员 2 没有评价样品 561)。因此,本示例采用的是不平衡设计试验。此外,表 B.12 中的数据缺失模式与表 B.8 有所不同。如样品 561,在表 B.12 中评价员 2 对应的位置存在一个空格,而在表 B.8 中相应位置则没有空格。

表 B.13 表 B.12 重新标度估计值的自然对数值

评价员	样品编号					
	561	274	935	803	417	127
1	2.301	2.994	3.553	3.687	4.246	4.940
2	—	2.908	3.550	3.696	4.355	4.987
3	2.077	—	3.582	3.687	4.315	5.009
4	2.044	2.806	—	3.709	4.346	5.003
5	2.483	3.217	3.636	—	4.315	4.975
6	2.485	3.091	3.555	3.689	—	5.075
7	2.195	2.888	3.553	3.687	4.302	—
小组均值	2.264	2.984	3.571	3.692	4.313	4.998

将该示例的数据文件(如位于 F 盘中“ME”文件夹中的文件)通过下列 3 个命令导入 R 中:

```
tableb13<-read.table("F:/ME/Table Annex B4.txt",header = T,sep="\t",dec=".")
attach(tableb13)
names(tableb13)
```

文件 tableb13 中包含 4 列(评价员、样品、评分、重新标度的评分自然对数值)和 36 行(评价员 1 有 6 行,评价员 2 到评价员 7 各 5 行,小计 30 行)数据。输入以下命令:

```
round (with(tableb13,tapply(LogScoreresc, list(Treatment),mean)),digits=3)
```


得到表 B.13 中的“小组均值”：

1T	2T	3T	4T	5T	6T
2.264	2.984	3.571	3.692	4.313	4.998

采用 R 软件进行方差分析，输入以下命令：

```
summary(aov(LogScoreresc~Treatment+Assessor,data=tableb13))
```

结果如表 B.14 所示。由于试验是采用不平衡设计，评价员因子的平方和不等于 0。样品因子具有极显著性。

表 B.14 表 B.13 的方差分析结果

方差来源	自由度	平方和	均方	F 值	P>F
样品	5	27.771	5.554	667.018	$<2\times 10^{-16}$
评价员	6	0.123	0.020	2.454	0.054
残差	24	0.200	0.008	—	—

如 B.1.2 中所述，6 个样品，每个样品有 6 个观测值，残差自由度为 24，当 α -风险=0.05 时，最小显著性差异计算如下：

$$4.37 \times \sqrt{0.5 \times 0.008 \times (\frac{1}{6} + \frac{1}{6})} = 0.160$$

样品组(935 和 803)间没有显著性差异，这一结论与 B.1 和 B.3 中的结论一致。

上述计算可通过下列命令直接输出：

```
TukeyHSD(aov(LogScoreresc~Treatment+Assessor,data=tableb13),"Treatment")
```

命令输出结果如下：

样品	评分差值	置信下限	置信上限	p_{adj}
2T-1T	0.7198333	0.55693836	0.8827283	0.0000000
3T-1T	1.3073333	1.14443836	1.4702283	0.0000000
4T-1T	1.4283333	1.26543836	1.5912283	0.0000000
5T-1T	2.0490000	1.88610503	2.2118950	0.0000000
6T-1T	2.7340000	2.57110503	2.8968950	0.0000000
3T-2T	0.5875000	0.42460503	0.7503950	0.0000000
4T-2T	0.7085000	0.54560503	0.8713950	0.0000000
5T-2T	1.3291667	1.16627169	1.4920616	0.0000000
6T-2T	2.0141667	1.85127169	2.1770616	0.0000000
4T-3T	0.1210000	-0.04189497	0.2838950	0.2339371
5T-3T	0.7416667	0.57877169	0.9045616	0.0000000
6T-3T	1.4266667	1.26377169	1.5895616	0.0000000

样品	评分差值	置信下限	置信上限	p_{adj}
5T-4T	0.6206667	0.45777169	0.7835616	0.0000000
6T-4T	1.3056667	1.14277169	1.4685616	0.0000000
6T-5T	0.6850000	0.52210503	0.8478950	0.0000000

相邻样品组间,仅样品组 4T-3T(803 和 935)间不存在显著性差异。

注 1: 由于本次试验采用了不平衡设计,在使用函数 `aov(LogScoreresc...)` 分析时,采用“评价员+样品”顺序得到的结果与采用“样品+评价员”顺序得到的结果存在差异。采用“评价员+样品”顺序时,评价员因子的 p 值 < 0.05 。但是,“样品+评价员”顺序是最为常用的设定顺序。

注 2: 当评价员因子设定为随机因子时,输入 R 命令:

```
library(lmerTest)
resb13<-lmer(LogScoreresc~Treatment+(1|Assessor),data=tableb13)
anova(resb13)
ranova(resb13)
```

得到结果中,样品因子的 F 值等于 653.29, p 值等于 2.2×10^{-16} ; 评价员因子的 p 值等于 0.141。在 $\alpha=0.05$ 时,评价员间的差异不显著(评价员因子设定为固定因子时, p 值等于 0.054,见表 B.14)。

B.5 有重复数据的数据分析

B.5.1 总则

表 B.15 展示了同一评价小组对表 B.1 中样品结果的重复评分结果。在完成测试样品的第一次评价后,评价员随即开始对同一组样品的第二次评价,但样品的评价顺序与第一次不同。评价员并未被告知这是对同一组样品的两次重复评价。

表 B.15 重复 1 和重复 2 的评分值

评价员	重复	样品编号					
		561	274	935	803	417	127
		评分值(自然对数值)					
1	1	10(2.303)	20(2.996)	35(3.555)	40(3.689)	70(4.248)	140(4.942)
	2	15(2.708)	25(3.219)	35(3.555)	38(3.638)	70(4.248)	135(4.905)
2	1	8(2.079)	20(2.996)	38(3.638)	44(3.784)	85(4.443)	160(5.075)
	2	8(2.079)	15(2.708)	35(3.555)	45(3.807)	90(4.500)	180(5.193)
3	1	8(2.079)	20(2.996)	36(3.584)	40(3.689)	75(4.317)	150(5.011)
	2	10(2.303)	20(2.996)	35(3.555)	35(3.555)	70(4.248)	145(4.977)
4	1	7(1.946)	15(2.708)	32(3.466)	37(3.611)	70(4.248)	135(4.905)
	2	10(2.303)	20(2.996)	35(3.555)	38(3.638)	65(4.174)	130(4.868)
5	1	12(2.485)	25(3.219)	38(3.638)	40(3.689)	75(4.317)	145(4.977)
	2	10(2.303)	25(3.219)	35(3.555)	40(3.689)	80(4.382)	150(5.011)
6	1	12(2.485)	22(3.091)	35(3.555)	40(3.689)	80(4.382)	160(5.075)
	2	10(2.303)	20(2.996)	35(3.555)	40(3.689)	80(4.382)	160(5.075)

表 B.15 重复 1 和重复 2 的评分值 (续)

评价员	重复	样品编号					
		561	274	935	803	417	127
		评分值(自然对数值)					
7	1	9(2.197)	18(2.890)	35(3.555)	40(3.689)	74(4.304)	145(4.977)
	2	10(2.303)	15(2.708)	35(3.555)	38(3.638)	70(4.248)	140(4.942)
自然对数的均值	—	2.277	2.981	3.563	3.678	4.317	4.995

将该示例的数据文件(如位于 F 盘中“ME”文件夹中的文件)通过下列 3 个命令导入 R 中:

```
tableb15<-read.table("F:/ME/Table Annex B5.txt",header = T,sep="\t",dec=".")
attach(tableb15)
names(tableb15)
```

文件 tableb15 包含 7 列(评价员、样品、评分、评分的自然对数值、重复、浓度、浓度的自然对数值)和 84 行(7 名评价员×6 个样品×2 次重复)数据。其中,评价员 1 的 12 个评分值在第 1 行~第 12 行,评价员 2 的 12 个评分值在第 13 行~第 24 行,以此类推。输入以下命令:

```
round (with(tableb15,tapply(LogScore, list(Treatment),mean)),digits=3)
```

得到表 B.15 中“自然对数的均值”:

1T	2T	3T	4T	5T	6T
2.277	2.981	3.563	3.678	4.317	4.995

B.5.2 当评价员因子设定为固定因子时,显著性分析的确定

试验设计中涉及有三个因素:样品、评价员和重复。通常情况下,样品因子和重复因子被设定为固定因子。因此,当评价员因子设定为固定因子时,整个试验设计中包含 3 个固定因子。此时使用 R 软件的相应命令执行三因素方差分析:

```
summary(aov(LogScore ~ Assessor * Treatment * Replicate-Assessor: Treatment: Replicate,
tableb15))
```

结果见表 B.16。

表 B.16 表 B.15 的方差分析结果

方差来源	自由度	平方和	均方	F 值	$P>F$
评价员	6	0.28	0.046	5.151	0.964×10^{-3}
样品	5	64.32	12.865	1 445.182	$<2 \times 10^{-16}$
重复	1	0.00	0.001	0.132	0.719
评价员×样品	30	0.55	0.018	2.047	0.027

表 B.16 表 B.15 的方差分析结果 (续)

方差来源	自由度	平方和	均方	<i>F</i> 值	<i>P</i> > <i>F</i>
评价员 × 重复	6	0.07	0.012	1.378	0.255
样品 × 重复	5	0.04	0.008	0.912	0.487
残差	30	0.27	0.009	—	—

评价员因子和样品因子均具有重要意义。此外,评价员 × 样品的交互效应也显著,表明不同评价员对咖啡因浓度变化的响应存在明显差异。即便在模型中不考虑重复因子时,所得结论仍然一致。使用以下命令进行方差分析:

```
summary(aov(LogScore~Assessor * Treatment,data=tableb15))
```

结果如表 B.17 所示。

表 B.17 表 B.15 双因素方差分析结果

方差来源	自由度	平方和	均方	<i>F</i> 值	<i>P</i> > <i>F</i>
评价员	6	0.28	0.046	5.035	0.000 6
样品	5	64.32	12.865	1 412.795	$< 2 \times 10^{-16}$
评价员 × 样品	30	0.55	0.018	2.002	0.019
残差	42	0.38	0.009	—	—

B.5.3 当评价员因子设定为随机因子时,显著性分析的确定

将评价员因子设定为随机因子,样品因子和重复因子设定为固定因子时,采用混合设计,使用 R 的 lmer 函数。输入下列 4 个命令:

```
library(lmerTest)
mod1<-lmer(LogScore~Treatment * Replicate+(1|Assessor)+(1|Assessor: Treatment)+(1|
Assessor:Replicate),tableb15)
anova(mod1)
ranova(mod1)
```

得到结果如表 B.18 所示,包含了 2 个部分。第一部分是模型的固定因子,由于本示例采用了平衡设计,固定因子的 *p* 值与表 B.16 中的结果一致,且提到 III 型和 Satterthwaite 法在本分析中均无效。第二部分是模型的随机因子。未采用 *F* 检验,而是运用似然比检验值(LRT 值)。似然比检验遵循卡方分布,其自由度已在表中给出。通过似然比检验值和自由度值,具体列在表中的“Pr(>Chisq)”列。

分析结果表明,评价员因子的 *p* 值等于 0.2699(表 B.16 中的 *p* 值等于 0.964×10^{-3}),表明评价员效应不显著;评价员与样品交互效应的 *p* 值接近 0.05(0.0513)(表 B.16 中的 *p* 值等于 0.027)。

表 B.18 评价员设定为随机因子时表 B.15 的数据分析结果

Type III Analysis of Variance Table with Satterthwaite's method						
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Treatment	6.395 5	1.279 11	5	42.978	143.688 9	<2e-16***
Replicate	0.000 9	0.000 85	1	6.000	0.095 7	0.767 6
Treatment:Replicate	0.040 6	0.008 12	5	30.000	0.912 0	0.486 5

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	16	39.505	-47.010			
(1 Assessor)	15	38.897	-47.793	1.217 3	1	0.269 9
(1 Assessor:Treatment)	15	37.606	-45.212	3.798 4	1	0.051 3
(1 Assessor:Replicate)	15	39.367	-48.734	0.276 2	1	0.599 2

由于重复因子自身以及重复因子与样品、评价员的 2 个交互效应均不显著,因此可将三因素模型简化为双因素模型。输入 R 命令:

```
mod2<-lmer(LogScore~Treatment+(1|Assessor)+(1|Assessor:Treatment),tableb15)
anova(mod2)
ranova(mod2)
```

结果如表 B.19 所示。

表 B.19 评价员设定为随机因子且忽略重复因子时的数据分析结果

Type III Analysis of Variance Table with Satterthwaite's method						
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Treatment	32.133	6.4266	5	29.998	705.86	<2.2e-16***

	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
<none>	9	49.401	-80.803			
(1 Assessor)	8	48.115	-80.230	2.572 8	1	0.108 72
(1 Assessor:Treatment)	8	47.225	-78.509	4.293 5	1	0.038 26*

评价员和样品间的交互作用显著($p=0.03826$),表明评价员对咖啡因含量变化的响应方式存在显著差异。

B.6 回归曲线斜率的确定(史蒂文斯幂函数,见 9.4)

B.6.1 总则

当样品中刺激物的浓度为已知(如表 B.1 和表 B.15 中的咖啡因浓度,以及表 B.4 中的蔗糖浓度)时,可对浓度的对数值与评分的对数值进行线性回归分析,并计算回归曲线的斜率。若评分未进行重新标度,有必要对每个评价员单独进行线性回归分析。此外,建议对评价员和浓度的交互效应进行分析,以检验每个评价员线性回归曲线的斜率是否相同。

本示例中,采用表 B.15 中的数据,为方便理解,对表中数据进行重新整理,只保留了均值数据。
输入下列 3 条 R 命令:

```
tableannex6<-read.table("F:/ME/Table Annex B6.txt",header = T,sep="\t",dec=".")
attach(tableannex6)
names(tableannex6)
```

文件 tableannex6 包含 11 列(浓度的自然对数值、评分自然对数小组均值、重复 1 的评分自然对数均值、重复 2 的评分自然对数均值和 7 个评价员各自评分的自然对数值)和 6 行(6 个浓度)数据。输出如下:

```
[1]"LogConc"          "LogScoregroup" "LogScoreRep1" "LogScoreRep2" "LogScoreAs1"
[6]"LogScoreAs2"      "LogScoreAs3" "LogScoreAs4" "LogScoreAs5" "LogScoreAs6"
[11]"LogScoreAs7"
```

采用评分自然对数值(非重新标度的评分自然对数值)进行计算,以突出评价员间的差异。

B.6.2 小组斜率的确定

回归曲线的斜率采用 R 命令的 lm() 函数计算获得,输入下列 3 条 R 命令:

```
model<-lm(LogScoregroup~LogConc,data = tableannex6)
coefficients(model)
confint(model)
```

得到结果如下所示:

```
coefficients (mode 1)

(Intercept)      LogConc
0.1210406        0.9758299

confint (model)

                2.5%          97.5%
(Intercept)    -0.085 199 07    0.327 280 3
LogConc         0.920 253 72     1.031 406 1
```

同理,将 R 命令中的“LogScoregroup”(小组)分别替换为“LogScoreRep1”(重复 1)和“LogScoreRep2”(重复 2),计算得到重复 1 和重复 2 的回归曲线斜率,分别为 0.99 和 0.96。

可绘制描述感官响应值与物质浓度值之间关系的图,输入下列 3 条 R 命令:

```
par (mfrow=c(2,4))
plot(LogConc,LogScoregroup,main="Group")
abline(lm(LogScoregroup~LogConc))
```

得到图 B.1 中“小组”的回归曲线(上排最左侧的图)。

注 1: 由于评分自然对数值和重复的交互效应不显著(见表 B.16),重复 1(0.99)和重复 2(0.96)的斜率非常接近。

注 2: 使用以下 3 条命令可以从 R 文件 tableb15 中获得前期的系数和近似舍入误差:

```
model<-lm(LogScoreg~LogConc,data = tableb15)
coefficients(model)
confint(model)
```

得到结果如下:

```
coefficients (mode 1)

(Intercept)      LogConc
0.1207476       0.9759245

confint (model)

                2.5%          97.5%
(Intercept)    0.006 962 75    0.234 532 5
LogConc        0.945 262 42    1.006 586 5
```

对于每个浓度,评分自然对数值和浓度自然对数值的图中都有 14 个数据(7 个评价员×2 个重复)。

B.6.3 评价员斜率的确定

由于评价员×样品的 p 值小于 0.05(见表 B.16),因此确定每个评价员的回归曲线斜率很有意义。输入评价员 1 的命令:

```
model<-lm(LogScoreAs1~LogConc,data=tableannex6)
coefficients(model)
confint(model)
plot(LogConc,LogScoreAs1,main="Ass1")
abline(lm(LogScoreAs1~LogConc))
```

得到结果如下所示:

```
(Intercept)      LogConc
0.5689183       0.8603457

confint (model)

                2.5%          97.5%
(Intercept)    0.183 316 2    0.954 520 4
LogConc        0.756 436 1    0.964 255 4
```

同时,得到图 B.1 中评价员 1 的回归曲线(上排左数第二个图)。

对于评价员 2,将评价员 1 的命令中的“Ass 1”替换为“Ass 2”,输入下列命令:

```
model<-lm(LogScoreAs2~LogConc,data=tableannex6)
coefficients(model)
confint(model)
plot(LogConc,LogScoreAs2,main="Ass 2")
abline(lm(LogScoreAs2~LogConc))
```

评价员 3 到评价员 7 的命令操作同上,最终得到表 B.20 中的结果和图 B.1(8 个图)。其中,评价员 1 的回归曲线斜率最小(0.860),与评价员 2 和评价员 4 的回归曲线斜率存在显著差异。评价员 2 的回归曲线斜率最大(1.116),与评价员 1、评价员 5 和评价员 6 的回归曲线斜率存在显著性差异。

表 B.20 评价员的回归曲线参数

评价员	截距	斜率	置信区间	
			下限	上限
1	0.57	0.860	0.756	0.964
2	−0.36	1.116	1.053	1.178
3	0.03	0.992	0.913	1.072
4	−0.04	0.993	0.973	1.012
5	0.43	0.911	0.773	1.049
6	0.22	0.965	0.851	1.078
7	0.00	0.994	0.929	1.060

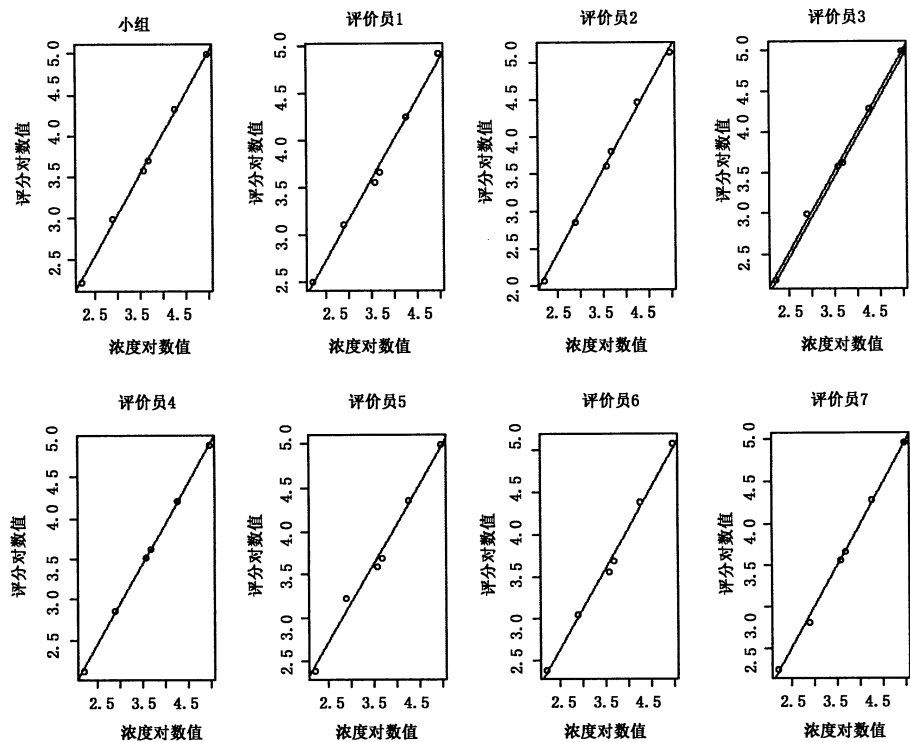


图 B.1 评价员的回归曲线(数据来自表 B.15)

参 考 文 献

- [1] Gacula M,C,Jr., Singh J., Altan S. Statistical methods in food and consumer research. Second edition. Academic Press, Orlando, USA, 2009.
- [2] Kramer C.Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*. 1956, 12, *pp.* 307-310.
- [3] Macfie H.J., Bratchell N., Greenhoff K., Vallis L.V. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*, 1989, 4 (2), *pp.* 129-148.
- [4] Moskowitz H.R. Product testing and sensory evaluation of foods. Food & Nutrition Press Inc., 1983, *pp.* 237-297.
- [5] Moskowitz H.R., Jacobs B.E. Magnitude estimation: scientific background and use in sensory analysis. In: Moskowitz H.R. (ed.) *Applied Sensory Analysis of Foods, Vol. 1*. CRC Press Inc., 1988, pp. 193-223.
- [6] Powers J.J., Ware G. O., Shinholster K.J. Magnitude estimation with and without rescaling. *Journal of Sensory Studies*. 1990, 5, *pp.* 105-116.
-