



中华人民共和国国家标准

GB/T 19547—2004/ISO 11056:1999

感官分析 方法学 量值估计法

Sensory analysis—Methodology—Magnitude estimation method

(ISO 11056:1999, IDT)

2004-06-11 发布

2004-12-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

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 原则	2
5 基本检验条件	2
6 评价员的选拔和培训	2
6.1 选拔和培训的基本条件	2
6.2 量值估计法的专项内容培训	2
7 评价员的数量	3
7.1 一般原则	3
7.2 分析和研究评价小组	3
7.3 消费者评价小组	4
8 程序	4
8.1 样品的提供	4
8.2 外部参比样	4
8.3 样品的提供顺序	4
8.4 量值估计	4
9 数据分析	5
9.1 数据分析方法的选择	5
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 重新标度方法	7
10 检验报告	7
附录 A(资料性附录) 数据分析示例	9
附录 B(资料性附录) 提问模式	15

前 言

本标准等同采用 ISO 11056:1999《感官分析——方法学——量值估计法》(英文版)。

本标准的附录 A、附录 B 均为资料性附录。

本标准由中国标准化研究院提出并归口。

本标准起草单位:中国标准化研究院、中国农业大学动物科技学院、中国肉类食品综合研究中心、中国食品发酵工业研究院。

本标准主要起草人:席兴军、刘文、张德福、张劳、熊正河、李气清。

引 言

量值估计法是一种心理物理学标度方法,是通过评价员对某一感官特性进行评分的一种评定方法。该方法要求评价员做出的评分要符合比例原则,即如果样品 B 某个特性的强度是样品 A 的两倍,则样品 B 的评分值应是样品 A 评分值的两倍。诸如强度、愉悦度和可接受性等特性均可用量值估计法进行评定。

当评价员的人数和培训评价员的时间有限时,量值估计法与其他标度方法相比有明显优点。量值估计法为评价小组组长和评价员提供较高的灵活性。在量值估计法中评价员经过培训后,再加上少量其他方面的培训,就能在更多的样品和特性上应用他们的技能进行评定。

应用量值估计法对边界效应的敏感性比用连续标度或间断标度法要小。当评价员对所要评定的感官指标不熟悉时,就会产生“边界效应”。于是评价员可能将某一初始样品的评分归到标度的边界。结果,由于没有足够的等级划分,只能将那些本该归到不同类别的样品归到了同一类别。使用量值估计法就不会产生这种情况,因为在理论上,各个类别都对应有不同的明确的数字。

允许评价员从任何数字(也就是用自己的标度)开始进行评分会产生一种很重要的“评价员”效应,但有不同的方法来解决这个问题:

- 方差分析(ANOVA)允许将评价员效应和他们的交互作用计算在内;
- 通过使用一个已赋值的参比样使评价员达到一个共同的标度范围;
- 通过应用多种重新标度方法中的一种,可以使每个评价员提供的数据纳入一个共同范围的标度。

由主试者根据具体条件选择最佳的分析方法。

在区别差异微小的刺激或者在刺激阈附近进行评估时,量值估计法并不是最有效的方法。

感官分析 方法学 量值估计法

1 范围

本标准给出了使用量值估计进行感官特性评价的方法,标准包括评价员的培训、评估值的获得和数据的数理统计解释。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版本均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB/T 3358.1 统计学术语 第一部分:一般统计术语(eqv ISO 3534-1)

GB/T 3358.3 统计学术语 第三部分:试验设计术语(eqv ISO 3534-3)

GB/T 10220 感官分析方法总论(neq ISO 6658)

GB/T 10221 感官分析 术语(neq ISO 5492)

GB/T 13868 感官分析 建立感官分析实验室的一般导则(idt ISO 8589)

GB/T 14195 感官分析 评价员选拔与培训 感官分析优选评价员导则(idt ISO 8586-1)

GB/T 16290 感官分析 方法学 使用标度评价食品(idt ISO 4121)

GB/T 16291 感官分析 专家的选拔、培训和管理导则(idt ISO 8586-2)

3 术语和定义

GB/T 3358.1,GB/T 3358.3 和 GB/T 10221 确立的以及下列术语和定义适用于本标准。

3.1

量值估计 magnitude estimation

对样品的特性强度或评价员喜好程度定值的过程,所定数值间的比例与评价员对特性感觉间的比例一致。

3.2

外部参比样 external reference

由评价小组组长预先给出数值并提供给评价员的样品。

注:该样品为与随后要评定的一系列样品相关的第一个样品。

3.3

模数 modulus

由检验负责人给定(固定模数)或由评价员选择(非固定模数)的外部参比样的数值。

3.4

重新标度 rescaling

将评价员给出的原始数据乘以一个因子后,将所有评价员给出的数据纳入相同标度范围的过程。

注:校正因子取对数后与原始数据取对数后再相加和上述过程是一个等价过程。

3.5

内部参比样 internal reference

插入系列检验样品中并作为参比样提供给评价员的检验样品。

注:对内部参比样给出的评分可校正某个评价员的数据。如果插入外部参比样,这个外部参比样的值与内部参比样的值原则上应相等。

3.6

斯蒂文心理物理功效函数 Stevens's equation psychophysical power function

其关系如下所示：

$$R = K S^n$$

式中：

R——评价员的反应(感觉强度)；

K——调整 S 和 R 所使用单位的常量；

S——刺激量(物理变量或化学物质的浓度)；

n——功效函数的指数,用对数 log 单位表示时, n 为 R 与 S 的回归曲线的斜率。

注：实际应用中,斯蒂文心理物理功效函数一般用自然对数表示, $\ln R = \ln K + n \ln S$ 。

4 原则

将样品连续提供给评价员,评价员将依据相同比例的原则,对每个样品的某一特性进行评分并记录。

参照同一类别的参比样的评分值进行评分。对这个参比样,评价员既可任意选一个值,也可以由检验负责人指定一个值,后一种情况称为“固定模数”。

5 基本检验条件

对于基本检验要求如设施、样品准备、提交和样品编号等参考 GB/T 10220 和 GB/T 13868 中的通用方法部分,或参考使用标度和分类方法方面的标准,如 GB/T 16290。

6 评价员的选拔和培训

6.1 选拔和培训的基本条件

选拔和培训基本条件应符合 GB/T 16291 和 GB/T 14195。

和其他的分析方法一样,评价小组组长应负责判断评价员具有的熟练程度。在制定培训计划时,应考虑检验的目标、评价员的可用性、召集新评价员所需经费以及评价员培训。评价员应经过 3 到 4 次检验培训后一般就可使用量值估计法进行评价。

6.2 量值估计法的专项内容培训

6.2.1 几何图形面积的评估已被证明特别适合于引导评价员掌握量值估计法的基本概念。下列的 18 个图形(表 1),包括 6 个圆形、6 个等边三角形和 6 个正方形,大小范围大致从 2 cm² 到 200 cm²,已经被成功地用来培训评价员。

表 1 培训图形的边长和面积

圆		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 引进两个边长为 11.1 cm 的正方形是为了评估评价员的可重复性能力。

6.2.2 将这些图形提供给评价员之前,告诉他们这种方法的基本原则。这些原则应包括(不必局限于)下列三点:

- 评分应按比例进行;如果某个特性的某一强度是另一强度的2倍,它的评分值也应是另一强度评分值的2倍;
- 使用的标度没有上限;
- 只有某一特性感觉不到时才能评为0分。

应告诉评价员,在培训时评分值通常用整数(如5、10、20、25等),但使用这种方法时,所有的数字都允许使用。由于评价员也容易受到培训中提到的比例的影响,所以建议他们使用不同的比例,例如3/1、1/3、7/5、5/6等比例,而不必仅限制于使用2/1或1/2。

6.2.3 在图上编号,并将这些图形分别放置在一张白色A4(21 cm×29.7 cm)纸的中央。指导每个评价员进行数字评估,从边长8.5 cm的正方形(外部参比样)开始,分别给出这一系列的正方形图形,记录反应。

根据检验阶段采用的程序,训练评价员用固定模数或非固定模数方法进行分析。使用固定模数进行培训时,检验负责人对边长为8.5 cm的正方形给出一个30到100之间的评分值。

用非固定模数进行培训时,让评价员自由选定第一个数值(参比样的值),但建议他们不要选择太小的值。

每次评分前,按随机的顺序给出样品,以保证对这些图形的形状和边长不形成特定的样式。

6.2.4 在完成一组图形的评定后,让评价员将各自评分结果与评价小组的平均值进行比较,如果这不可行,则将他们的评分结果与前一个评价小组的评分结果进行比较。

注:这样做的目的是提供正面的反馈,从而保证每一位评价员能够明白该培训的目的。注意不要造成一种有正确答案的印象。除非评分结果相差很大,否则应与评价小组平均评分结果有明显偏差的评分解释为顺序效应。

即由于评价员评估样品的顺序不同而产生了不同反应。应告诉他们,虽然个别评价员会受评分顺序的影响,但评价小组总体评分结果是准确的。

如果一些评价员的评分结果相差很大,应再次向这些评价员解释这种方法的原则。

6.2.5 当评价员成功完成面积估计的培训后,应根据在实际检验中待评价的产品和物质类型进一步培训。这可增加评价员在应用量值估计法区别待测物质特性方面的经验。评价小组组长可能需要设计一些练习,使评价员能正确认知待测物质的特性。这部分培训可根据GB/T 16291和GB/T 14195中的总体指南来设计。

7 评价员的数量

7.1 一般原则

与其他标度分析的检验技术一样,评价员的数量要根据:

- 待评价的不同产品的特性之间的接近程度;
- 评价员所接受的培训;
- 这个评价结果所得结论的重要性(见GB/T 16291和GB/T 14195);
- 可根据统计学的功效确定的目标。

在缺乏明显可确定目标时,参见本标准的7.2和7.3。

7.2 分析和研究评价小组

评价小组应由表2所列的成员组成。

表 2 评价小组的组成

评价员的类型	评价员最少数量	推荐人数
有经验的评价员,在所研究的产品及特性评估方面经过高度专业培训	5	10
有经验的评价员,在所研究的产品及特性方面经过专业培训	15	20 到 25
新培训的评价员	20	至少 20

统计学的功效应根据不同评分值之间的方差以及需要评价结果的精确度决定。

7.3 消费者评价小组

消费者评价小组和市场调查研究也可以使用量值估计法。消费者组的数量与检验类型所需的消费人群有关。就所需评价员数量方面而言,量值估计法并没有任何优势。其数目应与典型的消费者类型检验所需的数目相同,即至少 50 人,通常是更多。

注:对消费者而言,培训仅限于面积估计。

8 程序

8.1 样品的提供

所有样品用同一种方式提供(即相同的盛装容器、产品数量等)。盛装样品的容器尽量随机选用 3 位数字编号。

8.2 外部参比样

外部参比样对所有评价员应一样。研究样品的特性时,外部参比样的特性强度尽量与所有检验样品特性强度的几何平均值接近。

注:具有某个特性极值的参比样会引起偏差,即使产品间差异在允许范围内,这也降低这种方法的灵敏度。

一组检验样品中可以包括一个或多个已随机编号的参比样。这可以评估这一组评价员的可重复性能力。

8.3 样品的提供顺序

应将样品同时全部或依次提供给评价员。评价员应遵守指定的顺序进行评价。和所有感官分析一样,每个评价员的评价顺序是不一样的。理想的样品提供顺序是均衡的,评价小组组长可建议使用参考文献[1]中的拉丁方表综合考虑顺序设计和实行效果。如果这不可能,则使用随机顺序。

8.4 量值估计

8.4.1 一般要求

按 8.4.2 到 8.4.4 所描述几种方法中的一种进行检验。

8.4.2 无固定模数的外部参比样

每个评价员评估这个外部参比样,并给出评分值。建议评价员选值不宜太小。

接着,评价员将其他已编号样品与参比样依次比较,参照该评价员前面对该参比样所给评分值给出一个估计值。

8.4.3 有固定模数的外部参比样

评价小组组长要特别告诉评价员该参比样的值,如 30、50、100 或任何一个评价小组组长认为合适的值。

评价小组组长指导评价员参照对外部参比样给出的值对其他样品进行评分(固定模数)。

8.4.4 无外部参比样

无外部参比样时,也可使用量值估计法进行评价。由于感官系统适应性的限制,评价员很一致地参考第一个样品进行评分会很困难。无外部参比样时有两种解决方法:

a) 评价员在对后面的每个样品进行评分前,不必每次都重新对第一个样品进行评分。

鼓励评价员记住参比样的特性强度,必要时重新评价参比样的强度。

注 1: 这样产生的问题是,对某一个评价员而言,任何样品与参比样(第一个样品)之差的方差比任何两个待测样品之差的方差小 2 倍。

因此,可以通过:

- 在检验前,选择一种样品提供顺序的设计,使每个评价员的第一个样品不同;最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;这样才能使样品间的平均差异的方差相等。
- 在分析数据时,使用一个高权重的(理论上是无穷大)任意常数评估每个评价员对第一个样品的评分,以便正确估计差异的方差。

b) 评价员通过与紧邻的前一个样品比较对每个样品进行评分。

注 2: 这样产生的问题是,每一个评价员产生的评分误差自相关;两个相连样品之差的方差比两个不连续样品之差的方差要小。

因此,可以通过:

- 在检验前,选择一个样品提供顺序,这个顺序使样品的所有可能的排列能提供给相等数量的评价员;如果这不可能,尽量使所设计的提供顺序尽量接近理想模式;使样品间平均差的方差相等,或至少相当接近。

——数据分析时应用自相关误差模型,但这种方法稍为复杂。

必须说明的是,即使按照 a)(为进行评价而系统地与第一个样品即参比样进行比较)进行分析,仍要保留与前面样品的评价相关的自相关误差项,即使它可能很小(用 8.4.2 和 8.4.3 中所描述的参比样所进行检验结果也说明上述结论是对的)。因此,前面给出的样品提供的顺序要均衡的建议,在所有情况下都适用。

9 数据分析

9.1 数据分析方法的选择

见表 3。

表 3 数据分析方法的选择

重复次数	试验设计	数据处理	模型	产品间差异的比较			回归分析			
				方差来源	自由度	实例	方差来源	自由度	实例	
1	完全设计 $n=st$	取对数	正交	评价员处理 误差	$s-1$ $t-1$ $(s-1)(t-1)$	A. 1	评价员浓度的自然对数 评价员-浓度的对数 误差	$s-1$ 1 $s-1$ $s(t-2)$	A. 4	
≥ 2	完全设计 $n=str$	取对数	正交	简单的一种可能是首先计算每个评价员对同一处理的评分取对数的平均值。这就归结为在上面描述的 $r=1$ 的情况。(也可使用更复杂的模型考虑随机效果,如检查评价员与处理间的交互作用,或对同一个评价员的重复做出方差(协方差)矩阵的假设。)			A. 6	误差	$s(t-2)$	

表 3 (续)

重复次数	试验设计	数据处理	模型	产品间差异的比较			回归分析		
				方差来源	自由度	实例	方差来源	自由度	实例
1	不完全设计 $n \neq st$	取对数	非正交	评价员 处理 误差	$s-1$ $t-1$ $n-t-s+1$				
1	不完全设计 所有评价员评价相同的 子样品组	按全面重新标 度法校正后取 对数	正交	处理 误差(校正后)	$t-1$ $n-t-s+1$	A. 2			
1	不完全或完全设计 评价员对已确定参比 样或隐含参比样进行 评价	按参比样重新 标度后取对数	正交	处理 误差(校正后)	$t-1$ $n-t-s+1$	A. 4			
1	不完全设计 评价员用词汇描述的 评定进行赋值	外部重新标度 法重新标度后 取对数	正交	处理 误差(校正后)	$t-1$ $n-t-s+1$	A. 3			

注: t ——处理的数目; n ——总检验次数; s ——参加评定的评价员人数; r ——试验重复的次数。

9.2 原始数据的提供

结果可用双因素表格形式表示,行是经对数转换的评价员反应,列是不同样品。

当所有评价员对每个样品都做了相同次数的评分时,得到一个完全的均衡设计,这个具有评价员效应的模型是正交模型,如果某些样品没有被所有的评价员评估相同的次数,就是一个不完全设计,这个具有评价员效应的模型是非正交模型。

注:因为我们不能对 0 取对数,所以反应为 0 值会有问题。可以用不同的方法来处理 0 值。0 值应用很小的值代替。这个值的选择应考虑每个评价员所采用的标度(例如,评价员所评分最小值的一半)。

9.3 产品差异的确定

方差分析可以清楚解释所有的区组因素(包括不均衡或非正交因素),并用经对数转换的数据进行方差分析是最准确的方法。然而,不是所有的统计分析软件都可以进行不均衡设计的分析。实际上,并不总能保证所进行检验中所有的关键因素都是均衡和正交的完全设计。例如,当一个试验分几次进行时,每次试验的评价员组成可能不能都相同。建议统计员设计一个可能的最佳试验设计方案。在不能考虑试验设计的所有因素后再进行方差分析时,可以使用 9.5 中的一种重新标度方法进行校正。在那种情况下,把处理当作唯一的因素,对重新标度后的数据进行单因素方差分析,虽然不是最理想的方法,但也是可以解决问题的一种方法,可以计算出一个有整数的虽不准确但近似的结果。如果不是标准的试验设计,最好的解决方法是咨询统计学专家,以建立一个方差分析模型的适当格式。

当通过方差分析发现产品间有明显差异时,用平均值的多重比较方法中的一种进一步分析。

在完全设计(没有经过重新标度的数据)的产品特性差异比较的例子见 A.1。

9.4 回归分析

当变量 S (如浓度、物理量等)与反应 R 相关且值已知的情况下,如果它遵守斯蒂文法则,可以根据下列方程,对与这些物理和化学变量相关的感官反应进行线性回归分析估计它的参数:

$$\ln R = \ln K + n \ln S$$

在这个分析中,最重要的参数是与斯蒂文心理物理功效函数中 n 值所相对应的斜率。

也可以分析不同评价员间的回归曲线的斜率是否相等。

表 3 给出了不同情况下方差和回归分析的模式。

9.5 重新标度方法

注：如前所述，在不完全设计时，才会用这些方法。重新标度是评估评价员效应的一种方法。

9.5.1 全面重新标度

在均衡设计时，全面重新标度与常用的评价“评价员效应”的其他方差分析方法等效。

这种方法只有在所有的评价员评估了相同的一组样品(均衡的设计包括所有样品，不均衡的设计不包括所有样品)的情况下才能使用。

原理如下所述。因为每个评价员已经对相同一组样品进行评定，每个评价员对这一组相同样品的评定的总量级应相等，因此，每个评价员所使用的标度应调到使每一相同样品组的总量级相等。

如果采用这种重新标度方法，则误差的自由度应等于评价员人数减1，程序如下：

在被所有的评价员都评价过的所有样品里：

——计算每个评价员的评分值取对数后的平均值；

——计算所有评价员的平均值；

——计算出的每个评价员的校正正值，应使该校正值加上该评价员的平均值后，与这个评价小组的平均值相等；

——每个评价员对每个样品的评分值加上他(她)的校正正值。

全面重新标度的例子见附录 A 中的表 A.3 和表 A.4。

9.5.2 经参比样重新标度

如果一个或多个参比样(经随机编号)被插入到样品系列中，首先计算每个评价员对参比样(第一个样品或内部参比样，如果有的话)估计值的平均值。然后计算校正正值，这个校正正值应使参比样的平均值成为一个定值。将评价员的每个评分值乘以这个由参比样计算得来的校正正值，以校正由待测样品得到的数据。

注：值得说明的是，经全面重新标度程序比经参比样标度程序的方差综合分析得到的均方的误差要小。如 8.2 中所示，参比样的强度应与这个评价小组的所有样品强度的平均值接近。我们已经知道(见参考文献[3])，当参比样的强度与评价小组的几何平均值相等时，这个误差较小。参比样的评分值与实际几何平均值越接近越好。

9.5.3 外部重新标度

不同形式的外部重新标度方法已在参考文献中给出。在评定每个样品后，评价员会得到一个用词汇描述反应的标度，这个标度的范围可以包括 4 到 11 个等级。它包括下列表达，例如：

——极强；

——很强；

——中等；

——轻度，等。

评价小组组长要求评价员对这些用词汇描述的反应进行评分，这个评分与评价样品时使用的比例应一致。每个评价员给出的结果要用一个校正正值进行重新标度，通过全面重新标度法计算评价员对用词汇描述反应的估计值，得到该校正值。

见附录 A 的表 A.6、A.7 和 A.8 中外部重新标度的例子。

10 检验报告

检验报告包括以下内容：

——研究目的；

——检验结果；

——样品数量；

——所采用参比样(如果有的话)和这个样品特性(如果用的话)；

——检验重复次数；

- 评价员数量和水平；
- 检验一般条件，例如环境、日期和时间；
- 其他任何有助于评价检验可靠性的信息；
- 本标准参考文献的数量，以及对这种方法所做修改的说明；
- 检验负责人姓名。

附 录 A
(资料性附录)
数据分析示例

A.1 数据的方差分析及其解释

注：无重复，未重新标度的数据。

A.1.1 实例

表 A.1 为 7 个有经验的评价员对 6 种咖啡因含量不同的饮料中苦味强度的评价结果。括号内为估计值的自然对数。

表 A.1 6 个样品的数据

评价员	处理编号					
	561	274	935	803	417	127
	浓度/(mg/100 mL)					
	9	18	36	40	72	144
估计值(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)
算术平均值的自然对数	2.225	2.985	3.570	3.691	4.323	4.995

A.1.2 显著性差异分析

上表中的自然对数应用双因素方差进行分析结果见表 A.2。

表 A.2 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	0.244	0.046	4.35	0.0028
处理	5	32.905	6.581	704.71	0.0001
校正后的误差	30	0.280	0.009	—	—

方差分析显示，由于处理产生的影响差异显著。Tukey's 检验是用来检验每个样品间是否有显著差异的一种多重比较方法。在这个检验中，最小显著性差异按下式计算：

$$\sqrt{\left\{s^2 \times \frac{1}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)\right\}} \times c$$

式中：

s^2 ——校正后误差的均方；

n_i ——计算第一因素平均数时的观测数；

n_j ——计算第二因素平均数时的观测数；

c ——因子;在 a 水平上,校正后误差的自由度的函数,是总处理数的函数;由参考文献[2]的表 A.6 中获得。

在有 6 个处理样品和误差自由度为 30 的这个例子中,最小显著差异等于平均值的标准误差 ($\sqrt{\frac{0.009}{7}}$) 乘以 4.30,即 0.154。

任何显著性方式下都没有显著性差异的样品只有 803 和 935。他们的平均值仅相差 0.12。

A.2 在全面重新标度的情况下,数据的方差分析及其解释

在所有的评价员已经对样品 561、935、803 和 127 进行评价后,就可以在子分组中的样品上进行全面重新标度。

计算校正因子时,首先计算每个评价员在 4 个相同样品组中(表 A.3)估计数自然对数的平均值,接着计算整个评价小组估计数的自然对数平均值,最后,将整个评价小组的平均值减去每个评价员的平均值,计算出该评价员的校正因子。

表 A.3 评价员评分数的自然对数值和校正值的计算

评价员	处理编号						共同子分组的平均值	校正因子
	561	274	935	803	417	127		
	估计值(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
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
全组算术平均值							3.620	

校正后的数值的自然对数在表 A.4 中给出。

表 A.4 校正后估计值的自然对数值

评价员	处理编号					
	561	274	935	803	417	127
	估计值(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	3.571	3.704		4.992
算术平均值的自然对数	2.225	3.017	3.570	3.691	4.338	4.994

上述数据的方差分析结果见表 A.5。

表 A.5 方差分析结果

方差来源	自由度	平方和	均方	F 值	P > F
处理	5	30.417	6.083	768.01	0.0001
校正后误差	30	0.238	0.010		

对最小显著性差异的计算,必须考虑到不同处理的平均值不是通过对相同次数的观测计算获得(4个或7个评价员较合适)。可以使用 Tukey-Kramer 法进行计算(见参考文献[4])。

因此,样品(274,417)组是均有4个处理水平的一个组合,其最小显著差异等于:

$$\sqrt{\left\{0.010 \times \frac{1}{2} \left(\frac{1}{4} + \frac{1}{4}\right)\right\}} \times 4.37 = 0.219$$

[常量4.37由前面类似的一个表格(文献[2]中的表A.6)得到,该表格中自由度的数字经过校正]。

对(561,274),(935,274),(803,274),(127,274),(561,417),(935,417),(803,417)和(127,417)的各組,一个处理有7个水平(评价员),另一个处理只有4个水平(评价员),最小显著差异等于:

$$\sqrt{\left\{0.010 \times \frac{1}{2} \left(\frac{1}{7} + \frac{1}{4}\right)\right\}} \times 4.37 = 0.194$$

对所有有7个水平的其他处理组合,最小显著性差异等于:

$$\sqrt{\left\{0.010 \times \frac{1}{2} \left(\frac{1}{7} + \frac{1}{7}\right)\right\}} \times 4.37 = 0.165$$

因此,只有编号为803和935的两个处理之间没有显著性差异。

A.3 在外部重新标度情况下,数据的方差分析及其解释

在进行完主要试验后,评价员要对用词汇表达的衡量标度进行赋值。为了更形象地说明这点,建立了一个5点的评价范围(从“极苦”到“轻微苦”)。评价小组组长要求每个评价员对这些表达进行评分,这个评分与进行样品评分时所用的标度要一致。这个训练的假设结果如表A.6所示。

表 A.6 用词汇形容的评级的假设结果

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	-0.0025
2	5	30	60	100	160	-0.0880
3	5	25	50	100	150	-0.0025
4	5	20	45	90	140	+0.0980
5	5	25	50	100	150	-0.0025
6	3	30	55	110	170	0
7	5	25	50	100	150	-0.0025

^a 用全面重新标度法计算得到

首先用全面重新标度法计算出校正值,然后将每个评价员的评分值加上该评价员的校正值,来校正表A.7中的数据。并得到表A.8。

表 A.7 评价员评分的自然对数值

评价员	处理编号					
	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	

表 A.8 校正后估计值的自然对数值

评价员	处理编号					
	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	
全组算术平均值	2.264	2.984	3.572	3.693	4.313	4.998

表 A.9 方差分析结果

方差来源	自由度	平方和	均方	F 值	$P > F$
处理	5	27.771	5.554	516.74	0.0001
校正后的误差	30	0.323	0.011		

如 A.1.2 计算,自由度为 24 的 6 个处理样品之间最小显著性差异为 0.203,只有编号为 803 和 935 的两个处理之间没有显著性差异。

A.4 回归曲线斜率的分析

如果样品中的咖啡因含量增加,并且增加量已知,可对浓度(自然对数)和评分值(自然对数)进行线性回归分析,以评估回归曲线的斜率。如果评分值没有被重新标度,应为不同的评价员提供不同的纵坐标。

建议分析评价员和浓度自然对数之间的交互作用,以检验每个评价员的斜率是否相同。方差分析结果见表 A.10。

表 A.10 评价员与浓度自然对数两因素交互作用的方差分析结果

方差来源	自由度	平方和	均方	F 值	$P > F$
评价员	6	0.240	0.040	8.09	0.0001
浓度自然对数	1	33.129	33.129	6692.26	0.0001
评价员与浓度自然对数的交互作用	6	0.173	0.029	5.81	0.0005
误差	28	0.139	0.005	—	—

因为浓度的自然对数和评价员两因素的交互作用显著,有必要对每个评价员进行回归分析,以便得到各自的斜率。因此得到表 A.11。

表 A.11 回归曲线

评价员	斜率		r^2
	估计值	标准误差	
1	0.94	0.02	0.998
2	1.07	0.03	0.995
3	1.04	0.03	0.995
4	1.08	0.02	0.998
5	0.87	0.04	0.988
6	0.93	0.05	0.986
7	1.00	0.01	0.999

虽然两因素交互作用的影响显著,但它们的斜率之间没有显著差异。它们的平均值是 0.99,标准误差为 0.03。

A.5 在有重复的情况下,数据的方差分析及其解释

表 A.12 是由有相同数量的评价员的评价小组做出的重复估计数,其结果如表 A.1。

表 A.12 重复 1 和重复 2 的估计值

评价员	重复	处理编号					
		561	274	935	803	417	127
		估计值(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)

表 A. 12 (续)

评价员	重复	处理编号					
		561	274	935	803	417	127
		估计值(自然对数值)					
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)
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)
算术平均值的自然对数		2.277	2.981	3.563	3.678	4.317	4.995

一种方法是计算每个评价员和每个处理的平均值,并且如 A. 1 中的方差方法进行方差分析。另一种方法是使用一种将评价员效应和评价员与处理的交互作用看作是随机因素进行分析的模型。

表 A. 13 评价员和处理交互作用的方差分析结果

方差来源	自由度	平方和	均方	F 值	$P > F$
评价员	6	0.137	0.023	2.51	0.043
处理	5	32.167	6.433	705.84	0.000 1
误差	30	0.273	0.009	—	—

自由度为 30 的 6 个样品之间最小显著差异是 0.155。只有样品 803 和 935 之间没有显著性差异。

附录 B
(资料性附录)
提问模式

B.1 没有固定模数参比样的提问模式

姓名: 日期:

- 1) 提供一个编号为“R”的橙汁参比样给你
你品尝该参比样并根据酸度选择一个数值:

反应:

记住酸的强度。

- 2) 提供 6 个橙子饮料给你
你根据下列顺序评价它们。

参照参比样“R”所赋值,等比例对每个样品酸的强度给出一个评分值。
每次品尝样品之前,你必须重新品尝参比样。

样品 561

样品 274

样品 935

样品 803

样品 417

样品 127

B.2 有固定模数参比样的提问模式

姓名: 日期:

- 1) 提供一个编号为“R”的橙汁作为参比样品给你
对这个参比样的酸度给出的值是 50。

品尝样品记住它的酸度。

- 2) 提供 6 个橙子饮料给你

你要根据下列顺序来评价它们。

对每个样品酸的强度参照参比样“R”的值(50)等比例地给出一个评分值。
在品尝每个样品之前,你必须重新品尝参比样。

样品 561

样品 274

样品 935

样品 803

样品 417

样品 127

参 考 文 献

- [1] MACFIE, H. J., BRATCHELL, N., GREENHOFF, K. 和 VALLIS, L. V. 大厅试验中为消除提供顺序产生的影响和首次顺序执行的影响所进行的设计. 感官研究期刊, 4, 2, 1989
 - [2] GACULA, M. C Jr. 和 Singh, J. 食品和消费者研究中的数理统计方法. 美国奥兰多学术出版社, 1984
 - [3] POWERS, J. J., WARE, G. O. 和 SHINHOLSTER, K. J. 具有和缺少重新标度的量值估计法. 感官研究期刊, 5, 1999, 105~106
 - [4] KRAMER, C. Y. 具有不同重复次数的全组平均值的多领域试验的补充. *Biometrics*, 12, 1956, 307~310
-