



中华人民共和国国家标准

GB/T 19547—XXXX/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)

(征求意见稿)

202* - ** - **发布

20** - ** - **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 | I |
| 引言 | II |
| 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 | 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 评价员的数量 | 4 |
| 7.1 一般原则 | 4 |
| 7.2 分析和研究评价小组 | 4 |
| 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 | 5 |
| 8.3 样品的提供顺序 | 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 重新标度方法 | 7 |
| 10 检验报告 | 8 |
| 附录 A (资料性) | 9 |
| 附录 B (资料性) | 1 1 |
| 参考文献 | 27 |

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) 修改了“引言”，增加了 2 个示例和附录 B 中计算方法的解释；
- b) 增加了术语“内部参比样”、“模数”、“重新标度”（见 3.3、3.4 和 3.5）；
- c) 增加了“评价员的选拔和培训”中对消费者的培训（见 6.2.1，2004 年版的 6.2.1）；
- d) 增加了“评价员的数量”中统计学功效的决定要素，修改了消费者评价小组的人数要求（见 7.2 和 7.3，2004 年版的 7.2 和 7.3）；
- e) 更改了“评价员的数量”中消费者评价小组的人数（见 7.3，2004 年版的 7.3）；
- f) 修改了“数据分析方法的选择”中的内容（见 9.1，2004 年版的 9.1）；
- g) 修改了“重新标度方法”中的内容（见 9.5，2004 年版的 9.5）；
- h) 增加了“检验报告”中的内容（见第 10 章，2004 年版的第 10 章）；
- i) 修改了“附录 B”的内容（见附录 B，2004 年版的附录 A）。

本文件使用翻译法等同采用 ISO 11056:2021《感官分析 方法学 量值估计法》。

本文件由全国感官分析标准化技术委员会（SAC/TC566）提出并归口。

本文件起草单位：。

本标准主要起草人：。

本文件及其所代替文件的历史版本发布情况为：

——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方法，是通过评价员对某一感官特性进行评分的一种评价方法。该方法要求评价员给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 2 倍，则样品 B 的评分值应是样品 A 评分值的 2 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价。

相比专家定义的连续或非连续响应标度，量值估计法受“末端效应”的影响更小。当评价员对产品引发的感觉程度不熟悉时，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分配到接近标度边界一端的类别。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。

允许每位评价员从任何数字（即使用自己的标度）开始进行评分会导致“评价员”效应的影响增加，但有不同的方法来解决这个问题：

- 方差分析（ANOVA）将评价员效应和其交互效应考虑在内；
- 通过使用一个已赋值的参比样使评价员采用一个共同的标度范围；
- 通过使用一种重新标度方法，将每个评价员提供的数据纳入到一个共同的标度范围。

由主试者根据具体条件选择最佳的分析方法。

量值估计是确定斯蒂文斯心理物理学幂函数的特有方法，它也可用于解决具体问题。

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

示例 1：一家公司生产了一种市场销路很好的饮料，但最近竞争对手生产出更甜的产品抢占了该公司的市场份额。为了夺回部分市场损失，该公司决定将甜度提高三分之一。在配制新产品时，如果知道甜味剂的幂函数，就可以估算出甜度增加三分之一所需的甜味剂量。

示例 2：一款新研发的减肥饮料的配方中，所需甜味强度是已知的，但尚未决定使用阿斯巴甜或蔗糖作为甜味剂。已知每种物质的幂函数，可以绘制等甜度曲线，以确定达到所需甜度水平时所需每种甜味剂的浓度。这些信息与成本/产量信息相结合，可以帮助决定哪种甜味剂更具有成本效益。

附录 B 中的计算是使用 R 函数进行计算的。R 软件包可以免费访问。这些信息是为了方便本标准的用户而提供的，并不构成本文件对 R 包独家使用的认可或推荐。也可以使用其他软件来执行本标准所要求的计算。

R 软件包计算的相关文件位于 U 盘 F 中 ME 文件夹中（Text 格式（编号：附表））。

由于使用的软件不同，结果有时会因四舍五入的误差而有所不同。

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，包括评价员的培训、评估值的获得和数据的统计解释。

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 第一部分：一般统计术语（Statistics—Vocabulary and symbols—Part 1: General statistical terms and terms used in probability）

注：GB/T 3358.1—2009 统计学术语 第一部分：一般统计术语（ISO 3534-1:2006，IDT）

ISO 3534-3 统计学术语 第三部分：试验统计术语（Statistics—Vocabulary and symbols—Part 3: Design of experiments）

注：GB/T 3358.3—2009 术语 第三部分：试验设计术语（ISO 3534-3:1999，IDT）

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

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

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

注：GB/T 10221—2021 感官分析 术语（ISO 549:2008，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）

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 16291.1—2012 感官分析 选拔、培训与管理评价员一般导则 第1部分：优选评价员（ISO 8586:2023，IDT）

注：GB/T 16291.2—2010 感官分析 选拔、培训与管理评价员一般导则 第2部分：专家评价员（ISO 8586:2023，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 16290—2020 感官分析 定量响应标度使用导则（ISO 4121:2003，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：

——IEC电子百科：<http://www.electropedia.org/>；

——ISO在线浏览平台：<https://www.iso.org/obp>。

3.1

量值估计 magnitude estimation

对产品的某一特性强度或喜好程度进行赋值的过程，不同赋值之间的比例与被试对相应特性的感知强度间的比例一致。

3.2

外部参比样 external reference

提供给评价员的样品系列中的第一个样品作为参比样，随后待测的所有样品都依据该样品进行评价。

3.3

内部参比样 internal reference

提供给评价员的待测样品系列中的第一个测试样，随后的所有样品都依据该样品进行评价。

3.4

模数 modulus

由检验负责人给定（固定模数）或通过评价员选择（非固定模数）的外部参比样（3.2）的数值。

3.5

重新标度 rescaling

将每个评价员给出的原始数据乘以一个因子的过程，该因子将所有评价员的数据纳入到一个相同标度的范围。

注：将校正因子与原始数据分别取对数值后再进行加和，等同于上述过程。

3.6

斯蒂文斯心理物理幂函数 Steven's equation psychophysical power function

其关系如下所示：

$$R = K S^n$$

式中：

R ——评价员的响应（如感知的强度）；

K ——调整 S 和 R 所使用单位的常量；

S ——刺激量（物理量或化学物质的浓度）；

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

注：实际应用中，斯蒂文心理物理幂函数一般转化成自然对数表示： $\text{In}R = \text{In}K + n\text{In}S$ 。

4 原则

将样品连续提供给评价员，评价员采用比例原则对每个样品的某一特性强度进行赋值记录。

参照同一样品系列中的第一个样品（参比样）赋值进行评分。对于参比样，既可由每个评价员自由赋值，也可由检验负责人指定赋值，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和样品的准备、呈送和编码等应参照感官分析通用方法学的相关标准，如 ISO 6658 和 ISO 8589，或参考标度和分类相关方法标准，如 ISO 4121。

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 ISO 8586。

和其他的分析方法一样，评价小组组长负责判断评价员所需技能的熟练程度。在制定培训计划时，应考虑检验的目的、评价员的可用性以及召集和培训新评价员的所需经费。通常经过 3 到 4 次检验培训后评价员就可以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 已证实几何图形面积的评估特别适合引导评价员掌握量值估计法的基本概念。下列 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,尺寸范围约为 $2\text{ cm}^2\sim 200\text{ cm}^2$,目前已被成功用于评价员的培训。

对于消费者评价小组,可使用简化版培训内容,如培训内容可仅限于面积估计。

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 这里2个边长为11.1 cm的正方形是为了评估评价员的再现性。

6.2.2 将图形提供给评价员前,告诉评价员该方法的基本原则。这些原则应包括但不局限于下列 3 点:
 ——评分应按比例进行:如果某特性强度是另一个的 2 倍,评分值也应是另一个评分值的 2 倍;
 ——使用的标度没有上限;
 ——只有某一特性感觉不到时才能赋值 0 分。

告诉评价员,培训时的评分值通常使用整数(如 5、10、20、25 等)。但是使用这种方法时,所有的数字都是被允许使用的。

评价员也容易受到培训中提及的比例影响,因此建议他们使用不同的比例,例如 3/1、1/3、7/5、5/6 等比例,而不必仅限于使用 2/1 或 1/2。

6.2.3 对图形编号,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。

指导每个评价员对图形进行量值估计,从边长 8.5 cm 的正方形(外部参比样)开始,分别给出这一系列正方形图形的响应值并进行记录。

根据测试阶段采用的程序,对评估员进行有固定模数或无固定模数的培训。如果采用固定模数,评价员根据一个边长为 8.5 cm 的正方形对边长介于 30 cm~100 cm 之间的正方形进行赋值。

用非固定模数进行培训时,让评价员对第一个图形进行自由赋值,但建议评价员不要选择太小的数值。

每次评分前,样品的呈送遵循随机顺序,以保证这些图形的形状和边长不会形成特定的模式。

6.2.4 在完成一组图形的评估后，让评价员将各自评分结果与评价小组的平均值进行比较。如果实际操作中无法实现，则将评分结果与前一个评价小组的评分结果进行比较。

这样做的目的是提供正面反馈，保证每一位评价员能够明确培训的目的。宜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，否则宜将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顺序的影响。宜向评价员保证，尽管存在个体顺序影响，但评价小组的结果是准确的。

如果部分评价员的评分结果相差很大，应重新向这些评价员解释量值估计法的原则。

6.2.5 当评价员成功完成面积估计的培训后，宜根据在实际检验中待评价产品和物质类型开展进一步的培训。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，使评价员能正确识别待测特性。这部分培训可参见ISO 8586。

7 评价员的数量

7.1 一般原则

与其他标度分析方法一样，评价员的数量主要依据下列方面：

- 多个待测产品特性之间的相似程度；
- 评价员接受的培训；
- 这个评价结果所得结论的重要性（见 ISO 8586）；
- 根据统计学的效度能够确定的目标。

如果没有明确的目标时，参见本文件的 7.2 和 7.3。

7.2 分析和研究小组

评价小组成员的数量应满足表 2 中的要求。

统计学的效度应根据个体评分之间的方差以及需要检测结果的差异程度来决定。

表 2 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

消费者评价小组和市场调查研究也可使用量值估计法。消费者选择的数量应根据这些测试类型所需的消费者抽样要求来确定。就所需评价员到的数量而言，量值估计法并没有任何优势。消费者小组数量应与典型消费者类型检验所需的消费者数量相同，即至少 60 人，通常是更多。

8 程序

8.1 样品的呈送

所有样品应采用同一种方式提供（如相同的盛装容器和产品量等）。
盛装样品的容器应采用随机三位数字进行编码。

8.2 外部参比样

对于待测特性，参比样的特性强度尽量与所有待测产品特性强度的几何均值接近。

注：具有某个特性极值的参比样会引发失真。

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

8.3 样品的呈送顺序

应将样品一次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不同，理想的样品呈送顺序是均衡送样。

评价小组组长可参照参考文献中的拉丁方表的平衡顺序设计和实施效果^[3]。如果无法采用拉丁方设计，使用随机顺序。

8.4 量值估计

8.4.1 一般要求

按 8.4.2~8.4.4 所描述方法中的一种进行检验。
参比样的调查问卷模式见附录 A。

8.4.2 无固定模数的外部参比样

每个评价员评估外部参比样并进行赋值。建议评价员选值不要太小。

接着，评价员将其他已编码的样品与参比样依次比较，参照该评价员前面给参比样的评分值对样品进行赋值。

8.4.3 有固定模数的外部参比样

评价小组组长要向评价员指出该外部参比样的值，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的评分值（固定模数）对其他样品进行赋值。

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（如记忆）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 2 种解决方案：

a) 评价员对后面的每个样品进行评分前，无需再重新评价第一个样品。

鼓励评价员记住参比样的特性强度，必要时再重新评价参比样的强度。

因此，可通过下列 2 种方式：

——检验前：设计一个样品呈送顺序，使每个评价员的第一个样品都不同；最理想的设计宜为每个

样品都有相同数量的评价员将其作为参比样；这样样品间平均差异的方差才能相等。

——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) 评价员与紧邻的前一个样品比较，对每个样品进行评价。

注：这样产生的问题是，每一个评价员产生的评分误差自相关，同时两个相邻样品差异的方差小于两个不相邻样品差异的方差值。

因此，可通过下列2种方式：

——检验前：设计一个样品呈送顺序，这个顺序使样品的所有可能的排列方式都能提供给相等数量的评价员；如果无法实现，尽量使设计的样品呈送顺序接近理想模式；使样品间平均差异的方差相等，或尽量接近。

——数据分析时：应用自相关误差模型，但这种方法稍微更复杂。

必须说明的是，即使按照 a)（通过系统与参比样比较方式进行样品评价）进行分析，仍要保留与前面样品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 中所描述的参比样进行的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

分析方法取决于（见附录B）；

——试验设计：完全设计或不完全设计，

——是否存在重复，

——归因于评价员因子（固定因子或随机因子）的状态，样品因子通常设为是固定因子。

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响应值，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，得到一个完全平衡设计方案，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。

由于不能对0取对数，所以响应值为0时会有问题。有多种方法处理样品的0值。宜用很小的数值代替0值，这个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评分值的一半）。

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（包括不平衡设计或非正交因素），而且用经对数转换的数据进行方差分析是最准确的方法。实际上，并不总能保证所进行检验中所有的关键因素都是平衡和正交的完全设计。例如，当一个试验分几次进行时，可能无法每次试验时都是完全相同的评价员组成。推荐咨询统计学专家以建立一个最佳的试验设计。

当通过方差分析发现产品间存在显著差异时，采用均值多重比较方法中的一种进一步分析。一个采用完全设计（没有经过重新标度的数据）的产品特性差异比较的示例见B.1。

9.4 回归分析

在已知变量值（如 S 浓度和物理量等）与响应值 R 的情况下，如果遵循斯蒂文斯定律，可以根据公式（1）对与这些物理变量和化学变量相关的感官响应进行线性回归分析并估计模型参数：

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

在这个分析中，最重要的参数是与斯蒂文斯心理物理幂函数中 n 值相对应的回归曲线的斜率。也可以分析不同评价员间回归曲线的斜率是否相等。

9.5 重新标度方法

9.5.1 全面重新标度

原因如下所述。因为每个评价员都评价的是同一组样品，每个评价员对这一组样品响应的总量值宜相等。因此，对于同一组样品，每个评价员所使用的标度具有相同的总量值。

程序如下：

在被所有的评价员都评价过的所有样品中：

——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。

对于每一位评价员：

——计算出的每个评价员的校正值，该校正值加上评价员自身的均值后将与评价小组的均值相等；

——每个评价员对所有样品的评分值加上自身的校正值。

全面重新标度的示例见附录B.2。

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（第一个样品或可能存在的隐藏参比样）估计值的平均值。然后计算校正值，使参比样均值成为一个定值。将每个评价员的评分值乘以这个由参比样计算得来的校正值，对待测样品获得的数据进行校正。

值得说明的是，全面重新标度法的方差综合分析得到的均方误差要小于经参比样的重新标度法。如8.2中所示，参比样的强度宜与该评价小组所有样品强度的几何均值接近。当已知参比样的强度与评价小组几何均值相等时，这时的误差较小^[6]。参比样的评分值与实际几何均值越接近越好。

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。样品评价完后，评价员会得到一个包括4~11级测量范围的语义标度来描述响应，相关表述示例如下：

——极强；

——很强；

——中等；

——稍强，等。

评价小组组长要求评价员对这些用词汇描述的响应进行量值估计，这个评分与评价样品时使用的比例一致。每个评价员的评价结果都用一个校正值进行重新标度，该校正值是使用全面重新标度法计算评价员对语义响应标度的估计值得到的。

外部重新标度的示例见附录B.4。

10 检验报告

检验报告应包括以下内容：

- 研究目的；
- 检验结果；
- 样品数量和样品描述；
- 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；
- 评价员数量和资质水平；
- 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；
- 本文件的标准号，以及对这种方法所做修改的说明；
- 检验负责人姓名；
- 检验的时间。

附录 A

(资料性)

问卷模式

A.1 无固定模数参比样的问卷模式

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的强度给出一个数值：

响应值：

记住该样品酸的强度。

2) 提供6个橙汁饮料给你

请根据下列顺序依次评价6个样品。

参照参比样“R”的赋值，等比例对每个样品酸的强度进行赋值。每个样品品尝之前，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

A.2 有固定模数参比样的问卷模式

姓名：_____ 日期：_____

1) 提供一个编号为“R”的橙汁参比样给你

该参比样的酸度赋值是50。

品尝样品并记住它的酸度。

2) 提供6个橙汁饮料给你

请根据下列顺序依次评价6个样品。

参照参比样“R”的值（50），等比例对每个样品酸的强度进行赋值。每个样品品尝之前，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

附录 B
(资料性)
数据分析示例

B.1 无重复，无重新标度数据的数据分析——所有评价员一次评估完该系列所有产品

B.1.1 示例

表 B.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。对于评价员 1、2 和 3，样本 274 是外部参比样，强度赋值是 20。对于评价员 5、6 和 7，样本 803 是外部参比样，强度赋值是 40。最后，对于评价员 4，样本 935 是参比样，强度赋值是 32。对于每个评价员，外部参比样是待测样品系列中的第一个样品，其余 5 个样品以随机顺序呈现，评价员之间都不同。评价员并未对外部参比样（表 B.1 中外部参比样的赋值是由小组组长给出）赋值，但他们对参比样都至少品尝了 5 次（5 种待测产品品尝之前，每次都会先品尝参比样 1 次），并将其与小组组长给出的参比样赋值进行对应。

括号内为估计值的自然对数值，保留 3 位小数。

表 B.1 6 个样品的数据

| 评价员 | 样品编号 | | | | | |
|-------------|----------------|------------|------------|------------|------------|-------------|
| | 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 (5.010) |
| 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) |
| 自然对数值的均值 | 2.225 | 2.985 | 3.570 | 3.691 | 4.323 | 4.995 |

此文件存在 U 盘 F 中 ME (2019) 文件夹下，名字为：附表 B1。通过以下三个命令将其导入 R：

```
tableb1<- read.table ("F:/ME(2019)/Table Annex B1.txt", header = T, sep="\t", dec=".")
attach(tableb1)
names(tableb1)
```

该文件中表 b1 有 7 列（评价员、样品、评分、评分的自然对数值、重新标度评分的自然对数值、浓度的自然对数值）和 42 行（6 个样品×7 个评价员）。命令是：

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

从而得到表 B.1 中“自然对数值的均值”，具体如下：

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| 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 命令对表 B.1 的自然对数值进行双因素方差分析：

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

分析结果如表 B.2 所示，样品因子的效应非常显著。结果是很合理的，因为样品间咖啡因浓度存在很大差异（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 | 753.11 | < 2×10 ⁻¹⁶ |
| 残差 | 30 | 0.26 | 0.009 | — | — |

多重比较测试：图基（Tukey's）检验是检验两两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：

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

式中：

s^2 —表 B.2 中残差项的均方；

n_i —计算第一样品组均值时的观测值数量；

n_j —计算第二样品组均值时的观测值数量；

C —因子，根据残差的自由度、总样品量和所选的 α -风险计算的函数，如表 B. 6 给出了该值。学生化极差的临界值参见参考文献^[1]。

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

$$4.30 \times \sqrt{0.5 \times 0.009 \times \left(\frac{1}{7} + \frac{1}{7}\right)} = 0.154 \quad (\text{B.2})$$

任何显著性分析方法下均没有显著性差异的样品对只有 803 和 935 这一组。自然对数的均值仅相差 0.121。这个结论是合理的，样品 803 和 935 的咖啡因浓度很接近，浓度对数值仅相差 0.046，而其他相邻样品对之间的咖啡因浓度对数差分别为 0.255（样品对 417 和 803）、0.301（样品对 274 和 561、935 和 274 以及 127 和 417）。

上面的计算可通过使用 R 的 TukeyHSD 命令获得：

```
TukeyHSD(aov(LogScore ~ Assessor + Treatment, data= b1), "Treatment")
```

这个命令得到：

| 样品 | 差异 | 置信下限 | 置信上限 | P_{adj} |
|-------|-------|-------|-------|------------------|
| 2T-1T | 0.763 | 0.608 | 0.913 | 0.000 |
| 3T-1T | 1.345 | 1.193 | 1.498 | 0.000 |
| 4T-1T | 1.467 | 1.314 | 1.619 | 0.000 |
| 5T-1T | 2.098 | 1.945 | 2.250 | 0.000 |
| 6T-1T | 2.770 | 2.617 | 2.922 | 0.000 |
| 3T-2T | 0.585 | 0.432 | 0.738 | 0.000 |

| | | | | |
|-------|-------|--------|-------|-------|
| 4T-2T | 0.706 | 0.554 | 0.859 | 0.000 |
| 5T-2T | 1.338 | 1.185 | 1.490 | 0.000 |
| 6T-2T | 2.009 | 1.857 | 2.162 | 0.000 |
| 4T-3T | 0.121 | -0.031 | 0.274 | 0.182 |
| 5T-3T | 0.753 | 0.600 | 0.905 | 0.000 |
| 6T-3T | 1.424 | 1.272 | 1.577 | 0.000 |
| 5T-4T | 0.631 | 0.479 | 0.784 | 0.000 |
| 6T-4T | 1.303 | 1.150 | 1.456 | 0.000 |
| 6T-5T | 0.672 | 0.519 | 0.824 | 0.000 |

对于每一个样品对，该命令可以得出：

- a) 两个样品之间的差异值；
- b) 置信水平为 95%的置信下限 (lwr) 和置信上限 (upr)；
- c) *p* 值 (“adj”这个词是没用的)。

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

注 1：当评价员因子设为随机因子时，必须使用 lmer()命令 (来自 lmerTest R 包)。这些命令是：

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

由于采用平衡试验设计，因此样品因子的 *p* 值相同的 (与表 B.2 中的数值相同)。但“评价员”因子的 *p* 值 (0.006) 不同，这一数值高于表 B.2 中的数值 (0.002)。

当“评价员”因子设为随机因子而非固定因子时，“评价员”效应通常偏低 (因此 *p* 值较高)。

注 2：对表 B.1 每个评价员的 6 个评分值数据进行重新标度，如 B.2 所示。在文件中表 b1 中，重新标度后的评分值在 R 语言命令的列“重新标度评分的自然对数值”中。

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

结果如表 B.3 所示。重新标度后结果是 7 个评价员的总和均相同，等于 3.631，并且“评价员”的平方和为 0。

表 B.3 表 B.1 的双因素方差分析在重新标度后的结果

| 方差来源 | 自由度 | 平方和 | 均方 | <i>F</i> 值 | <i>Pr</i> (> <i>F</i>) |
|------|-----|-------|-------|------------|-------------------------|
| 评价员 | 6 | 0.00 | 0.000 | 0 | 1 |
| 样品 | 5 | 33.17 | 6.635 | 739 | < 2×10 ⁻¹⁶ |
| 残差 | 30 | 0.26 | 0.009 | - | - |

B.2 无重复，有内部重新标度数据的数据分析——所有评价员一次评估完该系列所有产品

表 B.4 是一个研究项目的结果，该项研究中感官小组成员给一系列蔗糖溶液的感知甜度进行赋值。使用 2%蔗糖溶液作为参比样，所有评价员给该参比样的赋值为 10 分。参比样的评分值不参与数据分析中。2%蔗糖参比样同时也作为隐藏样品进行评分。

表 B.4 6 个蔗糖样品的数据集

| 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：各评价员的几何均值计算如下： $GM = \sqrt[n]{(x_1) \times (x_2) \dots (x_n)}$

例如，小组成员 L.M.， $GM = \sqrt[6]{(5 \times 1 \times 10 \times 20 \times 50 \times 100)} = 13.08$

表 B.4 经过自然对数转化后得到表 B.5。

表 B.5 表 B.4 数据经过自然对数转化后的数据

| 蔗糖浓度% (对数值) | L.M. | B.W. | M.M. | H.H. | J.D. | J.F. | |
|----------------|--------|-------|-------|-------|-------|--------|-------|
| 0.5 (-0.693) | 1.609 | 0 | 0 | 0 | 0.693 | -0.693 | |
| 1 (0.000) | 0 | 1.386 | 0.916 | 0 | 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 | |
| 算术均值 | 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 的总均值等于 2.393。对于每个评价员，校正因子用于重新标度评分值。校正因子等于小组总均值与评价员均值的差值。例如，对于评价员 L.M.，校正因子等于 2.393-2.571=-0.178。因此，校正后 0.5%蔗糖浓度的评分值为 1.609-0.178=1.431，1%蔗糖浓度的评分值为 0-0.178=-0.178，2%蔗糖浓度的评分值为 2.303-0.178=2.125。

计算结果见表 B.6。

表 B.6 表 B.4 数据进行全面重新标度后的数据

| 蔗糖浓度%的对数值 | 评价员 | | | | | | 算术均值 |
|-----------|--------|-------|-------|--------|-------|--------|-------|
| | L.M. | B.W. | M.M. | H.H. | J.D. | J.F. | |
| -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 | 0.614 |
| 0.693 | 2.125 | 2.411 | 2.128 | 1.449 | 2.43 | 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 | 4.420 | 4.431 | 5.544 | 3.816 | 5.351 | 4.598 |
| 算术均值 | 2.393 | 2.393 | 2.393 | 2.393 | 2.393 | 2.393 | |

此文件存在 U 盘 F 中 ME (2019) 文件夹下，名称为：附表 B2。通过以下三个命令将其导入 R：

```
tableb6 <- read.table ("F:/ME(2019)/Table Annex B2.txt", header = T, sep="\t", dec=".")
attach(tableb6)
names(tableb6)
```

文件中表 b6 有 7 列（评价员、Concsu、浓度、浓度的自然对数值、评分、评分的自然对数值和重新标度评分的自然对数值）和 36 行（6 名评价员×6 个浓度）。命令：

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

得出表 B.6 中的平均值如下：

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| 1c | 2c | 3c | 4c | 5c | 6c |
| 0.268 | 0.614 | 2.034 | 3.053 | 3.789 | 4.598 |

评价员具有相同的自然对数值的均值 2.393，由命令得到的：

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

使用 R 命令对表 B.6 的数据采用双因素方差分析：

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

结果见表 B.7。评价员因子的平方和等于 0，因为数据已全面重新标度。

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 | 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。R 命令如下：

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

R 命令显示，具有相邻浓度的样品对 2c-1c 和 3c-2c 存在显著差异 ($p < 0.05$)，但是对于样品对 4c-3c、5c-4c 和 6-5c 之间不存在显著性差异 (p 值分别为 0.07、0.31 和 0.22)。

这个命令得到：

| 样品 | 差异 | 置信下限 | 置信上限 | p_{adj} |
|-------|-------|--------|-------|-----------|
| 2c-1c | 0.347 | -0.728 | 1.421 | 0.916 |
| 3c-1c | 1.766 | 0.691 | 2.841 | 0.000 |
| 4c-1c | 2.785 | 1.796 | 3.859 | 0.000 |
| 5c-1c | 3.522 | 2.466 | 0.460 | 0.000 |
| 6c-1c | 4.330 | 3.255 | 0.541 | 0.000 |
| 3c-2c | 1.420 | 0.345 | 2.495 | 0.005 |
| 4c-2c | 2.438 | 1.363 | 3.513 | 0.000 |
| 5c-2c | 3.175 | 2.100 | 4.250 | 0.000 |
| 6c-2c | 3.984 | 2.909 | 5.059 | 0.000 |
| 4c-3c | 1.018 | 0.057 | 2.093 | 0.071 |

| | | | | |
|-------|-------|--------|-------|-------|
| 5c-3c | 1.755 | 0.680 | 2.830 | 0.000 |
| 6c-3c | 2.564 | 1.489 | 3.639 | 0.000 |
| 5c-4c | 0.737 | -0.338 | 0.181 | 0.313 |
| 6c-4c | 1.546 | 0.471 | 2.621 | 0.002 |
| 6c-5c | 0.809 | -0.266 | 1.884 | 0.224 |

B.3 子组样品进行内部重新标度情况下的数据分析

假定表 B.1 示例中，所有评价员并没有对系列样品中的所有样品进行评价。评价员 2、3、4、5、6 和 7 都只评价了 6 个样品，试验采用非平衡设计。

实际上，表 B.4 与表 B.1 相同，只是没有删去了评价员 2、4 和 6 对样品 274 的评分值以及评价员 3、5 和 7 对样品 417 的评分值。

在所有评价员对样品 561、935、803 和 127 进行评价后，就能对该子组样品进行重新标度，步骤如下：

- 首先，计算每个评价员对子组中 4 个样品估计值自然对数值的均值（例如： $(2.303+3.555+3.689)/4=3.622$ ，见表 B.8），
- 其次，计算整个评价小组估计值自然对数值的均值（等于 3.620）
- 再次，将整个评价小组的均值减去每个评价员的均值，计算每个评价员的校正值（如评价员 1 的校正值： $3.620-3.622=-0.002$ ）
- 最后，用这个值修正每个评价员的评分，得到表 B.9。

表 B.8 评价员评分值的自然对数值和校正值计算

| 评价员 | 样品编号 | | | | | | 共同子组的平均值 | 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 |
| 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 | 3.604 | +0.016 |
| 全组均值 | — | — | — | — | — | — | 3.620 | — |

表 B.9 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 |
| | 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 | 2.213 | 2.906 | 3.571 | 3.705 | — | 4.993 |
| 全组均值 | 2.225 | 3.017 | 3.570 | 3.691 | 4.338 | 4.994 |

此文件存在 U 盘 F 中 ME (2019) 文件夹下，名称为：附表 B3。通过以下三个命令将其导入 R：

```
tableb9<- read.table ("F:/ME(2019)/Table Annex B3.txt", header = T, sep="\t", dec=".")
attach(tableb9)
names(tableb9)
```

文件中表 b9 有 5 列（评价员、样品、样品方法、评分的自然对数值、重新标度评分的自然对数值）和 36 行（评价员 1 有 6 行，评价员 2、3、4、5、6 和 7 各 5 行共计 30 行）。R 命令：

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

得出表 B.9 中“自然对数的算数平均值”的以下值：

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

使用 R 命令对表 B.9 的数据采用方差分析：

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

结果见表B.10。

表 B.10 表 B.9 的方差分析结果

| 方差来源 | 自由度 | 平方和 | 均方 | F值 | Pr> F |
|------|-----|--------|-------|---------|----------------------|
| 样品 | 5 | 30.416 | 6.083 | 641.543 | <2×10 ⁻¹⁶ |
| 评价员 | 6 | 0.010 | 0.002 | 0.173 | 0.982 |
| 残差 | 24 | 0.228 | 0.009 | — | — |

“样品”因子具有很高的显著性。由于仅对 4 个样品进行重新标度，因此评价员因子的平方和不等 于 0。R 命令给出每个评价员的均值：

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

获得

| As1 | As2 | As3 | As4 | As5 | As6 |
|-------|-------|-------|-------|-------|-------|
| 3.620 | 3.780 | 3.501 | 3.773 | 3.756 | 3.478 |

多重比较测试：对最小显著性差异的计算，必须考虑到不同样品的均值并非通过相同数量的观测值 计算获得（4 个评价员评价了 274、417 和 121，7 个评价员评价了 567、935 和 803）。可采用图基检验 进行计算^[2]。该示例中（6 个样品和 24 个残差自由度），当α-风险为 0.05 时，C 等于 4.37。

因此，样品组（274、417）是汇有 4 个处理的一个组合，其最小显著性差异等于：

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

样品组（561、274）的最小显著差异等于：

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

最后，对于样品组（561、935）的最小显著性差异等于：

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

最终，只有样品对 803 和 935 之间没有显著性差异。这一结论与 B.1 相同。

上述计算可通过使用 R 命令直接得到：

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

该命令得到：

| 样品 | 差异 | 置信下限 | 置信上限 | <i>p</i> _{adj} |
|-------|-------|--------|-------|-------------------------|
| 2T-1T | 0.792 | 0.603 | 0.981 | 0.000 |
| 3T-1T | 1.345 | 1.184 | 1.506 | 0.000 |
| 4T-1T | 1.467 | 1.306 | 1.628 | 0.000 |
| 5T-1T | 2.113 | 1.925 | 2.302 | 0.000 |
| 6T-1T | 2.770 | 2.609 | 2.931 | 0.000 |
| 3T-2T | 0.553 | 0.365 | 0.742 | 0.000 |
| 4T-2T | 0.675 | 0.486 | 0.863 | 0.000 |
| 5T-2T | 1.321 | 1.108 | 1.534 | 0.000 |
| 6T-2T | 1.978 | 1.789 | 0.166 | 0.000 |
| 4T-3T | 0.121 | -0.040 | 0.282 | 0.210 |
| 5T-3T | 0.768 | 0.579 | 0.957 | 0.000 |
| 6T-3T | 1.424 | 1.263 | 1.585 | 0.000 |
| 5T-4T | 0.647 | 0.458 | 0.835 | 0.000 |
| 6T-4T | 1.303 | 1.142 | 1.464 | 0.000 |
| 6T-5T | 0.656 | 0.468 | 0.845 | 0.000 |

注：当评价员因子设为随机因子时，函数 lmer()（来自 lmerTest R 包）：

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

得到样品因子的 *F* 值（768.74）和 *p* 值（ 2.2×10^{-16} ）。对于评价员因子，*p* 值等于 1：评价员之间的差值为零。

B.4 外部重新标度情况下的数据分析

在完成主要试验后，评价员对语义类项标度进行赋值。具体而言，建立一个 5 点标度（从“极苦”到“轻微苦”），评价小组组长要求每个评价员采用待测样品评价时所使用的标度对这些语义描述进行赋值。这个训练的假设性结果如表 B.11 所示。

小组组长要求评估员按照评估测试样本时所使用的标度，为这些表达式分配量级。与评估测试样本时使用的标度一致。

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 | -0.0025 |
| 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 计算同B.2中保持一致 | | | | | | |

对表 B.12 中数据进行重新标度处理，结果如表 B.13，具体步骤如下：

——首先，用全面重新标度法计算出校正值；

——其次，将每个评价员的评分值加上该评价员的校正值，校正表 B.8 中的数据。

需注意，表 B.12 中，只有评价员 1 评价了该系列中的所有 6 个产品，其他评价员只评价了 5 个样品（如评价员 2 没有评价样品 274）。因此，试验是不平衡设计。还要注意，表 B.11 中的空格与表 B.8 中的空格不相同。如样品 561，表 B.8 中有一个空格（评价员 2 的单元格），但表 B.11 中没有空格。

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 | — |

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 | 2.300 | 2.993 | 3.553 | 3.686 | 4.246 | 4.939 |
| 2 | — | 2.908 | 3.550 | 3.696 | 4.355 | 4.987 |
| 3 | 2.077 | — | 3.581 | 3.686 | 4.315 | 5.008 |
| 4 | 2.044 | 2.806 | — | 3.709 | 4.346 | 5.003 |
| 5 | 2.482 | 3.216 | 3.635 | — | 4.315 | 4.974 |
| 6 | 2.485 | 3.091 | 3.555 | 3.689 | — | 5.075 |
| 7 | 2.195 | 2.888 | 3.553 | 3.686 | 4.302 | — |
| 全组均值 | 2.264 | 2.984 | 3.571 | 3.692 | 4.313 | 4.998 |

此文件在 U 盘 FM 中 E（2019）文件夹下，名称为：附表 B4。通过以下三个命令将其导入 R：

```
tableb13<- read.table("F:/ME(2019)/Table Annex B4.txt", header = T, sep="\t", dec=".")
attach(tableb13)
names(tableb13)
```

文件中表 b13 有 4 列（评价员、样品、评分、评分的自然对数值）和 36 行（评价员 1 有 6 行和评价员 2、3、4、5、6 和 7 的 30 行）。命令：

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

得出表 B.13 中“全组均值”：

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

使用 R 命令对数据采用方差分析：

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

结果见表 B.14。由于试验是不平衡设计，“评价员”因子的平方和不等 于 0。样品因子具有非常高的显著性。

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 | 27.770 | 5.554 | 669.077 | <2×10 ⁻¹⁶ |
| 评价员 | 6 | 0.122 | 0.020 | 2.455 | 0.054 |
| 残差 | 24 | 0.199 | 0.008 | — | — |

如 B.12 中所述，6 种样品，每种样品有 6 个观测值，残差的自由度为 24，当 α 风险=0.05 时，LSD 按以下公式计算：

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

样品对 935 和 803 之间没有显著性差异。这一结论与 B.1 和 B.3 的结论相同。该结果也由 R 命令得到：

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

该命令得到：

```
Tukey multiple comparisons of means
 95 % family-wise confidence level

Fit: aov(formula = LogScoreresc ~ Treatment + Assessor, data = tableb13)
```

| | | | | |
|-------|-------|--------|-------|-------|
| 2T-1T | 0.720 | 0.557 | 0.882 | 0.000 |
| 3T-1T | 1.307 | 1.145 | 1.470 | 0.000 |
| 4T-1T | 1.428 | 1.266 | 1.159 | 0.000 |
| 5T-1T | 2.049 | 1.887 | 2.212 | 0.000 |
| 5T-4T | 0.621 | 0.459 | 0.784 | 0.000 |
| 6T-1T | 2.734 | 2.571 | 2.896 | 0.000 |
| 3T-2T | 0.588 | 0.425 | 0.750 | 0.000 |
| 4T-2T | 0.708 | 0.546 | 0.871 | 0.000 |
| 5T-2T | 1.330 | 1.167 | 1.492 | 0.000 |
| 6T-2T | 2.014 | 1.851 | 2.177 | 0.000 |
| 4T-3T | 0.121 | -0.042 | 0.283 | 0.234 |
| 5T-3T | 0.742 | 0.579 | 0.905 | 0.000 |
| 6T-3T | 1.427 | 1.264 | 1.589 | 0.000 |
| 6T-4T | 1.306 | 1.143 | 1.468 | 0.000 |
| 6T-5T | 0.685 | 0.522 | 0.847 | 0.000 |

两个相邻的样品没有显著性差异的是 4T（935）和 3T（803）。

注 1：由于试验是不平衡设计，函数 aov（LogScoreresc...）中的评价员+样品样品顺序给出的结果与“样品-评价员”的结果不同。对于“评价者+样品”顺序，评价员因子的 p 值<0.05。但样品-评价员顺序通常设为是最相关的顺序。

注 2：当评价员因子作为随机因子时，R 命令：


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

样品因子的 F 值等于 655.32, p 值等于 2.2×10^{-16} 。对于评价员因子, p 值=0.14; 在 $\alpha=0.05$ 时, 评价员之间的差异不显著(评价员因子设为固定时, $p=0.054$)。

B.5 有重复情况下的数据分析

B.5.1 总则

表 B.15 列出了同一评价小组对表 B.1 中结果进行的重复量值估计结果。在完成对重复样品的第一次评价后, 评价员立即开始重复样品的第二次评价, 但顺序与第一次评价顺序不同。评价员没有被告知包括两次重复评价。

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 |
| | | 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) |
| 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) |
| 自然对数值的均值 | - | 2.277 | 2.981 | 3.563 | 3.678 | 4.317 | 4.995 |

此文件在 U 盘 F 中 ME (2019) 文件夹下, 名称为: 附表 B15。通过以下三个命令将其导入 R:

```
tableb15<- read.table("F:/ME(2019)/Table Annex B5.txt", header = T, sep="\t", dec=".")
attach(tableb15)
names(tableb15)
```

表 B.15 有 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 \times 6 个样品 \times 2 次重复)。评价员 1 的 12 个评分值在第 1~12 行, 评价员 2 的 12 个评分值在第 13~24 行, 以此类推。命令:

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

得到表 B.15 中“自然对数值的均值”如下:

| | | | | | |
|----|----|----|----|----|----|
| 1T | 2T | 3T | 4T | 5T | 6T |
|----|----|----|----|----|----|

| | | | | | |
|-------|-------|-------|-------|-------|-------|
| 2.277 | 2.981 | 3.563 | 3.678 | 4.317 | 4.995 |
|-------|-------|-------|-------|-------|-------|

B.5.2 当评价员因子设为固定因子时，显著效应的确定

试验设计中有三个因素：样品、评价员和重复。样品因子和重复因子通常设为固定因子。因此，当评价员因子为固定因子时，试验设计中有三个固定因子。在这种情况下，使用 R 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 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 | 5.152 | 0.000962*** |
| 样品 | 5 | 64.32 | 12.864 | 1445.232 | < 2×10 ¹⁶ *** |
| 重复 | 1 | 0.00 | 0.001 | 0.133 | 0.718 |
| 评价员×样品 | 30 | 0.55 | 0.018 | 2.047 | 0.027 |
| 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 | — | — |

评价员和样品因子是非常重要的。此外，评价员×样品的交互效应也显著；表明评价员对咖啡因含量变化的响应方式不同。当模型不考虑重复因子时，这些结论得到了证实。使用以下命令进行方差分析：

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

结果见表 B.17。

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 | F值 | P > F |
|--------|-----|-------|--------|---------|----------------------|
| 评价员 | 6 | 0.28 | 0.046 | 2.28 | 0.0006 |
| 样品 | 5 | 64.32 | 12.864 | 1412.86 | < 2×10 ¹⁶ |
| 评价员×样品 | 30 | 0.55 | 0.018 | 200 | 0.19 |
| 残差 | 42 | 0.38 | 0.009 | — | — |

B.5.3 当评价员因子设为随机因子时，显著效应的确定

将评价员因子设为随机因子，样品因子和重复因子设为固定因子时，设计是交互的，正确的 R 函数是 lmer 函数。

有 4 个命令：

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

得到表 B.18，有 2 个部分。

第一部分是模型的固定部分。由于试验是平衡设计， p 值与表 B.17 的值相同，而且提到类型 III 和 Satterthwaite 法均无法使用。

第二部分是随机部分。没有采用 F 值检验，而是使用似然比检验值（LRT 值），其遵循具有表中给定的自由度数值的卡方分布。使用似然比检验值和自由度 2 列获得 p 值（列 Pr(>Chisq)）。

表 B.18—当评价员因子是随机因子时表 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 | 31.4157 | 6.2831 | 5 | 30.000 | 705.8649 | <2e-16 *** |
| Replicate | 0.0009 | 0.0009 | 1 | 6.000 | 0.0963 | 0.7668 |
| Treatment:Replicate | 0.0406 | 0.0081 | 5 | 29.999 | 0.9119 | 0.4865 |
| | npar | logLik | AIC | LRT | Df | Pr(>Chisq) |
| <none> | 16 | 39.508 | -47.015 | | | |
| (1 Assessor) | 15 | 38.896 | -47.792 | 1.2239 | 1 | 0.2686 |
| (1 Assessor:Treatment) | 15 | 37.609 | -45.217 | 3.7983 | 1 | 0.0513 . |
| (1 Assessor:Replicate) | 15 | 39.370 | -48.739 | 0.2760 | 1 | 0.5993 |

评价员因子不显著： p 值等于 0.2697（表 B.16 中的 $p=9.000$ ）。评价员与样品交互作用的 p 值与 0.027 相比，非常接近于 0.05（0.0513）。

由于重复因子和涉及重复因子的 2 个交互作用都不显著，因此可以将三因子模型简化为双因子模型。
R 命令：

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

得到表 B.19。

表 B.19 评价员因子是随机因子且忽略重复因子的数据分析结果

| Type III Analysis of Variance Table with Satterthwaite's method | | | | | | |
|---|--------|---------|---------|--------|---------|---------------|
| | Sum Sq | Mean Sq | NumDF | DenDF | F value | Pr(>F) |
| Treatment | 32.135 | 6.427 | 5 | 30 | 705.86 | < 2.2e-16 *** |
| | npar | logLik | AIC | LRT | Df | Pr(>Chisq) |
| <none> | 9 | 49.404 | -80.808 | | | |
| (1 Assessor) | 8 | 48.117 | -80.233 | 2.5745 | 1 | 0.10860 |
| (1 Assessor:Treatment) | 8 | 47.257 | -78.514 | 4.2935 | 1 | 0.03826 * |

评价员和样品交互作用显著（ $p=0.038$ ），表明评价员对咖啡因变量变异的响应方式不同。

B.6 回归曲线斜率的确定（斯蒂文斯方程指数，见表 9.4）

B.6.1 总则

当已知样品中刺激物浓度时（如表B.1和表B.15中的咖啡因浓度或表B.4中的蔗糖浓度），可对浓度自然对数值和评分的自然对数值进行线性回归分析，并评估回归曲线的斜率。如果没有重新标度评分，有必要为不同的评价员提供不同的纵坐标。建议进行评价员和浓度自然对数间的交互效应分析，以检验每个评价员的斜率是否相同。

本示例主要是表 B.15。为了便于阅读，该表重新进行组成且只包含平均数。采用以下 3 个 R 命令：

```
tableannex6 <- read.table("F:/ME(2019)/Table Annex B6.txt", header = T, sep="\t", dec=".")
attach(tableannex6)
names(tableannex6)
```

得到一个包含 6 行（6 个浓度）和 11 列的文件：1 列给出浓度的自然对数值，1 列是小组均值，2 列是重复的均值，7 列是每个评价员的均值：

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

采用原始评分的自然对数值（而不是重新标度评分的自然对数值）执行计算，以突出显示评价员之间的差异。

B.6.2 小组斜率的确定

斜率是通过 R 命令的 `lm()` 函数来确定的。有 3 个 R 命令：

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

lead to:

coefficients(model)

(Intercept)  LogConc
0.1210406    0.9758299

confint(model)

          2.5 %    97.5 %
(Intercept) -0.08519907 0.3272803
LogConc      0.92025372 1.0314061
```

有可能绘制一个描述感官响应与物质浓度之间关系的图。有 3 个 R 命令：

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

参见图 B.1 的“小组”（顶行，左侧）。

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

注2：使用以下三条命令可以从文件中表b15中获得先前的系数和近似的舍入误差：

```
modellbis <- lm(LogScore ~ LogConc, data = tableb15)
coefficients(modellbis)
confint(modellbis)
```

这些得到:

```
coefficients(modellbis)
(Intercept)    LogConc
  0.1208089    0.9759042

confint(modellbis)
                2.5 %    97.5 %
(Intercept) 0.007027373 0.2345903
LogConc     0.945243041 1.0065653
```

对于每个浓度而言, 响应值的自然对数值和浓度的自然对数值的图中有 14 个数据点 (7 个评价员 × 2 个重复)。

B.6.3 每个评价员斜率的确定

由于评价员 × 样品的 p 值接近 0.05 (见表 B.16), 因此确定每个评价员的斜率很有意义。

对于评价员 1 的命令:

```
model <- lm(LogScoreAs1~LogConc, data= tableannex6)
coefficients(model)
confint(model)
plot(LogConc, LogScoreAs1, main="Ass 1")
abline(lm(LogScoreAs1~LogConc))
```

得到:

```
(Intercept) LogConc
 0.5698973 0.8602127
confint(mod)
                2,5 %    97,5%
(Intercept) 0.1860820 0.9537125
LogConc     0.7567846 0.9636409

                2,5 % 97,5 %
(Intercept) 0.1860820 0.9537125
LogConc     0.7567846 0.9636409
```

和图 B.1“评价员 1” (顶行, 左边第二个)

对于评价员 2, 需要在评价员 1 的函数中, 将数字 1 替换为数字 2, 获得:

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

对于评价员 3 以及之后的评价员方法同上。

给出表 B.19 和图 B.1（包含 8 个图表）。

评价员 1 的斜率最低为 0.860，与评价员 2 和 4 有显著差异。

评价员 2 的斜率最高为 1.116，与评估员 1、4 和 5 有很大差异。

表 B.19 回归曲线

| 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 | 1.179 |
| 3 | 0.03 | 0.992 | 0.913 | 1.072 |
| 4 | -0.04 | 0.993 | 0.974 | 1.012 |
| 5 | 0.43 | 0.911 | 0.773 | 1.049 |
| 6 | 0.22 | 0.965 | 0.851 | 1.078 |
| 7 | 0.00 | 0.995 | 0.929 | 1.060 |

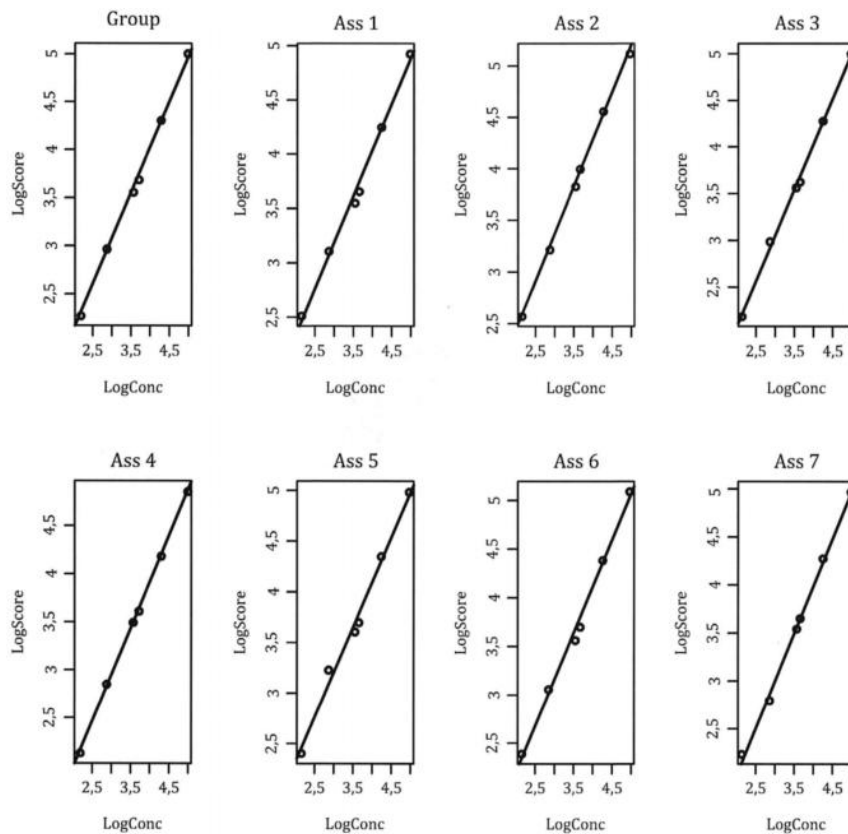


图 B.1 回归曲线（数据来自表 B.15）

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 rephcations. *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
-