

ICS 67.240
CCS X 04



中华人民共和国国家标准

GB/T 22366—202X/ISO 13301:2018
代替 GB/T 22366—2008

感官分析 方法学 采用三点强迫选择法（3-AFC）测定嗅觉、味觉和风味觉察阈值的一般导则

Sensory analysis—Methodology — General guidance for measuring odour, flavour and taste detection thresholds by a three-alternative forced-choice (3-AFC) procedure

(ISO 13301:2018, IDT)

（征求意见稿）

202X-XX-XX 发布

202X-XX-XX 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.....	II
引 言.....	IV
1 范 围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	2
4 基本原则.....	3
4.1 检验程序.....	3
4.2 数据处理.....	4
5 检验程序.....	4
5.1 样品制备.....	4
5.2 刺激浓度选择.....	5
5.3 样品提供.....	5
5.4 评价员训练.....	6
5.5 评价员选拔.....	7
5.6 实验设计.....	7
6 数据处理.....	10
6.1 数学和统计模型.....	10
6.2 数据预检验.....	10
6.3 Logistic 拟合模型的最大似然估计和误差限估计.....	11
6.4 结果解释.....	12
6.5 觉察概率 (p_d) 不等于 0.5 时的情况.....	13
6.6 最佳估计阈值 (BET)	13
6.7 结果表述.....	14
附 录 A 给定精度下所需评价员数量的估计.....	15
附 录 B 示例.....	16
参考文献.....	30

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 22366—2008。与 22366—2008 相比，除编辑性修改外，主要技术变化如下：

——修改了图 1 和图 2；

——修改了文中的“化学刺激物”为“刺激”；

——删除了“规范性引用文件”中的部分引用文件和注；

——修改了“术语和定义”中刺激物、待测样、3 点选配法检验、觉察阈值、个体阈值、平均阈值、合并数据的组阈值等术语的名称，并增加了“ISO 和 IEC 维护术语数据库同样适用于本文件”和“3 点选配法检验”定义中的注。

——删除了“检验步骤（见 4.1）”中“评价员能够识别刺激物是 3-AFC 检验的必要条件”；

——删除了“检验步骤（见 4.2）”中要求“评价员应提交 1 个结果和如何进行数据处理和统计分析”的内容；

——删除了“评价员的训练（见 5.4）”中“若评价员对样品比较熟悉，则这种偏差可通过强化训练来降低”；

——修改了“一般原则（见 5.5.1）”中“ISO 8586-1 和 GB/T 16291”为“ISO 8586”；

——删除了“合并数据的组阈值（见 5.6.3.2）”中“见 6.5 的讨论”；

——修改了“数学和统计模型（见 6.1）”中最后一句话的“ p_d ”为“ p_c ”；

——增加了“线性回归法（见）”B.2.4 中对回归模型使用的限定条件；

——修改了文中的“log”为“ln”；

——修改了附录 B 中表 B.4 中的内容；

——增加了附录 B 中“结果和计算（见附录 B.3.3）”中对 B.2 的说明，修改了“表 B.6”中“1 与 3”为“2 与 3”；

——修改了“表 B.7（见附录 B.4.4）”中第二列第 7 行和 18 行“-2.30”为“0.00”；

本文件使用翻译法等同采用 ISO 13301:2018《感官分析 方法学 采用三点强迫选择法（3-AFC）测定嗅觉、味觉和风味觉察阈值的一般导则》。

本文件做了下列编辑性修改：

修改了文中的“log_e”为“ln”；

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：

本文件主要起草人：

本文件及其所代替文件的历次版本发布情况为：

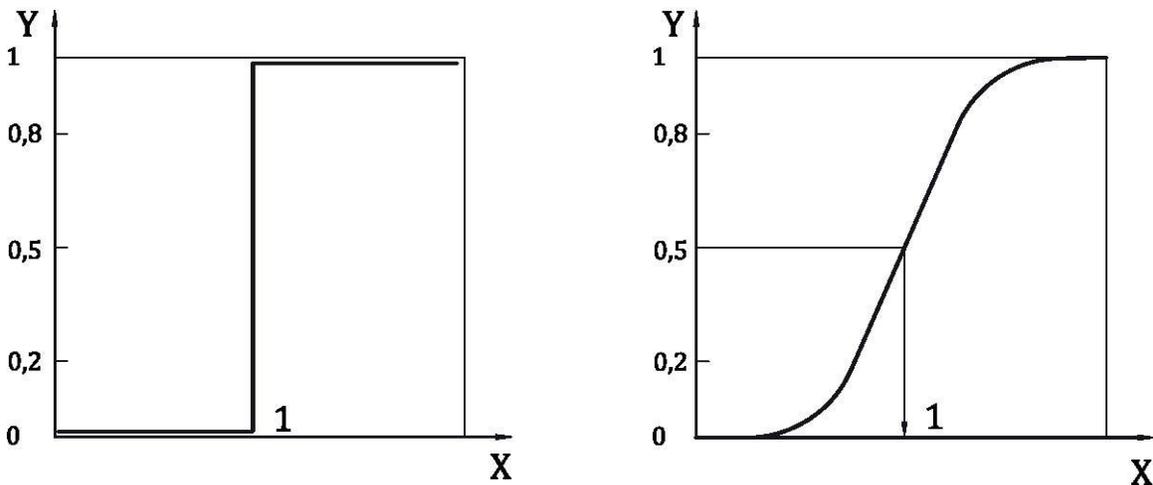
——2008 年首次发布为 GB/T 22366—2008；

——本次为第二次修订。

引言

阈值是感官分析中的重要术语，被广泛应用于食品和饮料的感官研究中。刺激的感官阈值数据常应用于 2 方面研究：测量评价员或评价小组对特殊刺激的敏感性，以及检验物质引起评价员产生感官响应的能力。前者用于判别评价员评价水平的高低，后者主要作为某种物质特性的度量指标。

术语——“阈值”是 19 世纪心理生理学家提出的，用来表示一个刺激的浓度：高于此浓度时刺激可被觉察，低于此浓度时则无法被觉察（图 1a）。



a 传统概念的阈值

b 概率特性的阈值

注释：

X——刺激浓度， x

Y——觉察概率， P_d

1——阈值

图 1 传统概念和概率特性的阈值

刺激浓度与对应的觉察概率¹之间的关系通常表现为 S 形曲线（图 1b）。由于特殊刺激的浓度有时会高于或低于阈值，为便于计算，假定阈值处于波动状态，通过估计瞬时值的中值得到阈值，即阈值是觉察概率为 0.5 时的刺激浓度。这种阈值定义类似于生物学中，如药理学、毒理学中化学物质对生物体半数有效量的定义。

¹本文件采用 3-AFC 法对刺激进行检验，觉察概率 $p_d = 1.5 \times p_c - 0.5$ 。其中， p_c 是做出正确选择的概率。严格来说，这是评价员行为的“推测模型”，并非评价员判断过程的心理测量学模型。信号检测模型同样适用，见参考文献^[13]。

多个实验室的检测结果表明，水或空气中某一物质觉察阈值的变化范围通常可达两三个甚至更多个数量级，大幅超出预期的实验误差或者数据处理造成的差异^{[6][10][14]}。推测原因可能与各实验室对阈值的定义不同或采用不同的检验程序存在差异有关。因此，Devos^[6]等给出了一个检测空气中阈值的标准化程序。

本标准使用时应明确觉察阈值的测定是需要大量的实验工作，而并非简单的表观描述。实验结果证实，反复多次测试后，有观测到评价员个体阈值趋于降低和个体间差异趋于减小。最初，评价员对阈值测定并不熟悉，但随着对被检物和检验程序的逐渐熟悉，灵敏度也逐步提高。3-AFC的程序要求评价员应识别出刺激。尽管训练程序很繁琐，但同时也提供了一些实验者所需的信息，如每位评价员的觉察浓度范围。实验者了解了评价员个体的觉察浓度范围，并给其提供适宜浓度范围时，最终检测结果将得到改进，见6.3中内容。

感官分析 方法学 采用三点强迫选择法(3-AFC)测定嗅觉、 味觉和风味觉察阈值的一般导则

1 范围

本文件给出了采用三点强迫选择法(3-AFC)测定刺激引起的嗅觉、味觉和风味响应数据,并通过数据处理进行阈值估计、误差估计以及与检测相关的其他统计方法的一般导则。

通常,本标准中给出的方法适用于下列2种情况中的1种(尽管实验能包括2种情况):

- 1) 测定评价员对特定刺激的敏感性:主要用于个体或特定群体中个体敏感性差异的研究,以及年龄、性别、生理状态、疾病、药物作用和环境条件对敏感性的影响研究;
- 2) 检测化学物质引起感官响应的特性:主要用于以下5个方面:
 - a) 风味化学研究以及特定化学物质对食品风味影响的研究;
 - b) 根据目前环境中存在的化学物质对人类的影响进行分类;
 - c) 化学物质的分子结构与其刺激能力之间的关联性研究;
 - d) 排放气体、水、食品和饮料的质量保证;
 - e) 嗅觉相关的生理反应研究。

上述情况中,由于评价员做出正确选择的概率随刺激强度变化而改变,所以除阈值测定外,对于这种变化的分析即剂量—响应(效应)关系曲线斜率也是重要研究内容。本文件中给出的数据处理程序中提供了该方面的相关信息。本文件核心是获取数据和计算程序,对于数据的有效性,本文件仅给出了一般规定和预先警告。本文件中,未严格区分“觉察阈”和“差别阈”。实际上,本文件给出的方法是测定差别阈,通过被检样与参比样对比进行测量。本文件不反对实验设计时参比样中含有刺激的情况,或者并不确定参比样中是否含有刺激。

本文件给出的方法不适用于ISO 5492中定义的识别阈的检验,也不涉及欧盟标准EN 13725中空气质量测定的标准化方法。

2 规范性引用文件

下列文件对于本文件的应用必不可少,其全部或部分在本文件中被正式引用。凡是注明日期的引用文件,仅注明日期的版本适用于本文件。凡是未注明日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

3 术语和定义

ISO 5492 中界定的术语和定义适用于本文件。

ISO 和 IEC 维护术语数据库同样适用于本文件:

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

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

3.1

刺激 stimulus

能或不能激发感觉的物质, 能否被一种或多种感受器感知主要取决于该物质的存在量。

3.2

介质 medium

阈值测定时, 用于溶解、稀释、分散或吸附刺激 (3.1) 的物质。

3.3

参比样 reference sample

不额外添加刺激 (3.1) 的介质 (3.2)。

3.4

测试样 test sample

含有已知浓度刺激 (3.1) 的介质 (3.2)。

3.5

三点强迫选择法 (3-AFC) three-alternative forced-choice test

提供给评价员 3 个样品, 其中一个是测试样 (3.4), 含有评价员熟悉的某种已知刺激 (3.1), 另外两个是参比样 (3.3), 要求评价员从中识别出测试样。

注: 标准3-AFC检验属于一种特定差别检验, 要求评价员根据某种指定特性进行判断。阈值测定时, 评价员对所关注特性的识别阈通常会偏低。因此, 阈值测试时通常要求评价员采用非特定3-AFC检验, 选择与另外2个不同样品进行比较, 而不是依据某种特定的特性。

3.6

提供样 presentation

一套用于三点强迫选择法 (3.5) 的 3 个样品。

3.7

阈值模型 threshold model

特定检验中, 检验刺激 (3.1) 能 (结果是做出正确选择) 或未能被觉察到 (结果是做出随

机选择)的感官觉察模型。

3.8

信号检测模型 signal-detection model

特定检验中,有证据表明刺激(3.1)存在的感官觉察模型。

注:这些证据有助于评价员判断刺激是否存在。

3.9

觉察阈 [限] detection threshold

引起感觉所需要的最小感官刺激(3.1)强度,即测试条件下,采用阈值模型(3.7)计算的觉察概率为0.5时激发感觉所需要的最低刺激浓度。

3.10

个体阈 [限] individual threshold

单个评价员的觉察阈(3.9)。

3.11

平均阈 [限] average threshold

个体阈值(3.10)的平均值(应明确平均值的类型,如算术平均值、几何平均值和中位数等)。

3.12

合并数据组阈 [限] group threshold from pooled data

将一组特定的评价员对所有刺激(3.1)浓度的评价结果输入到统计模型中,通过模型拟合得到的估计值。

4 基本原则

4.1 检验程序

用介质配制一定浓度的刺激样品,连同2个参比样一起提供给评价员,要求评价员从中选出含有刺激或者刺激浓度更大的样品。刺激应能溶解于空气中或水中。检测时,若提供给评价员的是气态介质,应使用空气而非其他气体作为介质;若提供的是液体或固体介质,则可使用液态或固态食品代替水介质来稀释或吸附刺激。无论采用何种介质,介质均应均匀一致,确保所提供的2个参比样的一致性和均一性。

配制一系列不同浓度的刺激样时,每个浓度点都应进行重复检验。确保检验次数足够多,以获得期望的阈值精度和数学模型参数。实验设计时,重复实验应包括评价员个体、评价员之间以及兼顾两者的重复检验。

4.2 数据处理

每次检验时评价员提交的是一个二选一的结果，即评价员指出的样品不是测试样（选择正确）就是参比样（选择错误）。统计模型判断依据是某一浓度下做出正确选择的次数是服从二项分布的。

3-AFC 检验中，通常认为阈值是正确选择次数的概率等于 2/3 时的刺激浓度，对应的觉察概率为 0.5（公式 1）。正确选择次数概率可简单地通过计算和内插而得到。但对阈值和阈值范围更准确的估计是通过对数据进行模型拟合获得的。许多导则中均采用 Logistic 模型，该模型是通过最大似然法或最小二乘法进行拟合。拟合过程主要是估计模型中的 2 个参数—位置参数和形状参数。前者确定拟合曲线的位置，后者决定曲线的陡峭性。通过拟合曲线，可导出正确选择次数概率在 50%以上的估计。

最简单的拟合模型是正确选择次数概率近似服从正态分布。该情况通常发生在数据来自于单一评价员重复评价的情形。通常简单 Logistic 函数就具有较好的拟合性，该函数含有曲线参数的一对数值。但在评价员对化学物质敏感性的评价中，更常见的情形是数据既不是正态分布，也非对称分布。部分刺激的分布并非正态分布，而呈现明显的双峰分布（除非检验样品数量为 100 以上）。当简单 Logistic 函数不很好拟合偏离正态分布的数据时，需建立其他拟合数学模型。

5 检验程序

5.1 样品制备

5.1.1 一般原则

依据 ISO 6658 中的规定，选用的刺激和介质应具有代表性，在整个检验过程中应保持恒定，且无毒性和无致敏性。确保选取的刺激和介质能代表检验目的，例如排出的气体可能会在处理过程中发生变化，化学物质宜先进行纯化处理去除异常气味或刺激物后再用于检验。准备的刺激和介质应足量且均匀，确保整个检验过程中无论刺激的浓度和位置如何变化，呈现给评价员时都能保持均匀一致。样品制备设施应符合 ISO 8589 中的规定。所使用的容器不应测试样有吸附作用，也不应产生任何气味和味道。除化学检验外，应确保评价员无法以视觉或其他任何方式觉察出介质中是否含有刺激。非检验期间，样品应避光、避热存放。

5.1.2 气体样品

将刺激和介质收集到具有聚四氟乙烯涂层的瓶或球状容器中。若刺激本身无味但含有其他气味杂质时，应用新鲜样品连续吹扫冲洗容器、阀门和管道数次，使气体样品在整个容器内达

到饱和。此外，应在近似恒定温度下向评价员提供气体样品，避免气体体积变化。应使用具有聚四氟乙烯涂层的光滑管道和阀门，避免压力的突然改变。

5.1.3 液体样品

用液态介质配制刺激时，应确保样品在整个检验过程中始终保持完全溶解的状态。不能完全溶于水的物质，可用经活性炭处理除味过的乙醇、乙烯或乙二醇溶解和稀释。由于蒸馏水和无水乙醇常带有强烈气味，可用食用级产品代替或使用前先用活性炭纯化。对于完全不溶于水的疏水性刺激，应用无臭的疏水溶剂配制，如液体石蜡、邻苯二甲酸二壬酯等。应避免接触塑料容器，防止容器中某些物质溶解析出。需注意，逐级稀释时随稀释概率增高，因器壁吸附造成的刺激损失也会增大。因此，应尽量采用微量注射器稀释，且用直接原液稀释代替逐级稀释。

5.1.4 固体样品

固体样品常用的典型介质包括奶酪、鱼和肉等食品。采用适宜技术将固体样品重新分散后再组合，样品精细切分或粉碎后加入到适合的介质中，充分混合均匀并确保样品在介质中充分扩散一段时间后，再提供给评价员。

5.2 刺激浓度选择

根据3-AFC实验需要设置系列浓度样品，样品浓度逐渐递增，浓度因子用 X 表示。 X 值依据可接受的阈值估计误差限来确定。对于精度要求不高的研究，通常选择 $X \approx 3 \sim 5$ ；对于精度要求较高的研究，一般选择 $X \approx 2$ 。应根据每位评价员的个体特性选择合适的检验方案，确保检验结果能够对应Logistic模型S形曲线的某个范围，分布在评价员的偏好性觉察阈范围内。最有效的数值点是对应于正确选择概率范围为45%~90%的点，即， $p_d = 0.18 \sim 0.85$ 。

为节约样品和评价员时间，通常先设置一个较大的 X ，确定每位评价员能感知的刺激浓度范围，阐明检验机理，并教会评价员如何在刺激浓度超过其偏好性觉察阈范围时识别刺激。

然后，再选取一个较小的 X 为每位评价员设置一套适合的3-AFC检验浓度梯度。检验完成后，若检验结果无法确定评价员觉察阈值时，负责人可通过增加刺激浓度水平继续检验，直到满足要求为止。应每隔几个样品就要求评价员描述其对刺激的感觉，防止对刺激的记忆消失。也可通过询问方式，测定评价员在没有觉察刺激时随机猜对的次数，如每27个三选一的系列检验中，会有1次猜测正确的机会。

5.3 样品提供

5.3.1 样品制备

制备好的测试样提供给检验区（检验区规定见ISO 8589）的评价员，评价员按照ISO 6658

规定的感官分析良好操作规范进行样品评价。测试样采用三位随机编码，或按一定方式摆放，如将第一个样品放在评价员前方最左边，然后从左到右依次摆放样品，回答表上也采取同样的排列方式记录。采取 AAB、ABA 和 BAA 方式，平衡 3 个样品在评价员面前的提供次序，避免位置偏差。对高于觉察阈值浓度的样品，评价员应只摄入最低摄取量。此外，2 个样品间的间隔时间应足够长，确保评价员的感官能力得到恢复，降低评价员的感官疲劳。

5.3.2 气体样品

气体样品的提供具体参见文献^[8]和^[12]，借助嗅觉测量装备进行样品评价。

5.3.3 液体样品

提供的液体样品是非挥发性化学物质时，先将非挥发性化学物质溶解于纯水或无味溶液中，再用不吸附该物质的容器进行存放，如 100 mL 玻璃烧杯（溶液量不超过容器 1/4 体积）。对于挥发性化学物质，应使用具塞、且适宜嗅闻和品尝的广口容器，如采用 250 mL 挤压瓶将标准测量体积的顶空气体或液体送入鼻腔或口腔中，参见文献^[4]、^[7]和^[11]。介质为饮料时，应使用符合日常习惯的感官评价容器。

5.3.4 固体样品

介质是为食品时，感官评价时应尽量保持食品的原有形状不变。

5.4 评价员训练

通常情况下，检验中所关注的阈值是有经验评价员的觉察阈值。只有参加多次待测物的相关训练，能够在任何时候只要刺激（例如空气或水中的某种污染物，或者是食品或饮料风味中的某种成分或杂质）存在就能觉察到的评价员才能称为有经验的评价员。因此，对检测物的熟悉度也是 3-AFC 检验中的一项要求。若评价员缺乏足量的训练，检测阈值范围的上限将人为的增加 1-2 个数量级。当需检测非专业检验者的觉察阈值时，如家用煤气泄漏的警告剂量，则应采用未经训练的评价员，且允许存在一定的干扰物（如噪音），同时采用三点检验或者成对比较检验代替 3-AFC 检验。

训练程序应先选择一个高浓度的刺激，再选择 2 种或更多的刺激浓度，要求评价员对其排序。确定评价员的偏好性觉察阈范围后，再采用 3-AFC 法进行检验。经过该训练程序后的评价员的阈值水平会降低，并在 3~5 次检验后逐步趋于稳定。不同评价员对同一物质的敏感性不同，评价员之间敏感性的差异可达到 2~3 个甚至更高的数量级。

5.5 评价员选拔

5.5.1 一般原则

依据研究目的，按照 ISO 8586 中的原则和规定进行评价员选拔。

5.5.2 个体阈值

可采用下列方法进行检验，如将个体的觉察阈值同已有报道中的阈值、不同环境下的测定值或评价员对其他物质的觉察阈值进行比较。该检验可用于嗅觉缺失或嗅觉过敏症诊断，以及味觉缺失或味觉过敏症诊断。

5.5.3 阈值分布

实验者有时希望获得群体中阈值的分布状况，该群体可能是人群抽取样本，或者是被选人群全体成员，如评价小组的所有成员。群体选择不属于本文件中的内容，但实验者应仔细研究群体及群体样本，结果表达应按照 6.7 中的规定进行。

5.5.4 刺激阈值测定

刺激的检验数据或平均阈值仅对检验中所使用的评价小组有效。对于评价小组以外的结果，实验者必须谨慎地进行推断。实验者应根据实验目标和目的选择评价小组。例如，对一组化学物质进行感官特性研究时，宜选择一个人数较少的评价小组；若对食品中潜在风味化合物特性进行研究时，应选择一个人数较多的评价小组，因其代表了一个特定群体。

要获取更加准确的估计值，应同时考虑评价员的数量和被检样品的数量。评价员人数较少时，应增加样品检验次数，以获得足够的数量。反之，对于人数较多的评价小组，每个评价员只需对每个浓度（样品）或部分浓度（样品）检验 1 次即可。

5.6 实验设计

5.6.1 个体阈值

评价员做出正确选择的概率为 45%~90%时所对应的刺激浓度，可有效估计 Logistic 模型参数。此浓度范围内，如果刺激浓度主要分布在阈值周围，参数的精度主要取决于样品数量。表 1 列出了样品数量与阈值估计误差限的关系，单位是原始浓度，见附录 A。

表 1 阈值估计值的期望精度所需的样品数量

样品数量	40	60	80	100	120	160	200
阈值估计值的误差限	2.5	2.2	2.0	1.8	1.7	1.6	1.5

估计阈值的变化范围是通过阈值除以和乘以表1中的误差限获得。例如样品量为80、检验阈值为2.4 mL/m³时，阈值范围为1.2 mL/m³~4.8 mL/m³。当样品量达到200以上时，精度提升缓慢。因此，通过增加样品量提高精度的方法并不适用。下列是一种有效提高精度的方法。在每个浓度点开展重复检测，再利用Logistic模型拟合估计阈值和误差限。通过Logistic模型获得有效的浓度范围，进一步对每个有效浓度点进行多次重复测试，直至达到所需的精度。

5.6.2 阈值分布

优选评价员按照 5.6.1 中的规定进行检验，结果以直方图或累积频率图呈现。平均阈值结果应以算术平均值、几何平均值、中位数等形式给出。若阈值呈现双峰或多峰分布时，宜采用降维处理，具体处理步骤按第 6 章中的规定进行。

5.6.3 评价小组组阈值的测定

5.6.3.1 一般原则

实验设计时，应注意评价员间的敏感度差异可能是评价员个体敏感度差异的数倍。因此，重复检验时应采用增加评价员数量的方式，而非增加每位评价员评价的样品数，这种检验结果更具有实用性和代表性。

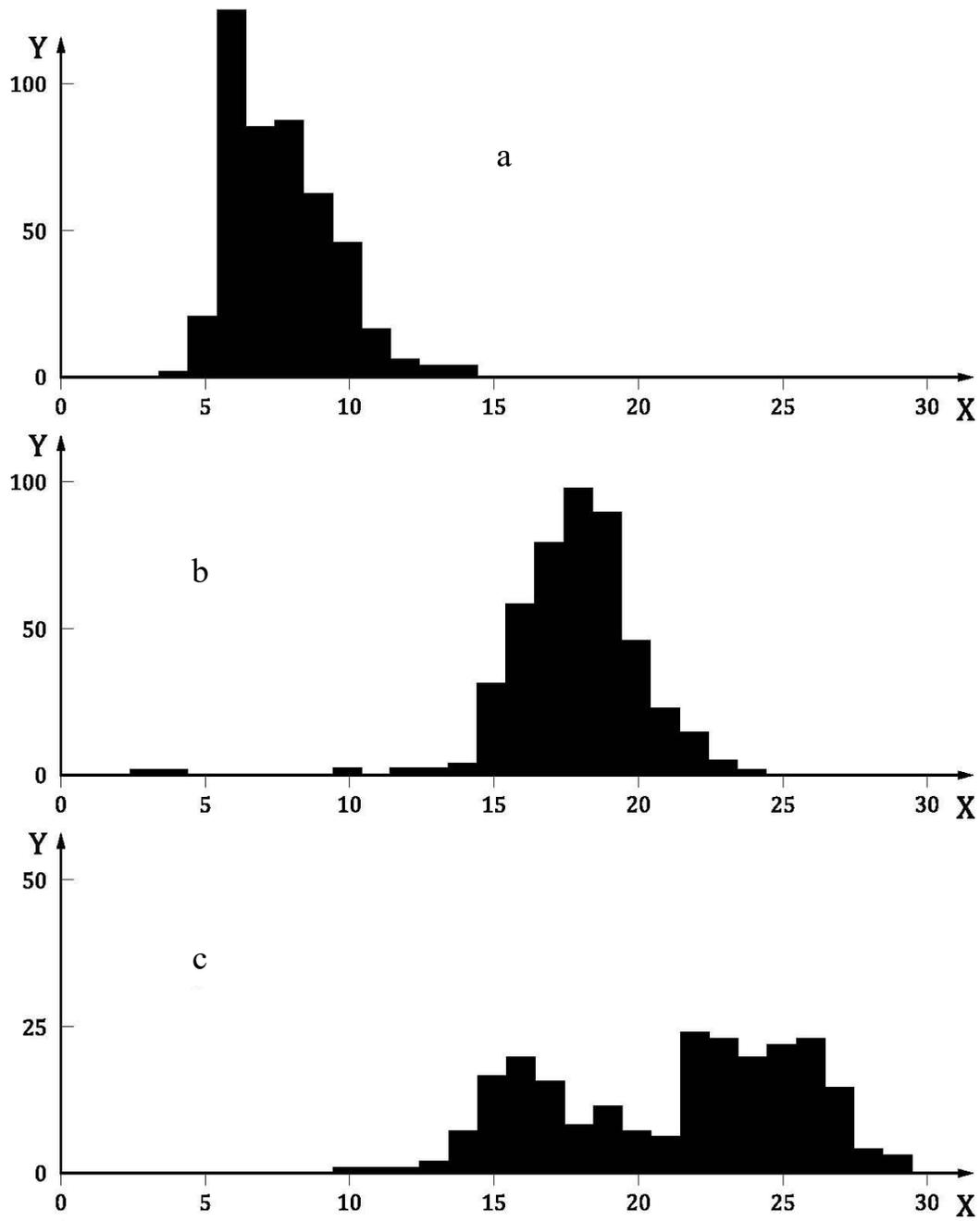
5.6.3.2 合并数据的组阈值

与个体阈值中将每位评价员数据单独进行Logistic模型拟合不同，组阈值是采用合并数据进行拟合，即将样品每个浓度下所有评价员结果一起进行Logistic模型拟合。数据量越多、分布越集中时，组阈值的拟合性就越好。3.9中定义的觉察阈值属于组阈值，具体见示例B2和B4。该方法适用于不把个体差异作为主要考虑因素的实验设计，例如依据化学物质污染或感官玷染严重程度进行分级的研究。

5.6.3.3 平均阈值

依据 5.6.1 中的规定采用优选评价员进行重复检验，结果以直方图（图 2）或累积频率图呈现。若研究目的中包括个体差异时，如风味化合物、污染物或感官玷染物对特定群体的影响研究，应采用此方法。

图 2a 和图 2b 中有 443 位评价员参与阈值测定，图 2c 中有 222 位评价员参与测定。为便于直观比较，图 2a 和 2b 的纵坐标是图 2c 的 2 倍。稀释步骤中，“0”代表每种气味的饱和浓度，为最高阈值^[3]。



注释:

X 稀释倍数

Y 个体数

1 异丁酸异丁酯

2 异戊酸

3 异丁醛

图 2 群体中嗅觉阈值的分布

6 数据处理

6.1 数学和统计模型

3-AFC检验中，评价员对3个样品随机选择，选择正确的概率是1/3，近似为0.33。根据阈值模型，选择正确的概率 p_c 与觉察概率 p_d 之间关系如式（1）：

$$p_c = p_d + \frac{1}{3}(1 - p_d) = \frac{2}{3}p_d + \frac{1}{3} \quad (1)$$

检验中，评价员能正确觉察出刺激的概率为 p_d ，剩余概率（ $1-p_d$ ）表示评价员不能觉察出刺激。此时，评价员需做出随机猜测，猜测正确的概率仅为1/3。由于阈值是 $p_d=0.5$ 时的刺激浓度，依据式（1），即 $p_c=0.67$ 时的数值。

p_c 是检测数据， p_d 是阈值模型的推算数据。反向推导出计算式，如式（2）：

$$p_d = 1.5p_c - 0.5 \quad (2)$$

通常设置一系列不同浓度的刺激，并对每个浓度进行重复检验，用评价员做出正确选择的概率 p_c 和刺激浓度作图，图形大致为S形曲线。根据式（2），将 p_c 转换成 p_d 后可得到另一条S形曲线（图1b）。当浓度足够低或足够高时，渐近线分别为0和1。在对化学刺激敏感的情况下，刺激的浓度通常以浓度对数值或稀释度表示。

P_c 与刺激浓度的S形曲线可用累积正态分布拟合，或更方便地用累积Logistic分布拟合，如式（3）所示：

$$p_c = \frac{\frac{2}{3}}{1 + e^{b(t-x)}} + \frac{1}{3} \quad (3)$$

式中， x 代表刺激浓度，系数 t 和 b 可由数据拟合求得。当 $x=t$ 时， $p_d=0.5$ ，此时 t 为刺激浓度阈值。参数 b 是 x 每变化一个单位时 p_c 的变化，决定曲线分布的陡峭度。

6.2 数据预检验

6.2.1 准备工作

采用直接观察法或通过 p_c 和浓度对数值作图法对数据进行初步检验。需注意，无论结果是否呈现S形曲线，以及待测浓度是高于或低于估计阈值浓度，用于阈值估计的数据都应准确。否则，应获取更多的检验数据。先直观地估计阈值，判断准确度是否能满足需要。若否，采用6.2.2的logit变化法，或者6.3中的最大似然法，将数据进行模型拟合处理，具体参考示例B.2至B.4。

6.2.2 采用Logit变化法预估阈值和斜率

Logit变化法等同于概率，是基于累积正态分布的概率。采用式（4）将 p_d 转变成对数形式：

$$L_d = \log_e \left(\frac{p_d}{1-p_d} \right) \quad (4)$$

结合式 (4) 和式 (2)，转换成式 (5)：

$$L_d = \log_e \left(\frac{p_c - \frac{1}{3}}{1-p_c} \right) \quad (5)$$

将式 (3) 带入式 (5)，得到式 (6)：

$$L_d = b(t - x) \quad (6)$$

p_c 服从Logistic曲线时， L_d 随着刺激浓度增大呈线性增加。此时的关键是决定采用浓度的转换数据还是原始数据进行计算。在采用原始数据的 p_c 与 $\log_e x$ 图中， p_c 中间范围的数据近似为线性曲线，阈值是 p_c 为2/3时所对应的浓度。但是，无论是直接观测的S形曲线或是 p_c 中间范围数据拟合的线性曲线，采用转换数据并未提供任何便捷。

对于转换数据，阈值是 $L_d=0$ 时的浓度自然对数值，可通过目视大致为一条直线时估计，或通过线性回归拟合估计。应注意，渐近线附近的转换值并不稳定，该区域范围内发生很小的改变都可能会对结果带来较大影响。因此，对于 $p_c < 0.43$ 或 $p_c > 0.9$ 的数据（即转换值 $L_d < -1.75$ 或 $L_d > 1.75$ ）应不予考虑。除非检验的关注点就在该区域，具体参见 6.5。应注意，转换数据作图可以直接解释式 (3) 中参数， t 对应于 $L_d=0$ 时的刺激强度（浓度自然对数值）， b 是直线的斜率。

6.3 Logistic 拟合模型的最大似然估计和误差限估计

6.3.1 总则

最大似然法（Maximum likelihood, ML）拟合 Logistic 模型数据的原理是获得具有最大可能导致样本结果出现的参数 t 和 b 。应采用专用软件程序²进行 ML 估计，或者更简单采用示例 B.2~B.4 中描述的电子表格程序。例如，ML 结果表明估计值大于 t 值上限的概率仅为 0.05。多数情况下，将数据 Logistic 转换后再进行线性回归拟合也是 ML 常用替代法，如 B.2.4 所示。

6.3.2 参数 b

参数 b 是拟合方程的斜率，代表拟合曲线的陡峭程度。 b 为正值时表明觉察概率会随着刺激浓度增加而增高， b 为负值时表明觉察概率和刺激强度之间存在负相关。用稀释倍数衡量刺激强度时， b 通常为负值（稀释度越高=强度越低=觉察概率越低）。对于评价员个体而言， b 值代表评价员对刺激强度变化的辨别能力，与韦伯常数和史蒂文斯幂律指数等指标相关。评价员个体在 b 值上的差异类似于阈值上的差异， b 值的大小或在群体中的分布情况同样值得关注。 b 值高的评价员（陡峭的斜率）对刺激强度的微小变化更为敏感，在涉及质量控制和监测的任务中就更

² 如 GLIM^[1]和 SAS 软件是目前比较适用的商业化产品。该信息只是给本文件的使用者提供一些便利，并不涉及推荐认可产品。

加有效。评价员筛选时， b 值与阈值同样重要。

6.3.3 参数估计值的置信区间

置信区间是参数估计的取值范围，通常认为此范围内的数值真实可信。置信区间越窄，估计值的可信度越高。增加数据总量可以提升估计值的精度，或采用在0.25倍~4倍阈值浓度范围内等间距设置浓度的方式实现。

6.4 结果解释

阈值测定结果可用于多种目的，本文件并未给出基于特定目的进行阈值检验的实验设计指导。结果解释和阈值比较时，应考虑数据的采集和分析方式，以及如何设置数据置信度。

单个评价员的数据结果最容易解释和比较。Logistic拟合模型是评价员的心理生理函数，统计分析的结果可用于评价员之间或者同一评价员对不同待测物的比较。

通过简单Logistic扩展模型，可以对来自不同评价员或不同待测物的数据进行比较。其他设计还包括选择不同评价小组代表不同群体，并在小组间重复提供样品。采用标准方差分析方法比较不同物质或不同评价小组对某一给定物质的数据。假设所有估计值是以相同的方法和样品数量方式获得，则评价员个人的 t 和 b 的估计值可作为输入数据。如果数据是5.6.3.2中所述的合并数据，则 t 和 b 估计值可用于描述物质间或小组间的差异。评价员间阈值的分布可能会偏离正态分布，可采用直方图（图2）、正态概率图或Logistic概率图检验。如果偏态明显或呈明显的双峰分布，则应取合适的平均阈值，如偏态数据的中位数、双峰或多峰分布的多重平均值。

估计合并数据的组阈值及其误差限的实验设计中，每位评价员对每个浓度仅检验1次。该方法获取的数据具有高度可变性，通常不推荐使用。多数情况下采用另外1种设计，即每位评价员在每个刺激浓度下至少进行1次检验，如示例B.1所示。实验者应计算平均阈值和合并数据的组阈值，并检验阈值的分布。但应注意，这种实验设计的估计个体阈值精度较差，但合并数据的组阈值估计具有较高精度。例如采用2组、每组50位评价员对5个刺激强度进行检验的实验设计中，有500个数据可用于组阈值估计，但只有10个数据用于评价员个体的阈值估计。

采用每位评价员在每个刺激浓度下仅做1次检验的实验设计，即可进行合并数据的组阈值及误差限估计。但更常用的实验设计是要求评价员在每个浓度下至少完成1次检验。群体对刺激通常表现出宽范围变化的个体阈值。对于单个评价员，一个100倍的浓度范围覆盖了 p_d 值从0.05到0.95的变化范围；但对于一组评价员，个体阈值的变化范围可达10 000倍。若仅向评价小组提供一个100倍浓度范围的刺激测试时，这意味着对于部分评价员（通常是相当数量的评价员），所给与的浓度范围可能是接近其心理物理学函数渐进线的浓度范围。渐进线附近的结果对于 t 和 b 值的估计几乎没有贡献，这表明敏感性较高或较低的评价员对合并数据的参数估计意义不大。若非必要，不应采用这些数据进行组阈值估计，或阈值估计时对于不同敏感性评价员赋予不相

等的权重。

数据拟合处理的前提是假定阈值分布服从Logistic模型，与该分布的任何偏差都表明数据和拟合曲线的不匹配。拟合度不够可采用拟合优度统计程序检出，但单次Logistic模型的偏差不太可能检测出，除非实验设计中考虑了10种以上浓度、大于500倍的浓度范围以及此范围内总共数百次的检验。如果拟合优度检验结果表明拟合度明显不够，则需考虑采用单次Logistic函数外的模型进行拟合。最简单方法是考虑对部分评价员采用二次Logistic函数建模，函数带有不同 t 值，也可能包括 b 值，适用于偏态分布和双峰分布的情况。

6.5 觉察概率 (p_d) 不等于 0.5 时的情况

管理者通常希望空气中的恶臭物质在5%概率时就能被觉察，风味学家则希望判断出食物中是否添加了一定浓度风味物质的概率是95%。该水平效应可通过Logistic曲线计算得出，即分别是 L_d 值为-2.94和2.94时所对应的刺激强度， L_d 值通过式(4)或(5)得到。 p_d 值过高或过低时，实验者应确保该区域（高值区或低值区）内有足够多数量的数据，以获得此范围相应的刺激强度。超出研究范围外的任何推断均不具有可信度。

6.6 最佳估计阈值 (BET)

BET快速测定法（参考文献^[2]和示例B.1）是基于阈值模型(3.7)获得的粗略组阈值估计，精度较低，且存在一定风险性。3-AFC检验中，若样品浓度低于阈值水平时，评价员做出正确选择的概率为1/3；若样品浓度高于阈值水平时，做出正确选择的概率为1.0。该方法耗时短，每位评价员对每个浓度水平下的样品仅评价1次，但对评价员数量要求较高。

为方便清楚观测到随浓度增大做出正确选择的分布情况，按照样品浓度递增顺序（或如示例B.1所示，按稀释度递减顺序）将数据制成表格，再计算BET值，即做出错误选择所对应的最高浓度和其相邻的更高浓度的几何平均值。如附录B.1中示例，评价员1的 $BET = \sqrt{135 \times 45} = 78$ 。

如果评价员对一组浓度全部做出了正确选择，或者最高浓度时就做出了错误选择，则该方法不再适用，如示例B.1中的评价员6和4。此时，宜适当扩展浓度梯度继续检验，或采用下列通用方法：若评价员在最高浓度时仍做出错误选择，则假设最高浓度的相邻高浓度选择正确，并由此计算BET值。如果是一组全部选择正确的数据，则假定最低浓度的相邻更低浓度选择错误，再进行BET值计算。

组BET可通过个体BET的几何平均值计算获得，一种快速计算评价员间差异的方法是计算BET对数值的标准偏差，如示例B.1所示。由于1次猜测正确的概率为1/3，连续2次或者3次猜对的概率分别为1/9和1/27，组BET结果会有偏差。该方法风险较大，因为每位评价员只

检验 1 次样品，由于样品混淆、对刺激或测试方法不熟悉等原因，会导致对高于阈值的样品做出错误选择。如果 BET 恰恰落在浓度系列的极值范围内，或者测试的系列浓度过少，则 BET 对数值的标准偏差可能会被低估。

6.7 结果表述

检验报告中应包括下列内容：

a) 所有的检验条件：包括样品性质及来源、取样方式、选取介质（稀释液），以及样品提供时所用的容器系列和检测设备等；

b) 样品浓度或所用流速、温度及其他条件；

c) 给评价员的检验步骤说明和评分表；

d) 每一步的稀释因子；

e) 每位评价员检验样品的重复次数；

f) 评价小组成员的相关信息，如年龄、性别、评价经验以及其他一些重要信息（如评价员对刺激的熟悉度、健康状况、吸烟史、是否使用假牙以及最后进食时间等）。不应靠姓名来区分评价员，报告中同时应避免某个熟悉评价小组的人对评价结果给出个人判断；

g) 数据量不大时，结果宜以列表形式给出，如表 B.1。表中应列出检验的浓度系列和该浓度下评价员做出正确选择的次数。报告中应给出阈值估计值，如个体阈值、组阈值和误差限；

h) 若已检验评价员的阈值分布情况，报告应给出评价员的个体阈值以及部分描述统计值，如均值、方差和数据不服从正态分布时采用的检验方法；

i) 采用合并数据时，应根据合并数据给出组阈值和误差限，以及斜率 b 和其误差限。报告中应指出，评价员内部差异（较小的）和评价员之间差异（较大的）对计算合并数据误差限的潜在影响。

附录 A

(资料性附录)

给定精度下所需评价员数量的估计

即使相同刺激，个体的响应同样存在差异。依据本文件程序进行阈值估计时具有高度不确定性，表 A.1 是一个典型情况下达到预期精度（±标准偏差）的示例。

表 A.1 评价小组样本量和提供样品量下的获得的检验精度

序号	评价小组	目的	3-AFC检验中的样品数	检测精度
1	1人	内科医生希望知道某人对物质 X 是否具有嗅觉缺失症	8个浓度，每个浓度检验1次 n=8	±200%~500%
2	1人	希望知道他/她对1瓶调配水中物质 X 的敏感度	6个浓度，每个浓度检验6次 n=36	±50%~100%
			8个浓度，每个浓度检验10次 n=80	±20%~50%
3	8人小组	实验者希望比较2个小组的敏感性	6个浓度，每个浓度检验1次 n=48	±100%~300%
			8个浓度，每个浓度检验6次 n=384	±20%~50%
4	4个小组， 每个小组 由挑选出 的8名具有 代表性 的人员组成	市政工程希望知道污染物质 X 的阈值水平	6个浓度，每个浓度检验1次 n=192	±50%~200%
			6个浓度，每个浓度检验6次 n=1152	±20%~50%
5		工程师希望确定总体95%概率下未检测物 X 的浓度水平	上述检验加上 4×8 人中最敏感的 25%人的重复 n=大约1800	±40%~100%

3-AFC 检验精度与样品数量间的关系与所测阈值类型无关。例如一个桔子汽水的生产商希望知道某种特定成份是否能以 95% 的概率被检出，这种情况下所需的待测样数量约为 1800。评价员的所需数量很大程度上取决于评价小组的总体差异。若评价小组是由差异较大的群体构成时（例如许多年轻人和老人），达到既定检验精度通常需要更多的评价员数量。

附录 B

(资料性附录)

示例

B.1 采用 BET 法快速测定烟囱气体样品的嗅觉阈值

B.1.1 总则

本示例给出了采用 BET 法估计阈值近似值的快捷程序。最简单操作程序如下：部分评价员（9 名）组成的评价小组对 6 种选定浓度样品进行 1 次检验，评价员 4 和 6 不继续对扩展浓度进行检验。更准确的方法是继续检测，例如将样品稀释 3 倍、9 倍、27 倍及更高倍数，让评价员 6 继续开展检验，直到出现错误选择为止。检测精度过低时，宜重复整个检验程序。

B.1.2 检验

以无臭空气为介质制备 6 个不同浓度的烟囱气体样品，每个浓度的样品都与 2 个无臭空气样品组合成 1 组。浓度梯度按照 3 倍的概率递增。从参与检验的人群中随机选择 9 位评价员，每位评价员均按照从低浓度到高浓度的顺序依次对样品进行检验，从每个浓度组的 3 个样品中选出 1 个不同于其他 2 个的样品，检验结果如表 B.1 所示。

表 B.1 烟囱气体的嗅觉阈值

评价员	稀释比 (浓度递增)						最佳估计阈值 (BET)	
	3645	1215	405	135	45	15	阈值	Log ₁₀ 阈值
1	0	+	+	0	+	+	78	1.89
2	+	0	+	+	+	+	701	2.85
3	0	+	0	0	+	+	78	1.89
4	0	0	0	0	+	0	9	0.94
5	+	0	0	+	+	+	234	2.37
6	+	+	+	+	+	+	6313	3.80
7	0	+	+	0	+	+	78	1.89
8	+	0	0	+	+	+	234	2.37
9	+	0	+	+	+	+	701	2.85
阈值对数值 (log ₁₀ 阈值) 总和 BET 阈值平均值=几何平均值 阈值对数值 (log ₁₀ 阈值) 的标准偏差							∑log ₁₀	20.85 2.32 0.81
“0”表示评价员在 3 个样品中选择错误； “+”表示评价员在 3 个样品中选择正确。								

B.1.3 计算

每位评价员的 BET 值由评价错误的最高浓度值和其紧邻判断正确的更高浓度值的几何平均

值确定，具体参见参考文献[2]。计算时，假定评价员 4 在接下来稀释比为 5 时做出正确选择，评价员 6 在稀释比为 10935 时的选择错误。BET 结果的对数值总和为 20.85，平均值约为 2.32，这与平均阈值（稀释比 209）的反对数值接近。阈值对数值的标准偏差表征阈值的离散程度，可根据表中右列计算得出。数值 0.81 表明评价员间评价结果存在较大差异，但实际值可能比 0.81 更大。由于评价员 4 和评价员 6 并未继续对扩展的浓度系列进行检验，因此浓度范围被缩小。

B.2 合并数据的阈值计算（以鳟鱼体内的柴油污染阈值测定为例，比较最大似然法拟合法与最小二乘法拟合法）

B.2.1 总则

本示例给出了最大似然法的工作原理，以及采用电子表格软件确定阈值参数和界限的方法。将该方法应用于 9 个非等间距浓度、18 个评价员采用 3-AFC 法检验结果的大数据集，参见 B.2.3。为了便于比较，B.2.4 中对相同数据同时采用最小二乘法进行拟合。

B.2.2 检验

测试样品为柴油，检验暴露在含有柴油水样中的鳟鱼是否存在潜在柴油污染（参见参考文献 5）。鳟鱼暴露在不同浓度柴油中，24 h 后收集这些鳟鱼，用 18 个有经验评价员构成的评价小组进行评价，结果与未经暴露的对照组进行比较。表 B.2 总结了电子表格软件中的公式，表 B.3 列出了实验数据和用于拟合 Logistic 模型的工作表。

表 B.2 电子数据表中用于估计 Logistic 参数和阈值估计误差限的公式

列	公式
B	@LN(A3)
E	+D3/C3
F	(2/3)/(@EXP(G\$18*(F\$18-\$B3))+1)+1/3
G	+\$D3*@LN(F3)+(\$C3-\$D3)*@LN(1-F3)
H	(2/3)/(@EXP(I\$18*(H\$18-\$B3))+1)+1/3
I	+\$D3*@LN(H3)+(\$C3-\$D3)*@LN(1-H3)
J	(2/3)/(@EXP(K\$18*(J\$18-\$B3))+1)+1/3
K	+\$D3*@LN(J3)+(\$C3-\$D3)*@LN(1-J3)

表 B.3 采用电子表格软件对数据进行 Logistic 模型拟合以及估计阈值误差限的示例

	A	B	C	D	E	F	G	H	I	J	K
1						最优模型		下限		上限	
2	C mL/m ³	ln(C)	n	r	P _{obs}	P _{est}	似然	P _{est}	似然	P _{est}	似然
3	0.010	-4.61	18	5	0.278	0.366	-10.951	0.414	-11.361	0.349	-10.844
4	0.028	-3.56	18	9	0.500	0.446	-12.583	0.538	-12.528	0.394	-12.892
5	0.032	-3.44	18	4	0.222	0.462	-11.762	0.558	-13.759	0.404	-10.865
6	0.060	-2.81	18	11	0.611	0.568	-12.097	0.671	-12.170	0.478	-12.672
7	0.095	-2.35	18	14	0.778	0.667	-10.072	0.755	-9.561	0.561	-11.382
8	0.285	-1.26	18	16	0.889	0.872	-6.304	0.903	-6.298	0.798	-6.812
9	0.324	-1.13	18	17	0.944	0.888	-4.206	0.914	-3.982	0.821	-5.067
10	0.673	-0.396	18	16	0.889	0.952	-6.865	0.959	-7.060	0.920	-6.389
11	0.992	-0.00803	18	17	0.944	0.970	-4.031	0.973	-4.075	0.951	-3.869
12							-78.871		-80.794		-80.792
13											
14								偏差	3.841		3.841
15											
16											
17						ln(t)	b	x	b	x	b
18						-2.3489	1.311	-2.8354	1.117	-1.8673	1.357
19					P						
20			浓度单位			0.09544		0.05870		0.1545	
21											
22						修正的阈值估计误差限		-2.8354		-1.8673	

B.2.3 基于最大似然法的结果和计算

最大似然法进行模型参数估计的原理是获得能使似然函数(式 B.1)达到最大值时的参数值。电子表格软件含有能计算函数最大值或最小值的功能函数，并方便运用 3-AFC 检验数据进行 Logistic 模型拟合。电子表格程序可广泛安装在实验室的个人电脑中，同时相比统计包具备更多的优点。操作者能很容易理解和熟悉电子表格软件的操作过程，且无须掌握编程语言相关的知识或掌握复杂统计包的工作原理。通过合理数量的观测数据（约 50 个以上，分布在阈值周围）即可顺利地获得参数和阈值的范围。与统计包或其他专用统计程序相比，电子表格软件的优点在于拟合程序的原理对操作者而言更加简单明了，尤其当拟合数据未能达到理想的数据量时，该方法的优势更加明显。

求解函数的最大值或最小值的命令在不同电子表格程序中的命名不同。本例中使用的是 QuattroPro³包中的 Optimizer 函数，等同于 Excel³ 和 Lotus1-2-3³ 中的 Solver 函数。

³ 文中所提及的电子制表软件是目前比较适用的商业化产品。该信息只是给本文件的使用者提供一些便利，

表 B.3 中，列 A~D 是检验数据，A 列是水中柴油浓度，以 C (ml/m^3) 表示；B 列是浓度对数值，表中采用自然对数 $\ln(C)$ （也可使用其他值为基数的对数）。刺激浓度无需为等比的几何序列或特殊间隔的浓度序列。为方便起见，通常按照浓度高低进行排序。

C 列和 D 列分别是每种浓度样品的检测数量 (n) 和该浓度下做出正确选择的数量 (r)。每种浓度的测试评价员数量无需一致。E 列是测试中做出正确选择的概率 (P_{obs})，该值在后面计算中并未出现。

F 列和 G 列是采用式 (3) 中 Logistic 函数拟合的数据。Logistic 模型由参数 b 和 t 确定，通过给定的参数值，即可计算出正确判断的概率估计值。假定给定参数值的模型，似然值是获得的检测比例的概率。所有数据的联合概率等于每一种浓度概率的乘积。当联合似然为对数似然值之和时，使用对数似然值更方便。概率对数值的计算如式 (B.1) 所示：

$$\sum_{i=1}^{i=N} r_i \ln(P_{\text{est}})_i + (n_i - r_i) \ln(1 - P_{\text{est}})_i \quad \dots\dots\dots \text{(B.1)}$$

式中：

N ——数据集（或浓度）的数量；

n ——检验的数量；

r ——判断正确的次数；

P_{est} ——估计比例。

第 12 行是似然值的总和，即联合概率的对数似然值。按照与 F 列和 G 列类似的公式，在 H 和 I 列、J 和 K 列中分别得到 t 和 b 值的下限值和上限值。

本例中，把数据输入到行 3 到行 11 中。引用表 B.2 中 F 列公式，在单元格 F18 和 G18 中分别输入阈值和斜率参数的初始估计值。近似分布在阈值附近的数据集对于初步估计非常可靠，而且初步估计值不需要特别接近最终值。阈值的初始估计值可选取浓度范围的中间值。B 列取浓度自然对数时，斜率初值为 1；若取以 10 为底的对数时，斜率初值为 0.5。优化函数或者具有相同功能的函数通常来自于工具箱或菜单。函数需要给定变量单元的地址（这些单元格包含的数据随着似然函数的最大值的变化而改变，如本例的 F18、G18 和结果单元的地址（该单元为最大似然值，如 G12。

函数有许多选项，其中很多选项设置为缺省值。检查函数是否使结果单元格中数值极大化。在本例中没有设置任何限制条件。函数停止迭代的标准是达到最大迭代次数和变量变化量达到最小，允许操作者根据缺省值重新设置迭代次数和最小变化量。对于提供多种求解方

并不涉及推荐认可产品。

法的电子表格，操作员应该选择广义简化梯度(GRG)非线性方法或适用于光滑非线性问题的等效方法。

开始进行迭代，通过个人计算机可在几秒钟内得到结果，但也有可能因为计算结果收敛太慢而停止迭代。F18和G18单元格中是所需的Logistic模型的参数估计值。

此时， $\ln(t) = -2.35$, $t = 0.095 \text{ ml/m}^3$, $b = 1.31$ 。

这两个参数就把 Logistic 模型定义为最大似然法的 Logistic 模型，即通常所述的最优模型。而参数的其他数值给出了较小的似然，模型之间似然的差异也就是偏差可定义为：

$$\text{偏差} = -2 (l_1 - l_2)$$

式中， l_1 和 l_2 分别是模型 1 和 2 的对数似然。本例中， l_1 和 l_2 分别代表最优模型和另外一个模型的决定值。模型之间的显著性差异由卡方检验进行衡量，自由度取适当的一个值。该关系用于确定估计阈值的误差限，目的是为了找到这样的一个模型，即其 t 值计算的偏差等于在确定的置信度水平和自由度为 1 时的卡方值。当自由度为 1，置信度 p 为 0.05 时，偏差为 3.841。

如果完全指定了适当的列，F3 和 G3 中使用的式可相应的复制到列 H 到 K 中。单元格 I14 的计算式为 $+2 \times (\$G12 - I12)$ ，表示拟合模型的下限和最优模型之间的偏差。把这个式对应复制到 K14 单元格内。

通过一次迭代过程得到误差限，在开始的时候有必要给定一个初始的误差限估计值。误差限的大小由数据集的一些特性所决定，其中最重要的一个特性就是检测总数。表 B.4 列举了阈值估计中步长增加的大小与不同检测总数的初始误差估计范围的关系。本例中，检测次数为 162，对应的初始边界的上下限分别为 $-2.35 + (-0.49) = -2.84$ 和 $-2.35 + 0.56 = -1.79$ 。这些值被输入到 H18 和 J18 单元格。

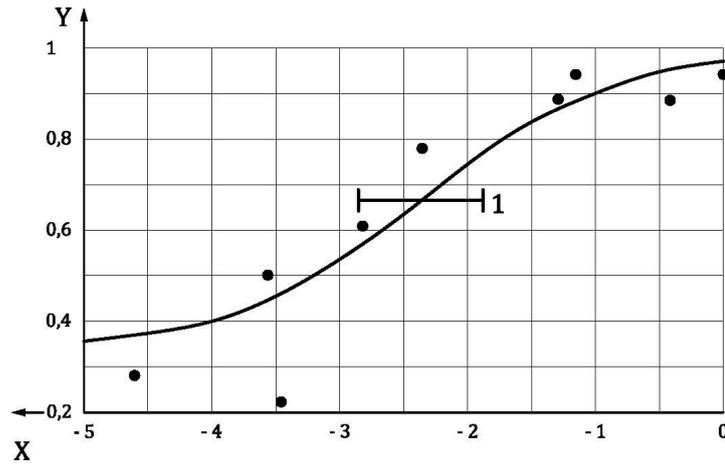
表 B.4 阈值估计中步长增加与初始的误差估计范围

测试总数	40	60	80	100	120	160	200	>200
步长下限	-1.02	-0.81	-0.76	-0.61	-0.58	-0.49	-0.45	-0.40
步长上限	1.39	1.05	0.84	0.78	0.69	0.56	0.49	0.45

单元格 I12 和 K12 是目标单元格，为似然值的总和。如果偏差不充分接近 3.84 这一目标值，应修正所估计的边界值。在数据集中 50 个以上的观测结果分布在阈值周围，这个偏差大致与距离阈值的步长平方根呈线性相关。修正的估计值通过简单的比值就可计算获得。单元格 H22 为计算边界的修正估计值式： $@SQRT(3.841/I14) * (H18 - \$F18) + \$F18$ ，该式被复制到单元格 J22。把修正值结果输入 H18 和 J18 单元格（或在此编辑相应的值），再一次运行优化函数。

当获得的偏差接近期望值时，迭代结束。在表B.3的示例中，三次迭代即足够。

操作员使用电子表格软件的绘图工具绘制测试的数据和 Logistic 拟合曲线，如图 B.1 所示。使用表 B.3 中 F 列数值绘制 Logistic 曲线。充分利用电子表格软件中的函数构造一组等间隔分布的浓度对数数据，并通过表 B.3 中 G 列公式计算概率和参数估计值即可得到一更光滑的曲线。



图中：

- X——浓度的自然对数；
- Y——选择正确的概率；
- 1——误差限；
- 检测值；
- 拟合曲线；

图 B.1—合并数据的组阈值—柴油污染的鱒鱼

参数拟合，尤其是寻找误差限的困难在于对非正常数据集的处理经验。非正常数据集包括小规模数据集、数据偏在阈值一侧的数据集、观测数据的比值随着浓度增加明显偏离平滑的单调递增曲线。优化函数就会无法收敛，给出错误提示后退出。通常是发现 b 值过高、过低甚至呈负值时程序退出（这里检验变量值随着浓度的升高而变大，例如，模型要求的觉察概率随着刺激的浓度增加而增加时， b 值呈正值。当测试不良的气味或水中、气体中的不良气味时，通常是稀释样品使其观测比与稀释程度相关联。若测试的变量值随着浓度的递减反而增加的话， b 值为负）。当拟合 Logistic 或确定边界范围遇到困难时，对实验者而言最好的解决办法是获取更多的数据，而操作员可重复拟合程序，包括调整优化函数的选项来限制 b 值。对 b 值的约束可以是取比 0 大点或者小点的一个数值。当象表 B.3 中列 B 那样取自然对数时， b 值可以小于 2；若取 10 为底的对数时， b 值可以在 1 左右。如果刺激浓度超过 20

倍，觉察概率范围在 0.05~0.95，对应的 b 值上限受到约束。如果未能找到一个合适的解决办法，操作员可指定一个 b 值来拟合阈值。在确定 b 值范围和极大似然函数后，该拟合程序可反复进行。检验极大似然值和 b 值的变化能够确定是否存在一个最优的目标值。

用单元格 H22 定义的式计算非正常数据集时，边界误差估计值有可能收敛很慢或者无法收敛。可在进行迭代时，在电子数据表格的相应列中记录 t 值和相关的偏差。在偏差为 3.84 时，经过三四次迭代，偏差与阈值与估计值之间的距离可绘制成图，从图上可直观地用外推或内插来得到步长，也即偏差减去 3.84 或者找到对应于偏差为 0 时步长。电子表格软件的回归函数也能给出偏差减少值与二次多项式步长的回归关系，解这个式可以在偏差减少值为 0 时得到 t 值。修订的边界估计值在下次迭代计算时使用。一个简单的策略是用新的修订值 x 和对应的偏差去替换存在的偏差和阈值向量，并重新计算表达式，提供一个新的修正值。优化函数中的 b 值应严格限制，以防超出正常范围。重复这一过程，如果存在一个合理的边界值时，计算很快聚合。检查偏差与阈值之间的距离变化可以提示是否能够快速找到一个边界值。

B 2.4 用线性回归法计算

该方法的原理是先采用 6.2.2 中描述的 logit 变换，再使用最小二乘法的计算机统计软件（如 SAS⁴等统计软件）确定与数据最匹配的直线。数据命令参见表 B.3，结果输出如图 B.2。回归模型仅适用于数据对数值在 -1.75~1.75 范围内的数据点。如 6.2.2 所述，超出此范围的值太过极端，无法进行模型拟合和曲线确定。

⁴ 该信息只是给本文件的使用者提供一些便利，并不涉及推荐认可产品。除 SAS 外，SPSS、S-PLUS 和 SYSTAT 软件也同样适用。

```

TITLE "Logistic Regression of Threshold Data";
DATA INPUT;
  INPUT CONC R N;          /* Input Data */
  P=MAX(R/N, 1/3);        /* Compute P(C) */
  LOGIT=LOG((P-1/3)/(1-P)); /* Perform Logit Transformation */
  LOG_C=LOG(CONC);        /* Convert to LOG Concentration */
CARDS;
.010  5 18
.028  9 18
.032  4 18
.060 11 18
.095 14 18
.285 16 18
.324 17 18
.673 16 18
.992 17 18
;
RUN;
DATA TRIMMED;             /* Trim data of extreme values */
  SET INPUT;              /* as per Section 6.2.1 */
  IF LOGIT GE -1,75 AND LOGIT LE 1,75;
RUN;
PROC REG DATA=TRIMMED OUTEST=EST; /* Fit the regression model */
  MODEL LOGIT=LOG_C;
RUN;
DATA EST;
  SET EST;
  LOG_T = -INTERCEP/LOG_C; /* Compute Threshold value */
  THRESHLD = EXP(LOG_T);

```

```

RUN;
PROC PRINT DATA=EST;          /* Output results          */
  VAR INTERCEP LOG_C LOG_T THRESHLD;
RUN;

```

1

Logistic Regression of Threshold Data

Model: MODEL1

Dependent Variable: LOGIT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5,16632	5,16632	26,367	0,0143
Error	3	0,58783	0,19594		
C Total	4	5,75415			

Root MSE 0,44265 R-square 0,8978

Dep Mean 0,49539 Adj R-sq 0,8638

C.V. 89,35510

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEP	1	2,368180	0,41498373	5,707	0,0107
LOG_C	1	0,900890	0,17544692	5,135	0,0143

Logistic Regression of Threshold Data

OBS	INTERCEP	LOG_C	LOG_T	THRESHLD
1	2,36818	0,90089	-2,62871	0,072172

图 B.2 logit 转化数据进行线性回归分析估计阈值示例

得出的回归模型为： $LOGIT=2.3682+0.9009 \log_e \text{浓度}$ ；阈值是 $LOGIT=0$ 时所对应的浓度，因此阈值（自然对数值）为

$\log_e \text{浓度}=-2.3682/0.9009=-2.6287$

或标准单位

$$\text{阈值} = \text{EXP}(-2.6287) = 0.072 \text{ mL/m}^3$$

B.3 以水中柴油为例，比较 2 位评价员的个体阈值

B.3.1 目的

目的是检验 2 位评价员在对某种物质的感知上是否存在差异。本示例给出了最大似然法分层统计模型的应用。

B.3.2 检验

将 1 种化学物质溶于水中，目的是测定和比较 2 位评价员对某种气味的觉察阈值。

B.3.3 结果和计算

方法一^[1]建立了较复杂的分层统计模型，检验数据符合模型的似然。当越多的参数增加到模型中时，似然也增大。随着参数增加而引起的似然函数变化，以偏差表示。偏差用于检验这种新模型是否能引起似然的显著增加。当 3-AFC 数据符合 Logistic 模型时，模型中增加的参数就是 t 值和 b 值。

表 B.5 列出了检验数据及计算结果，表 B.6 是三种统计模型拟合结果的总结。实验中，浓度梯度以及每个浓度的检验数量的设置均采用同种方式，即先在一个广泛浓度范围内初步小试，然后再在接近个体阈值的浓度范围内进行多次重复。

如 B.2 中的示例，重要的是确保在提供了多种求解方法的电子表格中，操作员应选择广义简约梯度(Generalized Reduced Gradient, GRG)法或适用于非线性问题光滑化的类似等效方法。

表 B.5 比较 2 位评价员响应的统计模型计算

C mL/m ³	Ln(C)	n	r	P _{obs}	模型 1		模型 2		模型 3	
					P _{est}	对数似然	P _{est}	对数似然	P _{est}	对数似然
评价员 1										
0.0015	-6.50	3	0	0.000	0.337	-1.23	0.334	-1.22	0.334	-1.22
0.0040	-5.52	3	0	0.000	0.347	-1.28	0.335	-1.23	0.335	-1.23
0.0046	-5.38	3	0	0.000	0.350	-1.29	0.336	-1.23	0.336	-1.23
0.0139	-4.28	3	0	0.000	0.404	-1.55	0.348	-1.29	0.349	-1.29
0.0200	-3.91	5	1	0.200	0.442	-3.15	0.360	-2.81	0.360	-2.81
0.0250	-3.69	7	1	0.143	0.473	-4.59	0.371	-3.78	0.372	-3.78
0.0417	-3.18	3	1	0.333	0.565	-2.24	0.415	-1.95	0.415	-1.95
0.0100	-2.30	13	8	0.615	0.759	-9.33	0.578	-8.70	0.578	-8.70

0.1250	-2.08	3	3	1.000	0.804	-0.65	0.636	-1.36	0.636	-1.36
0.5000	-0.693	5	4	0.800	0.961	-3.41	0.926	-2.91	0.925	-2.90
2.5000	0.916	2	2	1.000	0.995	-0.01	0.994	-0.01	0.994	-0.01
12.5000	2.53	2	2	1.000	1.000	0.00	1.000	0.00	1.000	0.00
评价员 2										
0.0015	-6.50	3	0	0.000	0.337	-1.23	0.344	-1.27	0.344	-1.27
0.0040	-5.52	3	2	0.667	0.347	-2.54	0.384	-2.40	0.383	-2.40
0.0046	-5.38	3	1	0.333	0.350	-1.91	0.395	-1.93	0.394	-1.93
0.0125	-4.38	8	4	0.500	0.395	-5.37	0.560	-5.60	0.559	-5.60
0.0139	-4.28	3	2	0.667	0.404	-2.33	0.586	-1.95	0.585	-1.95
0.0200	-3.91	5	4	0.800	0.442	-3.85	0.682	-2.68	0.682	-2.68
0.0417	-3.18	3	2	0.667	0.565	-1.97	0.855	-2.25	0.856	-2.25
0.0500	-3.00	7	6	0.857	0.604	-3.95	0.886	-2.90	0.886	-2.90
0.0100	-2.30	5	5	1.000	0.759	-1.38	0.958	-0.22	0.958	-0.21
0.1250	-2.08	3	3	1.000	0.804	-0.65	0.970	-0.09	0.970	-0.09
0.5000	-0.693	5	5	1.000	0.961	-0.20	0.997	-0.02	0.997	-0.02
2.5000	0.916	2	2	1.000	0.995	-0.01	1.000	0.00	1.000	0.00
12.5000	2.53	2	2	1.000	1.000	0.00	1.000	0.00	1.000	0.00
总和						-54.48		-47.76		-47.76

模型 1: 该模型中, 只有 1 个 t 值和 1 个 b 值用于估计表 B.5 中 P_{est} 列的值。 P_{est} 列中每个单元格的计算公式均涉及包含 t 和 b 的 2 个单元格, 这 2 个单元格构成电子表格软件中极大化程序的变量单元格。在目标单元格中, 两位评价员的值通过极大化程序得到对数似然值的加和。该函数的运行结果是给出 t 和 b 的最大似然估计值。模型中 t 和 b 的值如表 B.6 所示。

表 B.6 模型的分析表

模型	参数	评价员		对数似然	比较	偏离度	D.F.	χ^2 检验的 P 值
		1	2					
1	t	-2.72		-54.48				
	b	1.37						
2	t	-1.97	-3.97	-47.76	1 与 2	13.43	1	0.0012
	b	1.62						
	b	1.62						
3	t	-1.97	-3.97	-47.76	2 与 3	0.00	1	
	b	1.63	1.62					

模型 2: 数据检查结果表明 2 位评价员的阈值并不相同。模型 2 利用评价员各自的 t 值, 但只有 1 个 b 值。表 B.5 中 P_{est} 列模型 2 中的值就是采用评价员各自的 t 值和共同的 b 值计算得到。包含有 2 个 t 值和 1 个 b 值的 3 个单元格共同构成极大化程序的变量单元格, 目标单元格是该系列数据对数似然值的加和。函数运算结果给出了该模型 t 值和 b 值的最优值。此时模型数据拟合度较好, 且对数似然值较小。模型 2 和模型 1 的参数值和偏差如表 B.6 所示。

模型 3: 模型 3 是模型 2 的扩展, 包含 2 位评价员各自独立的 t 值和 b 值, 即 2 个 t 值和 2 个 b 值, 共计 4 个变量单元。而且, 目标单元格也包含上述数据的对数似然值的和。

B.3.4 结论和推断

表B.6列出了模型的参数和偏差的统计结果。模型每增加1个参数, 自由度就会减1。3 个模型中, 两两模型之间的偏差可采用1个自由度的 χ^2 检验。模型1和模型2之间的偏差极显著 ($p=0.0012$)。继续增加第3个参数, 即第2个 b 值后, 构成的模型3的拟合精度并未提高, 但模型3与模型2之间偏差变为0。

由此推断, 2位评价员对柴油样品气味的觉察阈值存在显著差异, 但对强度增加的敏感性并未表现出差异。

B.4 2种物质的阈值比较——水中的 α -萜烯和 β -萜烯

B.4.1 总则

例B.3中的应用模型同样适用于2种物质的阈值比较。多位评价员的数据合并后得到每种物质的组阈值数据。

B.4.2 检验

检验目的是测量和比较 α -萜烯和 β -萜烯的感官特性。评价小组由24名有经验的评价员组成, 每种物质设置10个浓度水平, 采用3-AFC方法对每个浓度进行检验。

B.4.3 结果和计算

检验数据和电子表格软件计算如表B.7所示, 表B.8中列出模型参数 t 和 b 的值及偏差的统计结果。

B.4.4 总结和推断

模型2对数据的拟合性显著优于模型1, 证实了2种异构体在阈值上存在差异的假设。模

型3有2个 t 值和2个 b 值，拟合度和对数似然值均更大。但模型2和模型3之间的偏差为3.37， χ^2 检验 $p=0.066$ ，与常用的0.05水平相比，并不显著。拟合曲线的斜率存在显著差异，建议进一步展开研究。对于 β -萘烯，4个低浓度数据点均靠近Logistic曲线的渐近线，不适宜用于曲线拟合； α -萘烯的浓度较均匀地涵盖了渐进线之间的浓度范围，适宜曲线的斜率估计。

模型2中的 b 值与模型3中 α -萘烯的 b 值更接近，表明模型3中 β -萘烯的 b 值估计精度低于异构体 α -萘烯。对于 β -萘烯，建议在该检验选取的最高浓度之上再设置2个或更多个浓度点，同时在 $5\text{ mL/m}^3\sim 25\text{ mL/m}^3$ 间选取1个浓度点，进行重复评价。

表 B.7 不同统计模型在计算 α -萘烯和 β -萘烯阈值上的应用

C mL/m ³	Ln(C)	n	r	P _{obs}	模型 1		模型 2		模型 2	
					P _{est}	对数似然	P _{est}	对数似然	P _{est}	对数似然
α -萘烯										
0.004	-5.52	24	11	0.458	0.371	-16.93	0.387	-16.80	0.412	-16.66
0.020	-3.91	24	8	0.333	0.408	-15.56	0.444	-15.89	0.473	-16.24
0.040	-3.22	24	13	0.542	0.432	-17.13	0.480	-16.73	0.508	-16.61
0.100	-2.30	24	14	0.583	0.473	-16.89	0.540	-16.39	0.563	-16.32
0.200	-1.61	24	16	0.667	0.511	-16.46	0.593	-15.55	0.609	15.45
0.500	-0.693	24	15	0.625	0.571	-16.02	0.669	-15.98	0.673	-16.00
1.000	0.00	24	20	0.833	0.622	-13.40	0.727	-11.57	0.722	-11.64
2.500	0.916	24	17	0.708	0.692	-14.50	0.797	-15.02	0.782	-14.85
5.000	1.61	24	19	0.792	0.743	-12.44	0.843	-12.50	0.823	-12.36
25.000	3.22	24	22	0.917	0.847	-7.41	0.920	-6.89	0.897	-6.94
β -萘烯										
0.004	-5.52	24	11	0.458	0.371	-16.93	0.349	-17.16	0.334	-17.35
0.020	-3.91	24	10	0.417	0.408	-16.30	0.368	-16.42	0.336	-16.64
0.040	-3.22	24	12	0.500	0.432	-16.86	0.381	-17.33	0.338	-17.96
0.100	0.00	24	8	0.333	0.473	-16.24	0.407	-15.55	0.346	-15.28
0.200	-1.61	24	11	0.458	0.511	-16.69	0.433	-16.58	0.358	-17.06
0.500	-0.693	24	9	0.375	0.571	-17.74	0.479	-16.40	0.393	-15.89
1.000	0.00	24	8	0.333	0.622	-19.35	0.523	-17.02	0.445	-15.90
2.500	0.916	24	13	0.542	0.692	-17.73	0.591	-16.67	0.558	-16.57
5.000	1.61	24	18	0.750	0.743	-13.50	0.648	-14.07	0.672	-13.85
25.000	3.22	24	21	0.875	0.847	-9.12	0.779	-9.77	0.894	-9.09
总和						-307.21		-300.32		-298.63

表 B.8 不同拟合模型在计算 α -蒎烯和 β -蒎烯阈值上的统计结果分析

模型	参数	α -蒎烯	β -蒎烯	对数似然	模型比较	偏差	自由度	χ^2 检验 P 值
1	t	-0.591		-307.2				
	b	0.459						
2	t	-0.723	-1.831	-300.3	1 与 2	13.79	1	0.000 20
	b	0.506						
3	t	-0.787	-1.581	-298.3	1 与 3	17.16	2	0.000 19
	b	0.426	1.016		2 与 3	3.37	1	0.066

参考文献

- [1] Aitkin M., Anderson D., Francis B., Hinde J. *Statistical Modelling in GLIM*. Clarendon, Oxford, 1989
- [2] ASTM E679-91(1997), *Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Series Method of Limits*
- [3] Amoore J.E. Specific anosmia and the concept of primary odors. *Chemical Senses and Flavor*. 1977, **2** pp. 267–281
- [4] Buttery R.G. Qualitative and sensory aspects of flavor of tomato and other vegetables and fruits. In: *Flavor Science: Sensible Principles and Techniques*, Acree T.E. and Teranishi R., Chapter 8, pp. 277–278, ACS Professional Reference Book, American Chemical Society, Washington, DC, 1993
- [5] Davis H.K., Geelhoed E.N., MacRae, A.W., Howgate P. Sensory analysis of trout tainted by diesel fuel in ambient water. *Water Sci. Technol.* 1992, **25** (2) pp. 11–18
- [6] Devos M., Patte F., Rouault J., Laffort P., van Gemert L.J. *Standardized Human Olfactory Thresholds*. IRL Press, Oxford, 1990
- [7] Doty R.L., Gregor T., Settle R.G. Influences of intertrial interval and sniff bottle volume on the phenyl ethyl alcohol olfactory detection threshold. *Chem. Senses*. 1986, **11** pp. 259–264
- [8] Doty R.L., Deems D.A., Frye R., Pelberg R., Shapiro A. Olfactory sensitivity, nasal resistance, and automatic function in the multiple chemical sensitivities (MCS) syndrome. *Arch. Otolaryngol. Head Neck Surg.* 1988, **114** pp. 1422–1427 [Description of the University of Pennsylvania's Dynamic Air-Dilution Olfactometer]
- [9] EN 13725, *Air Quality — Determination of odour concentration by dynamic olfactometry*
- [10] Fazzalari F.A. *Compilation of Odor and Taste Threshold Values Data*. American Society for Testing and Materials, Philadelphia, 1978
- [11] Guadagni D.G., Buttery R.G. Odor threshold of 2,3,6-trichloroanisole in water. *J. Food Sci.* 1978, **43** pp. 1346–1347
- [12] van Harreveld A., Ph., Heeres P., Harssema H. A review of 20 years of standardization of odor concentration measurement by dynamic olfactometry in Europe. *J. Amer. Waste Management Assoc.*, **49** (5)

[13] Macmillan N.A., Creelman C.D. *Detection Theory, A User's Guide*. Cambridge University Press, 2005, pp. 395

[14] van Gemert L.J., Nettenbreijer A.H. *Compilation of Odour Threshold Values in Air and Water*. Central Institute for Nutrition and Food Research TNO, Zeist, Netherlands, 1977. Supplement V, 1984

[15] ISO 6658, *Sensory analysis — Methodology — General guidance*

[16] ISO 8586, *Sensory analysis — General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors*

[17] ISO 8589, *Sensory analysis — General guidance for the design of test rooms*
