



中华人民共和国国家标准

GB/T 12311—2012
代替 GB/T 12311—1990

感官分析方法 三点检验

Sensory analysis method—Triangle test

(ISO 4120:2004, Sensory analysis—Methodology—Triangle test, MOD)

2012-06-29 发布

2012-11-01 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	I
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 检验条件和要求	2
6 评价员	3
7 程序	3
8 结果分析与表述	4
9 检验报告	4
10 精密度和偏差	4
附录 A (规范性附录) 三点检验所需正确答案数和评价员数查询表	5
附录 B (资料性附录) 示例	9
参考文献	14

前 言

本标准依据 GB/T 1.1—2009 给出的规则起草。

本标准代替 GB/T 12311—1990《感官分析方法 三点检验》。与 GB/T 12311—1990 相比,主要技术变化如下:

- 增加了“术语和定义”、“原理”、“评价员”(见第 3 章、第 4 章和第 6 章);
- 删除了“方法提要”、“设备”和“抽样”(见 1990 年版的第 3 章、第 4 章和第 5 章);
- “检验条件和要求”将具体条件逐项列出(见第 5 章),而不是引用其他标准(1990 年版的 6.1 引用 GB/T 10220);
- 有关评价员要求的内容作为单独条款进行详述(见第 6 章);
- “检验步骤”和“结果表达和解释”(见 1990 年版的第 7 章和第 8 章)更改为“程序”和“结果分析和表述”(见第 7 章和第 8 章);
- 增加了“精密度和偏差”(见第 10 章);
- 删除了附录 A、附录 B 内“答案格式”,增加了统计学表格和示例。

本标准修改采用 ISO 4120:2004《感官分析 方法论 三点检验》(英文版)。与 ISO 4120:2004 相比,本标准做了具有技术性差异的调整,具体调整如下:

- a) 关于“规范性引用文件”:
 - 用 GB/T 10221 代替了 ISO 5492:1992;
 - 用 GB/T 13868 代替了 ISO 8589:1988;
 - 增加了 GB/T 10220、GB/T 14195、GB/T 16291.2 和 GB/T 17321;
- b) 删除了 ISO 标准前言,增加了我国标准前言。

本标准由中华人民共和国农业部提出并归口。

本标准起草单位:中国农业科学院农业质量标准与检测技术研究所、农业部食品质量监督检验测试中心(济南)。

本标准主要起草人:张红、吕潇、钱永忠、王敏、任凤山、张丙春、朱志华、毛雪飞、吴伟。

本标准所代替标准的历次版本发布情况为:

- GB/T 12311—1990。

感官分析方法 三点检验

1 范围

本标准规定了确定两个产品的样品间是否存在可感觉到的感官差别或相似的方法。本标准方法为强迫选择程序。

本标准适用于一种或多种感官指标是否存在的差别的判定。

本标准适用于差别特性未确定时(即差别既不是由样品间差别的大小和范围确定,也没有代表差别的任何特征显示)。本标准仅适用于完全同类的样品。

本标准适用于下述情况:

a) 确定

——产生可感知的差别(三点差别检验);

——未产生可感知的差别(三点相似检验)。例如:当配料、工艺、包装、运输或贮藏有一项变化时。

b) 优选、培训和检验评价员

本标准在统计学上比二-三点检验(由 GB/T 17321 规定)更有效,但限制在余味强和(或)味觉持久的产品上使用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是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。

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

GB/T 10221 感官分析 术语(GB/T 10221—2012,ISO 5492:2008,MOD)

GB/T 13868 感官分析 建立感官分析实验室的一般导则(GB/T 13868—2009,ISO 8589:2007,IDT)

GB/T 16921.1 感官分析 选拔、培训和管理评价员一般导则 优选评价员(GB/T 16921.1—2012,ISO 8586-1:1993,MOD)

GB/T 16291.2 感官分析 选拔、培训和管理评价员一般导则 第2部分:专家评价员(GB/T 16291.2—2010,ISO 8586-2:2008,IDT)

GB/T 17321 感官分析方法 二-三点检验(GB/T 17321—2012,ISO 10399:2004,MOD)

3 术语和定义

GB/T 10221 界定的以及下列术语和定义适用于本文件。

3.1

α -风险 alpha-risk

当感官差别不存在时,推断感官差别存在的概率。

注:这也称为第 I 类错误、显著性水平或假阳性率。

3.2

β -风险 beta-risk

当感官差别存在时,推断感官差别不存在的概率。

注:这也称为第Ⅱ类错误或假阴性率。

3.3

差别 difference

根据感官特性,样品能被区分的状态。

注:能检测出样品间感官差别的评价员比例用符号 P_d 表示。

3.4

产品 product

被评价的原料。

3.5

样品 sample

检验过程中制备、呈送和评价的产品单元。

3.6

敏感性 sensitivity

用于评价检验性能特性的通用术语。

注:在统计学术语中,检验的敏感性用 α 、 β 和 P_d 值定义。

3.7

相似 similarity

样品间的感官差别小至产品可互换的状态。

3.8

三联样 triad

在三点检验中呈送给评价员的三个样品。

注:在三点检验中,用不同的编码标记每个样品。其中两个样品是相同的(即来自一个产品),另一个是不同的(即来自另一产品)。

4 原理

根据检验的敏感性要求选择评价员的数量(见 6.2 和 A.3)。

评价员接到一组三联样并被告知其中两个样品是相同的、另一个是不同的。评价员报出他们认为哪个是不同的,即使此选择仅凭猜测。

计算正确答案数,并根据统计学表确定显著性。

5 检验条件和要求

5.1 用书面形式明确检验目的。

5.2 设备和小隔间应符合 GB/T 13868 的要求。在完成所有评价前,应能防止评价员相互交流。

5.3 在评价员视野外以相同的方式制备样品(即相同器具、相同容器、相同数量的产品)。

5.4 评价员应不能通过样品的呈送方式鉴别出样品。例如,在品尝检验中,避免任何外观差别。用滤光器和(或)柔和灯光掩饰无关的色泽差别。

5.5 用统一的方式对盛样品的容器编码,宜在每次检验中使用随机选择的三位数。每组三联样由三个样品组成,每个样品用不同的编码。在一场检验中,每个评价员宜使用不同的编码。但若在一场检验中

每个评价员仅使用每个编码一次(例如,如果不同产品的几个三点检验在同一场次进行),则同样的三位数编码可以在一次检验中用于所有评价员。

5.6 每组三联样中,三个样品呈送的数量或体积应相同,这与给定产品类型的一系列检验中的其他所有样品一样,应规定被评价样品的数量或体积。否则,应告知评价员无论什么样品,取相似的数量或体积。

5.7 每组三联样中三个样品的温度应该相同,这与给定产品类型的一个检验系列中所有其他样品一样。呈送产品的温度宜与通常食用时一致。

5.8 应告知评价员是否吞咽样品或按其喜欢的方式自由选择。在后一种情况下,应要求评价员对所有样品按同样方式进行。

5.9 检验期间,在完成所有检验前,应避免提供有关产品特性、期望的处理结果或个人表现的信息。

6 评价员

6.1 评价员资格

所有评价员应具有相同资格等级,该等级由检验日的确定(见 GB/T 16291.1 和 GB/T 16291.2)。对产品的经验和熟悉程度可提高评价员的成绩,因而能增加发现显著差别的可能性。监测评价员一段时间内的成绩可能有助于提高检验的敏感性。

所有评价员应熟悉三点检验的技术方法(即形式、任务和评价程序)。

6.2 评价员数量

选择评价员数,以达到检验所需的敏感性(见 A.3)。用大量评价员能提高检出产品间微小差别的可能性。但实际上,评价员的数量通常取决于实际条件(如实验周期、评价员的人数、产品数量)。检验差别时,评价员数通常为 24 人~30 人。检验不显著差别时(即相似),达到同样的敏感性则需要两倍的评价员(即大约 60 人)。

尽量避免同一评价员的重复评价。但是,如果需要重复评价以产生足够的评价总数,应尽量使每位评价员重复评价的次数相同。例如,如果只有 10 位评价员,为得到 30 次评价总数,应让每位评价员评价三组三联样。

注:当用表 A.2 检验相似时,将 10 个评价员做的三次评价作为 30 次独立评价是无效的。但是,利用表 A.1 检验差别时,即使进行重复评价也是有效的。

7 程序

7.1 检验前,准备好工作表和评分表(参见 B.1 和 B.2),使 A 和 B 两产品的六种可能序列出现的次数相等:

ABB	AAB	ABA
BAA	BBA	BAB

六组样品随机分发给评价员(即在第一组六个评价员中使用每个序列一次;在下一组六个评价员中再一次使用每个序列,等等)。当评价员总数不是六的倍数时,会使结果的不平衡性降至最低。

7.2 如可能,向每个评价员按同样的空间排列同时提供每组三联样(例如,三角排列中,同一排通常是从左到右取样)。在一组三联样内,通常允许评价员按其需要对每个样品进行重复评价(当产品的特性允许重复评价)。

7.3 告知评价员按样品的呈送顺序评价。通知评价员两个样品相同,一个不同。评价员应指出三个样品中哪一个与另外两个不同。

7.4 每张评分表仅用于一组三联样。如果在一场检验中一个评价员进行一次以上的检验,在呈送后续的三联样之前,应收走填好的评分表和未用的样品。评价员不应取回先前的样品或更改先前的检验结论。

7.5 评价员做出选择后,不要问其有关偏好、接受或差别程度的问题。对任何附加问题的答案可能影响评价员刚做出的选择。这些问题的答案可通过独立的偏好、接受、差别程度检验等获得(见 GB/T 10220)。询问为何做出选择的陈述部分可以包含评价员的解释。

7.6 三点检验是强迫选择程序,不允许评价员回答“无差别”。当评价员无法判断出差别时,应要求评价员随机选择一个样品,并且在评分表的陈述栏中注明,该选择仅是猜测。

8 结果分析与表述

8.1 差别检验

用表 A.1 分析三点检验得到的数据。如果正确答案数大于或等于表 A.1 列出的值(符合评价员数和本检验选择的 α -风险水平),结论为:样品间存在感官差别(参见 B.1)。

如需要,计算能识别样品的人员比例的置信区间。见 B.3。

8.2 相似¹⁾检验

根据表 A.2,分析三点检验获得的数据。如果正确答案数小于或等于表 A.2 列出的值(符合评价员数、本检验选择的 β -风险水平和 P_a 值),则结论为:样品间不存在明显差别(参见表 B.2)。若结果用于一个检验与另一个检验的比较,则所有检验应选择相同的 P_a 值。

如果需要,通过计算可识别样品的人员比例的置信区间。该方法在 B.3 中规定。

当正确答案的最大值小于 $n/3$ 时,不应给出结论。

9 检验报告

报告本检验的对象、结果和结论,建议给出下列附加信息:

- 检验的目的和分析方法的性质;
- 样品的详细说明(即来源、制备方法、数量、形态、检验前的贮藏、呈送量、温度),如果仅由于关注点变化而生产不同样品,应告知样品所有贮藏、处理和制备的信息;
- 评价员人数、正确答案数和统计评价结果(包括用于本检验的 α 、 β 和 P_a 值);
- 评价员:经历(感官检验中对产品、对检验样品)、年龄和性别(见 GB/T 16291.1 和 GB/T 16291.2);
- 向评价员提供与检验有关的信息和具体建议;
- 检验环境(即所用的检验设备、同时或有顺序地呈送样品、检验后样品特性是否公开,若公开,以何种方式);
- 检验地点、日期和组长姓名。

10 精密度和偏差

由于感官鉴别检验的结果是个体敏感性的功能,因此不能做出适于所有评价员结果再现性的一般表述。对于一定数量的评价员,其精密度随小组规模的增加而提高,也随培训和对产品接触的增加而提高。

由于使用强迫选择程序,因此只有完全遵守第 7 章的规定,本方法获得的结果才无偏差。

1) 本标准中,“相似”不是指“相同”。确切地说,“相似”是指两个产品十分相似且可以互换。不可能证明两个产品是完全相同的。但能证实两个产品间存在的差别很小以致无实际意义。

附录 A
(规范性附录)

三点检验所需正确答案数和评价员数查询表

A.1 表 A.1 中列出的值是在一定 α -风险水平(列)上,达到显著性所需的最少正确答案数和相应的评价员数 n (行)。如果正确答案数大于或等于表 A.1 中的值,则拒绝“无差别”的假设。

表 A.1 三点检验确定存在显著性差别所需最少正确答案数

n	α					n	α				
	0.20	0.10	0.05	0.01	0.001		0.20	0.10	0.05	0.01	0.001
6	4	5	5	6	—	27	12	13	14	16	18
7	4	5	5	6	7	28	12	14	15	16	18
8	5	5	6	7	8	29	13	14	15	17	19
9	5	6	6	7	8	30	13	14	15	17	19
10	6	6	7	8	9	31	14	15	16	18	20
11	6	7	7	8	10	32	14	15	16	18	20
12	6	7	8	9	10	33	14	15	17	18	21
13	7	8	8	9	11	34	15	16	17	19	21
14	7	8	9	10	11	35	15	16	17	19	22
15	8	8	9	10	12	36	15	17	18	20	22
16	8	9	9	11	12	42	18	19	20	22	25
17	8	9	10	11	13	48	20	21	22	25	27
18	9	10	10	12	13	54	22	23	25	27	30
19	9	10	11	12	14	60	24	26	27	30	33
20	9	10	11	13	14	66	26	28	29	32	35
21	10	11	12	13	15	72	28	30	32	34	38
22	10	11	12	14	15	78	30	32	34	37	40
23	11	12	12	14	16	84	33	35	36	39	43
24	11	12	13	15	16	90	35	37	38	42	45
25	11	12	13	15	17	96	37	39	41	44	48
26	12	13	14	15	17	102	39	41	43	46	50

注 1: 因为表中的数值根据二项式分布求得,因此是准确的。对于表中未设的 n 值,根据下列二项式的近似值计算其近似值。

最小正确答案数(x)=大于式中最接近的整数: $x=(n/3)+z\sqrt{2n/9}$

其中 z 随下列显著性水平变化而异: $\alpha=0.02$ 时,0.84; $\alpha=0.10$ 时 1.28; $\alpha=0.05$ 时,1.64; $\alpha=0.01$ 时,2.33; $\alpha=0.001$ 时,3.09。

注 2: 当 n 值 ≤ 18 时,不宜用三点检验检验差别。

A.2 表 A.2 中所列出的值是在选定的 P_α 、 β 、 n 水平上,检验“相似”所需的最大正确答案数。如果正确答案数小于或等于表 A.2 中的值,则接受在 $100(1-\beta)\%$ 置信水平上“无差别”的假设。

表 A.2 根据三点检验确定两个样品相似所需最大正确答案数

n	β	P_d					n	β	P_d				
		10%	20%	30%	40%	50%			10%	20%	30%	40%	50%
18	0.001	0	1	2	3	5	66	0.001	14	18	22	26	31
	0.01	2	3	4	5	6		0.01	16	20	25	29	34
	0.05	3	4	5	6	8		0.05	19	23	28	32	37
	0.10	4	5	6	7	8		0.10	20	25	29	33	38
	0.20	4	6	7	8	9		0.20	22	26	31	35	40
24	0.001	2	3	4	6	8	72	0.001	15	20	24	29	34
	0.01	3	5	6	8	9		0.01	18	23	28	32	38
	0.05	5	6	8	9	11		0.05	21	26	30	35	40
	0.10	6	7	9	10	12		0.10	22	27	32	37	42
	0.20	7	8	10	11	13		0.20	24	29	34	39	44
30	0.001	3	5	7	11	14	78	0.001	17	22	27	32	38
	0.01	5	7	9	14	16		0.01	20	25	30	36	41
	0.05	7	9	11	16	18		0.05	23	28	33	39	44
	0.10	8	10	11	17	19		0.10	25	30	35	40	46
	0.20	9	11	13	18	21		0.20	27	32	37	42	48
36	0.001	5	7	9	11	14	84	0.001	19	24	30	35	41
	0.01	7	9	11	14	16		0.01	22	28	33	39	45
	0.05	9	11	13	16	18		0.05	25	31	36	42	48
	0.10	10	12	14	17	19		0.10	27	32	38	44	49
	0.20	11	13	16	18	21		0.20	29	34	40	46	51
42	0.001	6	9	11	14	17	90	0.001	21	27	32	38	45
	0.01	9	11	14	17	20		0.01	24	30	36	42	48
	0.05	11	13	16	19	22		0.05	27	33	39	45	52
	0.10	12	14	17	20	23		0.10	29	35	41	47	53
	0.20	13	16	19	22	24		0.20	31	37	43	49	55
48	0.001	8	11	14	17	21	96	0.001	23	29	35	42	48
	0.01	11	13	17	20	23		0.01	26	33	39	45	52
	0.05	13	16	19	22	26		0.05	30	36	42	49	55
	0.10	14	17	20	23	27		0.10	31	38	44	50	57
	0.20	15	18	22	25	28		0.20	33	40	46	53	59
54	0.001	10	13	17	20	24	102	0.001	25	31	38	45	52
	0.01	12	16	19	23	27		0.01	28	35	42	49	56
	0.05	15	18	22	25	29		0.05	32	38	45	52	59
	0.10	16	20	23	27	31		0.10	33	40	47	54	61
	0.20	18	21	25	28	32		0.20	36	42	49	56	63
60	0.001	12	15	19	23	27	108	0.001	27	34	41	48	55
	0.01	14	18	22	26	30		0.01	31	37	45	52	59
	0.05	17	21	25	29	33		0.05	34	41	48	55	63
	0.10	18	22	26	30	34		0.10	36	43	50	57	65
	0.20	20	24	28	32	36		0.20	38	45	52	60	67

注1: 表中的数值根据二项式分布求得, 是准确的。对于表中没有的 n 值, 根据以下二项式的近似值计算 P_d 在 $100(1-\beta)\%$ 水平的置信上限:

$$[1.5(x/n) - 0.5 + 1.5Z_\beta \sqrt{(nx - x^2)/n^2}]$$

式中:

x —— 正确答案数;

n —— 评价员数;

Z_β —— Z_β 的变化如下: $\beta=0.20$ 时, 0.84; $\beta=0.10$ 时, 1.28; $\beta=0.05$ 时, 1.64; $\beta=0.01$ 时, 2.33; $\beta=0.001$ 时, 3.09。

如果计算值小于选定的 P_d 值, 则表明样品在 β 显著性水平上相似。

注2: 当 $n < 30$, 不宜用三点检验检验相似。

A.3 表 A.3 列出确定评价员数的统计方法。检验统计学的敏感性是三个参数： α -风险、 β -风险和能识别样品人员的最大允许比例 P_d ²⁾ 的函数。检验之前，用下述准则选择 α 、 β 和 P_d 值。

作为经验法则，统计学的显著性结果在：

- 10%到5%(0.10到0.05)的 α -风险表示差别不显著；
- 5%到1%(0.05到0.01)的 α -风险表示差别中等显著；
- 1%到0.1%(0.01到0.001)的 α -风险表示差别显著；
- 0.1%以下(<0.001)的 α -风险表示差别极显著。

对于 β -风险，也用上述标准评价差别不明显的显著程度(用“ β ”代替“ α ”，用“不显著”代替“显著”)。

能识别样品的人员最大允许比例 P_d 在下列三个范围内：

- $P_d < 25\%$ ，代表小值；
- $25\% < P_d < 35\%$ ，代表中等值；
- $P_d > 35\%$ ，代表大值。

选择评价员人数以达到本检验所需的敏感性。从表 A.3 中，找到选定的 P_d 和 β 值，在选定的 α 值行中，查找所需的最少评价员数。反之，当使用的评价员数少于所需的限量值时，用表 A.3 可导出一组提供可接受敏感性的 α 、 β 、 P_d 值。该方法在参考文献中介绍。

表 A.3 列出的值，是由 α 、 β 、 P_d 值确定的敏感性一定时，进行三点检验所需的最少评价员数。从表 A.3 中，由选定的 P_d 、 β 和 α 值，查找最少评价员数。

表 A.3 三点检验所需的评价员数

α	P_d	β				
		0.20	0.10	0.05	0.01	0.001
0.20	50%	7	12	16	25	36
0.10		12	15	20	30	43
0.05		16	20	23	35	48
0.01		25	30	35	47	62
0.001		36	43	48	62	81
0.20	40%	12	17	25	36	55
0.10		17	25	30	46	67
0.05		23	30	40	57	79
0.01		35	47	56	76	102
0.001		55	68	76	102	130
0.20	30%	20	28	39	64	97
0.10		30	43	54	81	119
0.05		40	53	66	98	136
0.01		62	82	97	131	181
0.001		93	120	138	181	233
0.20	20%	39	64	86	140	212
0.10		62	89	119	178	260
0.05		87	117	147	213	305
0.01		136	176	211	292	397
0.001		207	257	302	396	513

2) 本标准中，正确答案的概率， P_c 由模型 $P_c = P_d + (1/3)(1 - P_d)$ 求得。其中， P_d 是能够区分两个样品的评价员的比例。评价员判定过程的心理学模型，例如 Thurstone-Ura 模型，也适用于三点检验。

表 A.3 (续)

α	P_d	β				
		0.20	0.10	0.05	0.01	0.001
0.20	10%	149	238	325	529	819
0.10		240	348	457	683	1 011
0.05		325	447	572	828	1 181
0.01		525	680	824	1 132	1 539
0.001		803	996	1 165	1 530	1 992

附录 B
(资料性附录)
示 例

B.1 示例 1: 三点检验确定存在差别

B.1.1 背景

某啤酒厂开发了一项工艺,以降低无醇啤酒中不良谷物风味。该工艺需投资新设备。在进行大规模的包括消费者的偏好检验之前,厂主想要确定研制的无醇啤酒与公司目前生产的无醇啤酒不同。当差别不存在时,厂主愿意冒点风险断定差别存在。而且,若有替代新工艺的方法,厂主乐意接受更大的风险忽略确实存在的差别。

B.1.2 检验目的

确定新工艺生产的无醇啤酒能区别于原无醇啤酒,以证明消费者检验的正确性。

B.1.3 评价员数

为避免厂主得到差别存在的错误判断,感官分析员建议 $\alpha=0.05$ 。为平衡样品的呈送顺序,分析员决定用 24 个评价员。根据表 A.3,选择用 24 个评价员,保证检验中检出差别的机会为 95% [即

表 B.1 示例 1 中三点检验检验差别的评分表

三点检验		
评价员编号	姓名	日期
说明: 从左到右品尝样品。两个样品相同;一个不同。在下面空白处写出与其他样品不同的样品编号。如果无法确定,记录你的最佳猜测;可以在陈述处注明你是猜测的。 与其他两个样品不同的样品是: _____ 陈述: _____		

B.2 示例 2: 三点检验确定两个样品相似

B.2.1 背景

一个糖果生产商想使用新包装材料,因为它使标签图案有更大的灵活性。但是,这种新材料应能提供相同的贮藏稳定性。生产商知道不可能证明两种产品相同,但想确定:贮藏三个月后,如果存在差别,只有非常小的评价员比例能检出。另外,由于原包装受欢迎,因此如果两种产品相同,生产商愿意冒相当大的风险,错误判定产品不同。对生产商来说,贮藏稳定性比图案吸引人更重要。

B.2.2 检验目的

确定用新包装材料贮藏三个月的产品是否与用原包装材料的相同。

B.2.3 评价员数

感官分析员和生产商一起确定适于本检验的风险水平,确定能够区分产品的评价者的最大允许比例为 $P_d=20\%$ 。生产商仅愿意冒 $\beta=0.10$ 的风险来检测评价者。当不存在差别时,生产商不太注重做出存在差别的错误判定。感官分析员选择 $\alpha=0.20$ 。当 $\alpha=0.20$ 、 $\beta=0.10$ 、 $P_d=20\%$,在表 A.3 中查到需要评价员 $n=64$ 个。

B.2.4 检验

感官分析员用表 B.2 所示的工作表和表 B.3 所示的评分表进行检验。分析员用六组可能的三联样: AAB、ABA、BAA、BBA、BAB 和 ABB 循环 10 次送给前 60 个评价员。然后,随机选择四组三联样送给 61 号至 64 号评价员。

B.2.5 结果分析与表述

检验中,在 64 个评价员中有 24 人正确辨认出检验的不同样品。查阅表 A.2,分析员发现没有 $n=64$ 的条目。因此,分析员用表 A.2 的注 1 中的公式,来确定能否得出两个样品相似的结论。分析员算出:

$$[1.5 \times (24/64) - 0.5] + 1.5 \times 1.28 \sqrt{(64 \times 24 - 24^2)/64^3} = 0.1787$$

即分析员有 90% 的置信度,小于 18% 的评价员能区分样品;分析员有 90% (即 $\beta=0.10$) 的确定性,判断新包装材料符合制造商的要求,而且不超过 $P_d=20\%$ 的评价员能检出差别。新包装材料可以代替原来的材料。

B.3 三点检验的置信区间

B.3.1 背景

如需要,分析员能计算出可识别样品的评价员比例的置信区间。计算公式见式(B.1)~式(B.5):

$$P_c = x/n \quad \text{.....(B.1)}$$

$$\hat{P}_d = 1.5P_c - 0.5 \quad \text{.....(B.2)}$$

$$S_d = 1.5\sqrt{P_c(1-P_c)/n} \quad \text{.....(B.3)}$$

$$\text{置信上限: } \hat{P}_d + Z_\alpha S_d \quad \text{.....(B.4)}$$

$$\text{置信下限: } \hat{P}_d - Z_\alpha S_d \quad \text{.....(B.5)}$$

式中:

x —— 正确答案数;

n —— 评价员总数;

P_c —— 正确率;

$\hat{P}_d$ —— 识别人员比例;

S_d —— $\hat{P}_d$ 的标准偏差;

Z_α —— 标准正态分布的临界值。对于 90% 的置信区间, $Z_\alpha = 1.28$; 对于 95% 的置信区间, $Z_\alpha = 1.64$; 对于 99% 的置信区间, $Z_\alpha = 2.33$ 。

表 B.2 示例 2 的工作表

日期:1993 年 10 月 4 日				检验编号:587-FF03			
三点检验样品顺序和呈送计划							
在样品托盘准备区张贴本表。提前将评分表和呈送容器编码。							
产品类型:糖块							
样品编码							
样品 1=包装 4 736(原来的) 样品 2=包装 3 987(新的)							
呈送容器编码如下:							
评价组成员		样品编码		评价组成员		样品编码	
1	1-108	1-795	2-140	16	2-582	2-659	1-486
2	1-189	2-168	1-733	17	2-429	1-884	2-499
3	2-718	1-437	1-488	18	1-879	2-891	2-404
4	2-535	2-231	1-243	19	1-745	1-247	2-724
5	2-839	1-402	2-619	20	1-344	2-370	1-355
6	1-145	2-296	2-992	21	2-629	1-543	1-951
7	1-792	1-280	2-319	22	2-482	2-120	1-219
8	1-167	2-936	1-180	23	2-259	1-384	2-225
9	2-689	1-743	1-956	24	1-293	2-459	2-681
10	2-442	2-720	1-213	25	1-849	1-382	2-192
11	2-253	1-444	2-505	26	1-294	2-729	1-390
12	1-204	2-159	2-556	27	2-165	1-661	1-336
13	1-142	1-325	2-632	28	2-281	2-409	1-126
14	1-472	2-762	1-330	29	2-434	1-384	2-948
15	2-965	1-641	1-300	30	1-819	2-231	2-674

表 B.2 (续)

日期:1993 年 10 月 4 日				检验编号:587-FF03			
三点检验样品顺序和呈送计划							
在样品托盘准备区张贴本表。提前将评分表和呈送容器编码。							
产品类型:糖块							
样品编码							
样品 1=包装 4 736(原来的) 样品 2=包装 3 987(新的)							
呈送容器编码如下:							
评价组成员 样品编码				评价组成员 样品编码			
31	1-740	1-397	2-514	48	1-574	2-393	2-753
32	1-354	2-578	1-815	49	1-793	1-308	2-742
33	2-360	1-303	1-415	50	1-147	2-395	1-434
34	2-134	2-401	1-305	51	2-396	2-629	1-957
35	2-185	1-651	2-307	52	1-147	2-395	1-434
36	1-508	2-271	2-465	53	2-525	1-172	2-917
37	1-216	1-941	2-321	54	1-325	2-993	2-736
38	1-494	2-783	1-414	55	1-771	1-566	2-376
39	2-151	1-786	1-943	56	1-585	2-628	1-284
40	2-423	2-477	1-164	57	2-354	1-526	1-595
41	2-570	1-772	2-887	58	2-358	2-606	1-586
42	1-398	2-946	2-764	59	2-548	1-201	2-684
43	1-747	1-286	2-913	60	1-475	2-339	2-573
44	1-580	2-558	1-114	61	1-739	1-380	2-472
45	2-345	1-562	1-955	62	1-417	2-935	1-784
46	2-385	2-660	1-856	63	2-127	2-692	1-597
47	2-754	1-210	2-864	64	1-157	2-315	1-594

表 B.3 示例 2 的评分表

三点检验		检验编号:587-FF03	
评价员编号:21	姓名:_____	日期:_____	
样品类型:糖块			
说明:			
从左到右品尝托盘中的样品。两个样品相同;一个不同。选出不同的样品,并且在对应的方格内划×。			
盘内样品	标出不同的样品	说明:	
629	☐	_____	
543	☐	_____	
951	☐	_____	
注:若解释选择的原因或样品特性,可写在说明栏中。			

B.3.2 结果分析与表述

以上述示例 2 的数据为例,当 $x=24$, $n=64$, 则:

$$P_c = 24/64 = 0.375$$

$$\hat{P}_d = (1.5 \times 0.375) - 0.5 = 0.0625$$

$$S_d = 1.5 \sqrt{0.375(1-0.375)/64} = 0.0908$$

$$90\% \text{ 的置信上限: } 0.0625 + (1.28 \times 0.0908) = 0.18$$

$$90\% \text{ 的置信下限: } 0.0625 - (1.28 \times 0.0908) = -0.05$$

分析员检验相似时,分析员能在 90% 的置信水平上确定,能区分样品的人员实际比例不大于 18%。当分析员检验差别时,由于 90% 的置信下限是负值,因此 $P_d = 0\%$ 在置信区间内,且是一个可能值,从而支持样品间无感官差别的结论。

总之,置信区间允许置信上限和下限有 10% 的误差。因此,感官分析员在 80% 的置信水平上确信:真正的识别人员比例是总评价人的 0%~18%。依据本检验的目的,研究人员可选择使用单边置信上限、单边置信下限或双边置信限。

参 考 文 献

- [1] ISO 3534-1 Statistics—Vocabulary and symbols—Part 1: Probability and general statistical terms
 - [2] BROCKHOFF, P. B., SCHLICH, P. Handling replications in discrimination tests. Food Quality and Preference, 1998, 9(5): 303-312.
 - [3] ENNIS, D. M., BI, J. The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. Journal of Sensory Studies, 1998, 13(4): 389-412.
 - [4] FRIJTERS, J. E. R. Three-stimulus procedure in olfactory psychophysics: An experimental comparison of thurstone-ura and three-alternative forced-choice models of signal detection theory. Perception & Psychophysics, 1980, 28(5): 390-397.
 - [5] KUNERT, J., MEYNEERS, M. On the triangle test with replications. Food Quality and Preference, 1999(10): 477-482.
 - [6] KUNERT, J. On repeated difference testing. Food Quality and Preference, 2001(12): 385-391.
 - [7] MEILGAARD, M., CIVILLE, G. V., CARR, B. T. Sensory evaluation techniques. 2nd Edition. CRC Press, Inc., Boca Raton, FL, 1991: 381.
 - [8] SCHLICH, P. Risk tables for discrimination tests. Food Quality and Preference, 1993(4): 141-151.
-