



中华人民共和国国家标准

GB/T 17321—2012
代替 GB/T 17321—1998

感官分析方法 二-三点检验

Sensory analysis method—Duo-trio test

(ISO 10399:2004, Sensory analysis—Methodology—Duo-trio 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 发布
中国国家标准化管理委员会

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..... 5

附录 A（规范性附录） 二-三点检验所需正确答案数和评价员数查询表 6

附录 B（资料性附录） 示例 10

参考文献 15

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准代替 GB/T 17321—1998《感官分析方法 二-三点检验》，与 GB/T 17321 相比，其主要技术变化如下：

- 增加了用于优选、培训和检验评价员的内容(见第 1 章)；
- “引用标准”改为“规范性引用文件”(见第 2 章)；
- 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 1998 年版的第 4 章、第 5 章和第 6 章)；
- “一般检验条件和要求”将具体条件逐项列出(见第 5 章)，而不是引用其他标准(见 1998 年版的第 7 章)；
- 删除了有关评价员要求，而是作为单独条详述(见第 6 章)；
- “检验步骤”和“结果表示”(见 1998 年版的第 8 章和第 9 章)更改为“程序”、“结果分析和表述”和“(见第 7 章和第 8 章)；
- 增加了“精密度和偏差”(见第 10 章)；
- 附录 A、附录 B 内删除“答案格式”，增加了统计学表格和示例。

本标准修改采用 ISO 10399:2004《感官分析 方法论 二-三点检验》(英文版)。与 ISO 10399:2004 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 12311、GB/T 14195 和 GB/T 16291.2。
- b) 删除了国际标准前言，增加了我国标准前言。
- c) 7.3 原文“注”的内容，编排为标准正文。

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：任凤山、张丙春、吕潇、钱永忠、王敏、朱志华、毛雪飞、王磊、赵平娟、吴伟。

本标准所代替标准的历次版本发布情况为：

- GB/T 17321—1998。

感官分析方法 二-三点检验

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程序。本标准方法为强迫选择程序。

本标准适用于单一或几种感官特性存在的差别。本标准也适用于差别特性未知时(即它既不确定样品间的差别程度也不确定差别范围,也没有任何特性差别迹象)。本标准仅适用于产品相当相似时。

本标准适用于下述情况:

a) 确定

——有感觉差别(二-三点检验检验差别);

——无感觉差别(二-三点检验检验相似),例如当配料、工艺、包装、处理或贮藏有一项改变时。

b) 用于优选、培训和检验评价员

本标准方法的两种形式:

——恒定参比技术,用于评价员对一个产品熟悉时(如样品来自固定生产线);

——平衡参比技术,用于评价员对一个产品不比另一个产品更熟悉时。

本标准统计有效性低于三点检验(见 GB/T 12311),但评价员较易实施。

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 12311 感官分析方法 三点检验(GB/T 12311—2012,ISO 4120:2004,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 16291.1 感官分析 选拔、培训和管理评价员一般导则 第1部分:优选评价员(GB/T 16291.1—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)

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:这也被认为是第I类错误、显著水平或假阳性率。

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:这也被认为是第Ⅱ类错误或假阴性率。

3.3

差别 difference

根据感官特性,样品能被区别的状态。

注:能检出产品间感官差别的评价员比例用符号 P_d 表示。

3.4

产品 product

被评估的原料。

3.5

样品 sample

检验过程中制备、呈送和评估的产品单元。

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

样品间的感官差别很小以致于产品可互换的状态。

3.8

三联样 triad

在二-三点检验中提供给评价员的三个样品。

注:在二-三点检验中,一个样品被标记为参比样,另外两个样品用不同编码标识。其中一个编码样品与参比样是同一产品,另外一个编码样品是检验中的其他产品。

4 原理

根据检验要求的敏感性选择评价员数(见 6.2 和 A.3)。

评价员得到一组三个样品(即三联样),一个样品被标记为参比样,另外两个样品编码不同。告知评价员其中一个编码样品与参比样相同,一个与参比样不同。根据检验前的训练和指导,评价员应报告哪个编码样品与参比样相同,或哪个编码样品与参比样不同。

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 设施和小隔间应符合 GB/T 13868 的要求。在完成所有评价前,应防止评价员相互交流。

5.3 在评价员视野外以完全相同的方式(即相同器具、相同容器、产品数量相同)制备样品。

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)柔和灯光掩饰任何不相干的色泽差别。

5.5 用统一方式对盛有样品的容器进行编码,宜使用为每个检验随机选择的三位数字。每组三联样一

个被标记为参比样,两个用不同编码标记。在一个会期期间,每个评价员宜使用不同编码。但若一个检验会期期间每个评价员仅使用每个编码一次(例如,若相同会期内实施了不同产品的几个二-三点检验),在一项检验内所有评价员可使用两个相同编码。

5.6 每组三联样内三个样品的呈送数量或体积应完全相同,对于一种规定的产品类型在一系列检验内的所有其他样品也相同。应规定被评估的数量或体积。若未规定,应告知评价员无论任何样品取相似的数量或体积。

5.7 每组三联样内三个样品的温度应完全相同,对于一种规定的产品类型在一系列检验内的所有其他样品也相同。宜在产品通常的食用温度呈送样品。

5.8 应告知评价员是否可吞咽样品或是否可按他们喜欢的方式随意去做。后一种情况,应要求评价员对所有样品以相同的方式进行。

5.9 检验会期内,完成所有检验前应避免给出有关产品特性、预期处理结果或独特特性的信息。

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,该等级根据检验目的确定(见 GB/T 16291.1 和 GB/T 16291.2)。对产品的经验和熟悉程度可改善一个评价员的成绩,因而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敏感性。

所有评价员应熟悉二-三点检验技术方法(即形式、任务和评价程序)。

6.2 评价员数

选择评价员数以达到检验所需敏感性(见 A.3)。使用大量评价员增加检出产品之间微小差别的可能性。但实际上,评价员数通常决定于具体条件(如试验周期、可利用评价员人数、产品数量)。当检验差别时,具有代表性的评价员数在 32 位~36 位之间。当检验无合理差别时(即相似),为达到相当的敏感性需要两倍评价员数(即大约 72 位)。

尽量避免同一位评价员的重复评价。若需要重复评价以得出产品足够数量的总评价,应尽量使每一位评价员评价次数相同。例如,仅有 12 位评价员可利用,应使每位评价员评价三组组合以得到 36 个总评价数。

注:当用表 A.2 检验相似时,由 12 位评价员得出的 36 个独立评估的三组评估论述无效。然而,当给出重复评估时用表 A.1 的差别检验是有效的。关于重复选择检验的出版物提出近似值二选一,用于选择检验中的平行评估分析。

7 程序

7.1 本标准方法有恒定参比技术和平衡参比技术两种形式。若评价员熟悉产品(如来自生产线的控制样),使用恒定参比技术。若对于两个产品都不太熟悉,使用平衡参比技术。

a) 恒定参比技术

检验前准备工作表和评分表(见 B.2),使用数目相同的 A、B 两个产品两种可能的序列。

$A_R AB$ $A_R BA$

在评价员之间两人一组随机分发样品(即在第一组两个评价员之间用一个序列,在下一组两个评价员中再使用这个序列,等等)。若评价员总数是奇数时,会使结果的不平衡性降至最低。

b) 平衡参比技术

检验前准备工作表和评分表(见 B.2),使用数目相同的 A、B 两个产品四种可能的序列。

$A_R AB$ $A_R BA$
 $B_R AB$ $B_R BA$

系列中前两个组合含有产品 A(即 A_R)作为对照,后两个组合含有产品 B(即 B_R)作为对照。在评价员之间四人一组随机分发样品(即在第一组四个评价员之间用一个序列;在下一组四个评价员中再使用这个序列,等等)。若评价员总数不是四的倍数时,会使结果的不平衡性降至最低。

7.2 若可能,应同时呈送每组三个样品,随后为每个评价员提供相同的空间排列(如总是从左到右直线排列、以三角排列呈送样品)。在三联样内,若愿意,一般允许评价员为每个样品给出重复评价(若产品性质允许做出重复评价)。

7.3 要求评价员首先评价参比样,然后按顺序评价呈送的两个编码样品。告知评价员其中一个编码样品与参比样相同,一个与参比样不同。要求评价员指出两个编码样品中与参比样相同的一个,或两个编码样品中与参比样不同的一个。

当要求评价员选择与参比样相同样品或选择与参比样不同的样品时,应考虑常规评价小组是否使用了其他鉴别检验方法。许多鉴别检验方法如三点检验,侧重于检验中鉴别“不寻常的”或“不同的”样品。要求评价员在一种方法中识别“不同”样品并在另一种方法中识别“相同”样品,这可能引起混淆并导致高概率的不正确回答。

7.4 应为每个三联样组合提供一个评分表。若一个评价员在一个会期内进行一项以上检验,在呈送随后的三联样前收集全部评分表和未用样品。评价员不能追溯到以前样品或改变以前检验的结论。

7.5 评价员做出选择后不要询问有关偏爱、接受或差别程度的问题。对任何附加问题的回答可能影响到评价员做出的选择。这些问题的答案可通过独立的偏爱、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包含评价员的陈述。

7.6 二-三点检验是强迫选择程序;评价员不允许回答“无差别”。应要求检验出样品之间无差别的评价员随机选择一个样品,并在评分表陈述部分内指明这项选择仅是一个猜测。

8 结果分析与表述

8.1 差别检验

用表 A.1 分析由二-三点检验获得的数据。若正确答案数大于或等于表 A.1 中给出的数(对应评价员数和检验选择的 α -风险水平),推断样品之间存在感官差别(见 B.1)。

若需要,根据能区分样品人数的比例计算出置信区间。方法见 B.3。

8.2 相似检验¹⁾

用表 A.2 分析由二-三点检验获得的数据。若正确答案数小于或等于表 A.2 中给出的数(对应评价员数和检验选择的 β -风险水平和 P_d 值),则推断出样品之间不存在有意义的感官差别(见 B.2)。

若一项检验与另外一项对照检验结果,则应为所有检验选择相同的 P_d 值。

若需要,根据能区分样品的人数比例计算置信区间。方法见 B.3。

9 检验报告

给出检验对象、检验结果和结论。建议给出以下附加信息:

1) 本标准中,“相似”不是指“等同”。更确切地说,“相似”指两个产品足够相似可用来互换。不可能证实两个产品完全相同,但可表明存在于两个产品之间的差别很小以致无实际意义。

- 检验目的和样品处理的特性；
- 样品的全部标识(即来源、制备方法、数量、状态、检验前的储藏、呈送的量、温度),样品信息应传达所有已进行的储藏、处理和制备,以这种方式生产的样品仅仅由于关注点的变化而不同,无论任何情况；
- 评价员人数、正确答案数和统计评价结果(包括检验使用的 α 、 β 和 P_d 值)；
- 评价员:经验(感官检验中、对产品、对检验中样品),年龄和性别(见 GB/T 16291. 1 和 GB/T 16291. 2)；
- 对评价员给出的有关检验的任何信息和明确建议；
- 检验环境(即所用检验设施、同时或连续呈送,检验后样品特征是否公开,若公开,以何种方式)；
- 检验地点、日期及小组组长姓名。

10 精密度和偏差

因为感官鉴别检验结果取决于个体敏感性,因此不能做出适用于所有评价员的结果再现性的一般陈述。关于特定数目评价员的精密度随小组规模的增加而增加,也随训练和对产品的接触而增加。

作为强迫选择程序使用时,若完全遵守第 7 章的注意事项,通过本标准方法得到的结果无偏差。

附 录 A
(规范性附录)

二-三点检验所需正确答案数和评价员数查询表

A.1 表 A.1 给出的值是在规定的 α 显著性水平所需的最少正确答案数(列),和相应的评价员数 n (行)。若正确答案数大于或等于表 A.1 中的值,则“无差别”的假设不成立。

表 A.1 二-三点检验推断感官差别存在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	5	6	6	—	—	26	16	17	18	20	22
7	6	6	7	7	—	27	17	18	19	20	22
8	6	7	7	8	—	28	17	18	19	21	23
9	7	7	8	9	—	29	18	19	20	22	24
10	7	8	9	10	10	30	18	20	20	22	24
11	8	9	9	10	11	32	19	21	22	24	26
12	8	9	10	11	12	36	22	23	24	26	28
13	9	10	10	12	13	40	24	25	26	28	31
14	10	10	11	12	13	44	26	27	28	31	33
15	10	11	12	13	14	48	28	29	31	33	36
16	11	12	12	14	15	52	30	32	33	35	38
17	11	12	13	14	16	56	32	34	35	38	40
18	12	13	13	15	16	60	34	36	37	40	43
19	12	13	14	15	17	64	36	38	40	42	45
20	13	14	15	16	18	68	38	40	42	45	48
21	13	14	15	17	18	72	41	42	44	47	50
22	13	14	15	17	19	76	43	45	46	49	52
23	15	16	16	18	20	80	45	47	48	51	55
24	15	16	17	19	20	84	47	49	51	54	57
25	16	17	18	19	21	88	49	51	53	56	59
<p>注 1: 因为是根据二项式分布得到,表中的值是准确的。对于不在表中的 n 值,根据下列二项式的正常近似值为遗漏的登记项计算近似值: 最少正确答案数(x)=大于下式的最近似整数:$x=(n/2)+z\sqrt{n/4}$ 其中 z 随以下显著水平不同而不同:$\alpha=0.20$ 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: n 值<24 时,通常不推荐二-三点检验检验差别。</p>											

A.2 表 A.2 给出的值是在选择的 P_d 、 β 和 n 水平检验“相似”所需的最大正确答案数。若正确答案数小于或等于表 A.2 中的值,则在 $100(1-\beta)\%$ 置信水平“无差别”的假设成立。

表 A.2 二-三点检验推断两个样品相似所需最大正确答案数

<i>n</i>	β	<i>P_d</i>					<i>n</i>	β	<i>P_d</i>				
		10%	20%	30%	40%	50%			10%	20%	30%	40%	50%
20	0.001	3	4	5	6	8	56	0.001	18	21	24	27	30
	0.01	5	6	7	8	9		0.01	21	24	27	30	33
	0.05	6	7	8	10	11		0.05	24	27	29	32	36
	0.10	7	8	9	10	11		0.10	25	28	31	34	37
	0.20	8	9	10	11	12		0.20	27	30	32	35	38
24	0.001	5	6	7	9	10	60	0.001	20	23	26	30	33
	0.01	7	8	9	10	12		0.01	23	26	29	33	36
	0.05	8	9	11	12	13		0.05	26	29	32	35	38
	0.10	9	10	12	13	14		0.10	27	30	33	36	40
	0.20	10	11	13	14	15		0.20	29	32	35	38	41
28	0.001	6	8	9	11	12	64	0.001	22	25	29	32	36
	0.01	8	10	11	13	14		0.01	25	28	32	35	39
	0.05	10	12	13	15	16		0.05	28	31	34	38	41
	0.10	11	12	14	15	17		0.10	29	32	36	39	43
	0.20	12	14	15	17	18		0.20	31	34	37	41	44
32	0.001	8	10	11	13	15	68	0.001	24	27	31	34	38
	0.01	10	12	13	15	17		0.01	27	30	34	38	41
	0.05	12	14	15	17	19		0.05	30	33	37	40	44
	0.10	13	15	16	18	20		0.10	31	35	38	42	45
	0.20	14	16	18	19	21		0.20	33	36	40	43	47
36	0.001	10	11	13	15	17	72	0.001	26	29	33	37	41
	0.01	12	14	16	18	20		0.01	29	32	36	40	44
	0.05	14	16	18	20	22		0.05	32	35	39	43	47
	0.10	15	17	19	21	23		0.10	33	37	41	44	48
	0.20	16	18	20	22	24		0.20	35	39	42	46	50
40	0.001	11	13	15	18	20	76	0.001	27	31	35	39	44
	0.01	14	16	18	20	22		0.01	31	35	39	43	47
	0.05	16	18	20	22	24		0.05	34	38	41	45	50
	0.10	17	19	21	23	25		0.10	35	39	43	47	51
	0.20	18	20	22	25	27		0.20	37	41	45	49	53
44	0.001	13	15	18	20	23	80	0.001	29	33	38	42	46
	0.01	16	18	20	23	25		0.01	33	37	41	45	50
	0.05	18	20	22	25	27		0.05	36	40	44	48	53
	0.10	19	21	24	26	28		0.10	37	41	46	50	54
	0.20	20	23	25	27	30		0.20	39	43	47	52	56
48	0.001	15	17	20	22	25	84	0.001	31	35	40	44	49
	0.01	17	20	22	25	28		0.01	35	39	43	48	52
	0.05	20	22	25	27	30		0.05	38	42	46	51	55
	0.10	21	23	26	28	31		0.10	39	44	48	52	57
	0.20	23	25	27	30	33		0.20	41	46	50	54	59
52	0.001	17	19	22	25	28	88	0.001	33	37	42	47	52
	0.01	19	22	25	27	30		0.01	37	41	46	50	55
	0.05	22	24	27	30	33		0.05	40	44	49	53	58
	0.10	23	26	28	31	34		0.10	41	46	50	55	60
	0.20	25	27	30	33	35		0.20	43	48	52	57	62

表 A.2 (续)

n	β	P _d					n	β	P _d				
		10%	20%	30%	40%	50%			10%	20%	30%	40%	50%
92	0.001	35	40	44	49	55	104	0.001	40	46	51	57	63
	0.01	38	43	48	53	58		0.01	44	50	55	61	66
	0.05	42	46	51	56	61		0.05	48	53	59	64	70
	0.10	43	48	53	58	63		0.10	50	55	60	66	71
	0.20	46	50	55	60	65		0.20	52	57	63	68	73
96	0.001	37	42	47	52	57	108	0.001	42	48	54	59	65
	0.01	40	45	50	56	61		0.01	46	52	57	63	69
	0.05	44	49	54	59	64		0.05	50	55	61	67	72
	0.10	46	50	55	60	66		0.10	52	57	63	68	74
	0.20	48	53	57	62	67		0.20	54	60	65	71	76
100	0.001	39	44	49	54	60	112	0.001	44	50	56	62	68
	0.01	42	47	53	58	64		0.01	48	54	60	66	72
	0.05	46	51	56	61	67		0.05	52	58	63	69	75
	0.10	48	53	58	63	68		0.10	54	60	65	71	77
	0.20	50	55	60	65	70		0.20	56	62	68	73	79

注 1: 因为是根据二项式分布得到,表中的值是准确的。对于不在表中的 *n* 值,根据下列二项式的正常近似值计算 100(1-β)%置信上限 *P_d* 近似值:

$$[2(x/n)-1]+2z_{\beta}\sqrt{(nx-x^2)/n^3}$$

式中:

x ——正确答案数;

n ——评价员数;

z_β ——随以下显著水平不同而不同:β=0.20 时,0.84;β=0.10 时,1.28;β=0.05 时,1.64;β=0.01 时,2.33;β=0.001 时,3.09。

若计算值小于选择的 *P_d* 限,则声明样品在 β 显著水平相似。

注 2: *n* 值<36 时,通常不推荐二-三点检验检验相似。

A.3 表 A.3 给出了确定评价员数的统计学方法。检验的统计学敏感性是三个值的作用:α-风险、β-风险和识别人员的最大允许比例 *P_d*²⁾。实施检验前,用下述导则选择 α、β 和 *P_d* 值。

作为一个经验法则,统计学显著结果在

- 10%~5%(0.10~0.05)的 α-风险表明差别不显著;
- 5%~1%(0.05~0.01)的 α-风险表明差别中等显著;
- 1%~0.1%(0.01~0.001)的 α-风险表明差别显著;
- 低于 0.1%(<0.001)的 α-风险表明差别非常显著。

对于 β-风险,差别不明显的显著程度用上述相同要求来评价(“β”代替“α”,“不明显”代替“明显”)。

识别人员的最大允许比例 *P_d* 在以下三个范围内:

- P_d*<25%,代表小值;
- 25%<*P_d*<35%,代表中值;
- P_d*<35%,代表大值。

2) 在本标准中,正确答案的概率 *P_c* 按模型 *P_c*=*P_d*+(1-*P_d*)/2 得出,其中 *P_d* 是能区分两个产品的评价员人数比例。一个评价员决策过程的心理学模型,如在二-三点检验中 Thurstone-Ura 模型也适用。

选择评价员数以得到检验所需敏感性水平。进入表 A. 3 选择的 P_d 值的相应部分和选择的 β 值的相应列。在选择的 α 值的相应行找到所需评价员的最少数。反之,表 A. 3 可用来导出当维持评价员数在实际限量内时提供的可接受敏感性的一系列 α 、 β 和 P_d 值。方法在参考文献中详述。

表 A. 3 给出的值是执行由 α 、 β 和 P_d 值决定的具有特定敏感性的二-三点检验所需评价员最小数。进入表中选择的 P_d 值的相应部分和选择的 β 值的相应列。在选择的 α 值的相应行读出所需评价员数。

表 A. 3 二-三点检验所需评价员数

α	P_d	β				
		0. 20	0. 10	0. 05	0. 01	0. 001
0. 20	50%	12	19	26	38	58
0. 10		19	26	33	48	70
0. 05		23	33	42	58	82
0. 01		40	50	59	80	107
0. 001		61	71	83	107	140
0. 20	40%	19	30	39	60	94
0. 10		28	39	53	79	113
0. 05		37	53	67	93	132
0. 01		64	80	96	130	174
0. 001		95	117	135	176	228
0. 20	30%	32	49	68	110	166
0. 10		53	72	96	145	208
0. 05		69	93	119	173	243
0. 01		112	143	174	235	319
0. 001		172	210	246	318	412
0. 20	20%	77	112	158	253	384
0. 10		115	168	214	322	471
0. 05		158	213	268	392	554
0. 01		252	325	391	535	726
0. 001		386	479	556	731	944
0. 20	10%	294	451	618	1 006	1 555
0. 10		461	658	861	1 310	1 905
0. 05		620	866	1 092	1 583	2 237
0. 01		1 007	1 301	1 582	2 170	2 927
0. 001		1 551	1 908	2 248	2 937	3 812

附录 B
(资料性附录)
示 例

B.1 示例 1:确定差别存在的二-三点检验——平衡参比技术

B.1.1 背景说明

一个番茄汤制造商想要推荐一个新的且价值更高的低盐配方以期获得市场收益。在将其提交品尝检验与旧配方对比之前,商家希望证实能通过感觉区分两个产品。选择平衡参比模式的二-三点检验,因为产品的复杂风味使得评价员的判断过程简单非常重要。当一个产品没有差别时,厂家领导乐意接受推断出仅一项差别存在的微小概率。但因为旧产品依然很受欢迎,他乐意接受一个弄错差别不存在的较高风险。

B.1.2 检验目的

确定新产品(B)能和传统产品(A)区分开,以验证消费者试验。

B.1.3 评价员数

为防止提供给厂家领导一个差别不存在的不真实推断,感官分析人员建议 $\alpha=0.01$ 。为平衡样品呈送顺序,分析人员决定采用 36 位评价员。

B.1.4 实施检验

制备样品(54 份 A 和 54 份 B)。其中 18 份样品“A”和 18 份样品“B”被标记为参照样。其余 36 个样品“A”和 36 个样品“B”用唯一性随机三位数进行编码。然后,全部样品分为 9 个系列,每个系列由以下 4 组样品组成。每组样品内呈送的第一份为参照样,标明为 A_R 或 B_R ,示例如下:

$A_RAB \quad B_RAB$
 $A_RBA \quad B_RBA$

每四个三联样组合被呈送 9 次,以使以平衡的随机顺序涉及 36 位评价员。工作表见表 B.1。评分表示例见表 B.2。

B.1.5 结果分析与表述

28 位评价员正确识别与参照相同样品。表 A.1 中对应 36 位评价员的行和对应 $\alpha=0.01$ 的列内,感官分析人员找出在 $\alpha=0.01$ 显著水平推断感官差别存在需要 26 个正确答案数。因此,28 个正确答案足以推断两个产品有感官差别。

推荐分析人员可根据能感觉出样品间差别的人数比例选择计算一个单边低置信区间。

计算结果为(亦见 B.3):

$$[2(28/36)-1]-2 \times 2.33 \sqrt{(28/36)[1-(28/36)]}/36=0.233$$

至少 23% 的人数能感觉样品间差别,分析人员可推断出 99% 置信水平。

B.1.6 报告和结论

感官分析人员报告在 1%($n=36, x=28$)显著水平,评价小组实际能区分传统产品的范例。使用新

工艺的生产试验应转到 B. 1. 2 提出的消费者试验。

表 B. 1 示例 1 工作表

日期:2003-09-20				检验编码:TX-0245			
二-三点检验样品顺序和呈送草案							
在样品盘制备区域公布本表格。预先将评分表和呈送容器编码。							
产品类型:番茄汤							
样品编码:							
A=传统的(编码 941 和 387) B=新的(编码 792 和 519)							
呈送容器编码如下							
专家小组成员	样品编码			专家小组成员	样品编码		
1	A _R	A-941	B-792	19	A _R	A-941	B-792
2	A _R	B-792	A-941	20	B _R	B-519	A-387
3	B _R	A-387	B-519	21	B _R	A-387	B-519
4	B _R	B-519	A-387	22	B _R	B-519	A-387
5	B _R	A-387	B-519	23	A _R	A-941	B-792
6	A _R	B-792	A-941	24	A _R	B-792	A-941
7	A _R	A-941	B-792	25	A _R	A-941	B-792
8	B _R	B-519	A-387	26	A _R	B-792	A-941
9	B _R	A-387	B-519	27	B _R	A-387	B-519
10	A _R	A-941	B-792	28	B _R	B-519	A-387
11	B _R	B-519	A-387	29	A _R	A-941	B-792
12	A _R	B-792	A-941	30	B _R	B-519	A-387
13	B _R	A-387	B-519	31	B _R	A-387	B-519
14	B _R	B-519	A-387	32	A _R	B-792	A-941
15	A _R	A-941	B-792	33	B _R	A-387	B-519
16	A _R	B-792	A-941	34	B _R	B-519	A-387
17	B _R	A-387	B-519	35	A _R	A-941	B-792
18	A _R	B-792	A-941	36	A _R	B-792	A-941
注 1: 用参照(Ref)或指定的随机三位数标记样品杯并按给每位评价员的呈送顺序排列。							
注 2: 在一个呈送盘内呈送、放置样品和一份编码评分表。							
注 3: 无论回答正确与否都回传涉及的工作表。							

表 B. 2 二-三点差别检验示例 1 评分表

二-三点检验			
评价员编码:_____姓名:_____日期:_____			
说明:			
从左到右品尝样品。左侧样品为参照,其他两个样品之一与参照相同。另外一个与参照不同。在与参照相同的样品框内标记“×”。若不确定,标记最好的猜测;也可在猜测的标记下作出标注。			
参照	941	792	
陈述:_____			
_____。			

B.2 示例 2:确定两个样品相似的二-三点检验——恒定参比技术

B.2.1 背景说明

一个软饮料公司希望证实申请的新包装不改变饮料的风味,消费者察觉不到差别。厂方管理者知道不可能证明两个产品完全相同,但他希望证实若差别存在,仅有一小部分评价员可能察觉到差别。另一方面,当产品相同时,他乐意接受错误推断出产品不同的相当大的概率,因为这仅仅意味着恢复满意的旧包装,可能改进新包装,然后再进行检验。

B.2.2 检验目的

确定填充并贮存在新包装内的产品与填充并贮存在传统包装内的产品是否非常相似。

B.2.3 评价员数

感官分析人员建议使用传统产品作为恒定参照的二-三点检验,因为评价员熟知这个产品,不需要花费时间或精力去熟悉产品本身的风味。分析人员然后与厂方管理者协商决定检验所用的风险水平。决定辨别人员的最大允许比例应为 $P_d=30\%$ 。制造商仅愿意采用不能检出的辨别人员 $\beta=0.05$ 的概率水平。应为检验招募 52 位感官分析评价员。

B.2.4 实施检验

感官分析人员使用表 B.3 和表 B.4 进行检验。分析小组由传统包装(A)制备 104 份呈送产品,并由新包装(B)制备 52 份呈送产品,以得到每两个可能的三联样组合的 26 组呈送样品: A_RAB 和 A_RBA 。

B.2.5 结果分析与表述

一位评价员缺席检验。51 位评价员参与检验,25 位正确识别出检验中与参照不同的样品。涉及表 A.2,分析人员发现没有 $n=51$ 的条目。则分析人员使用表 A.2 注 1 中的公式确定是否可推断出两个样品相似。分析人员发现:

$$[2(25/51)-1]-2\times 1.64\sqrt{(51\times 25-25^2)/51^3}=0.210$$

即,不大于 21% 的评价员能区分样品时,分析人员确定有 95% 的置信水平。分析人员推断新包装符合制造商以 95% 置信水平(即 $\beta=0.05$)确定不大于 $P_d=30\%$ 的人员能检出差别的判断标准。新包装可替代传统包装。

B.3 示例 3:二-三点检验置信区间

B.3.1 背景说明

若愿意,分析人员能根据区分样品人数的比例计算一个置信区间。如下计算:

$$P_c=x/n;$$

$$P_d=2P_c-1;$$

$$S_d=2\sqrt{P_c(1-P_c)/n}$$

表 B.3 示例 2 工作表

日期:2001-10-04				检验编码:587-FF03			
二-三点检验样品顺序和呈送草案							
在样品盘制备区域公布本表格。提前将评分表和呈送容器编码。							
产品类型: <u>软饮料</u>							
样品编码:							
A=包装 4736(传统的)				B=包装 3987(新的)			
呈送容器编码如下							
专家小组成员	样品编码			专家小组成员	样品编码		
1	A _R	A-795	B-168	27	A _R	A-795	B-168
2	A _R	B-168	A-795	28	A _R	B-168	A-795
3	A _R	A-795	B-168	29	A _R	A-795	B-168
4	A _R	B-168	A-795	30	A _R	B-168	A-795
5	A _R	A-795	B-168	31	A _R	A-795	B-168
6	A _R	B-168	A-795	32	A _R	B-168	A-795
7	A _R	A-795	B-168	33	A _R	A-795	B-168
8	A _R	B-168	A-795	34	A _R	B-168	A-795
9	A _R	A-795	B-168	35	A _R	A-795	B-168
10	A _R	B-168	A-795	36	A _R	B-168	A-795
11	A _R	A-795	B-168	37	A _R	A-795	B-168
12	A _R	B-168	A-795	38	A _R	B-168	A-795
13	A _R	A-795	B-168	39	A _R	A-795	B-168
14	A _R	B-168	A-795	40	A _R	B-168	A-795
15	A _R	A-795	B-168	41	A _R	A-795	B-168
16	A _R	B-168	A-795	42	A _R	B-168	A-795
17	A _R	A-795	B-168	43	A _R	A-795	B-168
18	A _R	B-168	A-795	44	A _R	B-168	A-795
19	A _R	A-795	B-168	45	A _R	A-795	B-168
20	A _R	B-168	A-795	46	A _R	B-168	A-795
21	A _R	A-795	B-168	47	A _R	A-795	B-168
22	A _R	B-168	A-795	48	A _R	B-168	A-795
23	A _R	A-795	B-168	49	A _R	A-795	B-168
24	A _R	B-168	A-795	50	A _R	B-168	A-795
25	A _R	A-795	B-168	51	A _R	A-795	B-168
26	A _R	B-168	A-795	52	A _R	B-168	A-795
注 1: 用参照(Ref)或指定的随机三位数标记样品杯并按给每位评价员的呈送顺序排列。							
注 2: 在一个呈送盘内呈送、放置样品和一份编码评分表。							
注 3: 无论回答正确与否都回传涉及的工作表。							

表 B.4 示例 2 评分表

二-三点检验		检验编码 587-FF03
评价员编码: 21 姓名: 日期:		
样品类型:软饮料		
说明: 在样品盘中从左到右品尝样品。左侧样品为参照,其他样品之一与参照不同。选择 <u>不同</u> 的样品并在相应的样品框内标记“×”。		
盘内样品	指明与参照不同的样品	陈述:
参照		
795	_____	_____
168	_____	_____。
注: 如果你希望说明选择的理由或样品特性,可在陈述栏内描述。		

置信上限: $P_d + Z_\alpha S_d$
置信下限: $P_d - Z_\alpha S_d$
式中:
 P_c —— 正确比例;
 x —— 正确答案数;
 n —— 评价员总数;
 P_d —— 识别人员比例;
 S_d —— P_d 标准差;
 Z_α —— 标准正态分布的临界值。对于 90% 的置信区间, $Z_\alpha = 1.28$; 95% 的置信区间, $Z_\alpha = 1.64$; 99% 的置信区间, $Z_\alpha = 2.33$ 。

B.3.2 结果分析与表述

假设示例 2 的数据, $x=25, n=51$ 。则:
 $P_c = 25/51 = 0.49$
 $P_d = 2 \times 0.49 - 1 = -0.02$
 $S_d = 2\sqrt{0.49(1-0.49)/51} = 0.14$
95% 置信上限: $-0.02 + 1.64 \times 0.14 = 0.21$
95% 置信下限: $-0.02 + 1.64 \times 0.14 = -0.25$
若分析人员检验相似,分析人员有 95% 的置信水平确定能区分样品人员的实际比例不大于 21%。反之,若分析人员检验差别,由于 95% 置信下限是负值, $P_d = 0\%$ 在区间内,并因此是一个可能的值,因而支持样品间无感官差别的结论。
总体来看,置信区间允许上下限 5% 的误差,因而区分人员的真正比例约在人数的 0%~21% 之间,感官分析人员有 90% 的置信水平。根据检验目的,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 and Psychophysics, 1980, 28(5): 390-397.
 - [5] KUNERT, J. , MEYNNERS, 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: 338.
 - [8] SCHLICH, P. Risk tables for discrimination tests. Food Quality and Preference, 1993(4): 141-151.
-

中 华 人 民 共 和 国
国 家 标 准
感官分析方法 二-三点检验
GB/T 17321—2012

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100013)
北京市西城区三里河北街16号(100045)

网址 www.spc.net.cn
总编室:(010)64275323 发行中心:(010)51780235
读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 880×1230 1/16 印张 1.25 字数 33 千字
2012 年 11 月第一版 2012 年 11 月第一次印刷

*

书号: 155066·1-45751

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107



GB/T 17321-2012